# Communication as Joint Action: The role of cognitive alignment and coupling

Adam Boncz

Central European University

Department of Cognitive Science

*In partial fulfillment of the requirements for the degree of*

*Doctor of Philosophy in Cognitive Science*

Primary supervisor: Günther Knoblich

Secondary supervisor: Thalia Wheatley

Budapest, Hungary

2019

## 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.

Adam Boncz

# Abstract

Human communication is a multi-faceted phenomenon. Here we focused on communication in a joint action framework and aimed to answer three questions. First, we asked if people communicate efficiently in helping situations where signals can have a direct effect on task performance. We tested this question in four experiments using a precueing version of a reaction time task, where a helper participant's action provided a cue for a helpee participant. We found that helpers communicated efficiently but helpees did not utilize helpers' signals as much as they could. While helpers traded their own effort for helpees' performance gain, helpees avoided relying on helpers' communication, leading to a tension on the pair level. Second, we tested if alignment in verbal interactions is modulated by interactivity and individual goals, contrasting predictions of the interactive alignment model and automatic imitation accounts. Interactivity and goal overlap were modulated in a joint storytelling scenario and alignment was captured at multiple linguistic levels. We found independent effects of interactivity and individual goals: prosodic alignment (in terms of temporal structure) was affected only by the goal manipulation, while syntactic, lexical and semantic alignment was mostly modulated by interactivity. Our results suggest that interactivity increases high-level linguistic alignment, but prosodic alignment is unaffected by it. Third, we tested if interactivity elicits stronger brain-to-brain coupling using an fMRI hyperscanning setup. Employing the joint storytelling task we found evidence for stronger predictive coupling in an interactive condition relative to a non-interactive condition, potentially linked to temporal predictive processes. In sum, our work emphasizes the importance of studying communication from a broad, integrative perspective and by employing a variety of techniques.

# Acknowledgements

First, I would like to thank Günther, Natalie and Thalia for their supervision. I am deeply grateful for your continuous support, encouragement and guidance throughout all these years. I feel very fortunate that I had the opportunity to work with you – you have taught me so much not only as researchers but also as fantastic human beings.

I would also like to thank all current and former members of the SOMBY lab at CEU and the Social Systems Lab at Dartmouth. The discussions, ideas and – above all – the company and friendship of all of you was an essential part of this journey. There are too many people I should mention here, but I am especially grateful to Cordula, Vero and Beau for their insights, help and overall attitude to research. Special praise is in order for Veronika, Annamari, Terry, Stephanie and Tammy, whose help in the labs was invaluable.

I would also like to thank the whole Department of Cognitive Science at CEU for its creative and productive atmosphere. Reka and Andi, you are simply fantastic. Further thanks go to the Hasson lab for their inspiration and warmheartedness. You turned a short visit into a rich and cheerful experience.

Of course, I am grateful to my family and friends for their support. You are the lively mix of joy, reassurance, love, tolerance and forgiveness in my life that makes anything possible.

I would also like to thank CEU, the institution. Without this unique university I would have never had the opportunity to meet my supervisors, fellow students, collaborators and friends.

Finally, but most importantly, thank you Brigi. This thesis is dedicated to you and Samuel.

# Table of Contents

# Chapter One. Introduction

The last conversation I had before writing this paragraph was with the clerk of a coffee shop. During our brief exchange we both performed numerous actions: he operated the coffee machine while I fidgeted with a cardboard sleeve; he handed over the cup of coffee; I took the cup and handed over a banknote; he offered the change and so on. Our interaction was highly scripted and we were both prepared for its stages – after all, in other circumstances, the same procedure is reduced to button pushes on a vending machine. Still, in order for both of us to achieve our respective goals we needed to match our behavior and we did so through a variety of means. We uttered a few sentences - we greeted each other, I stated the order, he acknowledged it, we exchanged pleasantries, etc, - but also used gaze, posture and movements to glue our interaction together, coordinating our individual actions.

Most everyday interactions are comparable to the one above: we coordinate individual actions and communicative acts so that they are integrated seamlessly into the larger unit of interaction (H. Clark, 1996; Dale, Fusaroli, Duran and Richardson, 2013; Pickering and Garrod, 2004). Such interactions  - dialogues in terms of language use – are also the most natural form of communication (e.g. Garrod and Pickering, 2004), the primary arena of language acquisition (e.g. Ninio and Bruner, 1978; Tomasello, 1988) and the most effective way of knowledge transmission (e.g. Fox Tree, 1999; Kraut, Lewis and Sqezey, 1982; Mercer and Littleton, 2007).

Traditionally, the study of interactions emphasizes communication, especially language use. A long list of scholars investigating human communication have chosen interactions as the basic unit of their inquiries (e.g. Brennan, Galati and Kuhlen, 2010; Grice, 1989; Krauss and Weinheimer, 1966; Levinson, 1983; Sacks, Schegloff and Jefferson, 1978; Clark, 1996) and have argued that communication – including, but not limited to language use - can only be understood when analyzed as social actions (termed the language-as-action tradition by H. Clark, 1992).

Another, more recent approach grounded in the broader frame of embodied cognition (e.g. A. Clark, 1999) has focused on understanding interactions as the product of basic perceptual, cognitive and motor processes, invoking language use and other higher-level processes only as the far end of the spectrum of coordination tools (e.g. Gallese, Keysers and Rizzolatti, 2004; Knoblich and Sebanz, 2008; Schmidt and Richardson, 2008; Sebanz, Bekkering and Knoblich, 2006; Sebanz and Knoblich, 2009). Empirical research from this approach has studied the “nuts and bolts” of joint actions (Knoblich, Butterfill and Sebanz, 2011, p.), identifying a number of coordination mechanisms behind a wide range of collective human behavior from dancing to teaching.

My aim in this chapter is to bring these approaches closer together and to identify the main interfaces between joint action research and communication. By doing so I will also introduce the wider theoretical background for the thesis, as my work looks at interactions from multiple viewpoints. In what follows, I will consider joint action research as the starting point, providing a brief overview (section 1.1), and then look for its connections with communication. I will do so by approaching communication both as a coordination device serving joint action (section 1.2), and as a form of joint action in and of itself (section 1.3).

## 1.1. Joint action

### 1.1.1. Interactions as joint actions

When the barista handed me the cup of coffee and I grasped it, we coordinated our actions in time and space rather precisely to fulfill a common goal. Arguably, many everyday interactions contain similar episodes that are cooperative in nature and require immediate (here-and-now) coordination – common examples include dancing, music duets, moving a box together, etc. These situations have been studied under the term joint action, “whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment” (Sebanz et al., 2006, p. 70) or a “…mutually desired outcome” (H. Clark, 1996, p. 61). The emphasis here is on the problem of coordination – joint action research is primarily concerned with the representations and processes behind the fine-grained coordination of actions in ensembles (Knoblich et al., 2011; Vesper et al., 2017a).

A complete review of joint action research is outside the scope of this chapter. In what follows, instead, we first illustrate the phenomena researchers of joint action have been concerned with through an example. Then we introduce the minimalist cognitive model by Vesper, Butterfill, Knoblich and Sebanz (2010) that specifies the building blocks sufficient for most joint actions.

### 1.1.2. Object passing - a case study in joint action

Consider the barista passing the cup of coffee again. Our movements were tightly coupled - we know that object passing happens quickly, taking a few hundred milliseconds of haptic contact between passer and receiver (Controzzi et al., 2018; Endo, Pegman, Burgin, Toumi and Wing, 2012; Mason and MacKenzie, 2005), and that passer’s grip force starts to decrease precisely around receiver contact (most often in a ±50 ms window, see Controzzi et al., 2018; Mason and MacKenzie, 2005).

To investigate the processes behind such precision, Kourtis, Sebanz and Knoblich (2013a) measured EEG correlates of action planning in a grasping task where participants were asked to either pass, receive, individually lift an object or observe the other participant lifting it. They found that joint actions (passing or receiving) corresponded to stronger planning demands than individual actions, and that receiving the object correlated with motor processes temporally locked to passer’s action. The motor correlates were also linked with the temporal precision of behavioral coordination. Overall, the results suggested that coordination in object passing is achieved by (1) co-representing the other participant’s task and (2) predicting the other’s action via motor simulation (for further evidence see Kourtis, Knoblich, Wozniak and Sebanz, 2014; Ménoret et al., 2014). The same processes are assumed to explain the known effects of social context on reaching and grasping kinematics (e.g. Becchio, Sartori, Bulgheroni and Castiello, 2008; Becchio, Sartori and Castiello, 2010), and are theorized as general coordination devices prevalent across many types of joint action (Knoblich et al., 2011; Vesper et al., 2017a). Besides task co-representation and action simulation, further important processes supporting coordination - without being exhaustive - include entrainment, perception-action matching and joint perception (Knoblich et al., 2011).

### 1.1.3. A minimal architecture of joint action

Vesper et al. (2010; see also Vesper et al., 2017a) spelled out a minimal architecture that could support most types of here-and-now coordination. We describe its components and the main sources of evidence behind them, and then use the model in the next section as a reference point for thinking about interfaces between joint actions and communication.

Vesper et al.’s (2010) cognitive model consists of dedicated representations, processes and so-called coordination smoothers, that is, various heuristics and behavioral modulations that simplify coordination. In the simplest case, an agent in a joint action only represents the common goal (achievable by the agents together) and her own task. However, in many cases where the agents’ tasks and corresponding actions are interlocked, it is beneficial or even needed to represent the other’s task as well. We have already seen in the object passing example that the passer’s task is also represented by the receiver, and there is ample evidence for co-representations from other tasks as well (e.g. Atmaca, Sebanz, Prinz and Knoblich, 2008; Kourtis, Sebanz and Knoblich, 2010; Loehr and Vesper, 2016; Novembre, Ticini, Schütz-Bosbach and Keller, 2012; Sebanz, Knoblich and Prinz, 2003).

The model further assumes two processes operating on the representations: monitoring and prediction. Monitoring here refers to a process that - in the minimal case - evaluates outcomes with regards to the common goal. Evidence for monitoring joint outcomes has been provided by an EEG study on piano duets, where manipulating joint outcomes modulated an ERP component independently from changes of individual action outcomes (Loehr, Kourtis, Vesper, Sebanz and Knoblich, 2013). In many cases, agents might also monitor their own and their partner’s individual action outcomes, probably relying on the same action monitoring system for the two cases (e.g. van Schie, Mars, Coles and Bekkering, 2004; Yu and Zhou, 2006). Prediction is also theorized to operate both on individual and joint levels. There is plenty of evidence that we predict our own, but also others’ unfolding actions using internal models of the motor system (see e.g. Kilner, Friston and Frith, 2007; Wilson and Knoblich, 2005; Wolpert, Doya and Kawato, 2003). There are also several studies supporting the idea that such predictions underlie successful temporal coordination between participants (e.g., Novembre, Ticini, Schütz-Bosbach and Keller, 2013; Vesper, van der Wel, Knoblich and Sebanz, 2013).

It is not entirely clear, though, how parallel predictive simulations of self and other interact to achieve the intended outcome, especially in the case of complementary actions where parallel simulations could – in principle – result in interference (e.g. Sacheli, Tidoni, Pavone, Aglioti and Candidi, 2013). A prominent idea is that the joint goal representation acts as an integrative platform, maintaining the separation of simulations (Sebanz and Knoblich, 2009; Clarke, Francova, Szekely, Butterfill and Michael, 2018). According to a recent proposal, a joint goal representation incorporating parallel action simulations might be instantiated as a hierarchical predictive model of the interaction (Pesquita, Whitwell and Enns, 2017).

Finally, coordination smoothers include all behavioral modulations and strategies employed for boosting coordination. There are at least three categories of smoothers identified in experiments: (1) reduced variability of some task-relevant aspect of the agents' behavior, corresponding to increased predictability (Vesper, van der Wel, Knoblich and Sebanz, 2011; Vesper, Schmitz, Safra, Sebanz and Knoblich, 2016); (2) redistribution or restructuring of individual tasks by exploiting opportunities afforded by the environment (e.g. division of search area in Brennan, Chen, Dickinson, Neider and Zelinsky, 2008); and (3) coordination signals (Brennan, 2002; H. Clark, 1996), including non-conventional signaling as in action modulations conveying information to the partner, usually termed sensorimotor communication (Pezzulo, Donnarumma and Dindo, 2013; Pezzulo et al., 2018; Schmitz, Vesper, Sebanz and Knoblich, 2018; Vesper and Richardson, 2014).

## 1.2. Communication as a coordination tool

At the risk of stating the obvious: people often coordinate their actions by communicating with each other. At the same time, we have seen that many interactions can be characterized as joint actions and that, despite their variety, we can analyze them in terms of a relatively simple cognitive machinery. So how do the systems, supporting communication and joint action, interface? Vesper et al.'s (2010) model provides two points of contact. First, the model includes coordination signals explicitly as coordination smoothers. The second connection is given by what is missing from the model - the preconditions of joint actions. We consider both of these connections in the current section. For the sake of moving from the general to the specific, we start with the preconditions.

### 1.2.1. Preconditions of joint actions

**Establishing goals and sharing action plans.** Note that in the cup-passing example we invoked earlier we took it for granted that the participants agreed on a common goal (i.e., passing the cup of coffee from the barista to me) and clearly defined individual roles (moving the cup above the counter, and taking hold of it). Similarly, while the architecture described by Vesper and colleagues (2010, 2017a) emphasizes the role of the common goal representation and the individual tasks, it does not consider *how* the joint goal and personal responsibilities are established in the first place.

In many cases, opportunities for joint actions are readily recognized as such. For example, we know that infants as young as 14-month-old engage in helping and cooperative games, that is, in joint actions (Warneken and Tomasello, 2006, 2007; Warneken, Chen and Tomasello, 2006). They do so on the basis of action observation alone (e.g. experimenter reaching towards out-of-reach object), by identifying the underlying intention and treating it as a joint goal. The same happens in everyday helping situations too - someone failing to carry a heavy object is perceived as an opportunity for joint action. Besides action observation, conventions, precedence and joint affordances can also trigger joint action. For example, in our cup-passing scenario, the conventions associated with coffee shops implied the common goal with appropriate action plans. However, it is also easy to imagine scenarios where joint goals need to be negotiated in some way. Imagine preparing dinner together with someone for the first time - selecting the dish and agreeing on meshing sub-plans is a coordination problem that is unlikely to be solved by precedence or conventions alone.

Fortunately, human communication, especially language use, is remarkably good at sharing complex mental representations, including action plans (Tylen, Weed, Wallentin, Roepstorff and Frith, 2010). This point was illustrated rather elegantly by Roepstorff and Frith (2004) who compared how macaque monkeys and human participants were instructed to perform the Wisconsin Card Sorting Test while undergoing fMRI in the study by Nakahara, Hayashi, Konishi and Miyashita, 2002. While humans were simply verbally instructed and received a training of 30-60 min, the two macaques were subjected to a year-long operant training regime. Verbally aligning the task representations of the experimenter and a human participant (Pickering and Garrod, 2004) was thus several thousand times faster than learning it from the bottom-up.

Given the power of communication for sharing mental constructs, in the minimal case, people should rely on it whenever the opportunity for joint action is cognitively opaque for one of the agents, that is, the goal cannot be readily inferred from action observation or the environment. While we do not consider this proposal as controversial, we do not know of any study directly testing how opacity of joint action evokes the need for communication to establish a common goal.

**Securing commitment.** While we see the same human propensity for cooperation in society at large as in infants (e.g. Hill, 2002), there are many norms governing its everyday forms (e.g. Torrey, Fussell and Kiesler, 2008). Due to these rules, communication is often needed to elicit someone's participation in a joint action even when the opportunity is clear for the agents. In such cases the role of communication is to secure commitment, establishing who the agents of a joint action are rather than the content of it. For example, in the infant study by Warneken and Tomasello (2007) when an adult withdrew from cooperation abruptly, infants engaged in communicative acts (pointing and vocalizations) aimed at restarting the joint action. We also know from experiments on social dilemmas that communication secures commitment effectively even in one-shot games, against agents' self-interest (e.g. Balliet, 2010; Charness and Dufwenberg, 2006). Interestingly though, there are also cases where explicit commitment is not needed for the agents to feel committed to a joint action, for example, if they already have a history of coordinated actions (Michael, Sebanz and Knoblich, 2016).

**Navigating the hierarchy of joint actions.** Let us consider the initial coffee shop interaction again. In H. Clark's (1996, 2005) terms, the interaction with the clerk as a whole constitutes a so-called joint activity ("ordering and serving coffee") and is composed of nested joint projects and sub-projects (often episodes we understand as joint actions, e.g. "handing over the cup", "passing the change", etc.). The hierarchical structure of projects is analogous to an action hierarchy in the individual case (e.g. Haruno, Wolpert and Kawato, 2003). As an interaction progresses, the agents need to propagate through stages in the hierarchy, requiring coordination not only in terms of the content of each basic project, but also with regards to their sequence and temporal boundaries.

Bangerter and H. Clark (2003; Bangerter, H. Clark and Katz, 2004) investigated this issue in verbal interactions, trying to describe how participants navigate across the different phases. They focused on corpora from complex tasks with well-defined hierarchical structures (e.g. LEGO building, H. Clark and Krych, 2004) and looked for project markers, i.e., words marking either continuation of projects or initiation of new ones. They found that certain markers indicated the same types of transitions both across different pairs and data sets. Words like *"uh-huh"* and *"yeah"* were used to ground contributions and continue the current project (horizontal transitions), while words like "*all right*" and "*okay*" marked either the end or the start of a phase (vertical transitions). These studies provide evidence that simple linguistic markers can be used for progressing complex interactions by chunking them into smaller segments in which more basic processes could be responsible for coordination.

### 1.2.2. Communication as coordination smoother

In many joint actions, monitoring task progress and predicting upcoming actions require information transfer between the agents (but see Vesper et al., 2013 for an interesting exception). In the default case, information transfer consists of perceptual access to the other’s action or action effects. For example, in the cup-passing scenario I could observe the barista’s movement from the start and could rely on that to fine-tune predictions about his trajectory, updating my own movement accordingly.

It is often the case, though, that perceptual access is either impeded or insufficient for precise prediction. One common scenario is when task-relevant information is distributed asymmetrically and the agent without access to the privileged information cannot accurately represent the other’s task. People in such situations have to rely on alternative strategies: the aforementioned coordination smoothers (Vesper et al., 2010). Consider a piano duet for example. If one pianist has knowledge of upcoming tempo changes but the other not, their coordination will be temporarily disrupted at each change unless they can share information. While they might solve the problem by restructuring the task (e.g. sharing scores), they can also rely on coordination signals, that is, communication, to prepare for tempo changes (Bishop and Goebl, 2018). Communication as a coordination smoother could be either non-conventional, embedded in the action itself, or conventional, relying at least partially on learned codes (see more about the distinction in Pezzulo et al., 2018). We elaborate on both cases below.

**Sensorimotor communication (SMC).** Recently, there has been a surge of interest in non-symbolic, non-conventional coordination signals called SMC that most often emerge in asymmetric collaborations (Pezzulo et al., 2013, 2018). As an example, consider the kinematics study by Sacheli et al. (2013). They asked pairs of participants to synchronously grasp a bottle placed between them. The bottle could be grasped in two locations, upper and bottom half. Only one participant (leader) was instructed with a target location, the other (follower) had to infer the complementary grasping location from the leader’s action. They found that leaders modulated their kinematics in a way that helped followers to predict the grasping location, helping synchronized arrival at the bottle. For example, leaders moved their arms higher and with smaller grip aperture when grasping the upper, thinner part of the bottle than in an individual control condition. Similar results have been obtained by Vesper and Richardson (2014) who asked pairs to synchronize their tapping on a range of locations, with only the leader aware of the targets. They also found that leaders modulated their movement kinematics (primarily the movement amplitude) to help followers to predict the targets and to synchronize their tapping. Several more studies have reported action modulations that boosted coordination by encoding information for the interaction partner since (Candidi, Curioni, Donnarumma, Sacheli and Pezzulo, 2015; McEllin, Knoblich and Sebanz, 2018; Schmitz, Vesper, Sebanz and Knoblich, 2018; Vesper, Schmitz and Knoblich, 2017b).

SMC is interesting for at least three reasons. First, a defining feature of SMC is that the coordination signals are embedded in instrumental actions, not following any conventionalized code. Computational analysis of SMC results (Pezzulo et al., 2013) suggests that signalhood is derived from the apparent inefficiency of the movements: the deviation from an individually efficient trajectory disambiguates the targets early, trading off personal effort for communication success and, in turn, for joint efficiency. Second, SMC is intentional, as evidenced by its flexible use. For example, in a follow-up to Sacheli et al. (2013), Candidi et al. (2015) demonstrated that leaders were sensitive to followers' uncertainty: they signaled target location only until followers learned an underlying rule dictating the locations (see also Pezzulo and Dindo, 2011; Vesper et al., 2017b). Also, participants rely on SMC only when their partners can readily observe the signals but not when their movements are occluded (Vesper et al., 2014; Vesper et al., 2016). Third, there is some evidence for SMC from conditions with symmetric information (Schmitz et al., 2018; Vesper et al., 2016), suggesting that SMC is not only elicited by pressures for information transfer but acts as a more general coordination process.

**Towards conventional communication.** SMC exists in an interesting middle ground between action observation and conventional, symbolic communication. On one hand, SMC is a simple mechanism manipulating predictability. When coordination requires increased spatial and/or temporal predictability of an agent's movements, kinematics is altered so that an observer can predict its trajectory more easily. Notably, SMC does not have to be recognized as communicative in order to be effective, avoiding the need for the complex machinery of ostensive-inferential communication in the observer (Scott-Phillips, 2014; Sperber and Wilson, 1995). Instead it can boost coordination by exploiting the observer's capacity for action understanding in terms of the immediate goal (Kilner, 2011), reducing her uncertainty about the communicator's action (i.e., providing information). While people can differentiate between instrumental and informative intentions from kinematics alone (McEllin, Sebanz and Knoblich, 2018) and could, in principle, recognize SMC as communicative, as of now, we do not know if they do so routinely.

On the other hand, SMC might only be a step away from code-like communication. If a kinematic modulation can be detached from the instrumental action, it could be used as a communicative signal on its own, bridging a main difference between SMC and conventional signaling (Pezzulo et al., 2018). Importantly, by separating the signal from the action, communication would not exploit action understanding mechanisms in the observer anymore but – assuming the source is manifest - would become a proper communicative signal. For example, in the study by Sacheli et al. (2013), had the leaders been allowed to make a quick up or down movement indicating target location before the grasping action (detachment), it would have quickly become a conventional coordination signal. Once detached, a signal might also be discretized, enabling the use of combinations (Verhoef, Kirby and de Boer, 2014). Moreover, we know from experimental semiotics (Galantucci and Garrod, 2011), that repeated use of signals on a separate communication channel leads to quick conventionalization and to the loss of iconic and indexical nature, yielding symbolic communication (e.g. Galantucci, 2005).

Would people prefer separate communicative signals over SMC? Vesper et al. (2017b) asked this question in a sequential target matching task where pairs of participants had to move to the same target location in each trial. Leaders had to move first and could indicate the correct target for the followers. In one experiment, leaders could communicate either through action modulation (SMC) or action effects (signals separated from the movement). They found that given the chance, leaders preferred to send separate signals. In a conceptually similar task, Schmitz et al. (2018) found that leaders only preferred separate, conventional communication over SMC when the signals had the same ordinal structure as the information they needed to transmit.

**Conventional communication.** Conventionalized, code-like communication (pointing, gestures, language use, etc) is flexible enough to transmit any type of information necessary for coordination. The most powerful system in that regard is language. For instance, it enables the transmission of metarepresentational content. Participants who could freely talk with each other during a joint perceptual decision task were found to share their confidence in their individual judgments, increasing their pair-level performance beyond their individual capabilities (Bahrami et al., 2010).

However, language use is also slow and resource-hungry (Roelofs and Piai, 2011). If coordination does not require sharing abstract knowledge, but mostly spatial referencing, assurances or indexing of objects, people prefer to mix linguistic signals with pointing and other gestures, and only switch to purely linguistic means when a shared visual space is not available (H. Clark and Krych, 2004). Language is also at a disadvantage under time pressure. In a visual search task, pairs of participants were less efficient when they could coordinate verbally, than in a condition where they shared gaze information only. Shared gaze even had an advantage over a condition with both shared gaze and verbal communication, suggesting that language might even be detrimental under time pressure (Brennan et al., 2008; Neider, Chen, Dickinson, Brennan and Zelinsky, 2010).

## 1.3. Communication as joint action

So far we have considered communication as a coordination device in the service of joint action. There is another side of the coin though, namely that we can construe communication as a type of joint action in and of itself. In the current section we consider two broad accounts that depict communication as a collaborative activity. Then we give a brief overview of neuroscientific research on communication that was inspired by the same premise: communication should be understood as a process over multiple brains. Unlike in previous sections, the work discussed here focuses almost exclusively on verbal communication.

### 1.3.1. Communication as joint action – a philosophical account

Verbal communication is coordinated. Each contribution constrains the next one: a greeting is acknowledged and reciprocated, a question is followed by an answer, a request is by granting or denying it, and so on (H. Clark and Schaefer, 1989). We also coordinate in turn taking, by limiting our contributions and allowing only minimal gaps and overlaps (Levinson, 2016). We tailor our words to our partners (Brennan and Hanna, 2009), and also develop fleeting conventions with them for references (Brennan and H. Clark, 1996) and procedures (Mills, 2014). Furthermore, we see convergence across agents on almost all levels of verbal and nonverbal behavior, in syntax, speech rate, dialect, posture, facial expressions, etc. (Chartrand and Bargh, 1999; Giles, Coupland and Coupland, 1992; Pardo, 2006; Pickering and Branigan, 1999; Shockley, Santana and Fowler, 2003).

What are the mechanisms that produce coordination in communication? Is there a common set of mechanisms at play here? A rather philosophical answer was provided by H. Clark (1996, 2005; H. Clark and Blye, 1995), who argued that communication at its core is a nested hierarchy of joint actions, producing coordination at any level of analysis. The basic idea, rather colloquially put, is that mutual understanding depends on both agents, as none of them by themselves can guarantee it individually. In H. Clark's view, there is a set of interdependent actions taken by the two agents behind any communication (the so-called action ladder, H. Clark, 1996). As a demonstration, consider a single communicative act of pointing. To produce the pointing behavior, agent A (Alice) has to assume that agent B (Bob) will attend to it; for Alice to use pointing as a communicative signal, she needs Bob to identify it as a signal; to mean something by pointing, Alice has to trust Bob to recognize the meaning; and so on. The argument might be extended to larger structures of dialogues as well. Take referring as an example: selecting the right word is a joint action where the agents coordinate on the expression to use, and such coordination necessarily results in convergence in terms of lexical units (Brennan and H. Clark, 1996).

### 1.3.2. Communication as joint action – a psychological account

While H. Clark's ideas have been very influential, their foundations lay in discourse analysis and pragmatics, offering little in terms of cognitive or neural mechanisms. A more specific theory defining mechanisms for coordination in dialogue was provided by Pickering and Garrod (2004, 2013; Garrod and Pickering, 2009). Their interactive alignment model describes communication as a joint action, where the common goal is the alignment of situational models across the agents. A situational model (Zwaan and Radvansky, 1998) constitutes the overall representation of the current context, including the agents, the location, major events, etc. The idea here is that coordination problems requiring communication usually occur when people have different understandings of the same situation. Communication is then used to align the situational models and, hence, solve the task. For example, in a maze game (Garrod and Anderson, 1987) where pairs of participants had to navigate using language while they both had incomplete information, pair success depended on whether participants could converge on a shared, aligned description of the maze or not. Importantly, Pickering and Garrod (2004, 2013) argue that alignment of the situational models across speakers is achieved to a great extent by general mechanisms supporting joint action, namely, perception-action matching and action simulation.

**Covert imitation.** Perception-action matching is the process of mapping an observed action to one's own motor repertoire. Due to associations between corresponding perceptual and motor representations (e.g. Hommel, Muesseler, Aschersleben and Prinz, 2001; Brass and Heyes, 2005), observing an action can induce covert imitation of the observed action in the observer. Perception-action matching is a standard explanation behind our tendency to unconsciously mimic others' behavior, like facial expressions or posture (Wilson and Knoblich, 2005). The interactive alignment model recognizes language production and comprehension as special cases of action and action perception. It points out that perception-action matching entails covert imitation in verbal communication as well. In other words, listening to others' speech activates the corresponding representations used in production, priming the subsequent repetition of lexical units, syntactic rules and overall semantics used by the speaker (see Gambi and Pickering, 2017 for a discussion on representational parity in language).

By itself, covert imitation can already account for a range of coordination phenomena in verbal interactions, explaining convergence both in terms of motion (e.g. gesture, body sway) and linguistic representations. Based on evidence for priming across linguistic levels (Branigan, Pickering and Cleland 2000; Cleland and Pickering, 2003), Pickering and Garrod (2004, 2013) further claim that alignment on a given behavioral or linguistic level propagates to other levels as well, ultimately contributing to the alignment of situational models, that is, to successful communication. Note that the argument applies to the general case of language comprehension (one-way communication), not only to interactions. However, the back-and-forth, interactive nature of dialogues should result in especially fast convergence due to both agents covertly imitating each other by taking turns.

**Action simulation.** Linguistic representations activated via covert imitation also provide the local linguistic context, which, supplemented by the non-linguistic context of the interaction (common ground, H. Clark, 1996), can be used to derive speaker's intention behind the utterance (Pickering and Garrod, 2013). This derived intention serves as the basis for action simulation, that is, a predictive simulation of the current speaker's upcoming utterance, using the listener's own production system (see Pickering and Gambi, 2018 for details and evidence). Such prediction-by-simulation is important for language comprehension in general (Pickering and Garrod, 2013; Pickering and Gambi, 2018), but it also has specific roles in verbal coordination. First, prediction-by-simulation projects the future time course of the ongoing utterance, explaining the precise temporal coordination of individual actions observed in turn-taking (Hadley, November, Keller and Pickering, 2015; Magyari and de Ruiter, 2012). Second, predicting the content and form of the speaker's utterance can help planning overt imitative (e.g. finishing the other's sentence) and complementary answers in advance (e.g. Bögels, Magyari and Levinson, 2015). Finally, in verbal communication both participants engage in predicting each other. Mutual predictions have been shown to support increased coordination in behavioral tasks (e.g. Konvalinka, Vuust, Roepstorff and Frith, 2010; Noy, Dekel and Alon, 2011), suggesting that they also support the rapid development of mutual understanding in verbal interactions.

Overall, the interactive alignment model depicts verbal interaction as a coordination problem where the joint goal is expressed as alignment of representations. Importantly, it accounts for alignment via basic processes supporting joint actions in general. While this picture is certainly not complete without adding further mechanisms, enabling theory-of-mind and grounding, the interactive alignment model argues for the relative importance of perceptual and motor processes (similar to e.g. Glenberg and Gallese, 2011; Pulvermüller, Moseley, Egorova, Shebani and Boulenger, 2014), expanding the scope of the joint action approach. While there are important criticisms of the account, both regarding its premises and the details of its processes (e.g. Fusaroli, Raczaszek-Leonardi and Tylén, 2013; Brennan et al., 2010), it nevertheless presents an intriguing case of bootstrapping. In section 1.2, communication was invoked to account for limitations of basic coordination processes, but the view presented above depicts communication as built on the same processes itself.

### 1.3.3. Neural alignment, coupling and hyperscanning

Notions of alignment and coupling have been cardinal in the neuroscientific study of verbal interactions as well. A prominent neural model of communicating agents is that of brain-to-brain coupling (Hasson, Ghazanfar, Galantucci, Garrod and Keysers, 2012; Hasson and Frith, 2016; Nummenmaa, Lahnakoski and Glerean, 2018). Coupling here refers to neural processes that exhibit some lawful relationship (e.g. temporal or pattern similarity, conditional relations, etc.) across brains, and alignment is the special case when neural activity is mirrored across agents. The model aims to account for the various ways two neural systems exhibit coupling during interactions. The premise of neural coupling is that perceptual and motor systems of individual brains couple through the signals generated in interactions. As an example, think of a simple pointing gesture: the motor system of the pointing agent generates a change in the shared environment, engaging the other agent's perceptual system. While there is neural alignment across agents in early sensory areas due to general brain-to-stimulus coupling (reflecting a shared sensory environment, e.g. Hasson, Nir, Levy, Fuhrmann and Malach, 2004), it is the perceptuo-motor link that induces alignment in socially relevant systems. First, action-based alignment across agents emerges in mid-level areas (Giese and Rizzolatt, 2015; Keysers and Gazzola, 2014), supporting matching action representations. Second, following integration of the action-based level with contextual information (e.g. Brass, Schmitt, Spengler and Gergely, 2007), intention-based alignment emerges in higher areas, reflecting shared understanding and intentions across agents (e.g. Yeshurun et al., 2017).

The brain-to-brain coupling model is supported by experiments on verbal communication. A series of fMRI studies investigated speakers' and listeners' brain activity in a narrative storytelling scenario (Silbert, Honey, Simony, Poeppel and Hasson, 2014; Stephens, Silbert and Hasson, 2010). They found widespread alignment (temporally coupled, locally specific responses) across speakers and listeners. Importantly, as follow-up studies showed (Honey, Thompson, Lerner and Hasson, 2012; Lerner, Honey, Silbert and Hasson, 2011), only in early sensory areas was alignment driven by the auditory properties of speech, in higher-level areas it reflected a shared understanding of the narrative, independent of even the particular language. Interestingly, while listeners' brain activity mainly lagged behind that of speakers', in medial prefrontal areas there was evidence for prediction, and its strength correlated with better narrative comprehension (Stephens et al., 2010). Conceptually similar results were also acquired with non-verbal communication (gestural communication: Schippers, Roebroeck, Renken, Nanetti and Keysers, 2010; facial communication of affect: Anders, Heinzle, Weiskopf, Ethofer and Haynes, 2011).

While most neuroscientific work on communicative interactions has focused on simple forms of coupling (alignment and synchrony) so far, behavioral studies have started to investigate questions of mutual adaptation, complementarity and synergy (e.g. Fusaroli and Tylén, 2016; Fusaroli et al., 2017). However, moving away from similarity-based neural measures is a challenge for at least two reasons. The first is a lack of appropriate models. Alignment is a crucial tool for tracking shared representations, but it is unclear what type of neural dynamics would truly capture realistic, back-and-forth communication. There are a few candidates though, mainly measures built around conditionally occurring neural responses (Hasson and Frith, 2016). The second is methodological, as advancing our knowledge on brain-to-brain coupling requires hyperscanning (Babiloni and Astolfi, 2014), that is, simultaneous and synchronized recordings from multiple participants. While there are very promising results from studies already applying hyperscanning techniques (e.g. Dikker et al., 2017; Piazza, Hasenfratz, Hasson and Lew-Williams, 2018; Spiegelhalder et al., 2014; Stolk et al., 2014), it is still a relatively rare and challenging approach.

## 1.4. Research aims

In the next three empirical chapters we focus on three separate questions regarding human communication. In Chapter 2 we investigate how efficiently people transmit information using simple forms of communication. It is generally assumed that signals are selected efficiently in terms of effort needed for comprehension (Sperber and Wilson, 1995) but it is an open question how signals relate to task performance. In Chapter 3 we turn to verbal interactions and study the role of interactivity and goal overlap in alignment. Realistic, back-and-forth interactions require mutual adaptations from participants, arguably leading to different patterns of synchronized behavior than non-interactive, observational scenarios (e.g. Gallotti, Fairhurst and Frith, 2017; Pickering and Garrod, 2004). While there is evidence for such differences from joint actions (e.g. Konvalinka, Vuust, Roepstorff and Frith, 2010), the nature of the difference is less clear in verbal communication. Finally, in Chapter 4 we extend the investigations in Chapter 3 to the neural domain and ask if the difference between interactive and non-interactive situations can be described in terms of brain-to-brain coupling.

### 1.4.1. Chapter 2

It is a central feature of many jointly performed tasks that effort is shared across participants and efficiency is understood on the pair or group-level. This is true for signaling as well. We described earlier that in SMC individual efficiency is sacrificed for helpful signaling and, ultimately, better joint performance (section 1.2.2, Pezzulo et al., 2013). Accordingly, in sequential versions of SMC (Vesper et al., 2017b; Schmitz et al., 2018), leaders' actions or action effects might be understood as cues for followers that leaders selected them by trading off their own efficiency for pair-level success. Our aim in Chapter 2 is to extend this framework to a common type of joint action, helping. Helping is an interesting scenario as it entails the prioritization of the helpee's goal over the helper's individual goals. In fact, the helpee's goal becomes the joint goal. If this analysis is correct, helpers, in general, should prioritize helpees' efficiency over her own. We tested this hypothesis in a novel pre-cueing task where a helper's action provided a cue for the helpee.

### 1.4.2. Chapter 3

Verbal communication is a highly coordinated activity (section 1.3.1), with findings of behavioral mimicry (alignment) on multiple levels. The interactive alignment model (section 1.3.2, Pickering and Garrod, 2004, 2013) treats alignment (linguistic alignment especially) as important for successful communication. In Chapter 3, we investigate the role of two important factors of everyday interactions in modulating alignment, namely, interactivity and individual goals. Interactivity is a defining feature of dialogues and the initial motivation behind the interactive alignment model is to explain language use in dialogues. However, the putative mechanisms behind alignment, perception-action matching and action simulation are not specific to interactions. The model does not fully specify whether and how these mechanisms are engaged differently in interactive and non-interactive scenarios. Moreover, the available evidence addressing this issue is contradictory. Similarly, it is poorly understood if the convergence or divergence of individual goals should affect alignment at all. In Chapter 3, we manipulate interactivity and whether participants have the same goals. We test the effects of these manipulation on multiple levels of alignment in Chapter 3, using a naturalistic storytelling scenario.

### 1.4.3. Chapter 4

Alignment has become a prominent description of the effects of communication on the neural level as well (section 1.3.3, cf. section 1.3.2). Alignment here refers to mirrored neural activity across interacting agents. Operationally, alignment is established by treating one brain as the (relatively simple, linear) model of the other brain. A number of studies have reported neural alignment across brains (e.g. Anders et al., 2011; Stephens et al., 2010; Silbert et al., 2014; etc.). These results have been integrated into the brain-to-brain coupling model (Hasson et al., 2012; Hasson and Frith, 2016) which argues that alignment of different systems correspond to different types of shared representations, going from alignment driven by stimulus features to alignment as an effect of mutual understanding.

Brain-to-brain coupling has generally been studied as one-way communication. In Chapter 4, our aim is to extend this coupling model to genuine interactions (two-way communication) using fMRI. We implemented a hyperscanning study using the same story-telling task as our study in Chapter 3 and investigated the effects of interactivity on brain-to-brain coupling.

# Chapter Two. Overachiever helpers and underachiever helpees: Helping as a case of joint action

## 2.1. Introduction

It is hard to overstate the importance of helping in human societies. Humans are thought to be unique in their propensity to help each other (Fehr & Fischbacher, 2003; Stevens & Hauser, 2004) and pro-sociality develops at a very early age (Warneken & Tomasello, 2006). Previous research on helping has addressed how our motivation to help plays out in everyday situations, prompting a great deal of research on the situational and personality factors modulating when humans engage in helping (Eisenberg, 2000; Penner, Dovidio, Piliavin & Schroeder, 2005). However, less is known about the specific processes that enable helpers to select the actions that effectively help others and the processes that enable helpees to effectively utilize helpful actions by a helper. The aim of this study was to ask how efficient people are in providing and accepting help.

Helping can be defined as a form of prosocial behavior, that is, any behavior that has the purpose of benefiting someone else, rather than the actor herself (e.g. Twenge, Baumeister, DeWall, Ciarocco & Bartels, 2007). This and other common definitions of helping focus on the motivation behind actions and thus encompass a wide range of different behaviors. In order to address the efficiency of providing and accepting help, a slightly different focus is needed. Hence, we define helping here as a form of social interaction where one individual treats the known or inferred goal of another individual as her own goal. Thus, we take the helper’s commitment for granted (Gräfenhain, Carpenter & Tomasello, 2013; Michael, Sebanz, & Knoblich, 2015) and focus instead on the question of how helper and helpee coordinate their actions so that the helper’s actions contribute efficiently to the realization of the helpee’s goal.

A good starting point for studying helping from the point of view of interpersonal coordination are cases where helping consists in providing useful information that make another’s task easier and thus requires helper and helpee to come up with a way of communicating that makes the helpee’s task easier. Accordingly, the present study conceives of helping as a form of joint action (Knoblich et al., 2011; Sebanz et al., 2006) where the helper communicates information that can improve the performance of a helpee.

We believe that the above characterization of helping captures common features of many everyday helping situations. For example, imagine a friend showing you how much space is left while parking a car into a tight spot. The friend’s communication serves to provide information that supposedly helps you to achieve your goal more efficiently than on your own. We think of these and similar scenarios as helping situations fitting our focus here: The goal of the helpee is also the goal of the helper and helping can be achieved by communication. However, the motivation to help and the ability to communicate do not guarantee that the resulting joint action will be efficient. For the helper, the challenge is to provide maximally beneficial signals to the helpee. For the helpee, the challenge is to make the best of these signals. In the following sections we discuss these challenges for helper and helpee in more detail.

### 2.1.1. Helper’s efficiency

In order to help efficiently helpers often need to choose which information to communicate to helpees. Staying with our car parking example, communicating the distance to the edge of the sidewalk or to the other car behind might both be helpful, but which information is most useful also depends on situational factors and the driver’s expectations. Of course, the type of information to be provided can be explicitly negotiated in advance (e.g. the driver in our example might ask about the distance to the car behind), but frequently helpers need to select themselves what to communicate to the helpee.

So far, research on communication in cooperative situations has almost exclusively addressed helping with experiments where either the context or the instructions defined which information was needed to be communicated. Studies on language use have often employed cooperation tasks where crucial information was only available for one of the participants in a pair (e.g. Brennan & Clark, 1996; Clark & Krych, 2004; Gergle, Kraut & Fussell, 2004; Wilkes-Gibbs & Clark, 1992). In these studies, typically, a communicator is provided with some information that is necessary for a performer to achieve her goal (e.g. the target configuration of a set of pictures is known only by the communicator, as in Brennan & Clark, 1996) and what is helpful is dictated by which information the performer is lacking.

In the present study, we asked how efficient helpers are in selecting the right information to communicate when it is not predefined or entirely transparent which information the helpee could most benefit from. To this end, we considered a situation where helpers received task-relevant information before it became available for helpees who were instructed to perform a speeded reaction time task. Helpers needed to decide which aspect of the task-relevant information to communicate to helpees in order to help them speed up their performance. What would constitute the most beneficial communication in this situation? In order to be efficient, a helper would need to communicate information that improves a helpee’s performance most even if this implies added costs for themselves. Because efficient helping is a function of jointly achieving the helpee’s goal, helpers should be insensitive to their own effort and their behavior should be solely guided by the attempt to improve the helpee’s performance.

There is some evidence for such behavior in joint actions even in the absence of explicit instructions to benefit the other participant. In a study by Ray & Welsh (2011) participants performed the first part of a sequential joint action (passing a jug). They could choose freely between movements requiring more or less effort, leading to a slightly easier or harder task for the second actor, respectively. Importantly, the performance benefit for the second actor in the easier task relative to the harder one was roughly equal to the extra effort required from the first actor. Overall, the first actors preferred the action requiring more effort on their part, making the second actors' task as simple as possible, in accordance with our definition of helping. Similar results have been obtained in further studies on object passing by Gonzalez, Studenka, Glazebrook & Lyons (2011) and Meyer, van der Wel and Hunnius (2013). Later studies have further verified the prevalence of this tendency to accommodate to the partner's efficiency and have also shown that it is modulated by the social context (Dötsch and Schubö, 2015; Scharoun, Scanlan and Bryden, 2016; Constable et al., 2016).

Alternatively, helpers may use a split-effort model to distribute effort between helper and helpee. Help can become "too much" when it minimizes the effort required from the helpee, representing a threat to the helpee's self esteem (Nadler and Fisher, 1986; Nadler, 2015) and inducing aversive feelings of indebtedness (Greenberg and Shapiro, 1971; Greenberg and Westcott, 1983). For example, in the car parking scenario the helper is not expected to offer to valet and park the car herself as that might imply that the helpee lacks competence. Instead, the helper should only provide the necessary information (cf. DePaulo, Brittingham and Kaiser, 1983), ultimately splitting the effort required to achieve the goal with the helpee. This argument is further supported by by fairness considerations (e.g. Bolton and Ockenfels, 2000). It has been found in behavioral economics that "too good" offers in an ultimatum game lead to higher rates of rejection from recipients than 50/50 splits (Bahry & Wilson, 2006; Bellemare, Kroger and van Soest, 2008; Hennig-Schmidt, Li and Yang, 2008). This behavior is a peculiar form of inequality aversion (Fehr and Schmidt, 1999; Tricomi, Rangel, Camerer & O'Doherty, 2010), inasmuch as people would rather not receive any payment than to receive more than their partners. Unequal offers benefiting the receiver might feel immoral as they violate our sense of fairness (Baumard, André and Sperber, 2013). We assume that this reasoning is applicable to helping in general, not only to monetary rewards. Hence, the split-effort model predicts that the helper will choose to communicate in a way that avoids unequal distribution of effort between helper and helpee. This account predicts a lower level of help than the assumption that the helper treats the known or inferred goal of another individual as her own goal.

### 2.1.2. Helpee's efficiency

As outlined above, our aim is to address helping situations where helpers may contribute to a helpees' performance by communicating information but where the helpee could also perform the whole task alone. Think of the car parking example again. Drivers are capable of parking their cars perfectly well without help. Sure, it might be less comfortable - they might have to step out for a moment if space is really tight – but having a friend around is not essential. In such situations helpees do not need to rely on the information communicated by the helper but have the alternative option of ignoring helpers and relying on their individual performance. Helpees may also choose to only partly rely on the helper.

In such situations, helpees' efficiency entails optimal use of the information communicated by the helper, leading to the best possible performance. That is, information from the helper should be treated as any other information that is relevant for task performance. In order to judge how useful information provided by helpers is, helpees need to be able to assess the reliability and usefulness of helpers' communication. There is reason to believe that helpees should be capable of doing so that comes from group-level signal detection experiments (Bahrami et al., 2010, 2012a, 2012b; Sorkin, Hays & West, 2001). In these experiments pairs of participants provide joint perceptual decisions by combining their independent individual decisions. The main finding from the Bahrami et al.

studies is that if pairs are allowed to communicate, their collective performance might be approximated by a confidence-sharing model. Thus, communication serves as a way of signaling confidence in individual judgments, leading to a performance improvement relative to the performance of the better individual. Importantly, when participants could only share their individual choices without further discussion, pairs performed on the same level as the better member of the dyad. This latter result suggests that participants in a pair learned about each other's performance levels and acted according to the relative usefulness of their individual guesses.

So how could helpees maximize the efficiency of using information provided by helpers? They would need to monitor the usefulness of the helper's communication and to rely on this information to the extent that it benefits their own performance. This is also in line with the idea that helper's communication can be treated as any other cue so that helpees' behavior is no different than that of a participant in any cueing task where participants are provided with useful advance information before performing a task (precuing, Rosenbaum, 1980, 1983; see also Zaki, 2013). Helpees may further improve overall efficiency by making it easier for helpers to provide the best possible information to support helpees' performance.

However, there are at least two alternative accounts that would both predict inefficient utilization of the helper's communication. First, helpees might follow the split-effort model we introduced earlier. The split-effort model predicts that helpees – as helpers – prefer to share the required effort equally between the helper and the helpee. Thus, even if the helper communicated maximally helpfully, helpees would not fully utilize her communication, but would partially perform the task themselves so that they can claim they contributed to the performance.

Second, the helpee might fail to rely on the helper's communication due to a lack of trust, as trust "consists of increasing one's vulnerability to the other" (Pearce, 1973, p. 3).

While there is evidence that people tend to trust others more than purely rational actors would (e.g. Evans & Krueger, 2009), the degree of such initial trust might still be less than what would be required for efficient utilization of the helpers' contributions. For example, if helpers are maximally efficient, helpees – in order to be efficient as well – would need to trust helpers completely, without a hint of doubt. Furthermore, even if helpers' willingness to help is explicitly established, helpees might still stay epistemically vigilant (Sperber et al., 2010) and compare (1) helper's cost-benefit ratio to the apparent strength of commitment or (2) the requirements of the situation to helpers' capabilities. Either of these comparisons could warrant cautiousness and result in a lack of trust in helpers' communication.

### 2.1.3. Summary of General Hypotheses

To summarize, if people help efficiently, helpers are expected to communicate in a way that enhances the helpee's performance even if this entails more effort. Helpees are expected to show better performance as a result of utilizing information communicated by helpers. Alternatively, the split-effort model predicts that both helpers and helpees avoid unequal contributions and instead prefer that the helper and the helpee spend equal amounts of effort. From an efficiency point of view, such behavior would also correspond to inefficient helping from the helper and suboptimal utilization of the helper's communication by the helpee. However, regarding the helpee, suboptimal performance might also be accounted for by assuming a lack of sufficient trust. We contrasted these hypotheses in four experiments.

## 2.2. Experiment 1/A

To probe helpers' and helpees' efficiency, we developed a two-person, cueing version of the visual Simon task (see Figure 2.1 for a layout of the task; see Hommel, 2011 on the Simon effect). The basic Simon task is a reaction time task where participants respond to a stimulus feature (e.g. color) and there is interference - called the spatial compatibility effect - between the stimulus location (e.g. left or right) and the spatial arrangement of responses (e.g. left or right key presses). In our cueing version we had two participants perform the task, a helper and a helpee. The helpee was instructed to perform the Simon task, with the difference that the stimulus was preceded by a cue selected by the helper. The cue established a controlled form of one-way communication from the helper to the helpee. In each trial, the helper was aware of the upcoming stimulus the helpee would receive and was instructed to support the helpee by selecting a cue. Cues could be chosen from a set of symbols identical to the helpee's response set. The spatial layout of responses used by the helper for cue selection showed the same overlap with stimulus locations as helpees' responses, thus, the helper was also susceptible to the Simon effect. Helpees were aware of helpers' instructions but could perform the task without utilizing helpers' communication. As baseline, all participants performed the same task without cues individually.

Regarding helpers, we were primarily interested in which cues they selected to improve helpees' performance. The set of cues helpers had at their disposal could be used to implement three kinds of cueing strategies: (1) Response-cueing, i.e., cueing the response required by the helpee (response-cueing also implied cueing the response relevant stimulus feature, that is, color); (2) Location-cueing, i.e., cueing the response-irrelevant stimulus feature, location ; and a broad (3) other category. If helping is driven by efficiency, helpers are expected to select response-cueing, practically performing helpees' task for them. If the split-effort account is correct, we expect helpers to engage in location-cueing as determining requires considerably less effort for the helper but still supports helpees by enabling them to fully focus on the response relevant stimulus feature.

We also measured helpers’ response times (RT) and error rates (ER). The degree of effort invested by helpers can be estimated by comparing their effort in selecting cues to their individual performance of the actual task in the individual baseline condition. Since helpers were not explicitly instructed to follow a particular cueing strategy, their RT and ER are only interpretable after determining which cueing strategy was used. The efficiency account predicts that helpers prioritize helpees’ performance over their own and, thus, trade-off speed for accuracy resulting in larger RT and lower ER in the cueing phase relative to the individual baseline. Because the efficiency account predicts response-cueing, it also predicts that helpers display a spatial compatibility effect during cue selection, similar to the baseline condition. The split-effort model, on the contrary, predicts no speed-accuracy trade-off. As the split-effort model predicts location-cueing and selecting responses congruent with stimulus location requires less effort than the Simon task, the split-effort account predicts lower RT and ER in the joint phase relative to the individual baseline and no spatial compatibility for helpers during cue selection.

Regarding helpees, we measured RT and ER to determine the effect of helping on their performance. Irrespective of helpers’ cueing strategies, but supposing mostly valid cues, the efficiency account predicts that helpees would utilize the cues to the fullest extent, relying on them whenever possible. Consequentially, they (1) would show lower RT and ER relative to baseline and (2) should be sensitive to helpers’ errors, displaying both post-error slowing after helpers’ errors and a dependence of their own errors on helpers’ errors. Furthermore, in the case of response-cueing helpers, the efficiency account also predicts that helpees’ spatial compatibility effect diminishes in the joint phase relative to baseline, as they would select responses based on cues, not based on the stimuli they perceive. For the same reason we would also expect that the RT benefit of receiving help (the RT difference between the cueing and baseline phases) is roughly equal to the cue-stimulus interval. If there is variability in helpers' strategies, we should see better performance in helpees receiving response cues than in helpees receiving location cues.

By contrast, both the split-effort assumption and the lack-of-trust assumption predict suboptimal cue-utilization by helpees in the case of response-cueing helpers. The split-effort model predicts that helpees would prefer to perform at least a part of the task themselves, letting cues bias but not fully determine their response. This predicts RT and ER benefits relative to baseline and a spatial compatibility effect. The lack-of-trust account predicts the same data pattern.

We also performed analyses to investigate the role of performance monitoring. Monitoring here refers to the idea that the helpers and helpees in a pair are sensitive to each other's performance, even when a partner's performance is irrelevant for their own part in the task. For example, if helpees actively try to estimate their partners' reliability and general capability – a kind of action-specific vigilance – they might take helpers' speed into account to come up with such an estimate. We see such examples in other domains. For instance, it has been shown that the speed with which someone makes a moral decision has an effect on how others judge that person (Critcher, Inbar & Pizarro, 2013; see also Robinson, Page-Gould & Plaks, 2017). In a similar vein, decision speed also effects partner choice for collaboration (Van de Calseyde, Keren & Zeelenberg, 2010). In our task, changes in the speed of helpers' decisions might serve as cues of response selection conflicts or loss of attention (in the case of slower reactions) but could also signal growing confidence (in the case of speeding up). Similarly, helpers might use helpees' RT and ER as signs of helpees' overall performance, even though such information is not necessary for their part of the task. Here we tested for the presence of two specific types of monitoring: (1) attention to the partner's RT, as manifested in helper-helpee RT cross-correlations; and (2) helpers' attention towards helpees' errors, manifested as post-error slowing in helpers RT data following helpees' errors.

Finally, we also took measurements of helpers' and helpees' general theory-of-mind (ToM) capabilities. On the one hand, it has often been postulated that ToM (or mentalizing) capabilities are important for successful communication and are related to individual differences in pragmatic tasks (e.g. Shany-Ur et al., 2012; Winner, Brownell, Happé, Blum & Pincus, 1998). On the other hand, our task employs a cueing paradigm which – in principle - can be solved without postulating beliefs about another's mental states. To test if ToM can explain some of the inter-individual or inter-pair variance in cue-selection and cue-utilization we correlated ToM scores with performance measures.

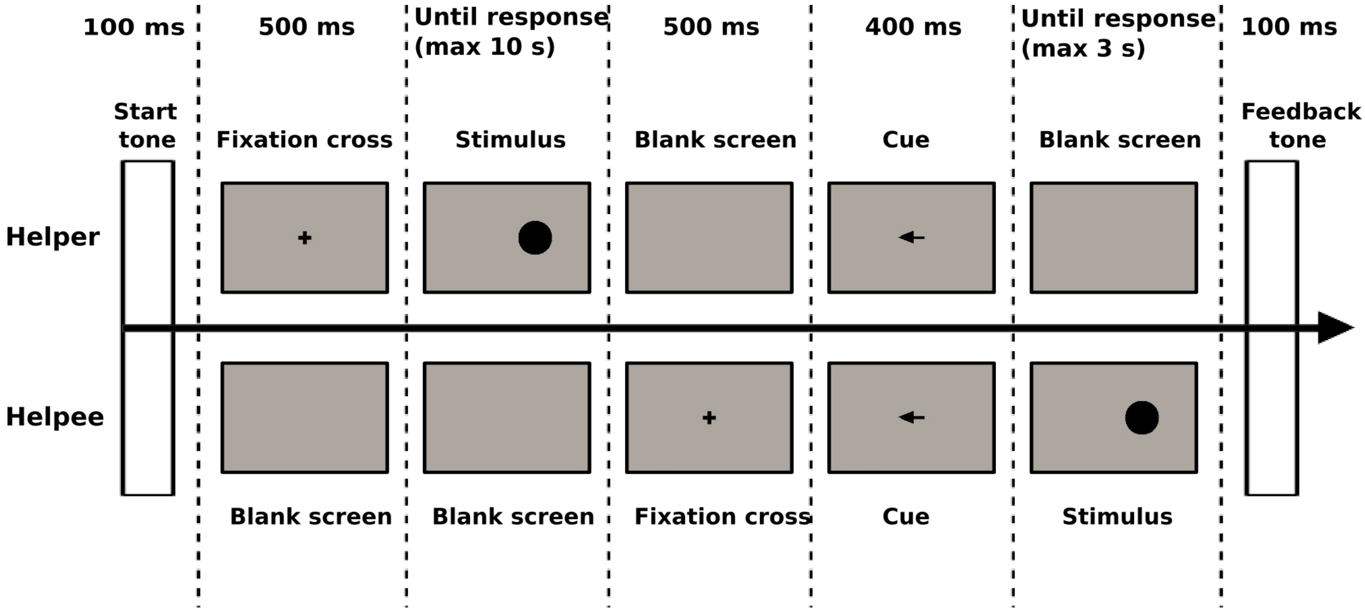


*Figure 2.1.* Trial structure in the joint condition in Experiment 1/A. After the trial start tone (100 ms), the helper's screen displays a fixation cross (500 ms) then the upcoming stimulus for the helpee (a disk or a box, either on the left or the right). Meanwhile, the helpee is presented with a blank screen. Following the helper's selection of a cue, the helpee's screen displays a fixation cross (500 ms), the cue selected by the helper (400 ms), and then the

stimulus (until response). The helper is shown the selected cue at the same time as the helpee. A tone at the end provides feedback about the helpee's response (right or wrong).

### 2.2.1. Methods

**Participants.** There were 56 participants in this experiment. Participants were run in pairs. Four participants (in three pairs) did not follow the instruction. The data from these three pairs were discarded, resulting in a final sample of 50 participants (25 pairs). The remaining participants (33 female, 8 left-handed) ranged in age from 18 to 27 years (M = 21.42 years, SD = 1.79). They all had normal or corrected-to-normal vision. The two participants in each pair were randomly assigned to the helper role and to the helpee role. Participants were recruited either through the SONA research participation system (http://www.sona-systems.com) or through a local student association and gave written consent. All participants were paid 2000 HUF corresponding to ca. 6.5 EUR. All experiments were approved by the ethics committee of the Hungarian Psychological Association (EPKEP).

**Stimuli and apparatus.** Participants performed a spatial compatibility task. During the individual familiarization phase, the stimuli consisted of black and white disks presented left and right. Color was the relevant stimulus feature and spatial position was the irrelevant stimulus feature. Disk diameter was 3.57 cm (2.55° of visual angle from a viewing distance of 80 cm). Stimuli appeared 4.91 cm (3.51°) to the left or the right of the screen center.

During the crucial joint phase of the experiment and the ensuing individual baseline phase (see Procedure), the stimuli were black squares (side length 3.57 cm, 2.55°) and disks (diameter 3.57 cm, 2.55°) with shape as the relevant feature. As during familiarization, stimulus position was the irrelevant stimulus feature. The cues the helper chose from were black left and right arrows (3.57 cm, 2.55°) with a maximum width at the arrowhead of 1.79

cm (1.28°). Arrows were presented centrally. All stimuli were displayed upon a uniform gray background.

Data collection and stimulus presentation were controlled by iMac computers (with 21.5 inch monitors) running MATLAB (version R2010b, The MathWorks, Inc.) with Psychophysics Toolbox extension (version 3, http://www.psychtoolbox.org, Kleiner et al., 2007). Response times were collected using the left and the right cursor keys on standard commercial keyboards. Participants completed a Hungarian translation of the 'Reading the Mind in the Eyes' test (revised version, Baron-Cohen et al., 2001).

**Procedure.** Participants were seated in separate rooms approximately 80 cm from their individual monitors. Each participant went through three consecutive phases: An individual familiarization phase, a joint phase where one participant acted as helper and the other acted as helpee, and a third phase that served to collect a baseline for individual task performance.

During the initial familiarization phase participants were asked to respond as quickly and accurately as possible to the color of a disk (black or white) by pressing the left or right arrow key. Key assignment to color was counterbalanced across participants. Each trial started with a short beep (1200 Hz, 100 ms), then a fixation cross appeared centrally (500 ms), followed by the stimulus. The maximum time allowed for a response was 3 s. After the response a blank screen was shown and a beep indicated whether the response was correct (1600 Hz, 100 ms) or incorrect (400 Hz, 100 ms). After an ITI of 1 s the next trial started. There were 80 familiarization trials, 40 spatially compatible trials where the response side was the same as the spatial position of the stimulus and 40 incompatible trials were response side and stimulus side were opposite. The order of compatible and incompatible trials was randomized.

In the second phase referred to as joint condition participants jointly performed a similar spatial compatibility task as in the familiarization block with shape as the relevant stimulus feature (disk or square). One participant acted as helper and one participant acted as helpee. Helpers knew the helpees' task and were instructed to provide cues to support the helpees' task performance. There was no further specification of what was considered helpful. Helpees were instructed to complete the task as quickly as possible. They were informed that helpers were instructed to provide helpful cues.

Each trial started with a short beep (1200 Hz, 100 ms) followed by a fixation cross (500 ms). Then the stimulus was displayed to the helper for a maximum of 10 s while the helpee looked at a blank screen (see Figure 2.1). As soon as the helper pressed the left or right arrow key to send a left or right cue to the helpee a blank screen was displayed for the helper and a fixation cross for the helpee appeared for 500 ms. Then the cue was displayed on both the helper's monitor and the helpee's monitor for 400 ms. Finally, the stimulus was shown to the helpee for a maximum time of 3 s while the helper's screen was blank. Helpees responded by pressing left or right arrow keys. The same feedback sounds as in the familiarization phase indicated whether the helpee's response was correct or incorrect. The sounds were audible for helper and helpee. The joint phase consisted of 3 blocks of 80 trials each (240 total). In each block half of the trials were compatible and half were incompatible. The trial order was randomized. Key assignment to shape was counterbalanced across pairs of participants.

The third phase referred to as baseline condition served to collect an individual baseline for the helpee's task during the second phase. All participants (helpers and helpees) were asked to perform this task but without cues (the stimulus immediately appeared after the initial beep and fixation cross). There were 80 trials, 40 compatible and 40 incompatible in random order. Key assignment to shape remained the same as in the joint phase. At the end of the experiment, participants were asked to complete the 'Reading the Mind in the Eyes' test.

### 2.2.2. Results

**Helpers.** For each helper, we first determined the cueing strategy they followed in the joint condition (Response-cueing, Location-cueing or Other). We tested if their responses were independent of the relevant and irrelevant stimulus features using Fisher's exact test of independence separately for each helper. If a helper's responses proved to be independent from the irrelevant stimulus feature (location) but dependent on the relevant feature (shape) we categorized that helper as following the Response-cueing strategy. If we observed the opposite pattern (dependence on irrelevant feature, independence from relevant feature), the helper was considered to fall into the Location-cueing group. The rest of the helpers were classified as Other-cueing.

According to our categorization procedure, 22 out of 25 helpers (88%) cued responses, meaning that their responses were dependent on the relevant but independent from the irrelevant stimulus feature. The rest of them (3 helpers, 12%) fell into the Other category, meaning that their responses were neither dependent on the relevant nor on the irrelevant dimension. The difference in frequencies is significant ($n = 22$ vs. $n = 3$, binomial test, $p < .001$). Post hoc analysis of the cueing behavior of helpers falling in the other category revealed that one of them used only one cue to signal responses (mostly correctly) but never the other one. The remaining two participants showed seemingly random cueing behavior.

Regarding the RT data, we first performed outlier rejection separately for each participant (+/- 3 SD, M = 2.01%, SD = 0.77 and M = 1.70%, SD = 0.91 for the joint and the baseline conditions, respectively). Then we compared helpers' RT across the joint and the baseline conditions and found a significant difference ($t(24) = 4.24$, $p < .001$), with larger values in the joint (M = 560 ms, SD = 136) than in the baseline condition (M = 456 ms, SD = 90).

For the Response-cueing group (n = 22) we could also look at compatibility effects for the helper. We performed a 2x2 repeated measures ANOVA with factors Condition (Joint vs Baseline) and Compatibility (Compatible vs Incompatible) on RTs (see Figure 2.2A). The interaction was not significant (F(1, 21) = .20, p = .66, η2p = .01). We found a significant main effect of Condition (F(1, 21) = 16.95, p < .001, η2p = .45), with larger average RT in the Joint (M = 576 ms, SD = 137) than in the Baseline condition (M = 464 ms, SD = 95). There was also a significant main effect of Compatibility (F(1, 21) = 19.36, p < .001, η2p = .48), with smaller average RT in Compatible (M = 507 ms, SD = 91) than in Incompatible trials (M = 533 ms, SD = 107). The interaction was not significant (F(1, 21) = .20, p = .66, η2p = .01). The same ANOVA performed on ER yielded similar results (see Figure 2.2B): significant main effects of Condition (F(1, 21) = 9.38, p = .006, η2p = .31) and Compatibility (F(1, 21) = 12.31, p = .002, η2p = .37) but no significant interaction (F(1, 21) = 1.22, p = .28, η2p = .06). Average ER were smaller in the Joint condition (M = 1.93%, SD = 2.01) than in the Baseline condition (M = 3.69%, SD = 3.31) and there were less errors in Compatible (M = 1.29%, SD = 1.73) than in Incompatible trials (M = 4.34%, SD = 4.08). Due to the small sample size (n = 3) and the heterogeneity of the strategies identified we did not analyze data from helpers in the Other-cueing category.

We also tested for correlations between helpers'  Reading the Mind in the Eyes' scores and RT and ER. Neither was significant (ps > 0.1).

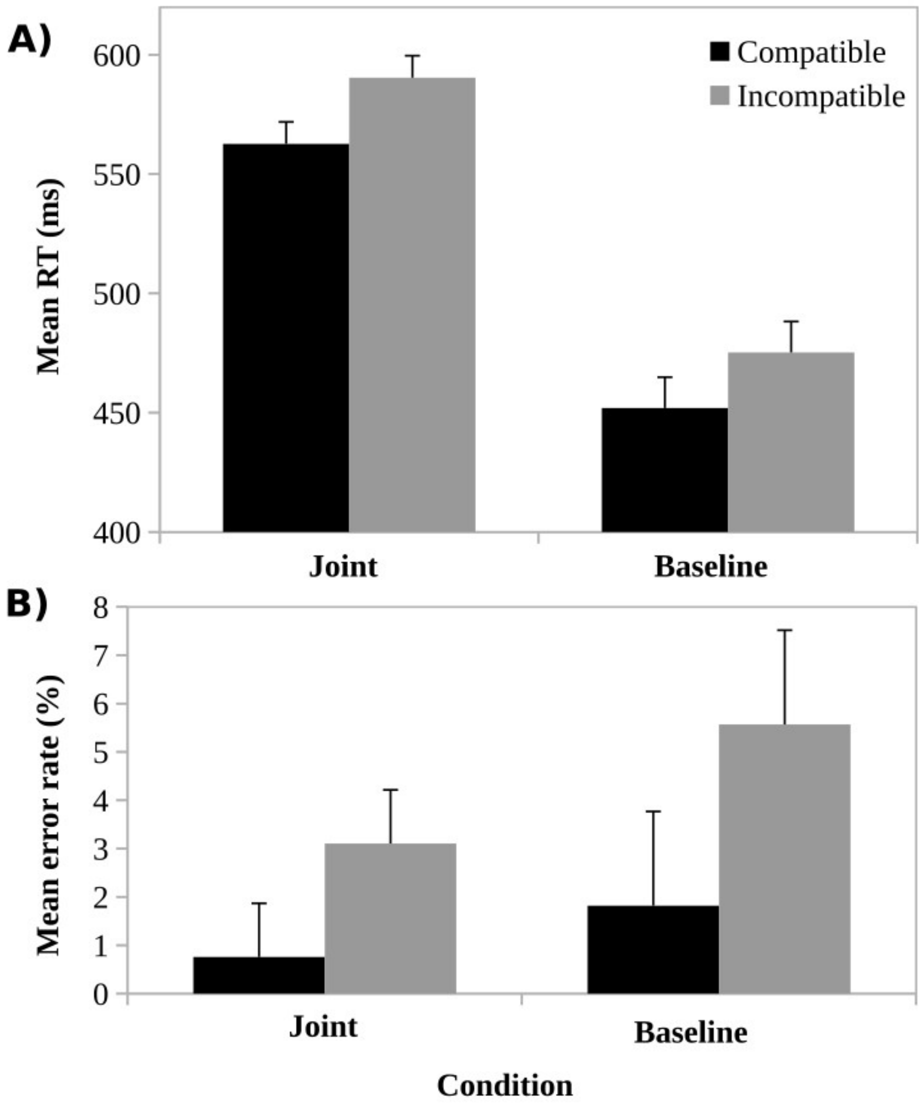


*Figure 2.2.* (A) Helpers' mean RT in Experiment 1/A. (B) Helpers' mean ER in Experiment 1/A. In both panels, black bars depict the compatible condition while gray bars depict the incompatible condition. The joint condition refers to the two-person, cueing phase of the experiment and the baseline condition refers to individual performance of the task without cues. Error bars show within-subject confidence intervals (95% CI, based on Loftus & Masson, 1994).

**Helpees.** As the different cueing strategies followed by the helpers in the joint condition determined how much helpees could benefit from cues, helpees' data were analyzed separately depending on which strategy the leader followed. Given that most helpees were in pairs where helpers cued responses (n = 22) we restricted our analyses to this group. For the helpees in this group, the cues signaled which response was required for the upcoming stimulus. Cue validity was close to perfect (M = 98.07%, SD = 2.01, ranging from 92.92% to 100%).

All response times reported below are relative to stimulus onset. Note that helpees could initiate responses following cue onset, occasionally leading to negative RT values relative to stimulus onset.

As with helpers' data, we first rejected outliers for each participant (+/- 3 SD, M = 1.33%, SD = 0.66 and M = 1.31%, SD = 0.90 for the joint and the baseline conditions, respectively). Then we performed the same 2x2 repeated measures ANOVA as for the helpers' data (with factors Condition: Joint vs Baseline; and Compatibility: Compatible vs Incompatible, see Figure 2.3A). The main effect of Condition was significant (F(1, 21) = 41.34, p < .001, η2p = .66), with smaller RT in the Joint (M = 209 ms, SD = 177) than in the Baseline condition (M = 422 ms, SD = 53). The main effect of Compatibility was also significant (F(1, 21) = 18.22, p < .001, η2p = .46), with smaller values in the Compatible (M = 306 ms, SD = 101) than in the Incompatible trials (M = 325 ms, SD = 110). There was no significant interaction (F(1, 21) = .13, p = .73, η2p < .01). For ER (see Figure 2.3B), the main effect of Condition did not reach significance (F(1, 21) = 3.77, p = .066, η2p = .15), although ER in the Joint condition (M = 3.28%, SD = 2.64) were somewhat lower than in the Baseline condition (M = 4.94%, SD = 3.91). The main effect of Compatibility was significant (F(1, 21) = 12.84, p = .002, η2p = .38). ). ER were smaller in Compatible (M = 2.63%, SD = 2.67) than in Incompatible trials (M = 5.59%, SD = 3.80). Similar to the RT results, the interaction was not significant (F(1, 21) = 2.56, p = .13, η2p = .11).

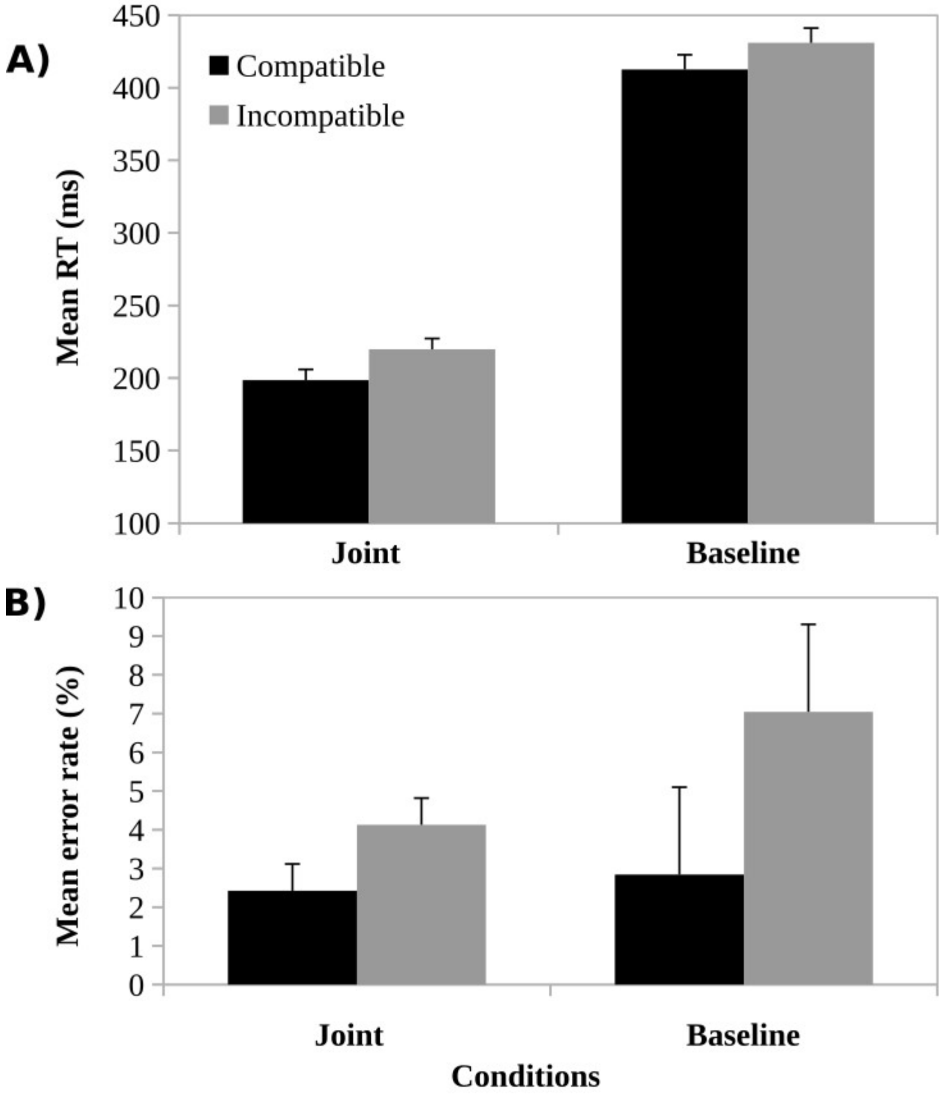


*Figure 2.3.* (A) Helpees' mean RT in Experiment 1/A. (B) Helpees' mean ER in Experiment 1/A. In both panels, black bars depict the compatible condition and gray bars depict the incompatible condition. The joint condition refers to the two-person, cueing phase of the task and the baseline condition is the individual version without cues. Error bars show within-subject confidence intervals (95% CI, based on Loftus & Masson, 1994).

**Relation between helper performance and helpee performance.** To investigate if the helpers' and helpees' performances were related, we first tested for a linear relationship between the compatibility effect in the joint condition and the cueing gain of the helpee (that is, the mean RT difference between the joint and baseline conditions). This correlation was significant: r(20) = -.68, p < .001. The more helpees' performance was boosted by helpers' communication, the smaller their spatial compatibility effect in the joint condition (see Figure 2.4). The correlation between cueing gain in terms of ER and spatial compatibility was in the same direction, but failed to reach significance: r(20) = -.45, p = .055.

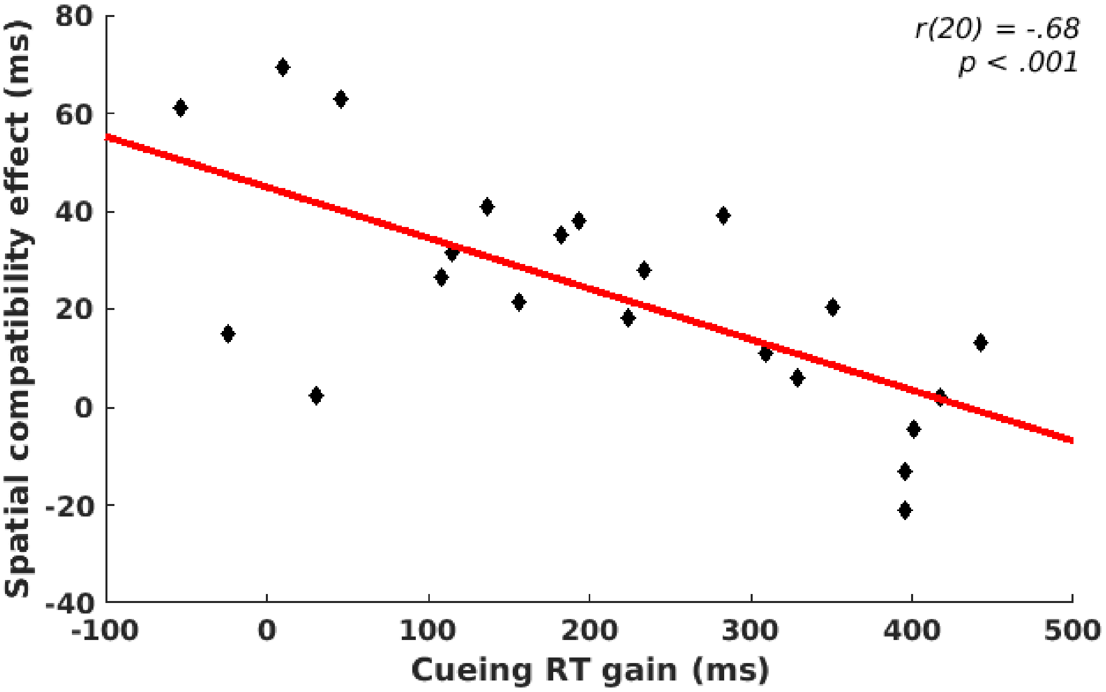


*Figure 2.4.* Negative relationship between RT cueing gain and spatial compatibility effect shown by helpees in the joint condition (Experiment 1/A). The more helpees benefited from helpers' cues the lower the spatial compatibility effect they exhibited.

To see if helpers' errors led to more errors in helpees we first tested for a link on the level of averaged ER and found a significant correlation between helpers' and helpees' ER: r(20) = .51, p = .015. To analyze this further we calculated correlation coefficients between helpers' and helpees' trial-to-trial accuracy data (represented as binary arrays) in each pair. That is, we tested if helpees' errors were more likely to occur in the same trials as helpers' errors. Correlations in this case are equivalent to phi coefficients. We included all pairs where both helper and helpee committed errors in the joint condition (n = 19, out of 22). Then we created pseudo pairs by creating all possible combinations of helpers and helpees from the real pairs that did not perform the task together (n = 342). We calculated error correlation values for the pseudo-pairs as well. To test whether real pairs were different from pseudo pairs, we performed a random permutation test with the difference of the mean correlation between real and pseudo pairs as the test statistic, with 10^6 iterations. The difference was significant (estimated p < .001, Cohen's d = 5.26), with a higher average correlation coefficient for real pairs than for pseudo pairs (M = .48, SD = .33 and M = -.003, SD = .05, respectively), see Figure 2.5.

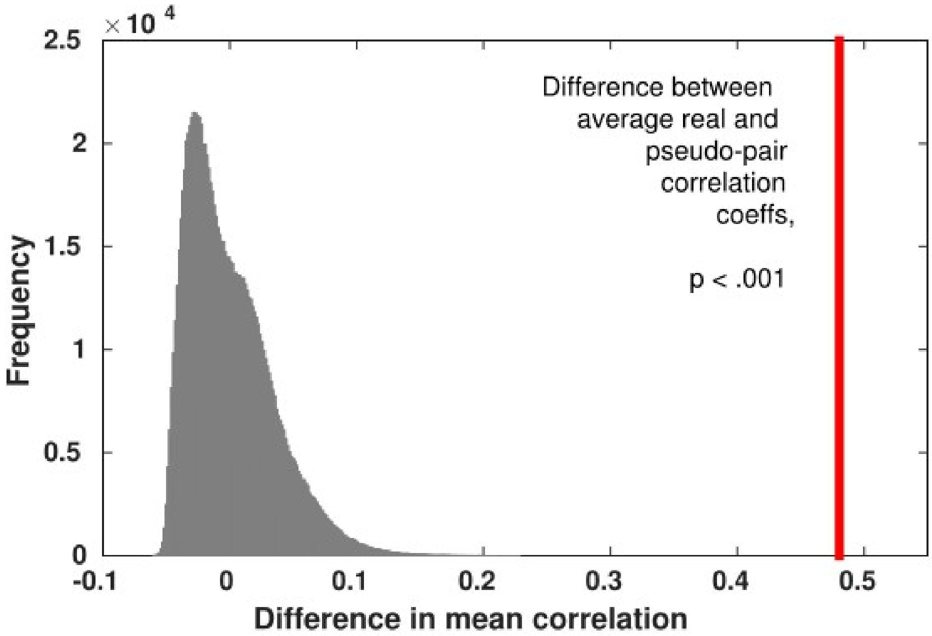


*Figure 2.5.* Random permutation test results for helper-to-helpee error correlation differences between real pairs and pseudo pairs. The x axis shows the difference of the mean correlation coefficient between real pairs and pseudo pairs. The red vertical line is the actually observed difference while the gray histogram displays the results of the random permutations.

To investigate helpees' sensitivity to helpers' errors further we investigated whether there were post-error slowing effects (Dutilh et al., 2012; Rabbitt, 1966) in helpees following helpers' errors. In order to differentiate between the effects of helpees' own errors and helpers' errors, we selected all error trials in the joint condition where either the helper or the helpee, but not both of them committed an error. We made sure that there were no other errors either before or after error trials included in the analysis (in a +/- 1 trials window). We tested if helpees showed a slow-down after errors committed by helpers, by comparing helpees' RT in trials preceding and following helpers' errors with a one-way ANOVA. There were 9 helpees with suitable data included in this comparison (with an average ER of M = 1.67%, SD = 1.92). The ANOVA yielded a significant effect, F(2, 16) = 3.92, p = .041, η2p = .33 (see Figure 2.6). Follow-up tests showed that helpees' RT were larger in trials with helpers' errors than in preceding trials (t(8) = 3.26, p = .012, Cohen's d = 1.09, with M = 396 ms, SD = 200 in the trials with errors and M = 548 ms, SD = 206 in the preceding trials). The other two comparisons (trials preceding errors vs. trials following errors and trials with errors vs. trials following errors) were not significant (ps > .1).

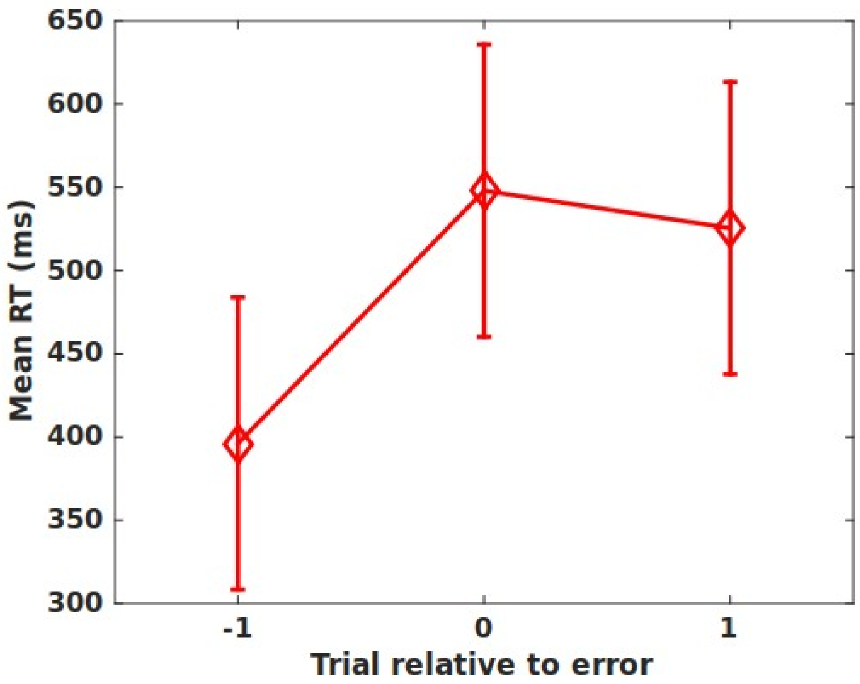


*Figure 2.6.* Helpees’ sensitivity for helpers’ errors (n = 9). Zero refers to the trial in which the error occurred and -1 and 1 refer to the preceding and following trial. The red line shows helpees’ mean RT across consecutive trials. The error bars depict the 95% within-subject CIs (based on Loftus & Masson, 1994), calculated for the ANOVA across the three trials.

We also tested for an effect of helpers’ ER on helpees’ RT and found a positive correlation that just failed significance: $r(20) = .42$, $p = .055$. As for helpers, we investigated whether there was a linear relationship between helpees’ ‘Reading the Mind in the Eyes’ scores and their RT and ER. Both correlations were non-significant (ps > 0.1).

**Monitoring results.** To investigate monitoring, we focused on participants from pairs with Response-cueing in the joint condition. We excluded the Other-cueing group from this analysis in order to focus on a homogeneous group in terms of cueing strategy. We looked for a putative link between the helper’s and the helpee’s RT time series. Because helpees heard the trial start tone and their task started as soon as helpers responded to the stimulus, they were aware of helpers' reaction times. Similarly, helpers knew that helpees' task in each trial started after their own response and ended when they heard the feedback tone. Thus, both helpers' and helpees' RT changes could – in principle – affect their partners' behavior.

We first tested for this by correlating helpers' and helpees' mean RT and found no significant effect ($r(20) = .15$, $p = .52$). Then we investigated monitoring at the pair level, that is, whether helpers and helpees adapted their behavior to one another on a trial-by-trial basis. To do so, we first detrended the RT time series of each pair and then calculated cross-correlation coefficients between reaction times of helpers and helpees. We were interested in the zero and plus one lag cross-correlation coefficients (similar to the approach in Konvalinka, Vuust, Roepstorff & Frith, 2010). Zero lag measures the linear relationship between the helper's and the helpee's RT in the same trials, while plus one lag measures the linear relationship between the helpee's RT and the helper's RT in the next trial. To test if the obtained values were statistically significant, we created pseudo pair data by pairing helpers and helpees from different pairs together (for a total of $n = 462$ pseudo pairs) and calculated cross-correlations for these pairs. In the second step, we used random permutation testing (10^6 samples) with the difference of the mean correlation coefficients between real and pseudo pairs as the test statistic.

The results are shown in Figure 2.7. For the zero lag we found a significant difference (estimated $p < .001$, Cohen's $d = 0.82$), meaning that real pairs showed larger RT correlations ($M = .095$, $SD = .11$) than pseudo pairs ($M = .026$, $SD = .076$). There was no significant difference, though, for the plus one lag (estimated $p = .14$, Cohen's $d = 0.37$; with $M = .047$, $SD = .12$, and $M = .018$, $SD = .087$ for the real and the pseudo pairs, respectively).

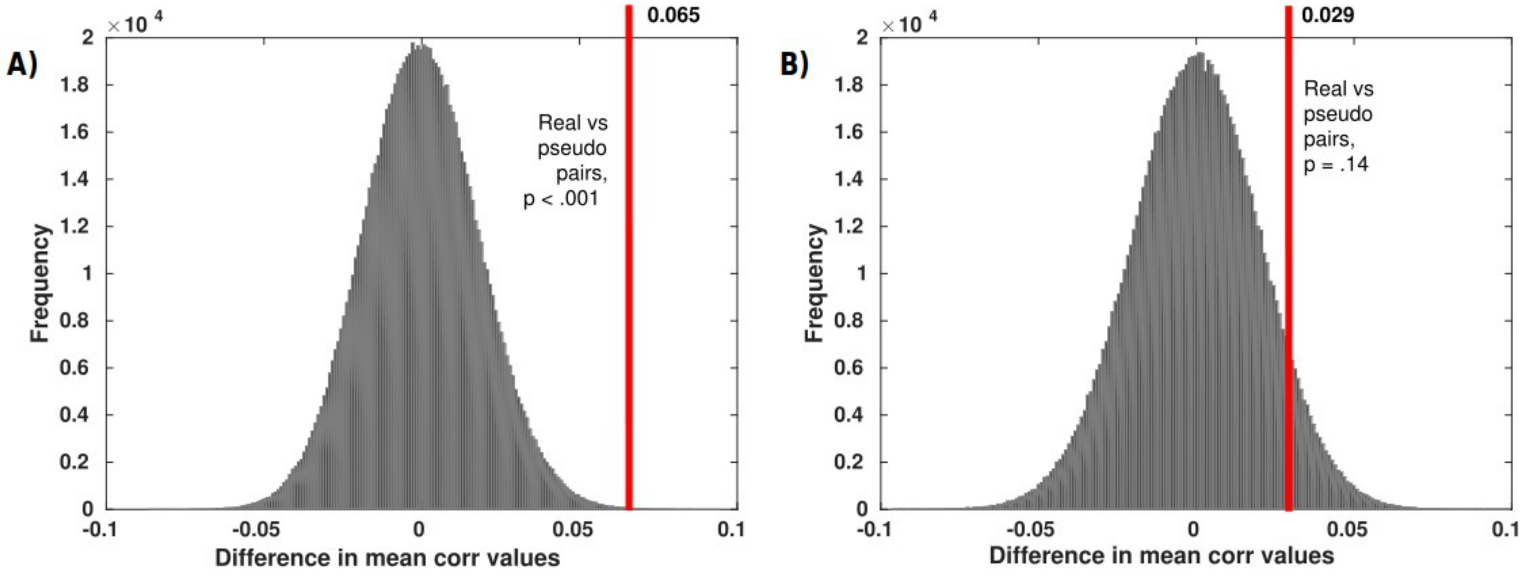


*Figure 2.7.* Random permutation test results for helper-helpee RT cross-correlation differences between real and pseudo pairs. The x axes show the difference of the mean correlation coefficient between real and pseudo pairs. The red vertical line is the actually observed difference, while the gray histogram displays the results of the random permutations. (A) Results for zero lag cross-correlation. The significant difference relative to the pseudo pairs suggests that on average, helpers' trial-to-trial RT affected the speed of helpees' responses in the same trials. (B) Results for plus one lag cross-correlation. Real pairs' were not significantly different from pseudo pairs'. Thus, there was no evidence that helpers' RTs in the current trial were affected by helpees' response times in the previous trial.

As a further measure of monitoring we analyzed post-error slowing in helpers in response to helpees' errors. To do so, we relied on the same error trials we selected for the analysis of the post-error slowing effect in helpees but focused on helpers' RTs in trials around helpees' errors using a one-way ANOVA. We could include data from 19 pairs in this analysis, with average helpee ER of M = 2.30% (SD = 1.92). There was no significant effect: F(2, 36) = 2.17, p = .13, η2p = .11. .

We tested for a relationship between helpers' and helpees' 'Reading the Mind in the Eyes' scores and the monitoring measures (RT cross-correlation at zero lag and post-error slowing in helpers). No correlations were significant (all ps > .1).

### 2.2.3. Discussion

As the efficiency account of helping predicted, we found that most helpers cued responses for helpees. Not a single helper cued the irrelevant response dimension (i.e., location). Response-cueing implied that helpers took over performing the helpees' task: to select the cue corresponding to the required response, helpers had to perform the whole task of responding to the relevant stimulus feature to ignore the irrelevant one. By doing so, helpers also provided the maximum potential performance benefit for their partners. In addition to their cueing choices, helpers' RT and ER results were also in line with the efficiency account. Both the whole group of helpers (*n* = 25) and those cueing responses (*n* = 22) were slower in the joint condition than in the baseline condition, suggesting that they sacrificed their own performance for the sake of boosting helpees' performance. This is further corroborated by the fact that Response-cueing helpers were not only slower relative to baseline, but also committed fewer errors, trading speed for accuracy, and thus, for a higher potential benefit for the helpees.

Regarding helpees, the results did not provide strong support for the efficiency account. Because helpers provided reliable cues of the response required, helpees' could simply follow the cues as soon as they appeared. The efficiency account predicted that they would do so, and indeed, helpees were significantly faster in the joint condition than in the baseline condition but the RTs seem too slow to reflect full reliance on the cues provided. The average performance gain relative to the baseline condition was about 200 ms on average. This is much less than expected following an efficiency account. Helpees had 400 ms to respond to the cue, a task that is considerably easier than performing the visual Simon task individually that took helpees about 400 ms to perform on average. Accordingly, had helpees chosen to let the cue guide their response one would have expected them to react in this interval before the stimulus appeared. This should have resulted in negative RTs or RTs close to 0. In fact, helpees took more than 600 ms after cue appearance to respond.

The analysis of errors showed that helpees' errors were linked to helpers' errors, confirming that helpees' relied on cues from helpers to some extent. However, the average performance boost in terms of RT (i.e, the gain relative to the baseline condition, *M* = 202 ms, *SD* = 159) was much less than the theoretical maximum (due to the simplification of the task, the potential gain is more than the interval between cues and stimuli, that is, 400 ms). Helpees also showed spatial compatibility effects in both conditions, suggesting that response selection was affected by the stimulus, even despite the presence of valid response cues from the helper. These latter results are in line with the split-effort and lack-of-trust accounts, as they support the idea that helpees utilized the cues to prepare their response but preferred to perform at least a part of the task themselves or to check the cue, even if such behavior was inefficient. The negative correlation between the size of helpers' spatial compatibility effects and helpers' gain from the cues further supports the idea that these two phenomena were rooted in the same relative lack of reliance on cues. Such a discrepancy between helpers and helpees has a potential to lead to differences in perceiving individual task contributions that might result in trouble: Whereas helpers rightly perceive themselves fully taking over the helpees' cognitive task, helpees perceive themselves as making a major contribution to task performance.

The idea that helpees did not make full use of the cues provided is further supported by the analyses of helpers' errors. First, trial-to-trial error correlations showed that whenever helpers committed errors, helpees were more likely to commit errors in the same trial as well. The strength of such correlations relative to the pseudo pair group describe the degree of absolute reliance on cues. While they were significantly larger than zero, they were not close to their theoretical maximum. Second, helpees' post-error slowing in trials with helper errors relative to the previous trial showed that (1) in some pairs, helpees occasionally corrected for helpers' errors; and (2) such corrections required effort. In other words, for a subgroup of helpees we can be certain that they did not fully rely on helpers' cues.

Regarding monitoring, we found that helpers' trial-to-trial RT was correlated with helpees' RT in the same trials (zero lag cross-correlation). That is, when helpers slowed down or sped up their responses, helpees tended to also slow down or speed up. This correlation may indicate that helpees payed attention to indirect timing information coming from the helper. Such information could be beneficial for the helpee, as helpers' RT could signal response conflicts, lapses in attention or – on the contrary – confidence in the selected responses. However, we did not find any link between the strength of this correlation and helpers' or helpees' performance. This could be due to the relative insignificance of the timing information for performance. Note also that while monitoring is a necessary precondition for the cross-correlation effect, this is not true the other way around. Helpers who had zero or weak cross-correlation effects may still have monitored their partners' RT.

We found no evidence for monitoring in the opposite direction. Testing for plus-one-lag cross-correlation yielded a null result. Helpers' post-error slowing data showed the expected reaction in the case of their own errors (marked slowing from error to subsequent-to-error trial), but a similar reaction was missing in the case of helpees' errors. The 'Reading the Mind in the Eyes' scores provided no indication of an effect of ToM on cue-selection or -utilization.

Helpees' results in Experiment 1/A can be explained by both the split-effort and the lack-of-trust hypotheses. Therefore, we conducted a further experiment to explicitly test that helpees may not have trusted the cues provided by the helpers. One potential reason for not trusting the cues is that helpers' cues were not perfectly valid on average, meaning that occasional errors still required vigilance from helpees (see e.g. Bestmann et al., 2008 about the effect of surprising cues on action preparation). Moreover, helpers' errors – however few there were - might have also served as strong negative signals about helpers' capability to fulfill their role. According to influential accounts of communication (theory of epistemic vigilance, Sperber et al., 2010, see also the developmental literature on selective trust, e.g. Mills, 2013), people are in constant need of evaluating the motivations and capabilities of their joint partners, due to the relative ease of deceit. Due to such heightened vigilance in case of human communication, even a few errors may become an obstacle for developing trust. We set out to test this idea in Experiment 1/B.

## 2.3. Experiment 1/B

There were two changes from Exp. 1/A to 1/B. First, cues were generated by a computer rather than a human helper and helpees' instructions were changed accordingly. The cues were always valid and cued the correct response. With these changes we aimed to eliminate the potential cause of helpees' putative lack-of-trust in the cues. If dedicated social cognitive processes (Sperber et al., 2010) evaluating a partner based on her errors led to a lack of trust in the previous experiment, eliminating human partners and any cueing errors should enable helpees to make full use of the information provided by the cues. If helpees still do not make full use of the cues and still exhibit a spatial compatibility effect, partial cue use would most likely reflect helpers' attempts to make a contribution to performing the task.

### 2.3.1. Methods

**Participants.** Of the 21 right handed participants in Experiment 1/B one had to be excluded due to excessive error rates following a misunderstanding of the instructions. For the remaining 20 participants (14 female), age range was 18-32 (M = 23.1 years, SD = 3.9).

They all had normal or corrected-to-normal vision. We recruited and compensated participants the same way as in Experiment 1/A. All participants gave written consent.

**Stimuli and apparatus.** Stimuli and apparatus were the same as in Experiment 1/A.

**Procedure.** The only change from Exp. 1/A to Exp. 1/B concerned the second phase of the experiment. While in Exp. 1/A half of the participants were assigned the role of helper and were instructed to help helpees by selecting cues for them, the cues in Exp. 1/B were generated by the computer (as in standard cueing tasks). All participants were assigned the role of helpee and performed the task individually. The instruction reflected this change and participants were told that the computer was selecting a helpful cue in each trial. Cues were 100% valid, pre-specifying the correct response side with a left or right arrow.

### 2.3.2. Results

We first performed outlier rejection separately for each participant (+/- 3 SD, M = 1.25%, SD = 0.89 and M = 1.25%, SD = 1.07 for the joint and the baseline conditions, respectively.

Then we analyzed helpees' RT and ER in the same way as in Exp. 1/A, by performing a 2x2 repeated measures ANOVA with factors Condition (Cueing vs. Baseline) and Compatibility (Compatible vs. Incompatible). Figure 2.8 depicts the results. For RTs, there was a significant main effect of Condition (F(1, 19) = 22.10, p < .001, η2p = .54), with larger RT in the Baseline (M = 436 ms, SD = 56) than in the Cueing condition (M = 208 ms, SD = 210). The main effect of Compatibility was significant as well (F(1, 19) = 22.47, p < .001, η2p = .54), with larger RT in Incompatible trials (M = 335 ms, SD = 119) than in Compatible trials (M = 309 ms, SD = 99). The interaction was also significant: F(1, 19) = 5.02, p = .037, η2p = .21. Follow-up tests showed that there was a compatibility effect in both the Cueing condition (t(19) = 5.15, p < .001, Cohen's d = 1.15, with M = 192 ms, SD = 198 in Compatible and M = 225 ms, SD = 222 in Incompatible) and Baseline conditions (t(19) = 2.86, p = .01, Cohen’s d = .64, with M = 427 ms, SD = 61 in Compatible and M = 445 ms, SD = 55 in Incompatible), but the effect was larger in the Cueing condition. For ER, there was no main effect of Condition (F(1, 19) = 2.12, p = .16, η2p = .1) but a significant main effect of Compatibility: F(1, 19) = 15.01, p = .001, η2p = .44. ER were higher in Incompatible (M = 5.02%, SD = 3.91) than in Compatible (M = 2.48%, SD = 2.46) trials. The interaction was not significant (F(1, 19) = .59, p = .45, η2p = .03).

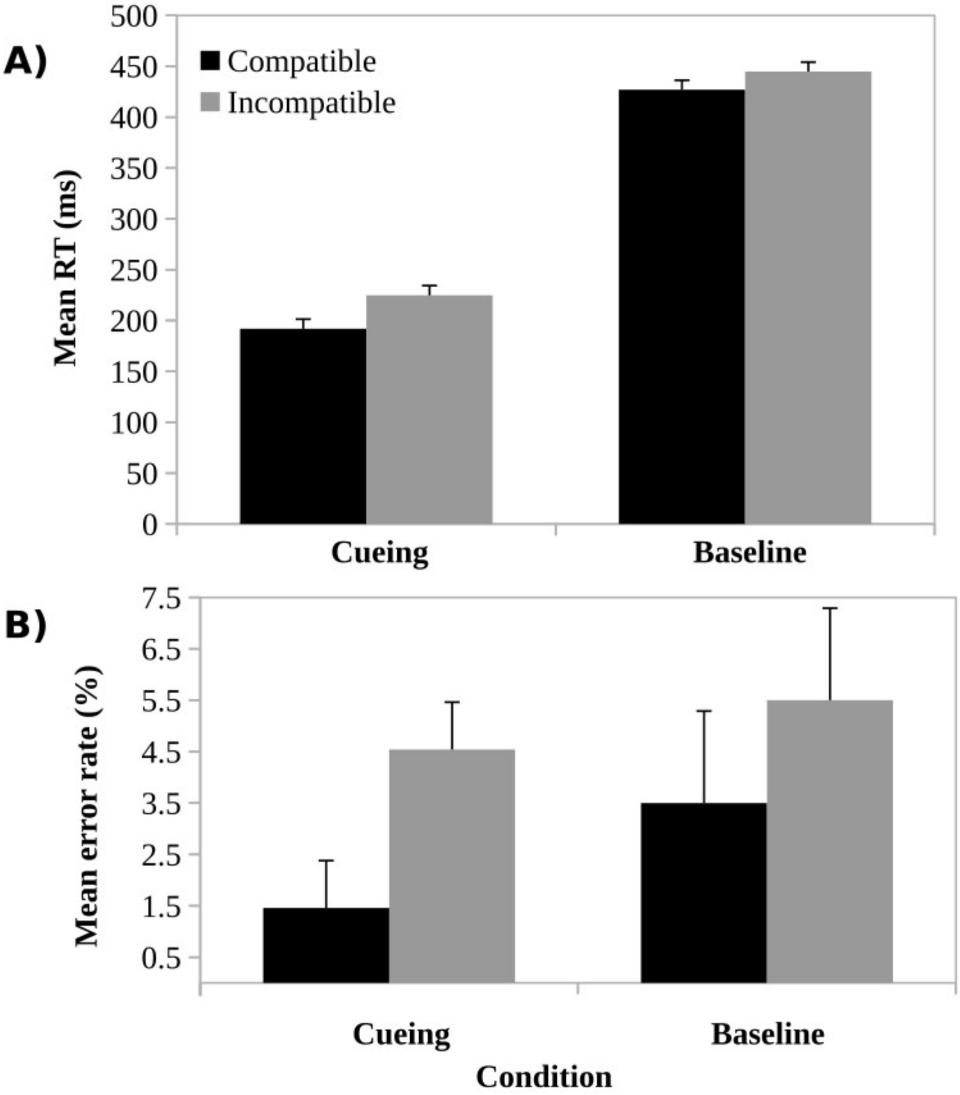


*Figure 2.8*. (A) Helpees' mean RT in Experiment 1/B. (B) Helpees' mean ER in Experiment 1/B. In both panels, black bars depict the compatible and gray bars depict the incompatible condition. In the cueing condition of Exp. 1/B, the cues were computer generated and helpees were instructed accordingly. The baseline condition was the same as in Exp. 1/A, it is the individual version without cues. Error bars show within-subject confidence intervals (95% CI, based on Loftus & Masson, 1994).

We also compared helpees' RT and ER across Exp. 1/A and 1/B. A 2x2x2 mixed design ANOVA on the RT data with between factor Cue Source (Human helpee vs. Computer), and within factors Condition (Joint vs. Baseline) and Compatibility (Compatible vs. Incompatible) yielded no main effect of Cue Source (F(1, 40) = .01, p = .94, η2p < .01) and no significant interactions involving the factor Cue Source (Cue Source X Condition X Compatibility: F(1, 40) = 1.17, p = .29, η2p = .03; Cue Source X Condition: F(1, 40) = .18, p = .67, η2p < .01; Cue Source X Compatibility: F(1, 40) = .13, p = .72, η2p < .01; Condition X Compatibility: F(1, 40) = 2.61, p = .12, η2p = .06). For ER the results were the same, no significant main effect of Cue Source (F(1, 40) = .10, p = .76, η2p < .01) and no significant interaction involving the factor Cue Source (Cue Source X Condition X Compatibility: F(1, 40) = 2.03, p = .16, η2p = .05; Cue Source X Condition: F(1, 40) = .01, p = .94, η2p < .01; Cue Source X Compatibility: F(1, 40) = .27, p = .61, η2p < .01; Condition X Compatibility: F(1, 40) = .19, p = .67, η2p < .01).

### 2.3.3. Discussion

Experiment 1/B asked whether helpees' would make better use of response cues if one eliminates potential reasons for a lack of trust. This was not the case. The data pattern obtained for helpees correspond to the data pattern obtained in Experiment 1/A and a cross-experiment analyses provided no indication there was a difference in the data pattern.

Unexpectedly, in Experiment 1B the Simon effect was somewhat larger in the cueing phase than in the baseline. One potential explanation is that helpees in the cueing condition of Exp. 1/B initiated their responses more consistently after stimulus onset than the helpees in Exp. 1/A, where the Simon effect had the same magnitude in the joint and baseline conditions. The Simon effect results from interference between the spatial features of the stimulus code and the spatial features of the response code (Hommel, Proctor and Vu, 2004; Hommel, 2011). Due to quick suppression of the irrelevant code feature following presentation, the activation of the response code and the stimulus code have to be temporally close for the Simon effect to manifest (Burle, van den Wildenberg and Ridderinkhof, 2005; Hommel, 1993). Thus, in the joint / cueing conditions of the task, the Simon effect is smaller if helpees either initiate their responses following cue onset but before stimulus onset (no overlap of response code and stimulus code) or after stimulus identification (temporal delay relative to stimulus detection), than if they initiate responses right after stimulus onset, as if the stimulus was a go signal. Following this line of reasoning, if helpees' RT distribution in the cueing condition of Exp. 1/B was narrower than in the joint condition of Exp. 1/A, this could explain why the Simon effect in the cueing condition of Exp. 1/B was larger than the cueing effect in the baseline. In terms of cueing effects, a narrower distribution could result from more consistent utilization of the cue for response preparation, a potential effect of 100% cue validity. An analysis of the RT quintiles across Exp. 1/A and 1/B showed differences in the expected direction (lowest quintile larger in Exp. 1/B than in Exp. 1/A, opposite for highest quintile), but the interaction of interest was only trending ($p = .057$), failing to reach significance (see the details in Appendix A, p.224) so this issue remains unclear.

Taken together the results of Experiment 1/A and 1/B indicate that occasional errors and the social nature of cueing were not the reasons behind helpees' failure to make full use of the cues. Thus lack-of-trust is unlikely to have caused helpees' inefficient use of cues in

Exp. 1/A. Rather, it seems that helpees seem to insist on doing a part of the task themselves or on checking cue validity even if they get ample evidence that cues are valid. Because there was no other participant in the task to split the effort with, it is not entirely clear how a split-effort model could explain the results. One hypothesis is that the computer cues were assumed to reflect some agency due to their highly conventionalized form as a communicative symbol, engaging the same mechanisms as cues generated by a partner. There are findings suggesting that cueing with culturally embedded symbols works in a similar vein to social cues. For example, centrally presented arrows have been shown to elicit similar attentional shifts as human gaze (at least in certain tasks, see e.g. Friesen, Ristic & Kingstone, 2004; Hommel, Pratt, Colzato & Godijn, 2001). Another possibility is that the cues were perceived as originating from the experimenter (that is, as generated by a rule set by the experimenter), and were essentially part of the instructions, constituting an explicit version of “top-top control” (Roepstorff & Frith, 2004, p. 195). However, in either case, the idea that computer-generated cues elicited agency attribution would contradict other experiments that used belief-manipulations to contrast the effects of social to non-social stimuli for the study of mentalizing (e.g. Gallagher, Jack, Roepstorff & Frith, 2002).

The simple task used in the first two Experiments has three limitations that restrict the generalizability of the findings. First, regarding the cue-selection strategies of helpers, it is important to remember the relationship between the cues (left-right arrows) and the required responses (left-right key presses). The cues match the responses in cardinality (that is, there are two distinct cues for two responses), allowing for simple conventions matching each cue to a response. There is also a cultural convention mapping arrows to spatial directions, which in turn correspond to the set of responses used in the task. Thus, when selecting a cueing strategy, helpers might have been driven by these potentially salient ways of mapping cues to responses and not by considerations of efficiency.

Second, the alternative to Response-cueing was Location-cueing, that is, communication of a helpful stimulus feature which is nevertheless irrelevant from a performance point-of-view. Moreover, cueing the only irrelevant stimulus feature risks driving helpees' attention towards the dimension of the task that is the source of interference in the Simon paradigm. Thus, Location-cueing might be considered an imperfect operationalization of the split-effort model for helpers.

Third, as evidenced by the low ER in the baseline condition of both experiments, the version of the Simon task used so far does not require a large amount of cognitive effort, making inefficient behavior "cheap". One might argue that we only failed to observe helpees' using cues in an efficient manner because the task did not provide a large enough motivation for relying on the partner. A more complex task would leave more room for cueing benefits and yield a more stringent measurement of cue utilization. Since these limitations offer alternatives to our conclusions so far, we set out to address them in Experiments 2/A and 2/B.

## 2.4. Experiment 2/A

To address the limitations mentioned above, we changed three aspects of the task used in Exp. 1/A (see Figure 2.9 for a layout of the version used in Exp. 2/A). First, we opted for a cue set that does not share any obvious links to the spatial arrangement or the cardinality of the response set. More specifically, instead of left and right arrows, the cue set in Exp. 2/A consisted of only one element, an exclamation mark. Helpers could choose between sending and not sending the exclamation mark to the helpee in each trial. Thus, the presence and absence of the cue could be used to indicate the correct response to the helpee. Importantly, helpees were presented with two distinct cues: when helpers chose not to send the exclamation mark (no cue provided), helpees still received ‘X’ as an explicit cue (see Methods for details) rather than receiving no cue.

Changing this cue set enabled us to test if helpers in Exp. 1/A were driven by salient cue-response mappings or by a more abstract notion of efficiency. If helpers in Exp. 1/A chose Response-cueing based on the salient mappings, we would expect a higher portion of helpers falling into the Other-cueing category in the present experiment.

Second, we addressed the potentially low efficiency of Location-cueing in Exp. 1/A by adding a further stimulus feature. In Exp. 2/A there were two stimulus features (shape and color) and we asked helpees to respond to feature conjunctions. Thus, both stimulus features were relevant, and this configuration made it possible to use more powerful cueing strategies corresponding to a split-effort model. Now the helper could cue one relevant stimulus feature, practically performing half of the helpee’s task, leading to a balanced sharing of the workload. If helpers cued responses in Exp 1/A due to the lack of adequate cueing strategies corresponding to distributing effort, we should see a higher portion of helpers cueing one stimulus feature in Exp. 2/A, and, thus, distributing effort equally between helpees and themselves.

Third, by introducing an additional stimulus feature and instruction for helpees, the task became considerably harder. Helpees needed to consider distinct features in each trial in order to successfully select the required response, increasing the potential benefits of cueing. If helpees’ level of cue utilization increases as a function of task difficulty, we would expect their performance to fit the predictions of the efficiency account in Exp. 2/A. More specifically, they should exhibit lower RT and ER in the joint phase relative to baseline, with the cueing benefit comparable to the cue-stimulus onset interval in terms of RT.

With the above changes in place, we tested the same predictions as in Exp. 1/A regarding helpers’ and helpees’ performance, namely whether an efficiency account or a split-effort/lack-of-trust account can better explain helpers’ choices and helpees’ performance. We performed the same analyses as in Exp. 1/A, focusing on helpers' and helpees' performance measures and helpees' sensitivity to helpers' errors. Because of the changes in the task, (lack of RT measures in about half of the trials from helpers) we could not test for this task whether helper and helpees monitored each other's behavior (note that there were hardly any effects in Experiment 1A).

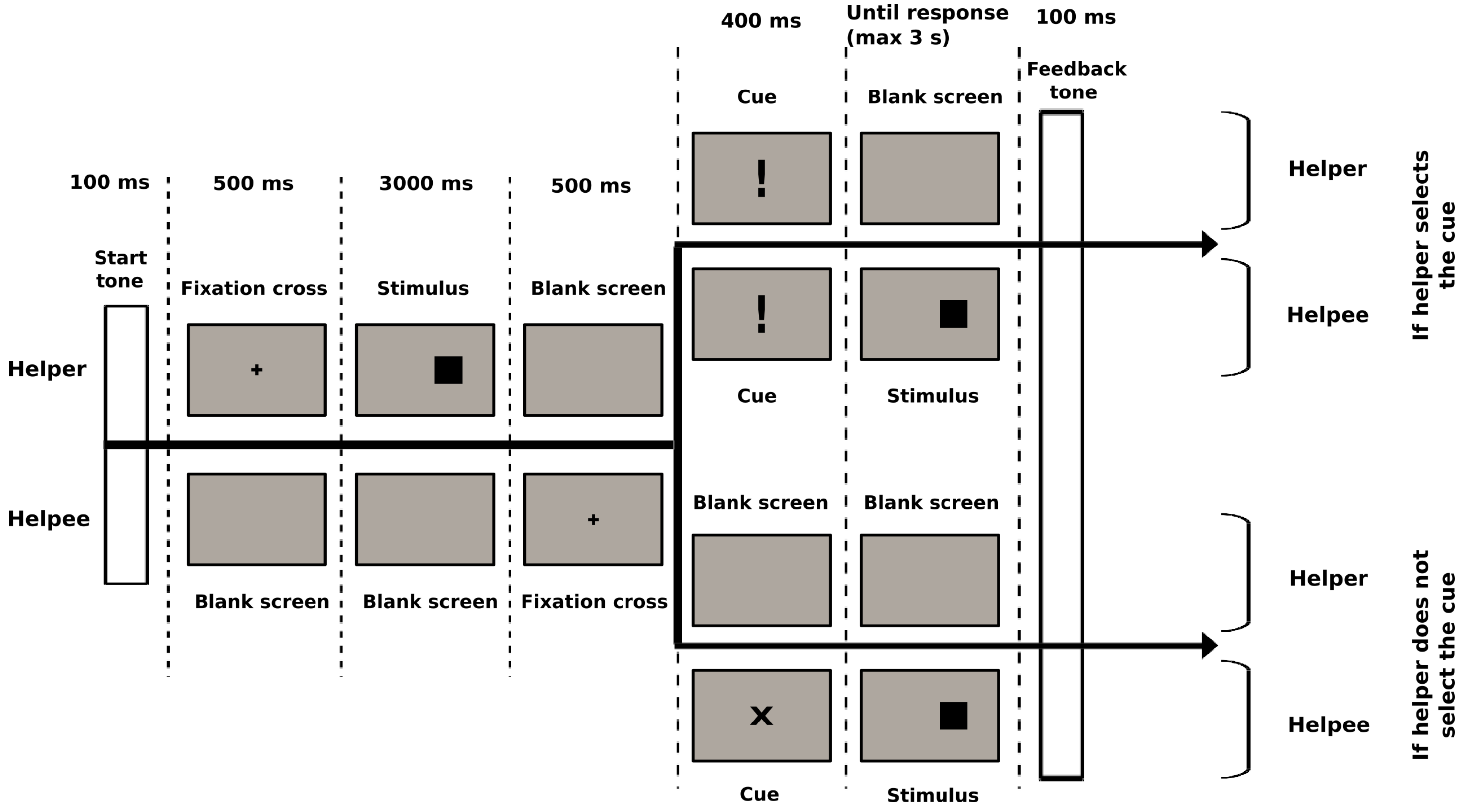


*Figure 2.9.* Trial structure in the joint condition of Experiment 2/A. After the trial start tone (100 ms), the helper's screen displays a fixation cross (500 ms) then the upcoming stimulus for the helpee (a black or white disk or box, either on the left or the right). Meanwhile, the helpee is presented with a blank screen. If the helper selects the cue, see upper half of figure, the helpee's screen displays a fixation cross (500 ms), the exclamation mark (the cue, for 400 ms) and then the stimulus (until response). The helper is also shown the cue at the

same time as the helpee. If the helper does not respond (no cue), the helpee's screen displays the alternative cue ('X'), while the helper is not shown anything. A tone at the end provides feedback about the helpee's response (right or wrong).

### 2.4.1. Methods

**Participants.** Of the 92 right handed participants in Experiment 2/A, 13 had to be excluded due to excessive error rates following a misunderstanding of the instructions. In each case, we discarded data from the whole pair, resulting in a final sample of 72 participants (36 pairs). For the final sample (44 female, 9 left-handed), age range was 18-35 (M = 22.34 years, SD = 3.20). All participants had normal or corrected-to-normal vision. We recruited and compensated participants in the same way as in Experiment 1/A. They all gave written consent.

**Stimuli and apparatus.** Stimuli and apparatus were the same as in Experiment 1/A and B, with the following changes.

In the first block of the individual familiarization phase, stimuli were identical to those used in Exp. 1/A and B (black and white disks with color as relevant feature). In the second block of the familiarization phase the stimuli were black squares and disks (with shape as relevant stimulus feature), just as in the second phase of Exp. 1/A and B. All stimuli appeared either on the left or on the right of the screen center. Location served as an irrelevant feature.

In the joint phase of the study stimuli varied both in color (black or white) and shape (square or disk). All four resulting combinations were used with equal frequency. Both color and shape served as relevant features, as helpees responded to the feature conjunction involving the two dimensions. Stimulus size and location were the same as in the familiarization phase. The set of cues seen by the helpees consisted of an exclamation mark

(‘!’ with height 1.96 cm, 1.40°, and width 0.4 cm, 0.29°) and an ‘X’ (height 1.96 cm, 1.40°, and width 1.61 cm, 1.15°), both shown centrally.

In the third, individual phase, the same stimulus set was used as in the joint phase.

**Procedure.** The procedure was the same as in Experiment 1/A with the following exceptions. The familiarization phase now consisted of two blocks of 40 trials and the relevant stimulus feature changed from the first to the second block of the familiarization phase. In the first block, participants responded to disk color (black or white) and in the second block they responded to stimulus shape (square or disk, always black). Stimulus location served as irrelevant feature throughout familiarization. The order of compatible and incompatible trials was randomized in each half.

The second, joint phase was the same as in the previous experiments, but helpers used different cues and the task to be performed was more difficult. Both the color and shape of the stimuli were relevant features and participants responded to orthogonal feature conjunctions. One of the response keys was assigned to black disks and white squares and the other response key to white disks and black squares.

Instead of having two different cue symbols at their disposal helpers could now only provide an exclamation mark as a cue and believed that not sending a cue would correspond to a blank screen for the helpees. Helpees knew that helpers were instructed to support them. However, they believed that helpers could choose between two cues, ‘!’ and ‘X’ in each trial. This mild deception was needed to create a situation that allowed us to study how helpers used a binary cue (presence/absence) while keeping the helpee’s response times comparable across different conditions.

Trial structure and feedback sounds were the same as in Experiment 1/A with the following exceptions. The stimulus was displayed to the helper for a maximum of 3 s while the helpee looked at blank screen. If the helper pressed the space bar to send the exclamation mark a blank screen was displayed on the helper’s monitor until 3 s elapsed from stimulus onset (see Figure 2.9). If the helper decided not to send a cue (space bar not pressed during 3 s interval) the stimulus remained on his monitor for 3 s. Then a blank screen was displayed for the helper while a fixation cross was shown for the helpee (500 ms). Following the fixation cross the cue for the helpee was presented. If the helper had chosen to send a cue both the helper and the helpee were shown a '!' for 400 ms. If the helper had chosen not to send a cue, the helpee was shown an 'X' and the helper was shown a blank screen for 400 ms. Then the stimulus was shown to both participants (until response from helpee or for 3 s). The joint phase had two blocks of 80 trials (160 trials total). In each block half of the trials were compatible and half were incompatible presented in random order. Key assignment was counterbalanced across pairs of participants. At end of the joint phase participants completed a short questionnaire that asked about their experience of trust towards the other participant while performing the joint task.

During the third phase all participants (helpers and helpees) were asked individually to perform the same task as the helpee in the second phase but without cues (the stimulus immediately appeared after the initial beep and fixation cross). There were 40 trials, 20 spatially compatible and 20 spatially incompatible in random order. Key assignment to shape remained the same as in the joint phase. At the end of the experiment participants were asked to complete the 'Reading the Mind in the Eyes' test as well.

### 2.4.2. Results

**Helpers.** We followed the same logic as in Exp. 1/A to identify helpers' cueing strategies. Helpers sending or not sending a cue was considered as communicating two alternative options, either referring to stimulus features or to the response that was required. By using Fisher's exact tests of independence on each helper's data, we established if helpers' cues reflected feature conjunctions dictating the response or referred to stimulus features (either color or shape). If a helper's cues were independent of stimulus features but depended on the color-shape conjunctions, we categorized that helper as engaging in Response-cueing. If a helper's cue depended on a particular stimulus feature and was independent of the other stimulus feature and the conjunctions defining the responses, the helper was considered to fall into the Stimulus-cueing group. The rest were classified as Other and we identified their exact strategies post hoc. Out of the 36 helpers, 20 cued responses, 1 fell into the Stimulus-cueing group (by cueing the color only) and 15 were in the Other category. Table 1.1 summarizes the distribution of cueing strategies in Exp. 1/A and 2/A. The difference between the experiments is significant (Freeman-Halton extension of Fisher's exact test, p = .011).

Table 1.1

*Helpers' cueing strategies in Exp. 1/A and 2/A.*

| | Response-cueing | Location / Singular – cueing | Other | Total |
|---|---|---|---|---|
| Exp. 1/A | 22 | 0 | 3 | 25 |
| Exp. 2/A | 20 | 1 | 15 | 36 |

Notes: The Location / Stimulus – cueing group refers to the number of helpers falling into the Location-cueing group in Exp. 1/A and the Stimulus-cueing group in Exp. 2/A.

The post hoc analysis of the Other group revealed further strategies we had not expected. 7 out of the 15 helpers in the Other category cued response-change, that is, when the required key press changed from the previous to the present trial for helpees. Furthermore, 2 helpers cued spatial incompatibility, that is, when the stimulus location appeared on the opposite side of the screen relative to the required response. There was also one helper who cued one specific stimulus out of the four possible feature arrangements. We verified that these helpers followed the above strategies by using post hoc Fisher's exact tests extended with the new possible dependencies (response-change cueing, spatial incompatibility cueing and cueing of specific stimulus). The rest of the helpers in the Other category (n = 5) used the cue seemingly randomly, though 3 of them used it hardly at all (in < 10% of trials) and 1 helper used it in 95% of the trials. This could be interpreted as a lack of understanding the task or a refusal to help in a minority of helpers (14%).

RTs for helpers could only be recorded from trials where they chose to show an exclamation mark. As one would expect the Response-cueing group (n = 20) showed the exclamation mark in half of the trials (M = 49.66%, SD = 1.6). Participants in the Other group (n = 15) chose the exclamation mark less often on average, with considerable variability across helpers: M = 34.33%, SD = 25, t(14) = 2.37, p = .033, Cohen's d = .94 (Welch's t-test was used because of unequal variances). The helpers in the Other group who cued response-change or spatial incompatibility came closer to showing an exclamation mark in half of the trials: M = 40.98%, SD = 11.91.

As the first step of RT analysis, we removed outliers separately for each participant (+/- 3 SD, M = 0.80%, SD = 0.77 and M = 1.46%, SD = 1.51 for the joint condition and the baseline condition, respectively. Then we tested if the task in Exp. 2/A still led to a spatial compatibility effect. There was no significant difference between compatible and incompatible trials in the baseline condition (t(35) = .57, p = .58, Cohen's d = .06). Therefore, we ignored spatial compatibility in the remaining RT analyses.

We compared helpers' RT in the joint and baseline conditions for all helpers (n = 36). The difference was significant (t(35) = 5.51, p < .001, Cohen's d = 1.24), with substantially larger RTs in the joint (M = 1189 ms, SD = 424) than in the baseline condition (M = 792 ms, SD = 162).

Next we investigated the effects of different cueing strategies. In addition to the Response-cueing group (n = 20), we also analyzed the RTs for a subset of participants in the Other-cueing group who signaled either response-change or spatial incompatibility (n = 9).

We had two reasons for involving this latter set and grouping them together. First, in principle, cueing response-change and spatial incompatibility provided very useful information about the required responses, and, thus, their effects could be comparable to each other and also to Response-cueing. Second, helpers in this subgroup had sufficient RT data and we could also define errors for them, enabling a full comparison. See Appendix A (p. 226) for RT and ER analyses involving the full Other group (n = 15).

First, we analyzed the helpers in the Response-cueing and in the subset of Other-cueing groups separately. For Response-cueing helpers we found that they had significantly larger RT in the joint (M = 1070 ms, SD = 368) than in the baseline (M = 808 ms, SD = 168) condition (t(19) = 3.41, p = .003, Cohen's d = .91), and also committed significantly less errors in the joint condition (M = 2.09%, SD = 4.27) relative to the baseline (M = 7.50%, SD = 6.18): t(19) = 3.02, p = .007, Cohen's d = 1.02. For the Other-cueing helpers we found a similar, significant difference in RT across conditions (t(8) = 5.08, p = .001, Cohen's d = 2.27), with larger values in the joint (M = 1224 ms, SD = 193) than in the baseline (M = 772 ms, SD = 205) condition. The numerical difference in ER was in the right direction, with higher ER in the joint condition (M = 23.68%, SD = 27.03) than in the baseline condition (M = 3.61%, SD = 3.09), but failed to reach significance (t(8) = 2.14, p = .065, Cohen's d = 1.04),

In order to contrast the performance of helpers across cueing groups, a 2x2 mixed design ANOVA was calculated both on RT and ER with the within factor Condition (Joint vs. Baseline) and the between factor Cueing (Response vs. Other). The results are shown in Figure 2.10. Regarding RT, the main effect of Condition was significant, F(1, 27) = 30.38, p < .001, η2p = .53, with larger RT in the Joint (M = 1118 ms, SD = 329) than in the Baseline condition (M = 797 ms, SD = 178). There was no main effect of Cueing (F(1, 27) = .51, p = .48, η2p = .02). and no significant interaction (F(1, 27) = 2.16, p = .15, η2p = .07). The same ANOVA on ER revealed significant main effects of Condition (F(1, 27) = 4.76, p = .038, η2p = .15) and Cueing (F(1, 27) = 8.18, p = .008, η2p = .23). Overall, ER were larger in the

Joint condition (M = 8.79%, SD = 18.01) than in the Baseline (M = 6.29%, SD = 5.65), and were smaller for helpers engaging in Response-cueing (M = 4.80%, SD = 3.48) than for helpers engaging in other cueing strategies (M = 13.65%, SD = 13.10). There was also a significant interaction: F(1, 27) = 14.36, p = .001, η2p = .35. Follow-up comparisons revealed that there was a difference between Response-cueing and Other-cueing helpers’ ER in the Joint condition (t(27) = 3.55, p = .001, Cohen’s d = 1.43), with Response-cueing helpers committing less errors (M = 2.09%, SD = 4.27) than Other-cueing helpers (M = 23.68%, SD = 27.03). The difference in the Baseline condition was not significant (t(27) = 1.78, p = .087, Cohen’s d = .71), although there were more errors in the Response-cueing group (M = 7.50%, SD = 6.18) than in the Other-cueing group (M = 3.61%, SD = 3.09).

We also tested for a difference in ‘Reading the Mind in the Eyes’ scores across the Response-cueing and Other-cueing groups, and found no significant difference: t(33) = .12, p = .91, Cohen’s d = .04. Correlations between ‘Reading the Mind in the Eyes’ scores and helpers’ performance measures were also non-significant (ps > .1).

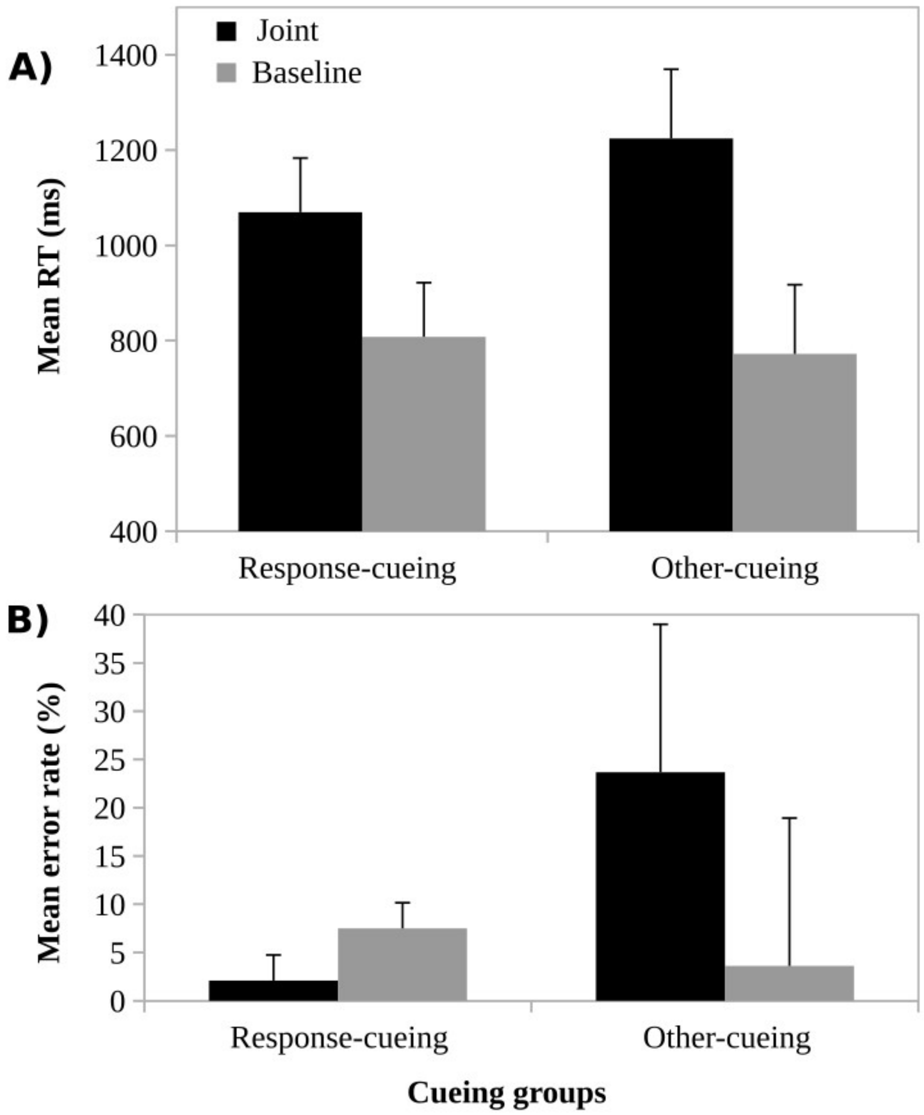


*Figure 2.10.* (A) Helpers' mean response times in Experiment 2/A. (B) Helpers' mean ER in Experiment 2/A. In both panels, black bars depict the mean values from the joint condition while gray bars depict mean values from the baseline condition. Response-cueing refers to the group of helpers signaling the required responses for helpees (n = 20) and Other-cueing refers to helpers who signaled either response-change or spatial incompatibility (n = 9). Error bars show within-subject confidence intervals, calculated separately for the two groups of helpers (95% CI, based on Loftus & Masson, 1994).

**Helpees.** As in Exp. 1/A, we analyzed helpees' performance data separately depending on helpers' cueing strategies. As for helpers we separately examined helpees in Response-cueing pairs (n = 20) and helpees in the Other category who received cues about response-change or incompatibility of the stimulus (n = 9). For analyses involving the full Other group (n = 15), see Appendix A (p. 227). As can be seen from helpers' ER, similar to Exp. 1/A, helpees in Response-cueing pairs received cues that had close to perfect validity (M = 97.91%, SD = 4.27). Cue validity was significantly lower in the Other-cueing group (M = 76.32%, SD = 27.03; and t(8) = 2.38, p = .044, Cohen's d = .14, using Welch's t-test).

For the RT analysis, we rejected outliers for each helpee separately (+/- 3 SD, M = 1.72%, SD = 0.98 and M = 1.47%, SD = 1.42). A test for spatial compatibility effects in the baseline condition showed no significant difference between compatible and incompatible trials: t(28) = 1.19, p = .23, Cohen's d = .17. Therefore, spatial compatibility was not included as a factor in the ensuing analyses.

Next, we looked at the effect of cueing in the two groups separately, by comparing helpees' RT and ER in the joint phase to the baseline. For helpees in Response-cueing pairs the RT difference across the conditions was significant (t(19) = 3.90, p = .001, Cohen's d = 1.09), with slower responses in the baseline (M = 801 ms, SD = 156) than in the joint condition (M = 529 ms, SD = 316). There was also a significant difference in ER (t(19) = 3.10, p = .006, Cohen's d = .99), with a lower error rate in the joint condition (M = 4.41%, SD = 3.29) than in the baseline condition (M = 8.75%, SD = 5.29). In the Other-cueing group, neither the RT (t(8) = .69, p = .51, Cohen's d = .19), nor the ER difference (t(8) = 1.03, p = .33, Cohen's d = .35) was significant.

Then we contrasted cueing effects (both in terms of RT and ER) across the different cueing groups using a 2x2 mixed design ANOVA with within factor Condition (Joint vs. Baseline) and the between factor Cueing (Response-cueing vs. Other-cueing), see Figure

2.11. For RT there was a significant main effect of Condition (F(1, 27) = 7.79, p = .010, η2p = .22), but no main effect of Cueing (F(1, 27) = 3.06, p = .091, η2p = .10). The interaction was significant: F(1, 27) = 4.64, p = .040, η2p = .15. Pairwise tests revealed that in the Joint condition, helpees in the Response-cueing group were faster than helpees in the Other-cueing group (M = 529 ms, SD = 316 and M = 778 ms, SD = 210, respectively, t(27) = 2.15, p = .040, Cohen’s d = .52). This difference was not significant in the Baseline condition (M = 801 ms, SD = 156 and M = 813 ms, SD = 145, for Response-cueing and Other-cueing, respectively, t(27) = .20, p = .84, Cohen’s d = .08). For ER, the main effect of Condition was significant (F(1, 27) = 6.33, p = .018, η2p = .19), with less errors in the Joint (M = 5.09%, SD = 3.34) than in the Baseline (M = 8.53%, SD = 5.11) condition. There was no significant main effect of Cueing (F(1, 27) = .33, p = .57, η2p = .01) and no significant interaction (F(1, 27) = 1.57, p = .22, η2p = .06).

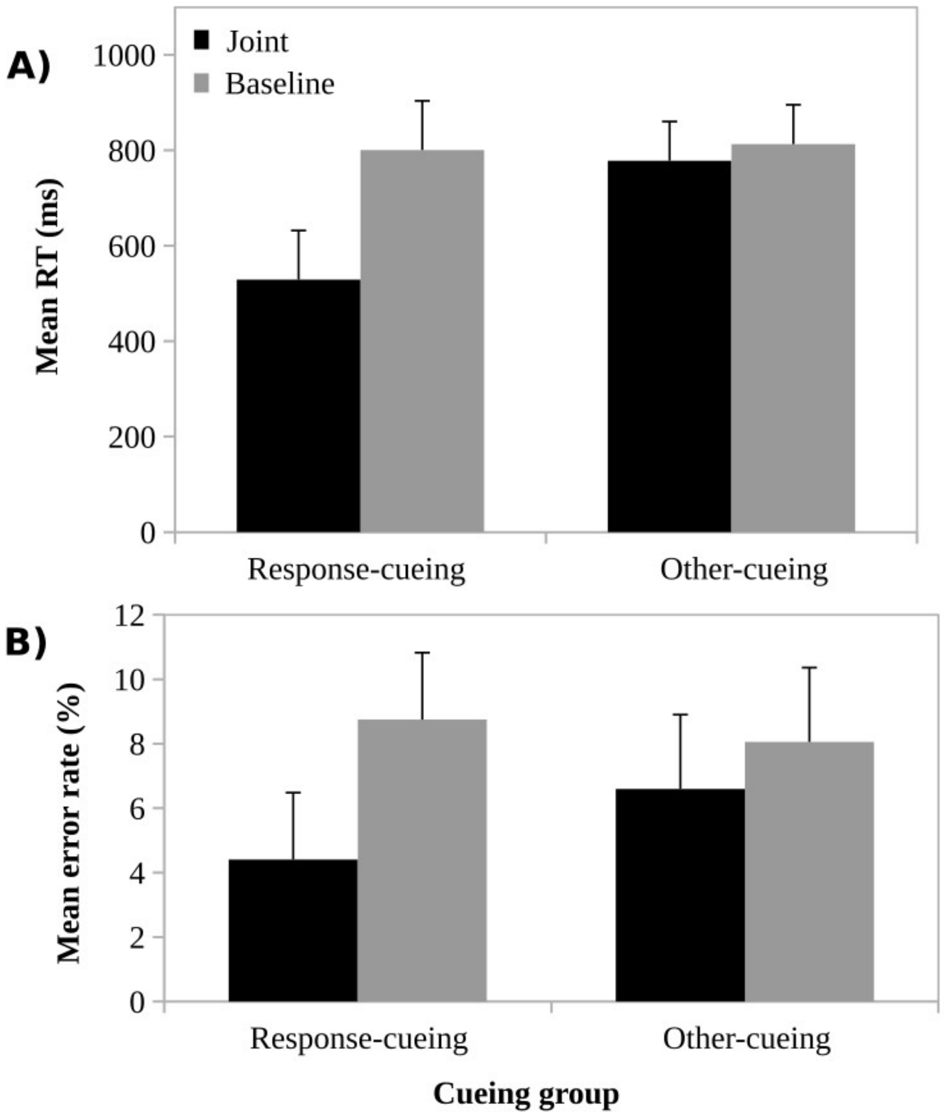


*Figure 2.11.* (A) Helpees' mean RT in Experiment 2/A. (B) Helpees' mean ER in Experiment 2/A. In both panels, black bars depict the means in the joint conditions while gray bars depict the means in the baseline conditions. Response-cueing refers to the group where helpers signaled the required responses for helpees (n = 20) and Other-cueing refers to the group where helpers signaled either response-change or spatial incompatibility (n = 9). Error bars show within-subject confidence intervals, calculated separately for the two groups of helpees (95% CI, based on Loftus & Masson, 1994).

**Relation between helper performance and helpee performance.** We tested if helpees were sensitive to helpers' errors. First we looked at the correlations between helpers' and helpees' ER, separately for the Response-cueing and Other-cueing groups. The correlation for the Response-cueing pairs was not significant: r(18) = .02, p = .93. For the Other-cueing group the correlation was numerically high but still not significantly different from 0, potentially due to the lack of power: r(7) = .64, p = .066.

Next, we performed the same helper-helpee error correlation analysis as in Exp. 1/A, for both groups of helpees (see Figure 2.12). For the Response-cueing group, there were n = 12 pairs where both helper and helpee committed errors in the joint condition (resulting in 132 pseudo pairs). As in Exp. 1/A, we performed 10^6 iterations for the permutation test with the difference of mean correlation coefficients as test statistic. We found a significant difference (estimated p < .001, Cohen's d = 1.36), with higher correlations in real pairs relative to pseudo pairs (M = .22, SD = .33 and M = .02, SD = .11, respectively). In the same analysis performed on the Other-cueing group (n = 9, with 72 pseudo pairs) we did not find a significant difference between real (M = -.04, SD = .03) and pseudo pairs (M = .01, SD = .08): estimated p = .104, Cohen's d = .77.

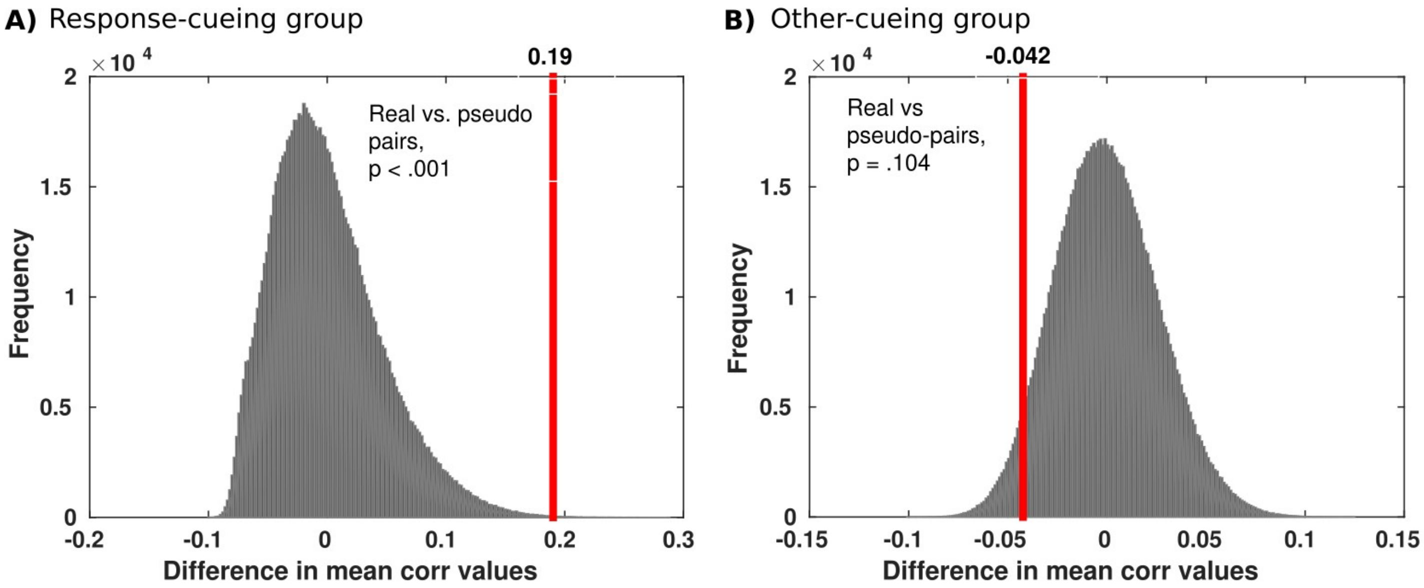


*Figure 2.12.* (A) Random permutation test results of helper-helpee error correlation, for helpees in the Response-cueing group. (B) Random permutation test results of helper-helpee error correlation, for helpees in the Other-cueing group with response-change or spatial incompatibility signaling. In both panels the x axis shows the difference of the mean correlation coefficient between real and pseudo pairs. The red vertical line is the actually observed difference while the gray histogram displays the results of the random permutations. The difference was significant for the Response-cueing group but failed to reach significance for the Other-cueing group.

We also investigated post-error slowing in helpees as a further measure of their sensitivity to helpers' errors. To this end, we selected errors in the same way as in Exp. 1/A (errors committed only by the helper with a +/- 1 trial window) and tested if helpees showed a reaction to helpers' errors using a one-way ANOVA on helpees' RT around the error trials. There were 11 helpees with suitable data in the Response-cueing group (number of errors: M = 3.90, SD = 6.18) and 8 in the Other-cueing group (M = 19.33, SD = 16.10). The results were non-significant for both groups (Response-cueing group: F(2, 20) = 0.68, p = .52, η2p = .06; Other-cueing group: F(2, 14) = 2.12, p = .16, η2p = .23).

As for helpers, we tested if there was a link between helpees' 'Reading the Mind in the Eyes' scores and their performance measures. There were no significant correlations (all ps > .1).

### 2.4.3. Discussion

**Cueing strategies.** In Exp. 2/A we tested if helpers' cueing strategies were determined by the cardinality and cultural relations between the cue set and the response set in Exp. 1/A. In Exp. 2/A we eliminated the cardinality match between cues and responses by using a cue set consisting of only an exclamation mark, while there were two responses. As the cues were reduced from left/right arrows to an exclamation mark, in Exp. 2/A we also broke up the cultural convention mapping cues to responses. As a result of this manipulation, the distribution of helpers' cueing strategies changed significantly from Exp. 1/A to Exp. 2/A, with a relative increase of helpers falling into the Other-cueing and a relative decrease of helpers in the Response-cueing category, proving that the saliency of cue-response mappings affects helpers' signaling behavior. However, the majority of helpers still used a Response-cueing category and most helpers in the Other-cueing group used cueing strategies that required them to effectively perform the task themselves (cueing response-change or spatial incompatibility, more on these cueing strategies later). Together, these results suggest that while salient cue-response mappings had some influence, helpers' cueing strategies aimed at providing maximally efficient information for the helper.

We also provided an opportunity for helpers in Exp. 2/A to split the effort with the helpee by cueing only one of two response-relevant features. Despite this change in the task only one person engaged in Singular-cueing (cueing only one of the response relevant stimulus features), signaling color in this case. This suggests that the lack of Location-cueing by participants in Exp. 1/A was not due to the lack of suitable strategies, further supporting that helpers are indeed better described by the efficiency account than by the split-effort model.

**Response-cueing group.** Helpers in the Response-cueing group traded speed for accuracy, providing valid cues for helpees in return for larger RT on their side. This is a robust result present both in Exp. 1/A and Exp. 2/A despite the differences in cognitive processing demands between the two experiments. In both cases, the low error rates suggested a floor effect in the joint condition, supporting the idea that helping was as efficient as possible irrespective of task complexity. Helpers' RT and ER results support the efficiency account.

Helpees in the Response-cueing group enjoyed the benefit of cueing (lower RT and ER relative to baseline) and were also sensitive to helpers' errors, often committing errors in the same trials as evidenced by the helper-helpee error correlation. However, helpees still gained less speed than suggested by the efficiency account, as the average RT relative to cue onset was around 950 ms, far greater than needed for acting solely on the cues. In fact, helpees' RT cueing gain was not significantly different across Exp. 1/A and 2/A ($p > .1$). This suggests that helpees let the cues bias their response selection but – at least on average – still processed the stimuli themselves. Importantly, this also means that the change in task difficulty from Exp. 1/A to 2/A did not result in helpees' total reliance on cues. Accordingly, our results regarding helpees in Exp. 1/A were not due to the relatively simple nature of the task. As in Exp. 1/A helpees' results support the split-effort and lack-of-trust models.

**Other-cueing group.** The post hoc analysis of cueing strategies in the Other-cueing group revealed that nine helpers employed strategies based on the cultural meaning of the exclamation mark. Seven participants signaled response-change, that is, sent the exclamation mark as cue when the required response was different from the one in the previous trial. Two participants signaled spatial incompatibility, cueing their partners when the stimulus was displayed on the opposite side relative to the required response. In both cases, helpers cued the trials requiring more attention and effort (as overcoming perseverance and spatial interference require cognitive effort). They did so in ways corresponding to the everyday ‘Alert’/’Attention’ meanings of the exclamation mark as if shouting at their partners in trials with higher processing demands.

At first glance the response-change and spatial incompatibility cueing strategies support the assumption that cue-response mappings are biased by cultural conventions. However, the nine helpers in the Other-cueing group who signaled response-change or spatial incompatibility also had to take the required responses into account. In order to select the cue for signaling response-change or spatial incompatibility the helper has to compute the response needed from the helpee and then needs one extra inferential step (comparison to last trial or to spatial position) for response selection. Accordingly, helpers cueing response-change or spatial incompatibility showed a larger effort spent in the joint condition than in the baseline condition. Interestingly though, the ER was also higher in the cueing phase than in the baseline. This lack of consistency in employing these strategies correctly might have been due to their difficulty or uncertainty about their efficiency. Either way, these helpers chose potentially efficient ways of signaling, conveying all the information necessary for the required response, in line with our definition of helping.

Despite the potentially helpful nature of response-change and spatial incompatibility cueing, helpees did not benefit from them as their RT and ER measures were not different from those in the baseline condition. Moreover, helpees’ errors in this group showed no dependence on helpers’ errors either. While this might have been due to the relatively low cue validity (~76%), post-experiment interviews suggested that helpees simply did not recognize these cueing strategies. While we cannot rule out the role of higher ER and the changes in the cultural meaning of the symbols, we suspect that the reason for not recognizing these signaling strategies was the discrepancy in the number of cues helpers and helpees were aware of. While helpers believed they either sent a cue or not, helpees always saw two distinct types of cues and had no reason to expect such elaborate ways of signaling as response-change or spatial incompatibility cueing. The production and comprehension of cues (signals) in this situation corresponds to a pure coordination game (that is, a Schelling game, see Mehta, Starmer and Sugden, 1994; Sugden and Zamarron, 2006) presented differently for the participants, denying them shared focal points.

While the case of response-change and spatial incompatibility cueing helpers nicely illustrates that helpers do not hesitate to exert effort in order to help, they do so to no avail as they do not boost helpees' performance. Assuming that helpers had indirect knowledge about cue utilization via helpees' response times, their behavior is a peculiar case of needless perseveration. While we do not have any data speaking directly to this possibility, it is tempting to understand the behavior of these helpers as serving self-presentation functions. If they cannot help, their costly signaling at least represents them as motivated cooperators (e.g. Jordan, Hoffman, Nowak and Rand, 2016; Roberts, 1998).

Similar to Exp. 1, we intended to test if the partial cue utilization exhibited by helpees was a result of helpers' occasional errors. We again ran a control experiment with computer generated cues, as in Exp. 1/B.

## 2.5. Experiment 2/B

The same change as from Experiments 1/A to 1/B was also applied from 2/A to 2/B. Specifically, helpees performed the same task as before, but there was no helper. Instead, 100% valid cues were generated by computer and helpees were instructed accordingly. Our reasoning was the same as in Exp. 1/B. If helpees did not utilize the cues fully in Exp. 2/A because of the effect of errors on social-communicative mechanisms aimed at evaluating partner reliability, we should see a performance gain in Exp. 2/B relative to 2/A.

### 2.5.1. Methods

**Participants.** Twenty (13 female, 1 left-handed) participants performed the tasks in Experiment 2/B. Their age range was 19-28 (M = 23.05 years, SD = 2.59). They all had normal or corrected-to-normal vision. We recruited and compensated participants the same way as in previous experiments. They all gave written consent.

**Stimuli and apparatus.** We used the same stimuli and apparatus as in Exp. 2/A.

**Procedure.** The procedure was the same as for helpees in Experiment 2/A. However, there was no helper and the cues in the second phase of the experiment were generated by a computer (as in standard cueing tasks). All participants were assigned the role of helpee and performed the task individually. The instruction reflected this change and participants were told that the computer would provide a helpful cue in each trial. Cues were 100% valid, pre-specifying the correct response side with an ‘X’ or ‘!’ symbol.

### 2.5.2. Results

First, we tested the effect of cueing in Exp. 2/B separately, by comparing RT and ER across conditions. There was a significant RT difference between the cueing and baseline conditions: t(19) = 5.27, p < .001, Cohen’s d = .1.18, with larger RTs in the baseline condition (M = 772 ms, SD = 111) than in the cueing condition (M = 379 ms, SD = 304). The difference was also significant for ER: t(19) = 2.16, p = .044, Cohen’s d = .48. ER were higher in the baseline condition (M = 6.75%, SD = 5.45) than in the cueing condition (M = 4.13%, SD = 3.52).

Then we compared helpees’ performance measures across Exp. 2/B and 2/A (using only data from Response-cueing pairs) in a 2x2 mixed design ANOVA with the between factor Cue Source (Human helpee vs. Computer) and the within factor Condition (Joint/Cueing vs. Baseline). The results are shown in Figure 2.13. For RT, the main effect of Cue Source was not significant (F(1, 38) = 2.56, p = .12, η2p = .06), but we found a significant main effect of Condition: F(1, 38) = 42.44, p < .001, η2p = .53, with slower responses in the Baseline (M = 786 ms, SD = 134) than in the Joint/Cueing condition (M = 454 ms, SD = 316). We found no significant interaction (F(1, 38) = 1.43, p = .24, η2p = .04). The results were similar with ER: there was no significant main effect of Cue Source (F(1, 38) = 1.12, p = .30, η2p = .03) but we found the main effect of Condition again (F(1, 38) = 14.09, p = .001, η2p = .27), with smaller ER in the Joint/Cueing condition (M = 4.27%, SD = 3.36) than in the Baseline condition (M = 7.75%, SD = 5.39). The interaction was not significant (F(1, 38) = .86, p = .36, η2p = .02).

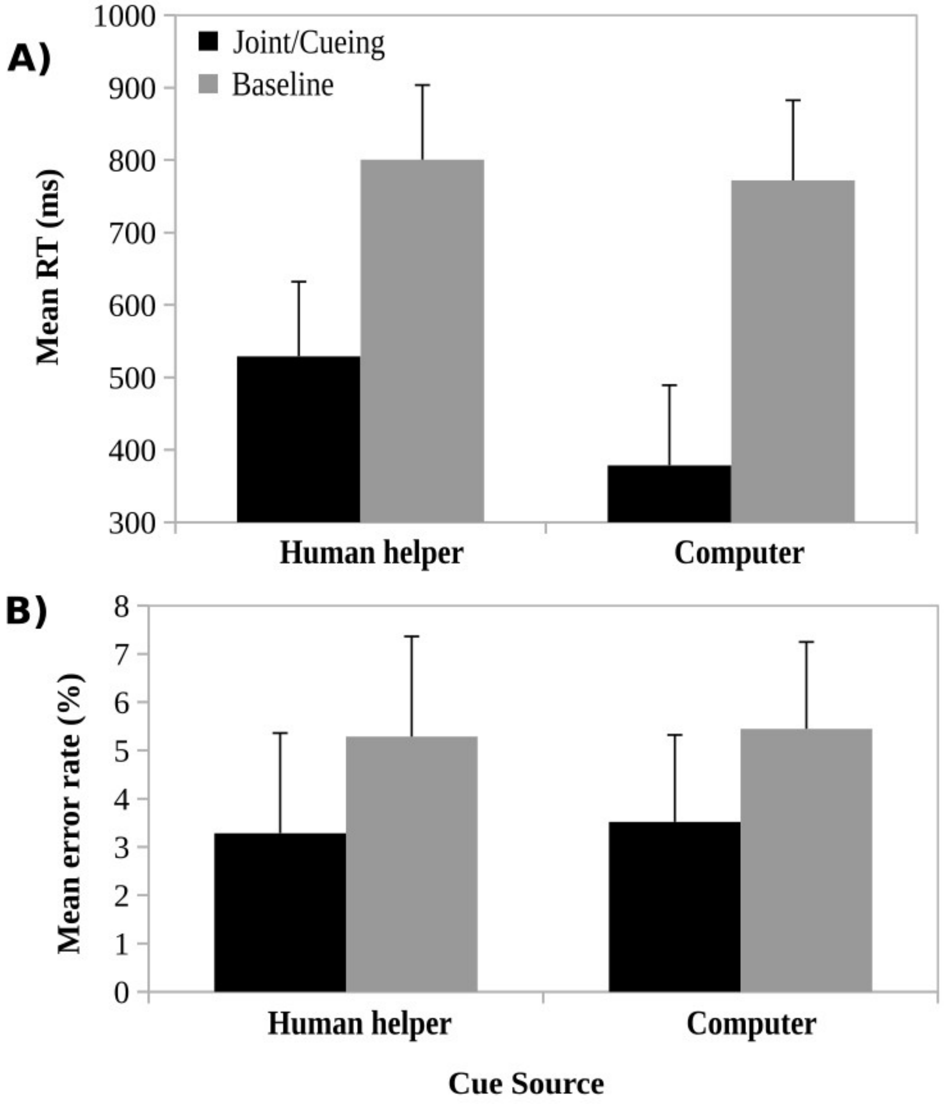


*Figure 2.13*. (A) Helpees' mean response times in Experiment 2/A and 2/B. (B) Helpers' mean ER in Experiment 2/A and 2/B. In both panels, black bars depict the joint/cueing conditions and gray bars depict the baseline conditions. In the joint condition of Exp. 2/B, the cues were computer generated and helpees were instructed accordingly. The baseline condition was the same as in Exp. 2/A, individual performance of the task without cues. Error bars show within-subject confidence intervals, separately for Exp. 2/A and 2/B (95% CI, based on Loftus & Masson, 1994).

### 2.5.3. Discussion

Just as in Exp. 1, we found no effect when contrasting mean RT and ER in the two experiments with cues provided by humans or computers, corroborating that helpees' reluctance to rely on the cues in Exp. 2/A was not due to errors committed by the helpers. In order to gain a better understanding of how helpees utilized the cues, we further analyzed helpees' RT data, using the data from all experiments.

## 2.6. Dynamics of cue utilization

The averaged performance measures showed that helpees in response-cueing pairs use the cues for selecting the correct response without relying on them completely. This finding held both for a simple and a more complex task, irrespective of the source of the cue (human helper or computer). To further investigate how helpees used cues over the course of the helping phase of the experiment we analyzed the time series of their responses. We asked (1) whether participants had different initial expectations about cue validity depending on the source of the cue and (2) whether the source of the cue had an effect on how cue utilization developed from earlier to later trials.

### 2.6.1. Initial expectations

It is important to realize that helpees could only utilize the cues after they understood which aspect of the task cues referred to. Thus, helpees may have needed several trials to discover the mapping between cues and cued properties/behaviors by monitoring the cues and testing plausible mappings consecutively, starting with the mapping deemed as most probable. The further the actual mapping was from helpees' initial expectation, the longer the identification of the correct mapping should take. Because testing different mappings takes time, cueing benefits (operationalized as a drop in RT) should occur only later if more mappings need to be tested. By measuring the rate of helpees' RT change in the early trials as a proxy for the distance between initial expectations and Response-cueing, we tested if helpees' expectations differed across experiments where human helpers or computers provided the cues. If helpees expected human partners, but not computers, to behave efficiently and, thus, to signal responses from the beginning, we should find a larger negative rate of RT change for experiments with human helpers than for their counterparts with computer generated cueing. Alternatively, if helpees expect more efficient cueing from computers, the rate of RT change should be more pronounced in the experiments where computers provided the cues.

**Results.** For the analyses of RT time series we relied on the raw data, because outliers could reflect searching for a correct mapping or evaluating cue validity. All helpees from the Response-cueing groups in the four experiments were included (*n* = 82).

Regarding initial expectations, we focused on the first ten trials. The number ten was chosen arbitrarily but the results were similar if more or fewer trials were included (see Appendix A, p. 229 for additional analysis). For each helpee we calculated the linear slope and performed a 2x2 factorial ANOVA on the coefficients with between factors Task (Exp. 1 vs. Exp. 2) and Cue Source (Human helper vs. Computer). The results are displayed in Figure 14. The main effect of Task was not significant ($F(1, 78) = 1.50$, $p = .22$, $\eta^2_p = .02$) but there was a significant main effect of Cue Source: $F(1, 78) = 6.73$, $p = .011$, $\eta^2_p = .08$. The mean slope value was smaller in Exp. 1/A and 2/A ($M = -31.54$, $SD = 41.19$) where humans provided the cues, than in Exp. 1/B and 2/B ($M = 10.62$, $SD = 31.15$) where computers provided the cues. There was no significant interaction ($F(1, 78) = .55$, $p = .46$, $\eta^2_p < .01$).

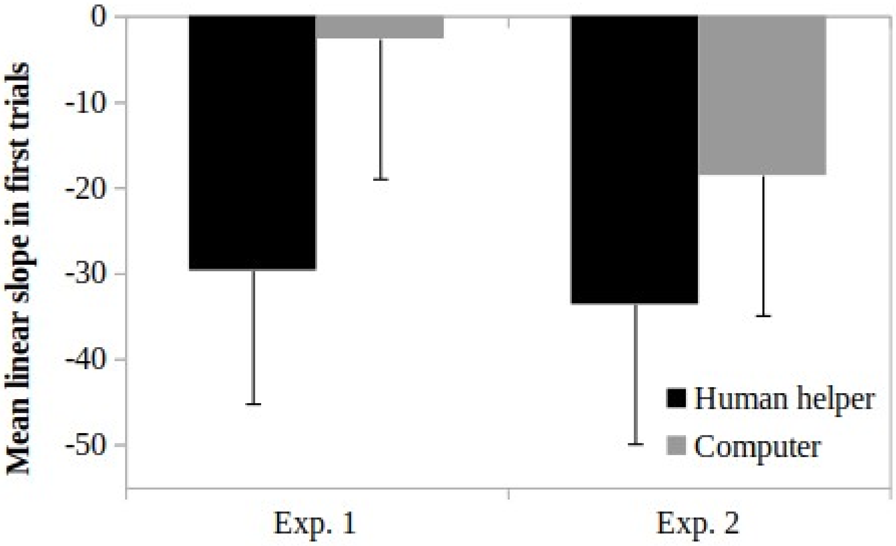


*Figure 2.14.* Mean linear slopes of helpees' RT data in the first 10 trials. All helpees from the Response-cueing groups in the four experiments (*n* = 82) were included in this analysis (Exp. 1/A: *n* = 22; Exp. 1/B: *n* = 20; Exp. 2/A: *n* = 20; Exp. 2/B: *n* = 20). The black bars depict data from the experiments with human helpers. |Gray bars depict data from experiments with computer generated cues. Error bars show 95% CIs.

### 2.6.2. Cue utilization

After the cue-response mapping was verified, helpees needed to evaluate cue validity and adjust their reliance on cues accordingly. One possibility is that the time course of cue utilization is not different from any other cueing task with gradual speedup where the learning effect in the RT time series is best captured by the standard "law of practice", that is, by a power function (Newell & Rosenbloom, 1981; but see Heathcote, Brown & Mewhort, 2000). Another possibility is that identification of the cue-response mapping opens up the possibility of responding purely on the basis of the cues, resulting in a new way of performing the task. Such a change is similar to an insight-moment in a problem solving task producing a qualitatively new solution (e.g. Knoblich, Ohlsson, Haider & Rhenius, 1999;

Rubin, Nakayama & Shapley, 1997). Because changing the solution (that is, helpees responding on the basis of the cues instead of identifying the stimulus) would mean switching to an easier task, a logistic (sigmoid) trend would capture such an insight-type change in terms of RT time series (Haider and Frensch, 1996; Haider and Frensch, 2002). There are at least two reasons to expect more people showing a stepwise RT change in the computer cueing experiments than in the ones with human helpers. First, human helpers occasionally committed errors, probably resulting in helpees dedicating more effort to evaluating cue validity. This would result in a gradual change in RTs rather than a sudden change. Second, because of their knowledge of how computers work participants may expect computer generated cues to be less variable and more rule-bound than cues provided by humans. Thus, they may more readily switch to relying on these cues once a mapping has been established.

**Results.** To test for the presence of specific trends in helpees' RT time series, we fitted two models using custom functions relying on the built-in fminsearch MATLAB function. To test for the abrupt, insight-type change, a logistic function was used (a sigmoid curve with the general form $f(t)=L/(1+e^{-k(t-t0)})$, where $t$ is the trial number). The model corresponding to the law of practice was the standard power law (a learning curve with the general form $f(t)=At^{B}+C$).

We categorized each helpee based on the goodness of the two fits, using adjusted $R^2$ as a general goodness of fit measure. If both $R^2$ values were below 0.1, there was no considerable change over time as captured by our functions and the helpee was excluded from this analysis (excluded $n$ = 29 across the four experiments). If at least for one fit $R^2$ was above 0.1 for a helpee, we labeled that participant with the function resulting in the best fit. See Figure 2.15 and 1.16 for examples of helpees falling into the power law and logistic categories. We summed the results of the categorization according to the different cue sources (Human helper vs. Computer). Table 1.2 displays the number of helpees falling into each function category. When helpees received computer generated cues they were more likely to show a better fit for the logistic functions whereas helpees receiving human generated cues were equally likely to show a better fit for logistic and power. A chi-square test showed that helpees receiving cues from the computer were more likely to show a better fit for logistic than helpees receiving cues from humans: $X^2(1) = 4.44$, $p = .035$. Note that the same results were obtained without excluding participants, simply categorizing them based on the better fit: $X^2(1) = 3.94$, $p = .047$.

### 2.6.3. Initial expectations and cue utilization

The initial rate of RT change and the categorization based on the trends in RT time series were not independent. A time series with less RT change in early trials more likely to have a better fit with a logistic function than with a power function. Accordingly, we also found a significant difference between groups in terms of the linear slope of early RT data between helpees falling into the logistic and power function categories: $t(80) = 2.04$, $p = .044$, Cohen’s $d = .45$. The mean slope was closer to zero in the case of computer cueing ($M = -12.74$, $SD = 34.68$) compared to experiments with human helpers ($M = -29.53$, $SD = 39.42$).

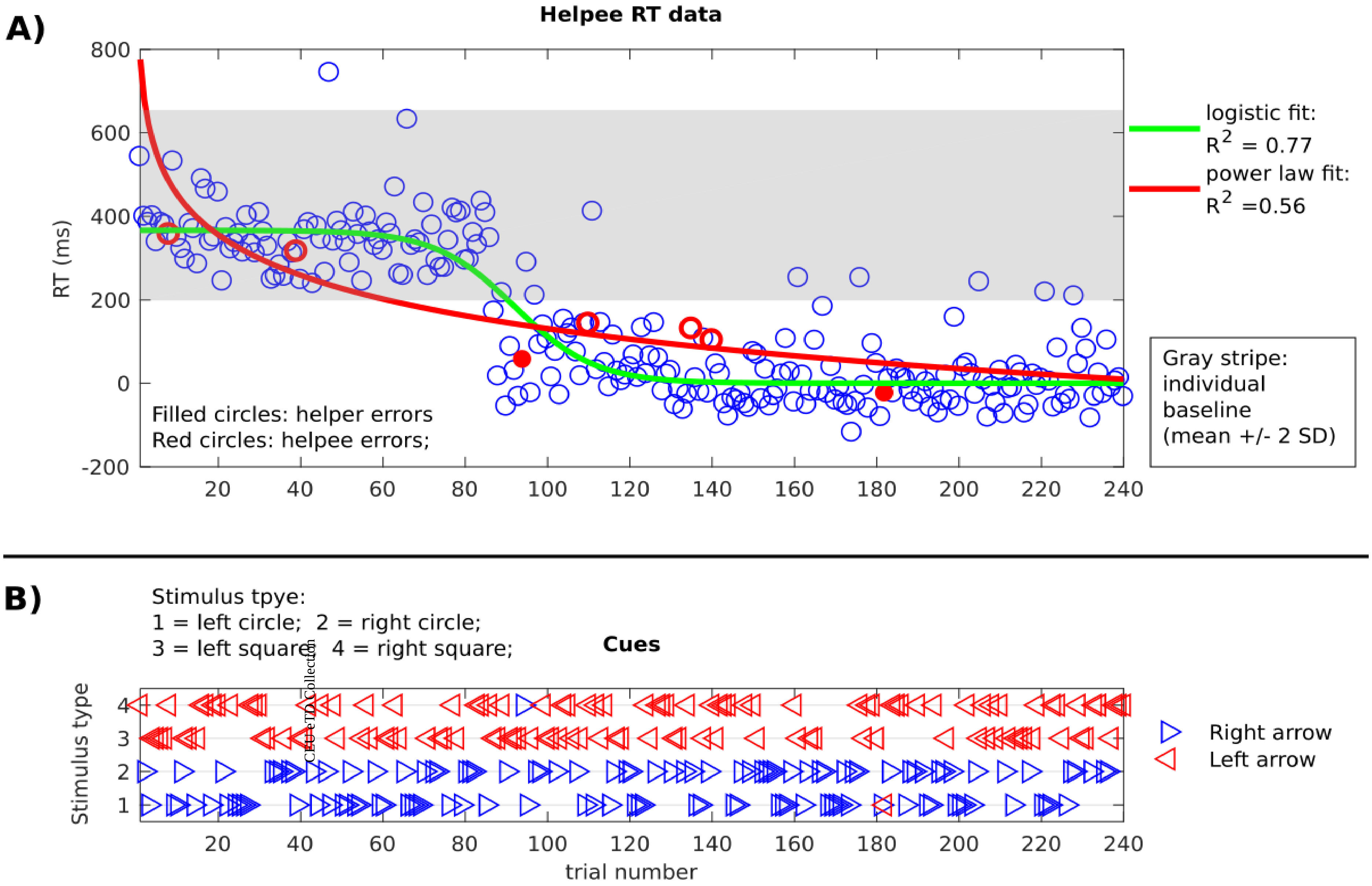

A)
Helpee RT data
800
600
400
200
0
-200
RT (ms)
20
40
60
80
100
120
140
160
180
200
220
240
logistic fit:
R$^2$ = 0.77
power law fit:
R$^2$ =0.56
Gray stripe:
individual
baseline
(mean +/- 2 SD)
Filled circles: helper errors
Red circles: helpee errors;
B)
Stimulus tpye:
1 = left circle; 2 = right circle;
3 = left square; 4 = right square;
Cues
Stimulus type
4
3
2
1
20
40
60
80
100
120
140
160
180
200
220
240
trial number
Right arrow
Left arrow

*Figure 2.15.* (A) Example data from Experiment 1/A where a logistic function provided the better fit to the helpee's RT time series. The gray horizontal stripe shows the range of this helpees' baseline data (M +/- 2 SD). Filled circles depict the helper's errors (invalid cues) while red circles correspond to helpee's errors. Note the change in RT around trial no. 90. The RT range of subsequent trials suggests that the helpee switched to responses based solely on the cues. Such a change is better modeled by the logistic function (green line) and we see this reflected in the R2 values on the right. (B) The cues produced by the helper in the same pair. In each trial, one of four possible stimulus was displayed, the values on the Y axis correspond to these stimulus types. The left or right arrows indicate helper's response and also the cue.

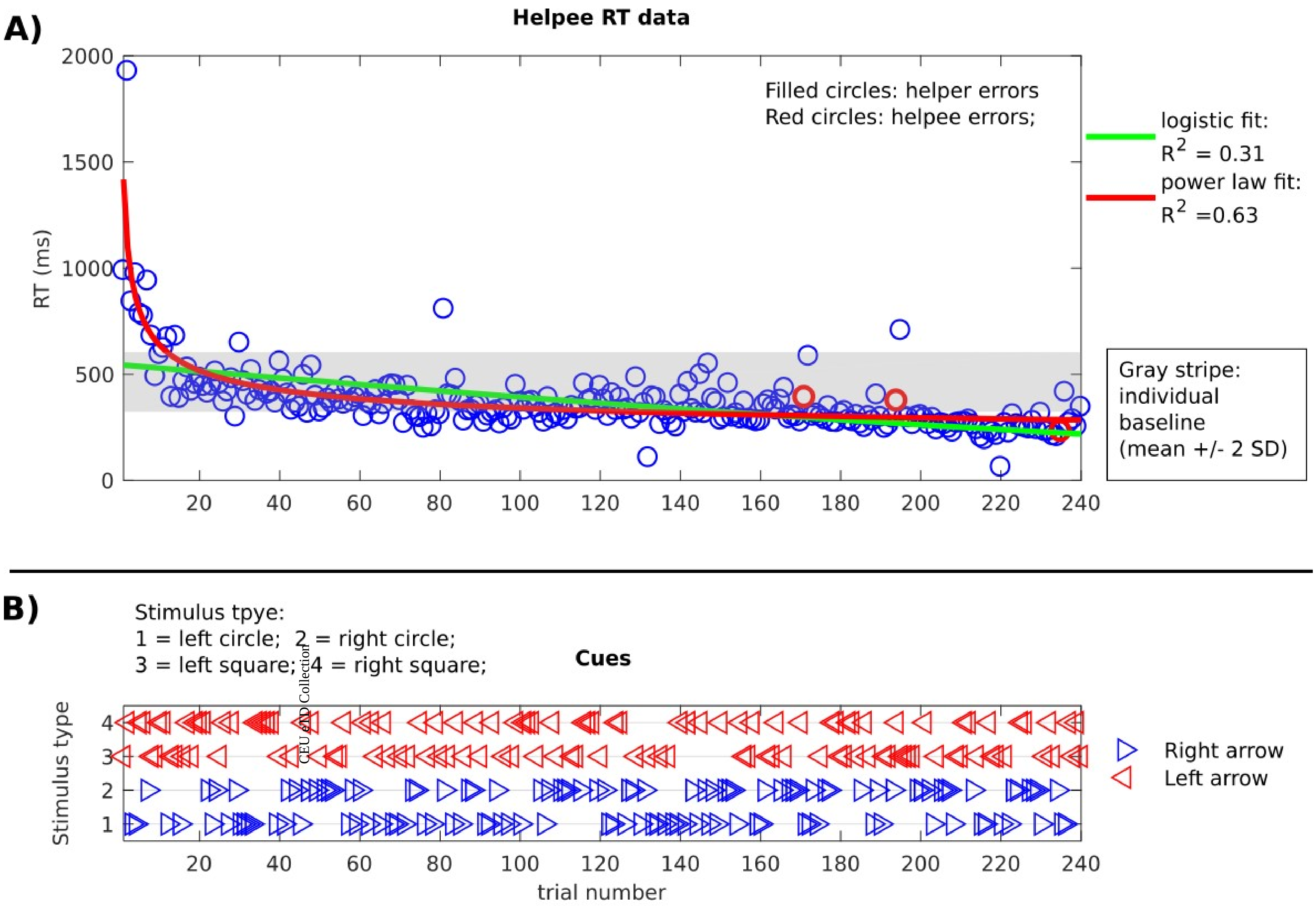

A)
Helpee RT data
Filled circles: helper errors
Red circles: helpee errors;
logistic fit:
$R^2$ = 0.31
power law fit:
$R^2$ =0.63
Gray stripe:
individual
baseline
(mean +/- 2 SD)
RT (ms)
2000
1500
1000
500
0
20
40
60
80
100
120
140
160
180
200
220
240
B)
Stimulus tpye:
1 = left circle; 2 = right circle;
3 = left square; 4 = right square;
Cues
Stimulus type
4
3
2
1
Right arrow
Left arrow
trial number

*Figure 2.16.* (A) Example data from Experiment 1/A where a power function provided a better fit for the helpee's RT time series. The gray horizontal stripe shows the range of this helpees' baseline data (*M* +/- 2 *SD*). Filled circles depict helper's errors (invalid cues) while red circles correspond to helpee's errors. Note the gradual change in RT, better captured by the power function (red line). (B) The cues produced by the helper in the same pair. In each trial, one of four possible stimuli was displayed, the values on the Y axis correspond to these stimulus types. The left or right arrows indicate helper's response and also the cue.

Table 1.2

*Categorization of helpees according to the trends in their RT time series.*

| | Power function fit | Logistic function fit | Total |
|---|---|---|---|
| Exp. 1/A and 2/A (Human cueing) | 13 | 13 | 26 |
| Exp. 1/B and 2/B (Computer cueing) | 6 | 21 | 27 |

Notes: *Categorization was based on goodness-of-fit (adjusted* $R^2$*)*

### 2.6.4. Discussion

The results show that helpees receiving cues from humans started to utilize the cues faster than helpees receiving cues from a computer, supporting the idea that helpees had different initial expectations in each of the cases. More specifically, we suspect that cues from a human helper were expected to correspond to response-cueing right from the first trials, while cues from a computer were initially treated more cautiously, indicating that helpees had less of a prior conviction about the meaning of the cues. A potential reason behind this discrepancy is that humans are routinely expected to use signaling based on inferences on the co-actor’s goals (relevance-driven communication, Sperber & Wilson, 1995), creating strong expectations based on common ground (Clark, 1996) or focal points (Sugden, 1995). For example, humans have been shown to flexibly change established signal-to-meaning mappings in a novel communicative situation based on changes in common ground, leading to successful communication overall (Misyak, Noguchi & Chater, 2016). The same considerations do not seem to apply to computer-generated cues.

In terms of the overall trends in RT time series, an equal number of helpees fell into the logistic and the power function groups in the experiments with human cueing, but significantly more helpees were better characterized with the logistic function in the computer cueing experiments. This shift suggests that helpees were more prone to a sudden change in their behavior when cues were generated by a computer, corresponding to a clear change in cue utilization at a given point during the task. Together with the results on initial expectations, the prevalence of logistic fits supports an evidence-accumulating view of cue utilization in the case of computer cueing. According to this view helpees lacked a clear expectation about the meaning of cues in the computer cueing experiments, requiring a longer initial period to establish the cue-response mapping.

## 2.7. General discussion

In the present study we defined helping as treating someone else’s goal as one’s own and set out to study helping behavior from the perspectives of both the helper’s and the helpee’s efficiency. In our investigation we focused on situations where helping was realized via communication and the helpee did not necessarily require help but could choose to resort to individual performance instead. Our aim was to establish (1) if helpers indeed prioritize helpees’ goals, even at a cost to themselves; and (2) if helpees exploit helpers’ contributions for their own performance. To this end, we conducted four experiments with a two-person sequential cueing task where a helper's response could potentially serve as a cue for a helpee's performance. Cues in this task served as one-way communication signals that could have a direct effect on helpees' performance in a reaction-time task. Helpers were free in mapping the cues to different aspects of the task, that is, to select mappings more-or-less in line with helpees' goals.

Regarding helpers' performance we found that they helped efficiently and in line with our initial definition. Their actions improved helpees' performance even when such improvement required greater effort. Two lines of evidence support this notion. First, helpers preferred response patterns (cues from helpees' viewpoint) that conveyed maximally task-relevant information over alternatives that would have led to a sharing of the task. While in Exp. 1/A this preference could be explained by the salient mapping between cues and responses, the same preference for efficient cueing was found in Exp. 2/A where the cues had an arbitrary relationship to the helpee's responses and were not even matched in their cardinality. Importantly, the selected cueing strategies required helpers to completely perform helpees' task. Because helpers chose to make this high investment, we could compare their performance in the joint condition to their performance when performing the original task alone (baseline conditions). In this second line of evidence helpers traded speed for accuracy in the joint conditions in both Exp. 1/A and 2/A. By doing so, they achieved close-to-zero error rates, providing reliable cues for helpees to follow. In other words, they did everything to ensure that helpees' could improve their performance.

Regarding helpees' performance our results show a different picture. On the one hand, the performance measures showed a marked improvement in the joint conditions relative to baseline, proving that helpees utilized the help (cues) they received. This effect was robust and present in all four experiments supporting the idea that helpees' use of cues was somewhat effective. Furthermore, helpees' errors tended to occur in the same trials as helpers', demonstrating that helpees at least partly relied on the helpers' cues. On the other hand, helpees avoided complete reliance on cues. First, their RT measures showed that their gains remained below the theoretical possible gain and did not scale with task difficulty (no significant differences across Exp. 1/A and 2/A). This latter point is important as it suggests that helpees did not use the cues to bypass the crucial stimulus identification step that could have provided a greater gain in Exp. 2. Second, helpees showed a spatial compatibility effect both in terms of RT and ER in the joint/cueing conditions of Exp 1. The effect was as large as in the baseline (even larger for RT in Exp. 1/B). Thus helpees processed the irrelevant spatial feature of the stimulus before responding. Third, although helpers' near-floor error rates made this analysis difficult, we found evidence that helpees occasionally corrected for helpers' errors in about half of the pairs in Exp. 1/A (n = 9). This indicated that many helpees identified and processed the stimuli themselves. Thus, helpees mainly used cues to support their response selection, but continued to process stimuli themselves. They were effectively verifying helpers' performance. This led to a suboptimal distribution of work, where the processes involved in response selection across helper and helpee were doubled unnecessarily. To the degree that our findings are generalizable, this result predicts suboptimal performance in many helping situations, especially in complex tasks where helpees' choices are either to follow blindly or to go through costly cognitive steps (e.g. think of getting help in solving a Rubik-cube).

Why did helpees not utilize helpers' cues more effectively? In line with a lack-of-trust account, this inefficiency on helpees' side could be the result of a need to monitor cue validity as a proxy for monitoring helpers' commitment or capabilities. However, the results of Exp. 1/B and 2/B showed that eliminating human helpers from the setup and using computer generated cues did not increase the cueing gain displayed by helpees. There were also no cueing errors in Exp. 1/B and 2/B. Thus, a lack-of-trust account could only explain the results by postulating a general mistrust in the cues used in our task, independent of the presence of invalid cues and task difficulty. Furthermore, when we analyzed the dynamics of cue utilization we found that if anything, helpees had more accurate expectations regarding cueing in experiments with human helpers, developing trust (that is, lowering their RT) quicker than with computer cueing.

Alternatively, helpees' behavior could be described with the split-effort model, according to which helpees seek a distribution of effort between helper and helpee in all situations, similar to general fairness considerations (Baumard et al., 2013; Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999). One prediction of the split-effort model is that helpees should avoid absolute reliance on helpers' communication when such communication would effectively relieve them from the task at hand and render them to followers of helpers' actions, just as people refuse unfair offers in an ultimatum game, even at a cost to themselves. Instead, helpees should perform the task themselves to some degree, just as we observe it in our experiments. To the degree that such a refusal is a general tendency and is independent of the source of unfair help (human or computer), the split-effort model can indeed explain helpees' results.

We described helpers as following the efficiency principle and helpees whose behavior is better captured by the split-effort model. We demonstrated that helpers do not try to share effort, instead they prefer to signal the required response, short-circuiting helpees' task, while helpees prefer to perform the task at least partially. This account of our results predicts tensions between helpers and helpees. Based on the efficiency and split-effort models, both the helper and the helpee could justifiably perceive their own efforts as efficient and necessary while the other's effort as inefficient and unnecessary. Consequentially, they could attribute positive and negative outcomes to themselves and to the other, respectively, leading to conflicts, negative affect, and, ultimately, to the collapse of the joint action. In line with these predictions, in situations where helpees have no control over the help they receive, helpees often experience negative affect and physiological stress (e.g. Deelstra et al., 2003; DePaulo et al., 1983).

A built-in aspect of our experiments was the need to develop a minimalistic communicative system on the basis of the symbols available for cueing and the known structure of the task. In that sense, our task tapped into the processes involved in building a novel communication system based on common ground (Clark, 1996). Unlike in many experiments on the emergence of non-conventional communication systems (e.g. Galantucci, 2005; Schmitz et al., 2018; Scott-Phillips, Kirby & Ritchie, 2009; Vesper et al., 2017b), the symbols we provided had well-established cultural meanings (left-right directions for the arrows in Exp. 1, attention capture for the exclamation mark in Exp. 2). Instead, we underspecified the task aspect the symbols had to refer to. This created a situation similar to a Schelling-type coordination game: participants had to arrive at a mapping between symbols and task aspects that they would both understand. While mapping arrows to the required responses in Exp. 1/A was simple as their relation was of iconic nature (Burks, 1949), such a clear focal point (Sugden, 1995; Sugden and Zamarrón, 2006) was missing in Exp. 2/A, with the added difficulty that only the helper was informed of a cardinality change. As a result, communication broke down relative to Exp. 1/A as evidenced by more pairs falling into the Other-cueing group.

Crucially, this break-down was due to participants' inability to solve the coordination problem, and not due to helpers abandoning the efficiency principle. The majority of helpers falling into the Other-cueing group followed unexpected ways of cueing, namely cueing response-change or incompatibility. Both could have made – in principle –helpees' task easier because they relied on the cultural meaning of the exclamation mark as a cue. The emergence of these strategies is a great demonstration of humans' flexibility in finding ways of communicating (e.g. Misyak et al., 2016).

By identifying sigmoid tendencies in helpees' RT time series corresponding to sudden changes in task performance and linking it to putative differences in expectations about the meaning of cues, our task also highlights the link between problem solving and communication games. The tasks used in studies addressing the emergence of communication systems (e.g. the tacit communication game in Newman-Norlund et al., 2009; the embodied communication game in Scott-Phillips et al., 2009; the figure-letter matching game of Selten & Warglien, 2007; etc) can be thought of as logical puzzles solved by theory-of-mind inferences (see e.g. de Weerd, Verbrugge & Verheij, 2015). Often a mutual, matching insight is required from the participants to solve the task, and such insight can be tracked by looking at its characteristic effects in performance (Haide and Frensch, 1996).

The results of our study are significant in at least two ways. First, we presented a novel approach to studying helping behavior that focuses on the process of helping instead of the motivational aspect. While studying the reasons why people help each other is obviously very important, little is known about how helping takes place and what cognitive and motor processes are involved in efficient helping. By defining helping as a form of joint action with absolute goal contagion from helpee to helper (cf. Paulus, 2014), we link helping behavior to the literature on coordination (e.g. Knoblich et al., 2011; Vesper et al., 2017a) and provide a framework for studying how social and cognitive factors might contribute to efficient help.

Second, our results show novel effects in how people coordinate their actions. By creating a helping framework we elicited a disregard to personal effort in helpers that proved to be independent of both task difficulty and complexity in signaling. While in many forms of joint action we observe some form of effort coordination, in the present example this was not the case. It would be interesting to study further if this effect was driven by an avoidance of communicating erroneous information (e.g. due to reputation concerns) or by a drive for precision, as these alternatives would have different predictions in a signal detection task similar to the one used in Bahrami et al. (2010). If helpers' behavior is driven by error-avoidance, it might even lead to suboptimal cues in certain cases.

The helping framework also elicited a discrepancy of efficiency in helpers and helpees, as helpees did not fully utilize helpers' communication. While helpees' results were mostly in line with the split-effort model, the role of different factors behind the split-effort account remains unclear. For example, we drew from two bodies of evidence to introduce the split-effort model, namely social psychology experiments and behavioral economics, that offer very different perspectives in terms of underlying mechanisms. Further experiments are needed to explore the factors eliciting such avoidance of relying on someone else's performance (e.g. cue validity, discrepancy in helper's and helpee's capabilities).

# Chapter Three. Alignment in verbal interaction as a function of interactivity and individual goals

## 3.1. Introduction

Interlocutors' speech gradually becomes more similar through the course of the linguistic interaction (Giles, Coupland and Coupland, 1992; Pickering and Garrod, 2004; Garrod and Pickering, 2004). This phenomenon, often termed alignment, has been shown to be present on multiple levels of the speech signal. According to the influential Interactive Alignment Model (IAM, Pickering and Garrod, 2004), such multi-level increase in similarity is functional, as it serves as the primary process matching interlocutors' underlying linguistic and situational representations. Given the putative key role of alignment, in the current paper we investigate two factors modulating its occurrence and strength: interactivity and individual goals. Our aim is to test two predictions of IAM, namely that (1) interactivity increases alignment and that (2) individual goals do not affect alignment. In what follows we first discuss IAM and the proposed processes behind alignment, and then review the questions of interactivity and individual goals in more detail.

### 3.1.1. Alignment and the IAM

Imagine yourself giving directions to a tourist on the streets of a city using a map, a situation studied in the well-known Map Task (e.g. Anderson et al., 1991). You have detailed knowledge of the street layout, transportation options and so on, and all these aspects are intertwined with autobiographical memories. Compared to you, the tourist has a rather abstract and incomplete representation of the city. In order to give directions to the tourist, the two of you need to develop a (partial) mapping between your respective representations, *aligning* them through verbal communication. Pickering and Garrod (2004) call this the alignment of situational models (Zwaan and Radvansky, 1988), that is, the matching of interlocutors' general understanding of the environment, events and further concepts relevant to the interaction. In IAM, situational alignment is treated as the general goal of communication, claiming that any other individual goal (e.g. the tourist needs to find the railway station) is achievable by first roughly equaling the situational models. In our example, successful communication would require that both parties agree on what is meant by the directions on the map, the current location, the distances represented, etc.

However, a direction-giver and a tourist would not only align at the conceptual level, but they would display convergence of other speech features too. It has been shown that people accommodate to others' prosodic features, including fundamental frequency (e.g. Gregory, 1990; Gregory and Webster, 1996), vocal intensity (e.g. Natale, 1975), speech rate (e.g. Levitan and Hirschberg, 2011; Street, 1984) and pauses (e.g. Edlund, Heldner and Hirschberg, 2009; Fusaroli and Tylén, 2016). Interlocutors' contributions also become more similar phonetically (e.g. Pardo, 2006), syntactically (e.g. Branigan, Pickering and Cleland, 2000; Weiner and Labov, 1983), lexically (e.g. Doyle, Yurovsky and Frank, 2016) and conceptually (e.g. Garrod and Anderson, 1987). A central tenet of IAM is that convergence on all levels of linguistic representation is important, because such alignment "percolates": alignment on one level strengthens alignment on other levels, eventually contributing to shared understanding and, thus, successful communication (for evidence, see e.g. Branigan et al., 2000, see also Mahowald, James, Futrell and Gibson, 2016).

As for the psychological process behind alignment, IAM originally claimed it is generated by automatic, effortless priming processes operating at multiple levels of linguistic representations, similar to covert imitation (Chartrand and Bargh, 1999; Heyes, 2011). In the direction-giving example above, an instance of such priming on the lexical level would be a tendency to repeat each other’s referring expressions used for certain landmarks, a process that has been demonstrated many times in experiments (e.g. Clark and Wilkes-Gibbs, 1986; Garrod and Doherty, 1994). This priming account presupposes the parity of representations across language production and comprehension, that is, that the representations activated during comprehension are also the representations used in production. Thus, if the activity of a certain representation is enhanced while listening to speech, it also remains more strongly activated during production, resulting in greater probability for its subsequent use - that is, in a priming effect.

From a broader theoretical perspective, such overlap of representations appears to be a special case of the perception-action links in common coding and related approaches (e.g. Prinz, 1997; Hommel, Musseler, Aschersleben and Prinz, 2001). Accordingly, newer versions of IAM have postulated (Garrod and Pickering, 2009; Pickering and Garrod, 2013, 2014; see also Gambi and Pickering, 2017) that priming effects (and thus, alignment) are the result of interwoven action and perception processes, with integrated speech production and comprehension being a special subset of a more general phenomenon. Moreover, they apply the idea that covert imitation involves motor processes emulating – and thus predicting - the perceived action(prediction-by-simulation, Elsner, D’Ausilio, Gredeback, Falck-Ytter and Fadiga, 2013 ; Csibra, 2008; Wilson and Knoblich, 2005). Speech comprehension is then defined as a form of action perception, and the processes behind production are also assumed to generate predictions of the incoming speech signal as emulators (similar to forward models in motor control, e.g. Wolpert and Ghahramani, 2000). While IAM argues that prediction-by-simulation is an important processing route facilitating comprehension, it is not described as an obligatory component of speech processing (cf. Menenti, Gierhart, Segaert and Hagoort, 2011).

Consider our direction-giving example again. If a local resident is about to specify a crucial left or right turn for the tourist, then - given the context - every other aspect of her utterance might be fully predictable ("And after walking along the canal there, you take a *left* at the statue."). As predictions could rely on a number of processes (for example on associative learning), the main claim here is that while listening to that sentence, the tourist relies on prediction-by-simulation (that is, predictions generated by her production system) as if she covertly imitated the local resident uttering that particular sentence. By doing so she could anticipate most words, their timings, prosodic features of the utterance, etc. - with the probable exception being the target word carrying new information ("*left")*.

What are the benefits of prediction-by-simulation? In the general case of action perception, simulation has been shown to facilitate perception by projecting the likely outcome slightly into the future and helping overcome noisy, missing or ambiguous sensory information, ultimately yielding more precise judgments (e.g. Knoblich and Flach, 2001; Springer, Parkinson and Prinz, 2014; Wilson and Knoblich, 2005). Similarly, in verbal interactions, such facilitation could help with noisy data, pre-focus attention to new information, and overall accelerate interlocutors' understanding of each other, speeding up successful cooperation. Furthermore, some form of prediction is needed for effective coordination during the dialogue itself as turn-taking behavior is considered too fast for sequential (comprehension - response preparation – production) processing (e.g. Levinson, 2016). There is evidence that prediction-by-simulation is involved in rapid turn-taking (Hadley, Novembre, Keller and Pickering, 2015) . Besides turn-taking, prediction-by-simulation is assumed to be the main process leading to alignment (Pickering and Garrod, 2013).

We do not review here the evidence supporting prediction-by-simulation in speech processing, but would like to note that the idea of prediction-by-simulation has been a component of other models of the language system as well (see e.g. Galantucci, Fowler and Turvey, 2006; Hickok, Houde and Rong, 2011; D'Ausilio et al., 2009).

### 3.1.2. The role of interactivity in alignment

In IAM, alignment is described in the context of back-and-forth, reciprocal communication, with the recurring claim that there is less and less alignment as an interaction deviates from unconstrained dialogue towards monologue (Pickering and Garrod, 2004, 2013). In the current paper we will refer to the key distinction between dialogues and monologues as *interactivity*, that is, the process of communicating partners *mutually influencing each other's contributions* while engaged in turn-taking. The general idea from IAM (see also Jaeger and Snider, 2013 ; Gallotti, Fairhurst and Frith, 2017) is that interactivity requires continuous mutual adaptation, which in turn is achieved by both interlocutors trying to predict each other. As their predictions are based on their own production systems, mutual prediction yields strong alignment on all levels of prediction and quick convergence on a shared understanding (cf. Friston and Frith, 2015).

However, the evidence for a positive link between interactivity and alignment is not conclusive. The account that interactivity predicts alignment is supported by early experiments on conceptual alignment. For example, Schober and Clark (1989; see also Wilkes-Gibbs and Clark, 1992) found that addressees in a picture matching task performed better compared to overhearers of the interaction. These results, however, can also be interpreted outside of IAM, as evidence for the benefit of backchannels and clarification questions in grounding (Clark and Brennan, 1991). More direct support comes from a priming study by Branigan, Pickering, McLean and Cleland (2007) who found that addressees displayed larger syntactic priming effects than side-participants (overhearers) in a subsequent speaker role. There is also evidence for the contrary though, as Schoot, Menenti, Hagoort and Segaert (2014) found no difference in syntactic priming between a communicative and non-communicative condition, neither in behavioral nor in neural (fMRI) data. To our knowledge, no other study investigated alignment or priming effects directly as a function of interactivity. Furthermore, the experiments by Branigan et al. (2007) and Schoot et al. (2014) focused only on syntax while the alignment claim covers all levels of linguistic representation, and they also employed tightly controlled laboratory studies far from the domain of naturalistic communication IAM is hypothesized to apply to.

There are good theoretical arguments both for the account that interactivity predicts alignment in general and for its alternative claiming the independence of these concepts. On one hand, verbal interactions might be considered a form of joint action (Clark, 1996; Garrod and Pickering, 2009) as the parties have a shared goal and coordinate their actions in order to achieve that goal (Sebanz et al., 2006; Sebanz and Knoblich, 2009). One of the main mechanisms identified in joint action research enabling fine-grained coordination is also prediction-by-simulation (Knoblich et al., 2011), a theoretically very similar process to the one argued by IAM. There is evidence from joint action experiments that prediction-by-simulation is stronger in back-and-forth, interactive situations than in non-interactive conditions. For example, in a series of EEG studies Kourtis and colleagues (Kourtis et al., 2010, 2013a, 2013b) have demonstrated that in a simple giver-receiver task participants simulated their partners’ actions (as evidenced by motor-related responses). Importantly, they showed weaker evidence for simulation towards the actions of a third-party with whom they had no history of interactions. Moreover, in Kourtis et al. (2013a) such motor simulation was correlated with increased temporal coordination over time. In a similar vein, Ménoret et al. (2014, see also Ménoret, Bourguignon and Hari, 2015) found that participants observing an actor moving an object showed larger mu suppression in an interactive than in a non-interactive condition, interpreted as motor simulation processes engaged more during the interaction. Furthermore, using a dual-EEG setup they could show that this simulation was reciprocal and present in the actor too. To the degree that these findings are generalizable to verbal interactions, they support the account that interactivity evokes stronger prediction-by-simulation processes and, as a result, predicts stronger alignment.

On the other hand, alignment might also be described as the result of automatic, covert imitation as originally proposed by IAM as well (Pickering and Garrod, 2004). As covert imitation is assumed to be a bottom-up process (Heyes, 2011), it should operate equally in interactive and non-interactive situations. Furthermore, for syntactic, lexical and semantic priming at least, alternative models have been put forward that construe alignment as an effect of implicit learning, spreading activation and cue-based memory retrieval processes (e.g. Chang, Dell and Bock, 2006; Cree, McRae and McNorgan, 1999; Reitter, Keller and Moore, 2011). These accounts do not make different predictions in interactive and non-interactive situations.

In our study we contrast these alternatives by manipulating interactivity in a naturalistic verbal interaction and measuring alignment between pair members at multiple paralinguistic and linguistic levels. If IAM is correct, interactivity should elicit stronger alignment at prosodic, syntactic, lexical and semantic levels, presumably driven by prediction-by-simulation as proposed by Pickering and Garrod (2013). Alternatively, if alignment is fully a result of bottom-up, automatic processes, we should find similar levels of alignment in interactive and non-interactive conditions.

### 3.1.3. The role of individual goals in alignment

Let us consider our direction-giving example again. In that scenario both parties have the same shared goal, namely that the tourist reaches her destination. Not all verbal interactions are built around a mutually desired outcome though. Intended end states might differ considerably across interlocutors even in everyday scenarios (e.g. bargaining, negotiation, competition, etc). How do such differences in individual goals affect alignment?

IAM (Pickering and Garrod, 2004; Garrod and Pickering, 2009) proposes that alignment of the situational models is a must for successful communication irrespective of particular individual goals. Intuitively, this makes a lot of sense - if the direction-giver in our example was to deceive the tourist by sending her in the wrong direction, they would still need to build-up a shared understanding of the situation first. From this point of view, individual goals are an extralinguistic component of communication, in the sense that they do not directly affect the linguistic processes enabling us to understand each other.

There is another line of argument to make though for the importance of shared goals. We know that automatic imitation of movements, body posture and gestures is positively linked with liking, perceived similarity and cooperation (e.g. Hove and Risen, 2009; see Chartrand and Lakin, 2013 for a review). Similar results have been obtained with linguistic alignment as well. For example, prosodic alignment has been linked with positive attitudes (Lee et al., 2010; De Looze, Scherer, Vaughan and Campbell, 2014; but see McGarva and Warner, 2003), syntactic alignment with initial impressions (Balcetis and Dale, 2005) and overall affective alignment (del Prado Martin and Du Bois, 2015). There is some evidence that the relationship between automatic imitation and social context is bidirectional, suggesting that certain features of the social interaction act as moderators of imitation, facilitating or attenuating its strength (Chartrand and Lakin, 2013). One of the inhibiting factors is related to people's personal goals. Several studies have manipulated whether interacting individuals agreed or disagreed on the topic under discussion (that is, were instructed with the same or different goals) and found that bodily synchrony (a form of automatic imitation) was reduced when the partners had different goals (Paxton and Dale, 2013, 2017). Similar results have been obtained regarding prosodic alignment (Abney, Paxton, Dale and Kello, 2014; Duran and Fusaroli, 2017; Riordan, Kreuz and Olney, 2014). For example, Riordan et al. (2014) manipulated agreement (disagreement versus neutral conditions) in text-to-text conversations and found that disagreement decreased alignment in terms of turn length. Based on these results one would expect that manipulating individual goals (same vs. different goals) would affect alignment at multiple levels. We will refer to this hypothesis as the facilitation/inhibition account.

Here we test the role of goal difference by manipulating the individual goals of communicating partners. IAM predicts no effect of the goal difference manipulation, while the facilitation/inhibition account predicts that shared goals would facilitate and different goals would inhibit alignment in general.

### 3.1.4. Alignment effects across levels

An important claim of IAM is that alignment at one linguistic level boosts alignment at other levels as well. This effect has mostly been demonstrated with syntax and word recurrence, with many studies reporting that lexical repetition increases syntactic priming effects (Mahowald et al., 2016; Fusaroli et al., 2017; but see also Healey, Purver and Howes, 2014). A number of studies attempted to measure alignment on levels other than syntax and word choices, either by including prosodic features or movement synchrony measures (De Looze et al., 2011; Levitan and Hirschberg, 2011; Louwerse, Dale, Bard and Jeuniaux, 2012; Bonin et al., 2013; Fusaroli and Tylén, 2016). While these latter studies are observational as they did not manipulate linguistic alignment on any specific level directly, their results show no clear pattern emerging for the covariance of alignment across different linguistic levels either. While not the primary purpose of the current study, we also explore if alignment strength varies together across speech features. Our aim is to find evidence for or against the “percolation” claim of IAM. If IAM is correct, we should find that alignment measures vary together across different speech features .

### 3.1.5. Present study

In our experiment, we employed a semi-naturalistic communicative setting where pairs of participants improvised narrative stories in the lab. We manipulated interactivity and individual goals across stories but kept the turn-taking structure and the amount of individual contributions constant.

Alignment was measured on multiple levels: for prosodic features (pitch, speech rate and pause length), syntax, lexical recurrence and semantic similarity. We selected speech features that covered the range of paralinguistic and linguistic measures often reported in the literature. We had two reasons for including multiple dependent variables. First, the theoretical predictions are not fine-grained enough to point to particular speech features. Second, results from the field are often difficult to compare across studies, due to measurements restricted to particular features or alignment definitions.

We manipulated interactivity by instructing participants to invent one (joint) storyline together or two (individual) storylines separately. The distinction is analogous to the difference between talking *with* someone versus talking *next* to someone.

Individual goals were manipulated via valence instructions: each participant was instructed to invent a positive or negative storyline. Pair members received the same valence instructions in half of the trials (Same goal condition) and different valence instructions in the other half (Different goal condition). An important feature of storytelling in everyday interactions is the speaker’s affective stance that characterizes both the function of the storytelling episode and the intended reaction from the listener (e.g. Ruusuvuori, Johanna, Peräkylä and Anssi, 2009; Voutilainen et al., 2014). Therefore, valence instructions constitute a goal manipulation relevant specifically for storytelling. By choosing a valence-based manipulation instead of a content-based one (e.g. instructions specifying endings or themes), we could also preserve the open-ended nature of the storytelling task, an important feature of naturalistic communication.

## 3.2. Methods

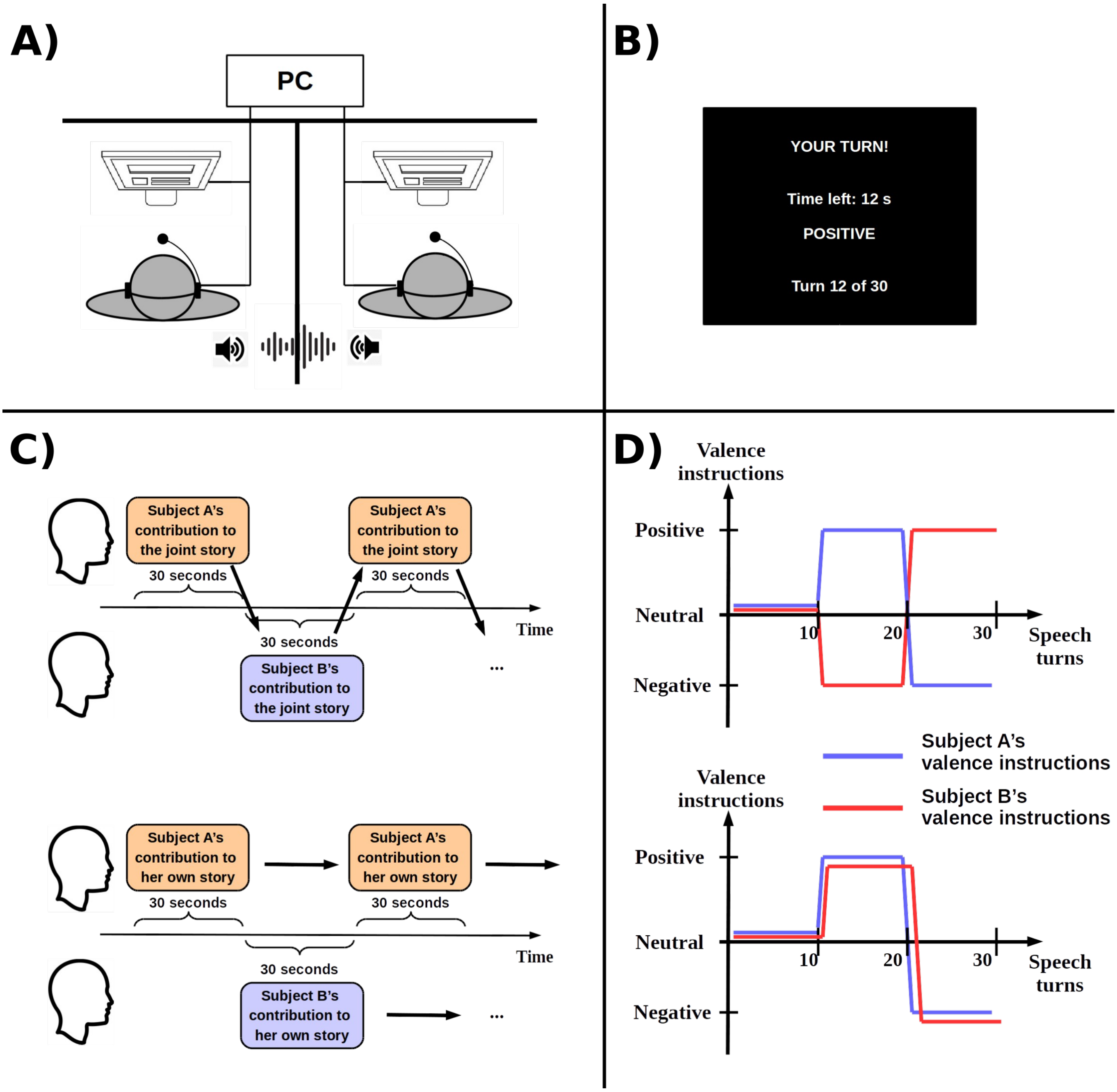


*Figure 3.1.* Study setup and the main manipulations. A) Pairs were seated in separate rooms and could speak to each other using headsets. B) A sample screen shot. Throughout each trial, participants were informed whether it was their turn to speak, the time left from the current turn, the valence instruction and the turn number relative to the full length of the trial. C) The interactivity manipulation. The upper part depicts the structure of a Joint trial, the

lower that of an Individual trial. Arrows indicate storyline continuations. D) The goal manipulation. The upper half depicts valence instructions to pair members in a Different trial, the lower displays valence instructions in a Same trial.

### 3.2.1. Participants

There were 104 participants (in 52 pairs) in the experiment. Participants were paired up randomly, based on availability. We did not control for the degree of acquaintanceship but asked participants to indicate it on a Likert scale. Due to technical issues and participant confusion about the instructions, 7 pairs were excluded, resulting in 90 participants (45 pairs) in our final sample (58 females, age *M* = 20.21 years, *SD* = 2.12, 20 female-male couples, 19 female-female, 6 male-male). Study participants were recruited through the Sona research participation system (Sona Systems, Ltd., www.sona-systems.com) of the Department of Psychological and Brain Sciences at Dartmouth College. Most participants were undergraduate students. They all had normal or corrected-to-normal vision and hearing. All participants gave written informed consent. Study procedures were approved by the Committee for the Protection of Human Subjects at Dartmouth College. Participants were reimbursed either financially (20 USD) or with course bonus credits.

### 3.2.2. Stimuli and Apparatus

The two members of each pair were seated in separate rooms, in front of identical computer monitors (see Figure 3.1A). Each participant sat cc. 60 cm from his or her respective monitor. They wore identical stereo headsets with microphone arms (Plantronics PLNAUDIO478). For the duration of each trial, an audio channel was open between the two participants. The audio channel was always one-way, transmitting only the voice of the person currently speaking according to the turn-taking structure of the task. For playing the audio we used 200 ms long buffers, introducing a latency of equal length. Both monitors, the audio channel, and the headsets were controlled by the same computer running Psychtoolbox 3 (Kleiner et al., 2007) on MATLAB (version 2012b, The MathWorks, Natick, MA). Speech was recorded throughout the experiment from both members of the pair.

Members of the pair performed a joint storytelling task. In each trial, they were asked to improvise either one story or two separate stories (see Procedure) to a given topic by taking turns. We used a set of five topics: “Alien encounter”, “A child who becomes POTUS”, “A family with child in trouble”, “Genie in a lamp” and “A student having a strange day”. The last topic was only used for the familiarization phase (see Procedures). The two participants in a pair always received the same topic in a trial. During each speech turn, the time left from the current turn in seconds was displayed centrally for both participants (see Figure 3.1B). The current role of the participant (“Your turn” or “Not your turn”), the turn number (e.g. “Turn 10 of 30”) and a short instruction regarding the valence trajectory to follow (“Positive” or “Negative”) were also shown on the screen.

### 3.2.3. Procedure

Each pair first went through a familiarization phase of the storytelling task and then went on to perform four trials. The task in each trial was to improvise one or two fictional stories (see below) for the topic provided. Participants were instructed to develop the storyline in any ways they wished to, with two constraints. First, a pilot study indicated that sometimes participants - presumably to save effort - told a well-known story (e.g. a movie plot) or a specific childhood memory during the task (e.g. first school day). To keep the stories open-ended, and to avoid the variability in the nature of the stories told, we asked participants to avoid retelling a known story or autobiographical memory. Second, to avoid potentially uncomfortable situations and distress for the participants, they were also asked to avoid references to each other and the experimental situation in the stories.

Each trial started with instructions displayed on the screen for 40 seconds. Then the topic for the trial was presented for 10 seconds after which the first speech turn started, as indicated on the screen (Figure 3.1B, see Stimuli and Apparatus). In each turn, there was a designated speaker and a listener, and these roles changed from speech turn to speech turn. A trial consisted of 30 speech turns, each 30 sec long, lasting a total time of 15 mins. At the end of each trial, participants were prompted by a message presented on their monitors to retell the story or stories they had just created. Participants were given 5 minutes for retell and they were instructed to try to utilize the full 5 minutes. They could not hear each other during retell. Each participant performed the retell individually. Retell ensured that participants paid attention to their partner’s speech. Following retell, we asked pair members to fill out a six-item questionnaire about the preceding trial. Participants received short breaks (2-5 mins) between trials as needed. After the fourth trial, an eight-item questionnaire was answered regarding the participants’ experience of the whole experiment. The familiarization phase consisted of one shorter storytelling trial with 10 secs long speech turns (with also 30 turns, for a total length of 5 mins) and added instructions.

Across the four trials we manipulated both the interactivity of the task (Interactivity: Joint vs. Individual, see Figure 3.1C) and the desired emotional valence trajectory of the story or stories (Goal: Different vs. Same, see Figure 3.1D), resulting in a within-pair 2 × 2 design. In Joint trials pair members were instructed to improvise one story together. We asked members of the pair to build upon the content from each other’s speech turns and to create a coherent storyline where individual contributions are hard to tell apart. In Individual trials we instructed both members of a pair to develop their own individual story in their own speech turns and to simply listen to the other person’s story when it was the other person’s turn to speak. We also asked participants to keep the two storylines apart as two separate narratives.

The factor Goal was operationalized as the degree of overlap in the desired valence of improvised stories. To that end, pair members were asked to follow specific valence trajectories over their speech turns. More specifically, after the first 10 speech turns (one third of a trial), a specific valence instruction (“Positive” or “Negative”) was displayed for the next 10 turns, and changed to its opposite (e.g. from “Positive” to “Negative”) for the last 10 turns in a trial. Participants were instructed to develop their stories in ways reflecting the valence instructions: “Negative” stories were to progress towards sad and tragic outcomes, while “Positive” referred to happy and joyful developments. Thus, participants were asked to follow the trajectory “Neutral – Positive – Negative” or the trajectory “Neutral – Negative – Positive” in a given trial. The combinations of these individual valence trajectories yielded two types of trials on the pair-level. One where the pair members followed the same trajectories (Same trials) and one where they followed different trajectories (Different trials).

Given the number of independent variables (Interactivity, Goal, valence trajectory combinations behind Goal, and Topic), we could not fully counterbalance the order of the trials. Instead we chose a slightly unbalanced design and mitigated potential order effects by using appropriate statistical procedures. We used two trial orders regarding Interactivity: Joint-Individual-Individual-Joint (21 pairs) and Individual-Joint-Joint-Individual (24 pairs). We selected these two orders because any linear order effect would cancel out for the two levels of Interactivity when averaged over levels of Goal. We crossed these two Interactivity orders with the four  possible Goal combinations, resulting in 8 different orderings that we counterbalanced across pairs. For the order of the four topics we used four possible combinations from a Latin square and counterbalanced them across pairs as well.

## 3.3. Feature extraction

### 3.3.1. Preprocessing and audio transcription

Audio was down-sampled to 16 kHz offline. For transcriptions we used Google's automated speech-to-text service to acquire a crude first transcript for each speech turn. Next we recruited users of Amazon's Mechanical Turk to verify and correct these initial transcripts by comparing it to the original audio recordings. Each transcript corrected by Mechanical Turk users was further verified by a research assistant who (1) read through it looking for inconsistencies and (2) compared two randomly selected turns from the transcript to the original audio. Mechanical Turk users were paid $7.5 for correcting the transcript of a 15 mins long recording (data from one trial). Participants consented to sharing the audio recordings for transcription purposes. In the next step we extracted six speech features: pitch, pause length, speech rate, syntactic similarity, lexical recurrence and semantic similarity. Speech features were selected to cover a the most frequently used prosodic and linguistic levels. Unless otherwise noted, feature extraction steps were implemented in Python 3.6.

### 3.3.2. Prosodic features

**Pitch.** We extracted the pitch contour of each speech turn using the detectpitch method from the Modulation Toolbox (v2.1, Atlas, Clark and Schimmel, 2010) in MATLAB with a moving window of 50 ms (25 ms overlap). The method first estimated voicedness (if the frame contained speech or pause) by fitting a two-component Gaussian mixture model to the log energy distribution of all frames for a given turn and applying the optimal decision boundary (Gerven and Xie, 1997). Then, a pitch estimate was derived for each voiced frame by fitting a harmonic model to the frame data in the least squares sense (Abu-Shikhah and Deriche, 2001), with the initial pitch values obtained by the autocorrelation method. Finally, a median filter of length five was applied to the contour. Because of the subharmonics and period doubling problem in pitch detection (see e.g. Gerhard, 2003), the pitch contour of each speech turn was visually inspected against the spectrogram, and – whenever necessary – we adjusted the voicing threshold and median filter length. For further analyses, we characterized each speech turn with the median pitch value in Hz (see e.g. De Looze et al., 2014).

**Pause length.** Pitch extraction also provided us with a voicedness marking for each 25 ms frame. In order to account for unvoiced speech segments, we filtered out very short pauses (<= 50 ms, similar to Fusaroli and Tylén, 2016). Then we described each speech turn by the median pause length .

**Speech rate.** We defined speech rate as the syllable rate in each speech turn, measured in syllable/second. Syllable counts were derived from the transcripts on the basis of the CMU pronunciation dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict). For implementation we relied on the built-in methods of the NLTK package (v3.3, Bird, Klein and Loper, 2009). In case of multiple pronunciations, the first (dominant) one was selected. For words not found in the dictionary we relied on a simple, rule-based approach counting vowels and applying common rules for pronunciation.

### 3.3.3. Syntactic, lexical and semantic measures

**Syntactic similarity**. We derived syntactic similarity values between speech turns using the CASSIM package (Boghrati, Hoover, Johnson, Garten and Dehghani, 2017). As a preprocessing step, we parsed each speech turn into sentences with the help of the automatic punctuator developed by Tilk and Alumae (2016). Then each speech turn was fed to the CASSIM routine, which first generated constituency parse trees for all sentences using the Stanford parser (Chen and Manning, 2014). In the next step, CASSIM calculated a graph-theoretical measure, Edit Distance for all sentence pairings across subsequent speech turns. Edit Distance is based on the minimum number of elementary parse tree editing steps (adding, deleting and renaming of nodes) needed to transform one syntactic structure into another and provides similarity scores for sentence pairings. Finally, those sentence pairings were selected that maximized overall similarity across speech turns.

Note that by using CASSIM, we did not focus on the recurrence of specific syntactic rules as in most priming studies, but on the overall similarity of parse trees. The present approach is conceptually very close to that of Reitter, Moore and Keller (2006; see also Reitter and Moore, 2014; Fusaroli et al., 2017) who measured recurrences for all phrase structure rules present in utterances. However, CASSIM is not only sensitive to repetitions of phase structures but goes further and takes into account the relations between those rules as embedded in parse trees. CASSIM has been validated both on an artificial dataset and against other measures of syntactic similarity (Boghrati et al., 2017).

**Lexical recurrence rate.** To characterize turn-to-turn lexical repetitions, we calculated the recurrence rate of words across subsequent speech turns. First we removed stop-words from the text (using the NLTK stop-word list), then lemmatized the transcripts to capture repetitions across different word forms too (with the Python package Pattern, v3.6, De Smedt and Daelemans, 2012). We defined the lexical recurrence rate simply as the number of lemmas in a given turn that were also uttered in the previous turn, divided by the average number of lemmas across the two speech turns. This approach is conceptually similar to the one employed by Fusaroli et al. (2017).

**Semantic similarity.** We aimed to capture semantic alignment beyond simple lexical repetitions, so we deleted the words repeated from the last speech turn from the transcripts before we calculated semantic similarity. As our method (see below) is based on a bag-of-words approach, deleting parts of the text is a valid step for our purposes. By controlling for lexical repetitions, simple imitation of the other’s words would not result in increased semantic similarity. Instead, this measure describes how conceptually close people’s contributions are, similar to the notion of conceptual alignment (e.g. Garrod and Anderson, 1987).

We measured semantic similarity between consecutive speech turns by combining word2vec embeddings (Mikolov, Chen, Corrado & Dean, 2013; Mikolov, Sutskerev, Chen, Corrado & Dean, 2013) trained on part of the Google News dataset (https://code.google.com/archive/p/word2vec/) with the Word Mover's Distance (WMD) document distance function, as described by Kusner, Sun, Kolkin & Weinberger (2015). In this approach, the first step is to transform the transcript of each speech turn to a list of word2vec vectors. The embedded vector space provides a semantic distance metric for all possible word pairings across written or spoken texts, that is, speech turns in the present context. The idea behind WMD is to minimize the overall "travel" needed in the vector space from the set of words in one speech turn to the other set of words in the subsequent speech turn. Solving this optimization problem yields an overall semantic distance metric across speech turns. WMD has been shown to provide state-of-the-art performance across a set of semantic classification tasks (Kusner, Sun, Kolkin & Weinberger, 2015). WMD is a special case of the more general class of Earth Mover's Distance metrics which have specialized algorithmic solutions (Pele & Werman, 2009). For implementation we relied on the PyEMD package (v0.4.4, Pele & Werman, 2008, 2009). As WMD is a distance value on a standardized scale from zero to one, for measuring similarity we used its inverse, $1 - WMD$, so that a value of 1 marked identical semantic content across speech turns and a value of 0 marked no semantic overlap, at all.

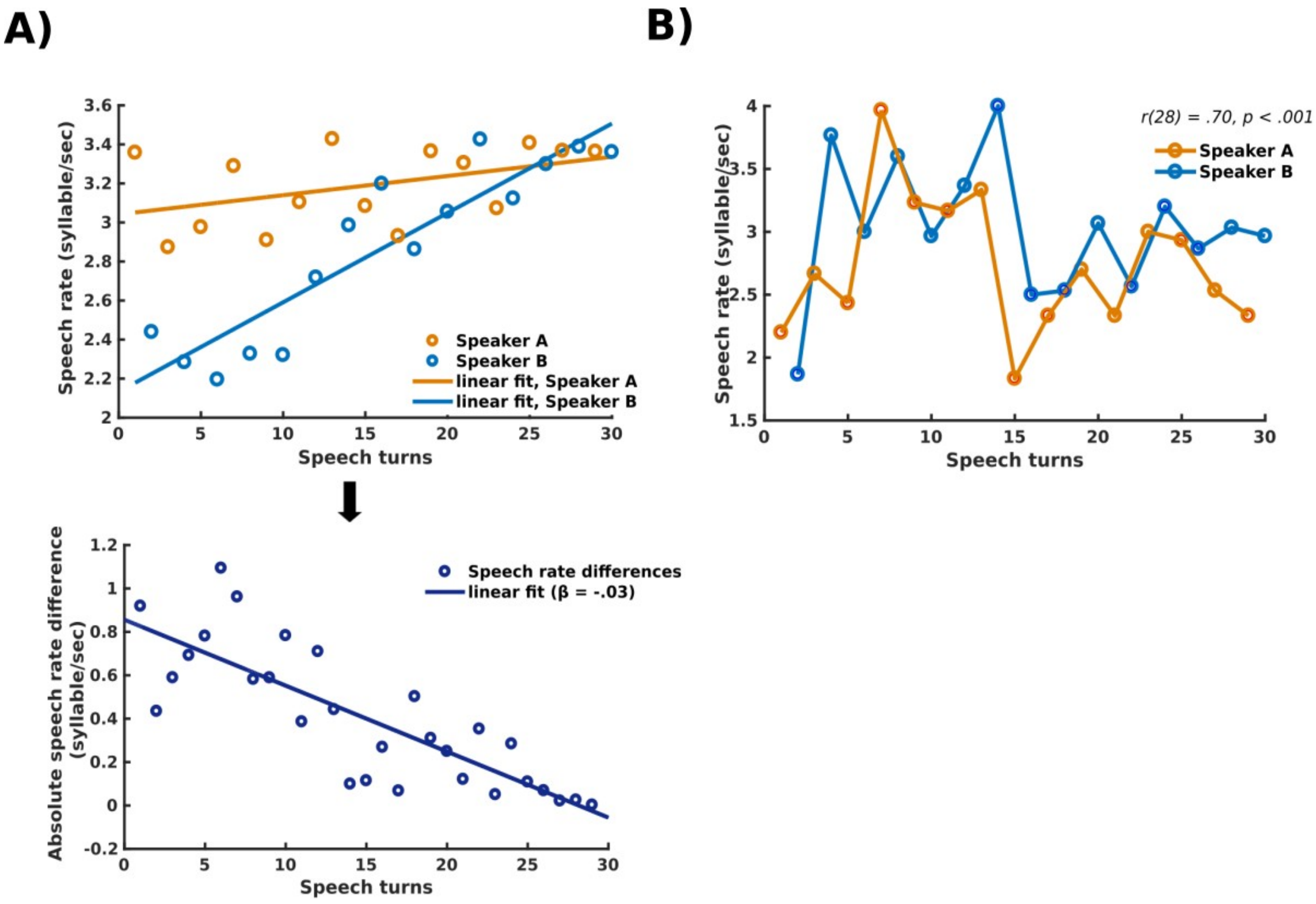


*Figure 3.2.* Operationalizations of alignment. We use speech rate here as an example, but the same measures apply to other features as well. In both panels of the figure, data points correspond to speech rate values (in syllable/s) of speech turns in a trial, from two example pairs selected for demonstrative purposes. A) Convergence. The top half depicts speech rate values of speakers across turns. in one trial (example selected for illustration). The bottom half depicts the operationalization of convergence: the values are the absolute differences between consecutive speech turns. Convergence is a negative linear trend of differences. B) Correlation between speakers’ speech rate values.

## 3.4. Statistical analyses

### 3.4.1. Types of alignment

In the literature on dialogue the term alignment is used rather freely to describe any measurable form of accommodation or temporal coordination. In the present paper we measure two general forms of alignment, convergence and correlation (Fig. 2, see also Edlund, 2009; Levitan and Hirschberg, 2011).

By convergence we refer to increasing similarity over the course of an interaction. Many studies reporting alignment have measured convergence of some feature (e.g. Garrod and Anderson 1987; Gregory and Webster, 1996; etc.). In the framework of IAM, such incremental build-up of similarity might correspond to the relatively slow alignment of higher-level representations (including the putative situational model). To calculate convergence for a given speech feature, we first derived absolute differences across consecutive speech turns in each trial. For example, in the case of pitch we first characterized each trial by the series of absolute differences in the pitch median values of consecutive turns. Then, on the group level, we quantified convergence by testing for an effect of Turn number on the difference values (see the details of models below). More importantly, we could test if our manipulations (Interactivity and Goal) interacted with Turn number, that is, whether they affected convergence in a particular feature. We repeated this process separately for all six speech features. Implicitly, investigating convergence also provided tests for similarity, that is, testing if our manipulations increased or decreased the overall level of differences between consecutive speech.

In contrast to convergence, correlation refers to relatively fast, turn-to-turn adaptations in terms of some speech feature. Correlation has also been frequently employed as the indicator of alignment (e.g. Bonin et al., 2013; De Looze et al., 2014; etc). Theoretically, such adaptations might be a signature of imitation-like coordination, with clear analogues in sensorimotor coordination (e.g. Konvalinka, Vuust, Roepstorff and Frith, 2010). In our study, we could calculate correlations only for prosodic features (pitch, pause length and speech rate) as our other measures (syntactic and semantic similarity, lexical recurrence) captured turn-to-turn differences and not features of individual speech turns. For each prosodic feature, we calculated the cross-correlation of pair members' turn-level values in each trial. As we had no hypothesis regarding leader-follower roles, we were equally interested in the zero-lag and plus-one-lag cross-correlation coefficients. Therefore, we used their average Fisher-Z transformed value in further analysis. Again, we were interested if the factors Interactivity or Goal affected the average cross-correlation coefficient.

### 3.4.2. Linear-mixed models

**Convergence.** In order to test the effect of Interactivity and Goal on convergence, we used separate linear-mixed effects models (LMMs) for each speech feature we measured. The absolute differences across speech turns formed the dependent variable. The structure of each LMM was determined using a two-step approach. We first determined the "maximal" model as advised by Barr, Levy, Scheepers & Tily (2013). The maximal model included Interactivity (with levels: Joint vs Individual), Goal (levels: Same vs. Different), Turn number (as continuous variable) and their interactions as fixed effects. Pair number, Topic, Acquaintance and Trial number (sequential order) were treated as random effects. Besides random intercepts, all possible random slopes with within-pair factors were also added for Pair number, Trial number and Acquaintance. Only random intercept was specified for Topic as there was no reason to expect different fixed effect estimates along different topics. When the model did not converge, we deleted the random effects starting from the highest-order effect.

As a second step we performed limited model-selection while maintaining the maximal nature of the model with regards to our factors of interest: Interactivity, Goal, Turn number and their interactions. To preserve generalizability, the random slopes of Interactivity, Goal and Turn number with regards to Pair number were also retained whenever possible. With regards to random effects including Topic, Acquaintance and Trial number we followed the procedure outlined by Matuschek et al. (2017, see also Bates, Kliegl, Vasishth & Baayen, 2015) and only included them if a likelihood-ratio test with threshold α = 0.2 suggested that the added complexity is balanced by better fit to the data. As Matuschek et al. (2017) showed, the maximal structure advocated by Barr et al. (2013) often reduces Type I error below its nominal level while also reducing power. In contrast, their approach yielded more parsimonious models while also balancing power and Type I error (similar to the step-procedure advocated by Kuznetsova, Brockhoff & Christensen, 2017). Please note that the model selection method followed here had no room for post-hoc experimenter decisions and the maximal structure regarding the main factors (random slopes of Interactivity, Goal and Turn number for Pair number) was retained whenever possible (that is, if the model converged). We report the final LMM for each feature in the Results section.

All models were fitted using restricted maximum likelihood as implemented in the lme4 package (v. 1.1-18-1, Bates, Mächler, Bolker & Walker, 2015) in R (v. 3.4.4, R Core Team). For significance testing, *p*-values were obtained via Satterthwaite approximations for the degrees of freedom (Luke, 2017), as provided by the lmerTest R package (v. 3.0-1, Kuznetsova et al., 2017). Homogeneity of variance and normality of residuals were evaluated using residual and QQ plots. Significant interactions were followed up by pairwise comparisons based on estimated marginal means, using the emmeans R package (v. 1.2.4). All reported marginal means are estimated means from the models.

**Correlation.** The same procedure as above was repeated with the correlation coefficients as dependent variable. Since there was only one datum per trial for correlation, Turn number was not included as a fixed effect. Also, due to the limited data relative to convergence measures, no random slopes were included in models of correlation.

**Pseudo pairs.** In principle, any convergence or correlation effect could be the result of general, not interaction-specific processes. For example, a general tendency to start slow but speak faster at the end of trials could result in a significant correlation effect without any pair-specific alignment. To control for this possibility and to acquire a baseline level of alignment, for any convergence or correlation effect identified by an LMM we we ran a further LMM that also included pseudo pairs constructed from members of real pairs. An additional fixed effect of Pair type (levels: Real vs. Pseudo) and its interactions with all other fixed effects were added to the final LMM used on real pairs. We expected that effects of Interactivity and Goal reflect pair-specific alignment and, thus, any significant effect in Real pairs would show an interaction with the factor Pair type in the extended LMM.

Pseudo pairs respected participants' roles, that is, if a participant was always starting the trials during the task (role of speaker A), in a pseudo pair she was only paired with members in the opposite role (speaker B). Furthermore, due to the potential combinations of valence trajectory and Interactivity instructions, not all possible pseudo pairs' data could be organized according to the 2 × 2 independent factor structure (Interactivity: Joint vs. Individual × Goal: Same vs. Different). To avoid potential confounds, only those pseudo pairs were included in any analyses whose trials fitted the 2 × 2 structure (n = 926, from a possible total of n = 1980).

## 3.5. Results

### 3.5.1. Pitch

The grand mean of pitch difference values across trials and pairs was *M* = 50.12 Hz (*SD* = 38.07). While a logarithmic scale is common for pitch values, in our case it led to a violation of the assumption of homoscedasticity of residuals in the LMM. We used instead a linear Hz scale for the dependent variable as this satisfied model assumptions.

**Convergence.** Convergence results are summarized in Fig. 3A and 3B. The final model included the random slopes of Interactivity, Goal and their interaction for Pair number (see ANOVA-type outputs for all LMM models in the Appendix B, from p. 231 on). There was a main effect of Turn number: $\beta = -6.7 \times 10^{-2}$, $SE = 1.4 \times 10^{-2}$, $p < .001$, proving the presence of an overall convergence effect. The interaction of Interactivity and Turn number was also significant ($\beta = 2.9 \times 10^{-2}$, $SE = 1.4 \times 10^{-2}$, $p = .037$), meaning that the marginal trend associated with Turn number was more negative in the Joint (marginal slope $\beta = -9.7 \times 10^{-2}$, $SE = 1.9 \times 10^{-2}$) than in the Individual condition (marginal slope $\beta = -3.8 \times 10^{-2}$, $SE = 1.9 \times 10^{-2}$). In other words, all other things being equal, pair members' median pitch values moved 2.83 Hz closer during Joint trials, while the average convergence was 1.10 Hz in Individual trials.

Importantly, when the model was extended to include pseudo pairs, the interaction of Interactivity, Turn number and Pair type was not significant ($p = .218$), suggesting that the significant Interactivity × Turn number interaction in Real pairs was not due to pair-specific effects. However, the interaction of Turn number and Pair type was significant ($\beta = 2.1 \times 10^{-2}$, $SE = 0.7 \times 10^{-2}$, $p = .006$), with a more negative marginal trend of Turn number in Real pairs ($\beta = -6.7 \times 10^{-2}$, $SE = 1.4 \times 10^{-2}$) than in Pseudo pairs ($\beta = -2.6 \times 10^{-2}$, $SE = 0.3 \times 10^{-2}$), corresponding to an average overall convergence of 1.94 Hz and 0.75 Hz in Real and Pseudo pairs, respectively.

**Correlation.** The final model included random intercepts for Pair number. No interaction or main effect of the fixed effects (Interactivity and Goal) was significant ($ps > .41$). When the model was extended to include pseudo pairs as well, we found a main effect of Pair type ($\beta = -2.34 \times 10^{-2}$, $SE = 1.09 \times 10^{-2}$, $p = .032$), with larger marginal means ($M = 0.06$, $SE = 0.021$) in Real than in Pseudo pairs ($M = 0.008$, $SE = 0.005$).

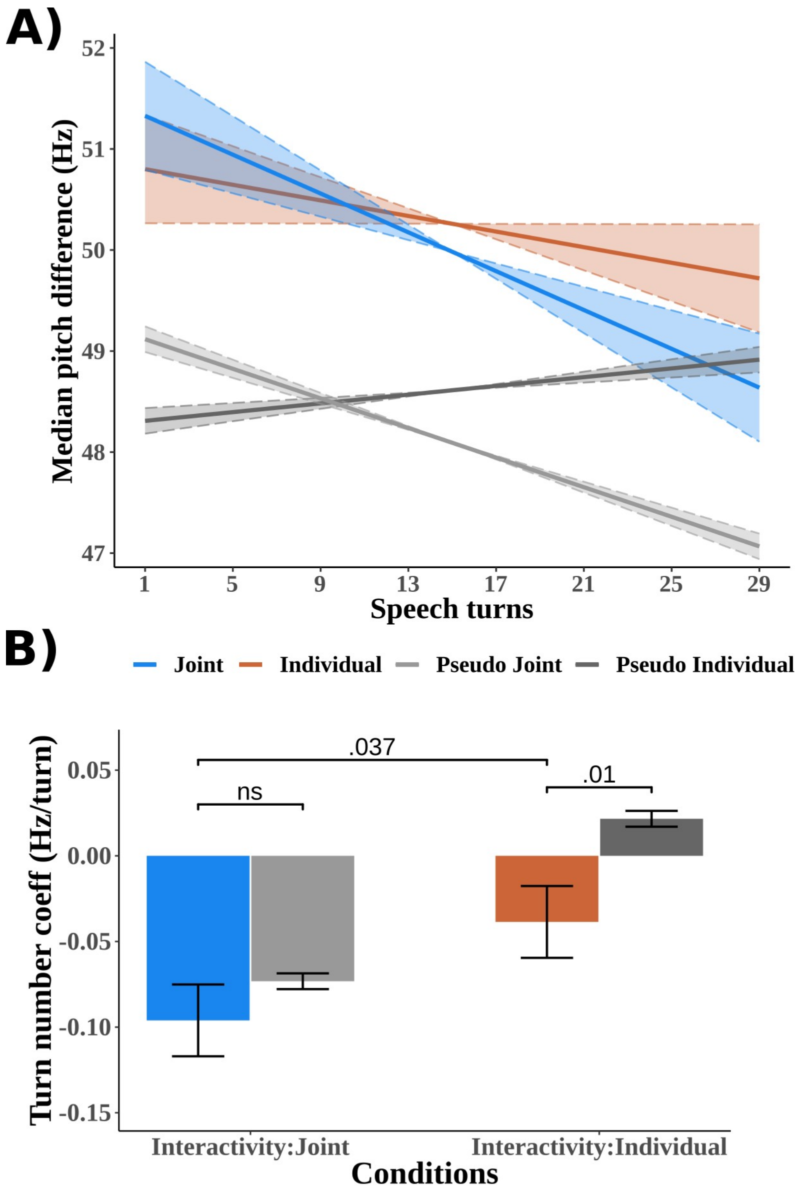


*Figure 3.3.* Pitch convergence results. Both panels show Turn number trends in the two levels of Interactivity (Joint vs. Individual) and their Pseudo pair counterparts, averaged over levels of Goal. A) The panel depicts Turn number trend effects over speech turns, with the shaded areas displaying 95% CIs for the coefficient of Turn number. B) The lower panel presents pairwise comparison results for the marginal Turn number coefficients depicted in panel A. Error bars represent SE.

### 3.5.2. Pause length

The grand mean of the median pause length values was $M$ = 139.91 ms ($SD$ = 66.46), with the mean of the difference values at $M$ = 57.09 ms ($SD$ = 67.12).

**Convergence.** Convergence results are summarized in Fig. 4A and 4C. The final model included random slopes of Interactivity, Goal and their interaction for Pair number. A significant interaction of Goal × Turn number was found (β = -0.237, $SE$ = 9.64 x $10^{-2}$ , $p$ = .014), with a negative marginal trend of Turn number in the Same condition (marginal slope β = -0.319, $SE$ = 0.14) and with a positive trend in the Different condition (marginal slope β = 0.154, $SE$ = 0.14). The latter though was not significantly different from zero (95% CIs: -0.113, 0.421). On average, the pause length difference was reduced by 9.25 ms over speech turns in the Same condition.

Then we extended the model to include Pseudo pairs as well and found the expected three-way interaction of Goal × Turn number × Pair type (β = 0.113, $SE$ = 5.16 x $10^{-2}$, $p$ = .029). Pairwise comparisons showed that Goal only had an effect on the slope of Turn number in Real pairs ($p$ = .014) but not in Pseudo pairs ($p$ = .62, with marginal slopes $\beta$ = 0.102, $SE$ = 3.14 x $10^{-2}$; and $\beta$ = 0.125, $SE$ = 3.14 x $10^{-2}$). Overall, there was convergence only in the Same condition in Real pairs.

**Correlation.** Correlation results are depicted in Fig. 4E. The final model included random intercepts for Pair number and yielded a main effect of Goal (β = -0.045, $SE$ = 0.020, $p$ = .025), with larger values in Different (marginal $M$ = 0.25, $SE$ = 0.029) than in Same trials (marginal $M$ = 0.16, $SE$ = 0.029). Importantly, when the model was extended to Pseudo pairs, we obtained a significant interaction of Goal × Pair type (β = 0.024, $SE$ = 0.01, $p$ = .019). Pairwise comparisons revealed that there was no effect of Goal in Pseudo pairs ($p$ = .73, with marginal $M$ = 0.180, $SE$ = 0.006; and $M$ = 0.177, $SE$ = 0.006 for Same and Different trials, respectively).

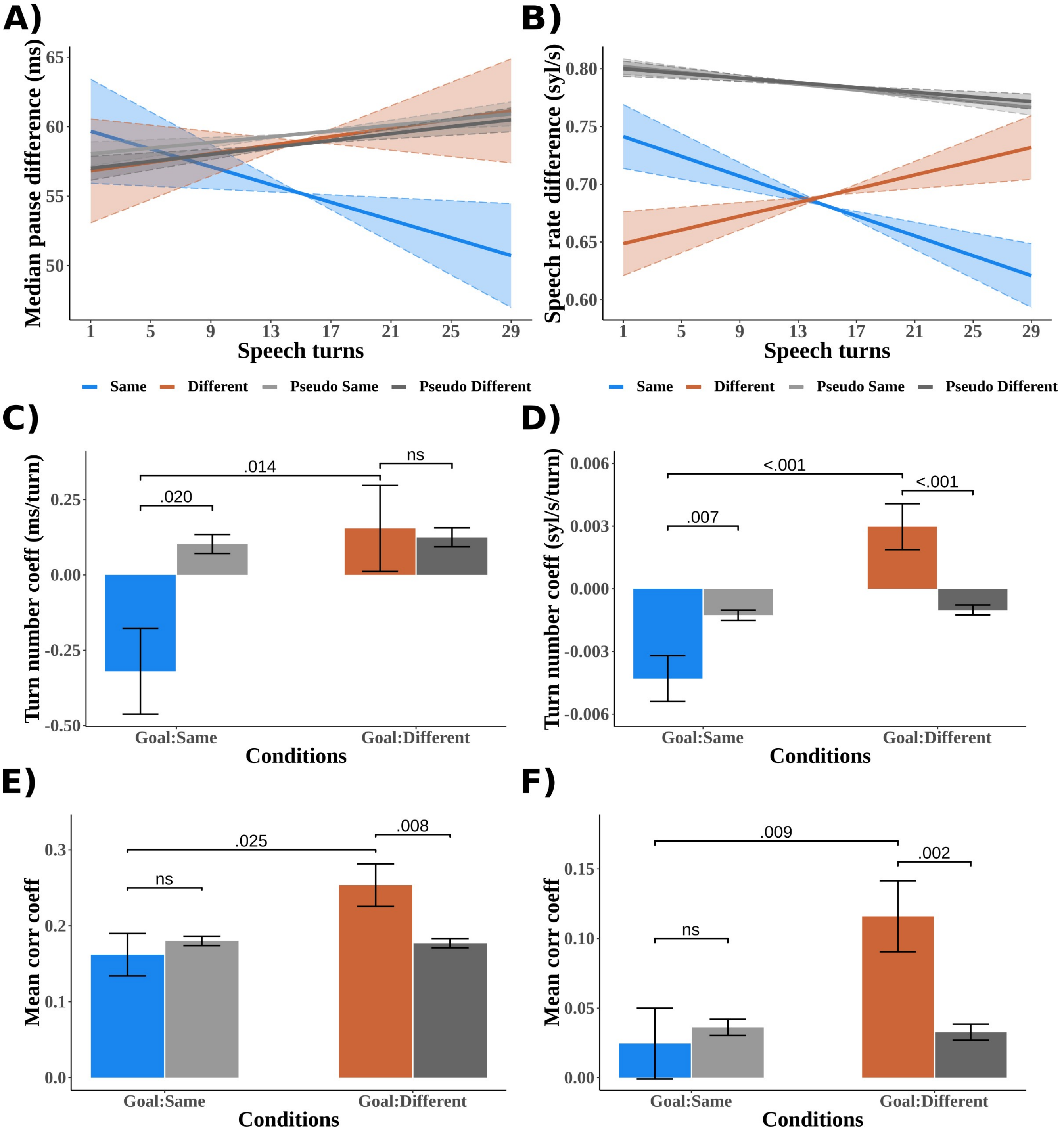


*Figure 3.4.* Convergence and correlation results for pause length and speech rate. A) - B) Convergence results for pause length and speech rate, respectively. Both panels depict Turn number trend effects over speech turns, in the two levels of Goal (Same vs. Different) and

their Pseudo pair counterparts, averaged over levels of Interactivity. The shaded areas display 95% CIs for the coefficient of Turn number. C) – D) Convergence results for pause length and speech rate, with both panels presenting pairwise comparison results for the marginal Turn number coefficients depicted in panels A and B. Error bars represent SE. E) – F) Correlation results for pause length and speech rate. Bars depict marginal mean cross-correlation of median values in the two levels of Goal and their Pseudo pair counterparts, averaged over levels of Interactivity. Error bars represent SE.

### 3.5.3. Syllable rate

The mean syllable rate across all trials and pairs was *M* = 3.05 syl/s (*SD* = 0.70) with a mean syllable rate difference of *M* = 0.69 syl/s (*SD* = 0.53).

**Convergence.** Convergence results are presented in Fig. 4B and 4D. The final model included the random slopes of Interactivity, Goal and their interaction for Pair number. We found a significant interaction of Goal × Turn number (β = 3.6 x $10^{-3}$, *SE* = 0.76 x $10^{-3}$, *p* < .001), with a negative marginal slope of Turn number in the Same (marginal slope β = -4.3 x $10^{-3}$, *SE* = 1.0 x $10^{-3}$) and a positive slope in the Different condition (marginal slope β = 3.0 x $10^{-3}$, *SE* = 1.0 x $10^{-3}$). On average, pairs reduced their initial syllable rate differences by -0.124 syl/s over the course of the interaction in Same and increased it by 0.086 in Different trials.

When the model was extended to include pseudo pairs, we found a main effect of Pair type (β = 5.6 x $10^{-2}$, *SE* = 2.6 x $10^{-2}$, *p* = .031), with smaller difference values in Real (marginal *M* = 0.67, *SE* = 0.05) than in Pseudo pairs (marginal *M* = 0.79, *SE* = 0.01). More importantly, there was also a significant interaction of Goal × Turn number × Pair type (β = 3.6 x $10^{-3}$, *SE* = 0.77 x $10^{-3}$, *p* < .001). Pairwise comparisons showed that Turn number slopes were not different across Same and Different conditions in Pseudo pairs (*p* = .47), supporting the idea that syllable rate convergence effect in Real pairs was specific to the interactions themselves.

**Correlation.** Correlation results are shown in Fig. 4F. The final model included random intercepts for Pair number. We found a significant main effect of Goal ($\beta = -4.6 \times 10^{-2}$, *SE* = $1.7 \times 10^{-2}$, *p* = .009), with larger values in Different (marginal *M* = 0.12, *SE* = 0.02) than in Same trials (marginal *M* = 0.02, *SE* = 0.02). Follow-up tests showed that this effect primarily originated from the zero-lag correlations, but was present as a tendency in plus-one-lag as well (Goal effect in zero-lag: $\beta = -4.7 \times 10^{-2}$, *SE* = $2.1 \times 10^{-2}$, *p* = .025; and in plus-one-lag: $\beta = -4.5 \times 10^{-2}$, *SE* = $2.5 \times 10^{-2}$, *p* = .069). Importantly, the extended model including Pseudo pairs yielded a significant Goal × Pair type interaction ($\beta = 2.4 \times 10^{-2}$, *SE* = $0.86 \times 10^{-2}$, *p* = .006). As expected, the difference between Same and Different trials was not significant in Pseudo pairs (*p* = .64).

### 3.5.4. Syntactic similarity

Please note that for syntactic, lexical and semantic alignment our results are based on similarity scores, not difference scores. The grand mean of syntactic similarity across all trials and pairs was *M* = 0.652 (*SD* = 0.032).

**Convergence.** Convergence results are depicted in Fig. 5A-C. The final model included random slopes of Interactivity, Goal and their interaction for Pair number, and also random intercepts for Trial number. The model showed a significant main effect of Interactivity ($\beta = -3.2 \times 10^{-3}$, *SE* = $1.45 \times 10^{-3}$, *p* = .033), with larger values in the Joint condition (marginal *M* = 0.6554, *SE* = 0.0014) than in the Individual condition (marginal *M* = 0.6522, *SE* = 0.0014). The main effect of the covariate Turn number was significant as well ($\beta = -8.5 \times 10^{-5}$, *SE* = $3.9 \times 10^{-5}$, *p* = .036), describing an overall decrease of similarity over the speech turns. The Turn number effect was modulated by Interactivity, as evidenced by the significant interaction Interactivity × Turn number interaction ($\beta = -9.9 \times 10^{-5}$, *SE* = 4.1 x

$10^{-5}$, *p* = .015), with a negative marginal slope of Turn number in Individual (marginal slope β = $-1.85 \times 10^{-4}$, *SE* = $0.58 \times 10^{-4}$), and with a positive marginal slope in Joint trials (marginal slope β = $1.4 \times 10^{-5}$, *SE* = $5.8 \times 10^{-5}$). The marginal slope for Joint trials was not reliably different from zero though (95% CIs: $-9.9 \times 10^{-5}$, $1.3 \times 10^{-4}$). In other words, syntactic similarity was reduced by -0.054 (standardized unit) over the course of Individual trials on average, while it remained virtually unchanged in Joint trials.

Including Pseudo pairs in the model corroborated the role of Interactivity. We found a main effect of Pair type (β = $4.3 \times 10^{-3}$, *SE* = $1.0 \times 10^{-3}$, $p < .001$), with larger values in Real (marginal *M* = 0.6538, *SE* = 0.0010) than in Pseudo pairs (marginal *M* = 0.6495, *SE* = 0.0002). The interaction of Turn number and Pair type was not significant (*p* = .29). We found a significant interaction of Interactivity and Pair type (β = $3.9 \times 10^{-3}$, *SE* = $1.6 \times 10^{-3}$, *p* = .014) and pairwise comparisons showed that while in Real pairs there was larger similarity in Joint trials than in Individual (*p* = .004), there was no such difference in Pseudo pairs (*p* = .18). Finally, there was a significant Interactivity × Turn number × Pair type interaction (β = $4.7 \times 10^{-5}$, *SE* = $2.0 \times 10^{-5}$, *p* = .019). Pairwise comparisons revealed that the marginal slope of Turn number was significantly different across levels of Interactivity in Real pairs (*p* = .011) but not in Pseudo pairs (*p* = .46).

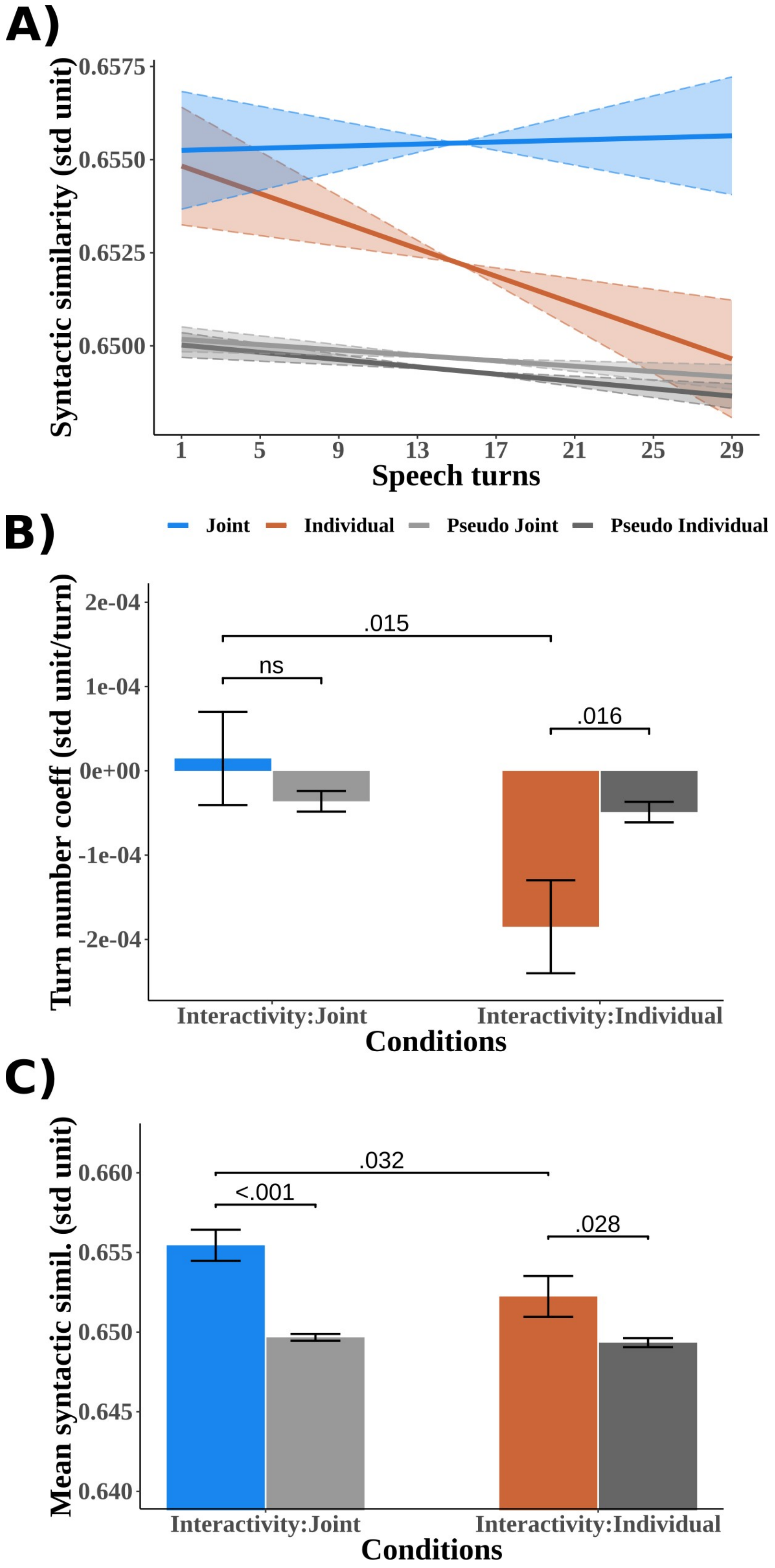
*Figure 3.5.* Effects of the factor Interactivity on syntactic similarity. A) Convergence results for syntactic similarity. The panel depicts Turn number trend effects over speech turns in the two levels of Interactivity (Joint vs. Individual) and their Pseudo pair counterparts, averaged over levels of Goal. The shaded areas displaying 95% CIs for the coefficient of Turn number. B) The panel presents pairwise comparison results for the marginal Turn number coefficients depicted in panel A. Error bars represent SE. C) Main effect of Interactivity on the overall similarity scores. Bars depict marginal mean syntactic similarity in the two levels of Interactivity and their Pseudo pair counterparts, averaged over levels of Goal. Error bars represent SE.

### 3.5.5. Lexical recurrence

The overall mean recurrence rate was *M* = 11.47% (*SD* = 7.99%).

**Convergence.** Lexical recurrence results are summarized in Fig. 6A, 6C and 6E. The final model included random slopes of Interactivity, Goal and their interaction for Pair number, and also random intercepts for Topic and Trial number. The model showed main effects of Interactivity (β = -3.44, *SE* = 0.25, *p* < .001) and Turn number ($\beta = -6.1 \times 10^{-2}$, *SE* = $1.1 \times 10^{-2}$, *p* < .001), while none of the interactions were significant (*p*s > .17). Regarding Interactivity, pairs displayed an overall larger recurrence rate in Joint (marginal *M* = 15.14%, *SE* = 0.63) than in Individual trials (marginal *M* = 7.81%, *SE* = 0.59). Turn number had an overall negative slope, with recurrence rate decreasing with 1.76% over speech turns on average.

With the model extended to Pseudo pairs, we found a significant main effect of Pair type (β = 3.83, *SE* = 0.11, *p* < .001), with larger recurrence rate in Real (marginal *M* = 11.47%, *SE* = 0.20) than in Pseudo pairs (marginal *M* = 4.30%, *SE* = 0.05). We also obtained significant interactions of Interactivity × Pair type (β = 1.80, *SE* = 0.08, *p* < .001) and Turn number × Pair type ($\beta = 1.16 \times 10^{-2}$, *SE* = $0.36 \times 10^{-2}$, *p* < .001). Regarding


A)
Syntactic similarity (std unit)
0.6575
0.6550
0.6525
0.6500
1
5
9
13
17
21
25
29
Speech turns
Joint
Individual
Pseudo Joint
Pseudo Individual
B)
Turn number coeff (std unit/turn)
2e-04
1e-04
0e+00
-1e-04
-2e-04
.015
ns
.016
Interactivity:Joint
Interactivity:Individual
Conditions
C)
Mean syntactic simil. (std unit)
0.660
0.655
0.650
0.645
0.640
.032
<.001
.028
Interactivity:Joint
Interactivity:Individual
Conditions

Interactivity, the difference between Joint and Individual conditions was only significant in Real ($p < 0.001$) but not in Pseudo pairs ($p = .17$). Interestingly, the pairwise comparisons revealed that the lexical recurrence rate was larger in Real pairs than in Pseudo pairs even in Individual trials ($p < .001$, with $M = 7.74\%$, $SE = 0.24$; and $M = 4.33\%$, $SE = 0.06$, respectively). Regarding Turn number, the marginal slope was more negative in Real (marginal slope $\beta = -6.08 \times 10^{-2}$, $SE = 0.69 \times 10^{-2}$) than in Pseudo pairs (marginal slope $\beta = -2.9 \times 10^{-2}$, $SE = 0.15 \times 10^{-2}$). Importantly, when the analysis was repeated including only the pseudo pairs where members were paired up from real pairs with corresponding topics across trials (n = 96), we observed the same general pattern of results (all reported effects significant in either cases).

The same general pattern of results was obtained when we considered recurrences of short phrases (2- and 3-gram recurrence rates).

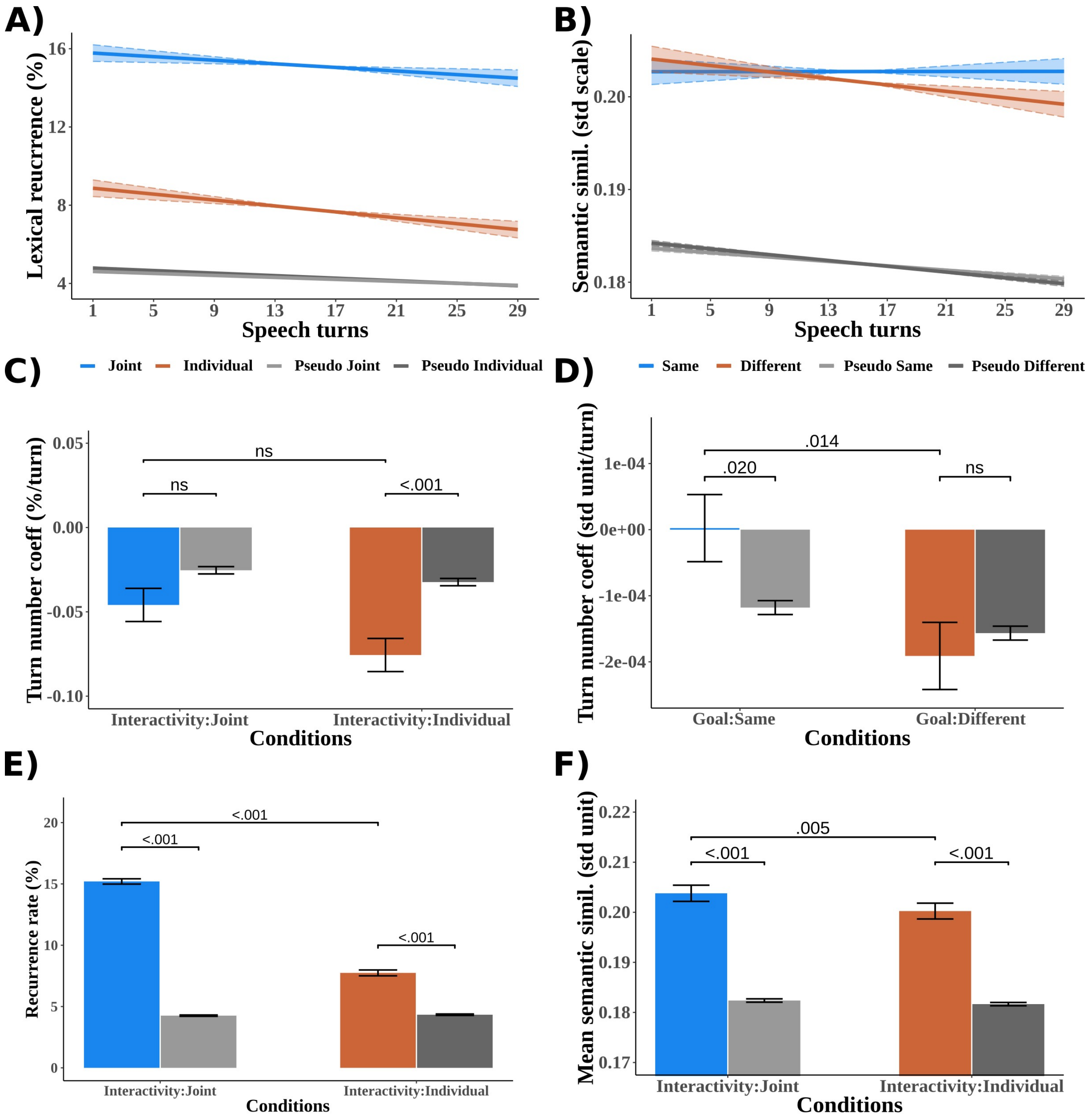


*Figure 3.6.* Effects of Interactivity and Goal on lexical recurrence and semantic similarity. A) Convergence results for lexical recurrence. The panel shows Turn number trend effects over speech turns in the two levels of Interactivity (Joint vs. Individual) and their Pseudo pair counterparts, averaged over levels of Goal. The shaded area displays 95% CIs for the

coefficient of Turn number. B) Convergence results for semantic similarity. The panel shows Turn number trend effects over speech turns in the two levels of Goal (Same vs. Different) and their Pseudo pair counterparts, averaged over levels of Interactivity. The shaded area displays 95% CIs for the coefficient of Turn number. C) Convergence results for lexical recurrence. The panel presents pairwise comparison results for the marginal Turn number coefficients depicted in panel A. Error bars represent SE. D) Convergence results for semantic similarity. The panel presents pairwise comparison results for the marginal Turn number coefficients depicted in panel B. Error bars represent SE. E) - F) Main effect of Interactivity on the overall recurrence rate and semantic similarity scores, respectively. Bars depict marginal mean lexical recurrence and syntactic similarity in the two levels of Interactivity and their Pseudo pair counterparts, averaged over levels of Goal. Error bars represent SE.

### 3.5.6. Semantic similarity

The overall mean similarity value was *M* = 0.2022 (*SD* = 0.002).

**Convergence.** Semantic similarity results are summarized in Fig. 6B, 6D and 6F. The final model included random slopes of Interactivity, Goal, and their interaction for Pair number, and also random intercepts for Topic. We found a main effect of Interactivity (β = -2.74 x $10^{-3}$, *SE* = 0.96 x $10^{-3}$, *p* = .005, see Fig. 6F), with larger semantic similarity values in Joint (marginal *M* = 0.205, *SE* = 0.002) than in Individual trials (marginal *M* = 0.200, *SE* = 0.002). We also found a main effect of Turn number (β = -8.65 x $10^{-5}$, *SE* = 3.57 x $10^{-5}$, *p* = .015), meaning an overall negative trend of the covariate (with an average decrease of 2.51 x $10^{-3}$ in semantic similarity over the course of an interaction). The interaction of Goal × Turn number was also significant (β = 8.76 x $10^{-5}$, *SE* = 3.57 x $10^{-5}$, *p* = .014, see Fig. 6B and 6D), suggesting a more negative slope of Turn number in Different (marginal slope β = -1.74

$\times 10^{-4}$, *SE* = 0.50 x $10^{-4}$) than in Same trials (marginal slope β = 1.22 x $10^{-6}$, *SE* = 5.49 x $10^{-5}$). The latter was effectively zero ( 95% CIs: -9.78 x $10^{-5}$, 1.0 x $10^{-4}$).

When we extended the model with Pseudo pairs, we found a significant effect of Pair type (β = -9.66 x $10^{-3}$, *SE* = 0.79 x $10^{-3}$, *p* < .001), with larger similarity values in Real (marginal *M* = 0.202, *SE* = 0.001) than in Pseudo pairs (marginal *M* = 0.182, *SE* = 3.0 x $10^{-4}$). There was also a significant interaction of Interactivity × Pair type (β = -8.26 x $10^{-4}$, *SE* = 3.48 x $10^{-4}$, *p* = .018), meaning that the Interactivity effect found in Real pairs was not present in Pseudo pairs (*p* = .39), as expected. Similarly, we found a significant interaction of Goal × Turn number × Pair type (β = -3.75 x $10^{-5}$, *SE* = 1.84 x $10^{-5}$, *p* = .041). Pairwise comparisons on the Turn number effect showed that in the Different condition the convergence effect was not significantly different across Real and Pseudo pairs (*p* = .52, with marginal slopes β = -1.91 x $10^{-4}$, *SE* = 0.51 x $10^{-4}$; and β = -1.58 x $10^{-4}$, *SE* = 0.11 x $10^{-4}$ for Real and Pseudo pairs, respectively), while there was a significant difference in the Same condition (*p* = .024, with marginal slopes β = 2.34 x $10^{-6}$, *SE* = 5.09 x $10^{-5}$; and β = -1.14 x $10^{-4}$, *SE* = 0.11 x $10^{-4}$ for Real and Pseudo pairs, respectively).

Again, as with lexical recurrence, we repeated the analysis with only including the Pseudo pairs with matching original topic lists across members and found the same pattern of results regarding the effect of Interactivity.

### 3.5.7. Relationship between convergence and correlation

We also assessed the relationship between the two types of alignment we analyzed. For all three prosodic features for which we could calculate both convergence and correlation, we re-ran the correlation LMMs with trial-level convergence added as a fixed effect covariate. We found that convergence did not have any predictive value regarding correlation, that is, neither the main effect of Convergence, nor any interactions involving

Convergence reached significance for any of the three features (pitch: *p*s > .12; pause length: *p*s > .13; speech rate: *p*s > .18).

### 3.5.8. Effects across features

We also tested if alignment at one level leads to alignment at other levels. To detect such effects, we re-ran all convergence LMMs reported above, but also including difference or similarity scores from other features as predictors. For example, we included as a predictor the lexical recurrence rate in the LMM of semantic similarity to test for the putative link between them. We did so for all pairings of the extracted features (15 pairings in total), testing for linear effects between any two of them. To simplify the interpretations, we did not include Turn number as a fixed effect in these models, only Interactivity, Goal, another speech feature and their interactions. In terms of random intercepts and slopes we followed the same procedure as for the all other models. As this analysis included many different models testing the same hypothesis, namely, that alignment is not independent across speech features, we applied Bonferroni correction. Hence $p$ values were compared against $\alpha = .0033$ instead of $\alpha = .05$. To determine if significant results were reflecting percolation effects specific to interactions, we ran additional, extended models for them with Pair type added as a fixed factor. The same Bonferroni correction applied to the results of the extended models as well.

**Syntactic similarity.** We found an effect of speech rate on syntactic similarity ($\beta = -1.87 \times 10^{-2}$, $SE = 7.24 \times 10^{-4}$, $p < .001$), meaning that smaller speech rate difference corresponded to larger syntactic similarity. Importantly, when the model was extended with pseudo pairs the Speech rate × Pair type interaction was significant ($\beta = -9.01 \times 10^{-3}$, $SE = 0.38 \times 10^{-3}$, $p < .001$), with a stronger (more negative) marginal trend of Speech rate in Real than in Pseudo pairs (marginal slopes: $\beta = -1.87 \times 10^{-2}$, $SE = 7.39 \times 10^{-4}$; and $\beta = -6.80 \times 10^{-4}$, $SE = 1.46 \times 10^{-4}$ , respectively).

**Recurrence rate.** We further found an effect of syntactic similarity on recurrence rate on the edge of significance: β = 0.11, *SE* = 0.037, *p* = .0036. When pseudo pairs were included there was a significant interaction of Syntactic similarity × Pair type (β = -5.79 x $10^{-2}$, *SE* = 1.22 x $10^{-2}$, *p* < .001), with larger effect of Syntactic similarity in Real than in Pseudo pairs (marginal slopes: β = 0.11, *SE* = 0.024; and β = -0.0044, *SE* = 0.0054 , respectively).

**Semantic similarity.** There was also a positive effect of recurrence rate on semantic similarity (β = 4.00 x $10^{-2}$, *SE* = 0.46 x $10^{-2}$, *p* < .001). When the model was extended with pseudo pairs, we found the expected interaction of Recurrence rate × Pair type (β = 8.09 x $10^{-3}$, *SE* = 2.51 x $10^{-3}$, *p* = .001). However, the effect of Recurrence rate was larger in Pseudo pairs than in Real pairs (marginal slopes β = 6.29 x $10^{-2}$, *SE* = 0.15 x $10^{-2}$; and β = 4.67 x $10^{-2}$, *SE* = 0.48 x $10^{-2}$ for Pseudo and Real pairs, respectively). There was also a Recurrence rate × Pair type × Interactivity interaction (β = 1.01 x $10^{-2}$, *SE* = 2.50 x $10^{-3}$, *p* < .001). Pairwise comparisons revealed that the coefficient of Recurrence rate was smaller in the Real Joint condition (marginal slope: β = 3.02 x $10^{-2}$ , *SE* = 5.62 x $10^{-3}$) than in the Real Individual, Pseudo Joint and Pseudo Individual conditions (β = 6.31 x $10^{-2}$ , *SE* = 7.72 x $10^{-3}$, p = .0032; β = 6.67 x $10^{-2}$ , *SE* = 2.17 x $10^{-3}$, p < .001; β = 5.90 x $10^{-2}$ , *SE* = 2.13 x $10^{-3}$, p < .001, respectively).

## 3.6. Discussion

Our study was motivated by IAM which asserts that accommodation on multiple linguistic levels plays a key role in successful communication. We tested two predictions of IAM, namely that interactivity (mutual, back-and-forth influences) increases alignment across multiple speech features and that individual goals have no effect on alignment. We employed a 2 × 2 repeated measures design with a naturalistic communication task (storytelling) and manipulated both interactivity (joint v individual) and individual goals (same vs different) across the four trials. Our dependent variables were alignment measures (convergence or correlation) across six separate speech features. We found that the two factors have independent effects and modulate alignment at different levels. Individual goals affected alignment mainly in terms of prosodic features, namely characteristic pause length and speech rate, both linked to the overall speed and rhythm of speech, with an effect on semantic convergence too. Manipulating interactivity led to separate changes in syntactic, lexical and semantic alignment.

### 3.6.1. Effects of the goal manipulation

**Convergence.** In the same goal condition participants showed a larger convergence effect both in terms of pause length and speech rate compared to the different goal condition. Furthermore,  semantic similarity decreased over speech turns in the different goal condition while it remained unchanged in the same goal condition. These effects were only present in real pairs, not in pseudo pairs, proving that the effects were specific to real interactions.

The convergence effects in pause length and speech rate support facilitation/inhibition account and are in line with earlier results showing that incompatible individual goals in a discussion (disagreement) decreases behavioral (Paxton and Dale, 2013; Paxton and Dale, 2017; but see also van Swol and Drury, 2017) and prosodic (Abney et al., 2014; Duran and Fusaroli, 2017; Riordan et al., 2014) synchrony or alignment. Moreover, the convergence results regarding speech rate and pause length replicate the findings of Abney et al. (2014), Duran and Fusaroli (2017) and Riordan et al. (2014), who all found an effect of different individual goals on alignment in terms of the temporal structure of communication. Several dependent variables showing an effect of individual goals in Duran and Fusaroli (2017) were calculated from speech rates: they used a moving window approach to extract speech rate time series for the two participants and then derived measures of alignment from cross-recurrence quantification analysis. More similar to our speech rate convergence measure but focusing on a different modality, Riordan et al. (2014) calculated absolute temporal and word number differences across consecutive turns in text-to-text communication, and tested for a linear trend of the absolute differences over turns. Our pause length measure was conceptually closer to the dependent variable in Abney et al. (2014), who computed complexity matching on speech-pause time series. Our measure of median pause length captures a characteristic value of speech-pause time series.

The goal manipulation also affected semantic convergence. Across all conditions we found a slightly negative trend of semantic similarity over speech turns (we address this negative trend later), but this slow divergence was modulated by individual goals: when pair members had the same goal, their semantic similarity stayed around the same level over the course of the interaction, but decreased in the different goal trials. The comparison against pseudo pairs showed that the effect was specific to the same goal condition, as there was no difference between real and pseudo pairs in the different goal condition. In other words, when people have the same goals, their stories tend to stay similarly closer, irrespective of interactivity. According to our hypotheses, this result suggests that having the same goals can facilitate imitation on the semantic level. However, as our semantic similarity measure captures alignment beyond lexical repetitions, this convergence effect cannot be based on faithful repetition, but suggests emulative verbal mimicry (see Csibra, 2008 on imitation/emulation). A potential explanation is that the valence instructions determined the sentiment of each speech turn to some degree and, as a result, in same goal trials participants might have used more synonyms (semantically close lemmas) for affective descriptions, balancing the general negative trend. It is interesting though that we have not found an effect of goal on lexical alignment, suggesting a rather subtle effect independent of one-to-one lexical imitation.

**Correlation.** Besides the convergence effects, the goal manipulation also had an impact on the correlation results of pause length and speech rate. Interestingly, the effect was the opposite of the convergence effect: pause length and speech rate correlations were larger when (real) pair members' goals were different than when they were the same. Furthermore, the convergence and correlation effects were linearly independent. Together, these results mean that the goal manipulation influenced faster, turn-to-turn adaptation and slower, long-term convergence separately and in opposite directions. The independence of the results across different timescales is not surprising in itself. In corpus studies for example, both Levitan and Hirschberg (2011) and Schweitzer and Lewandowski (2013) found a small correlation result for speech rate but no long-term convergence. We do not know though of any study reporting contradictory effects of a manipulation on different timescales of prosodic alignment. We can only offer a speculative explanation. In the different goal condition the partners' speech turns were incompatible with participants' own goals, with the conflict eliciting stronger attention. Attention in turn could have led to a stronger imitation effect in terms of temporal features of speech. However, as we discussed earlier, long-term convergence is negatively affected by different goals but facilitated by matching goals.

### 3.6.2. Effects of the interactivity manipulation

Interactivity influenced syntactic convergence and overall syntactic, lexical and semantic similarity. Generally, these results are in line with the predictions of IAM: interactivity increases overall alignment (in terms of convergence or similarity) at multiple levels.

Regarding syntactic convergence, we found that the Turn number effect was around zero in Joint trials but negative for the Individual trials. The average size of the drop over the course of individual trials was about as large as the overall difference between real and pseudo pairs, suggesting that (1) participants started with similar levels of syntactic similarity in both joint and individual conditions, and that (2) by the end of individual trials syntactic similarity between pair members was at chance level. In other words, by telling separate stories, participants' contributions gradually became independent. This could also explain the overall syntactic similarity effect of interactivity. These findings are in line with the predictions of IAM, and generally agree with the results of Branigan et al. (2007), who observed stronger syntactic priming in a picture matching task when a speaker was the addressee in the previous trial than when she was a side-participant. Also, Branigan et al. (2007) reported priming in side-participant (non-interactive) trials as well, similar to our findings in individual trials (see Pickering and Ferreira, 2008 about structural repetitions in non-interactive experiments).

Regarding lexical and semantic alignment we found larger average similarity for joint than individual trials, showing that interactivity modulates alignment at these levels according to the predictions of IAM. For both features, we also observed larger similarity in the individual condition in real pairs than in pseudo pairs, proving the presence of alignment across all conditions and thus replicating a wide range of studies reporting lexical repetitions and semantic alignment in a variety of tasks and corpora (e.g. Doyle, Yurovski and Frank, 2016; Garrod and Anderson, 1987; Fusaroli et al., 2012; Fusaroli et al., 2017; Watson, Pickering and Branigan, 2004; etc.).

### 3.6.3. Percolation across different features.

We also tested percolation, that is, if there are positive relations between alignment measures of different features. We found percolation from lower to higher linguistic levels. Speech rate predicted semantic similarity, which in turn predicted lexical recurrence. Lexical recurrence further predicted semantic similarity, however, the extended model revealed that the coefficient of lexical similarity was the smallest in the joint condition with real pairs. This result suggests that the link between lexical recurrence and semantic similarity is not interaction-specific, but is inherent in the close link between these features. In other words, we would expect any two stories with high rates of lexical overlap to also overlap semantically. The relatively weaker effect in real pairs in joint condition might be due to a ceiling effect, as that condition already showed the highest levels of lexical and semantic similarity.

The percolation effects from speech rate through syntax to lexicon were specific to real pairs, verifying an important claim in IAM. Our results are novel as existing studies have focused on the relation between syntactic and lexical alignment (Mahowald et al., 2016 ).

### 3.6.4. Relative lack of results

Contrary to the claim of IAM, we did not find overall convergence in real relative to pseudo pairs in any speech feature except pitch. Moreover, instead of convergence, we observed general divergence in lexicon. Furthermore, while we found overall similarity in terms of syntax, lexicon and semantics, there was no overall similarity result in any prosodic feature. Regarding the correlations, there was a larger correlation across conditions in real relative to pseudo pairs only in pitch, but not in speech rate or pause length. Our study supports the findings of other experiments with similarly mixed results (Bonin et al., 2013; De Looze et al., 2014; Howes, Healey and Purver, 2010; Levitan and Hirschberg, 2011; Schweitzer and Lewandowski, 2013). While our analyses yielded many significant results, the relative lack of alignment somewhat questions its generality as a description of verbal coordination.

Furthermore, at least for the prosodic features, the size of convergence effects were rather small. For pause length, the pair-level difference was reduced by ~10 ms in the Same goal condition. For speech rate, convergence meant a change of ~0.12 syllable/sec over the whole trial. We suspect that such changes are below detection thresholds in the case of natural speech (for speech rate at least, see Pfitzinger, 1998; Street, Brady and Putman, 1983).

### 3.6.5. Summary

Overall, we found that the factor goal mainly had an effect on prosodic convergence and correlation, independent of interactivity, supporting that similar to behavioral mimicry effects, prosodic imitation is also influenced by the social context. This finding is also in line with speech accommodation theory (Giles et al., 1992) as sharing the same goals is a form of “likeness” and prosodic variations reflected that relationship between pair members.

On the other hand, the predictions of IAM were verified at linguistic levels. Interactivity elicited syntactically, lexically and semantically more similar contributions from participants, supporting the notion that in realistic interactions, people mutually predict each other and as a result, they become (or stay) more similar to each other linguistically.

Tentatively, the dissociation of effects at the border of prosody and syntax suggests that the processes behind *what* is said in a conversation are different from the processes influencing *how* we say it.

# Chapter Four. Neural coupling in verbal interaction

## 4.1. Introduction

Brain-to-brain coupling refers to a linear dependence of listeners' neural activity on speakers’ neural activity, as measured in one-way communication (Stephens et al., 2010; Silbert et al., 2014; Hasson et al., 2012). Here we extend the neural coupling framework to two-way communication using an fMRI hyperscanning setup. The aim of our study is to contrast coupling between one-way and two-way communication and capture the effects of interactivity, a defining difference between the two scenarios. In what follows, we first review existing findings and then discuss the challenges that led to the current study.

### 4.1.1. Brain-to-brain coupling in one-way communication

Most neuroscientific research on verbal communication has focused on identifying the neural correlates of linguistic categories and computations, usually treating speech perception and production separately (see the reviews by Hickok and Poeppel, 2007, 2015; Friederici, 2012; Hickok, 2014). While this research program has been very successful in identifying the details of the language system, by focusing on only one linguistic process at a time it has precluded the study of information transfer across agents, that is, the core feature of verbal communication.

An alternative approach considers verbal interaction as an inherently joint activity (e.g. Garrod and Pickering, 2009; Hasson et al., 2012), where the speech acts of an agent (i.e., production processes) directly effect the other’s neural state (comprehension processes), eliciting a response (production processes), and so on and so forth, establishing and maintaining a causal link across brains via a set of shared language codes. In other words, there is lawful relationship between the neural processes of communicating agents, often denoted as neural, or brain-to-brain coupling (Hasson and Frith, 2016). As the agents in the interaction continuously influence and constrain each other, any description including only one side is necessarily incomplete. Therefore, in order to treat the interaction itself as the primary element of interest (De Jaegher, Di Paolo and Gallagher, 2010), we need to extend our scope from individuals to multiple brains, in line with the call for a “two-person” (Hari and Kujala, 2009; Hari, Henriksson, Malinen and Parkkonen, 2015) or “second-person” neuroscience (Schilbach et al., 2013). Furthermore, as interactions are usually open-ended, their study requires a move towards naturalistic scenarios allowing relatively unconstrained behavior (Hasson and Honey, 2012).

Seminal experiments by Stephens et al. (2010) and Silbert et al. (2014) applied the brain-to-brain coupling framework to one-way verbal communication. They utilized a naturalistic storytelling scenario where first a speaker (multiple speakers in Silbert et al., 2014) told a real-life story while undergoing scanning. Then the recorded story was played to a separate group of listeners who were also scanned. Coupling across speakers and listeners was characterized by applying inter-subject correlation (ISC) analysis (Hasson, Nir, Levy, Fuhrmann and Malach, 2004; Mukamel et al., 2005) that uses the brain responses from one person to model the responses of another person. As idiosyncratic responses and noise are independent across brains, ISC detects only shared neural activity and has proved to be a useful technique in scenarios with rich, complex stimuli (Ben-Yakov, Honey, Lerner and Hasson, 2012; Kauppi, Jaaskelainen, Sams and Tohka, 2010; Lankinen, Saari, Hari and Kokinen, 2014). Employing ISC, Stephens et al. (2010) and Silbert et al. (2014) found that neural activity in speakers and listeners was coupled across a network of motor, linguistic and extralinguistic areas, including the superior temporal gyrus (STG), the temporal poles (TP), the temporo-parietal junction (TPJ) and the inferior frontal gyrus (IFG) bilaterally, and the precuneus (PC), posterior cingulate (PCC) and medial prefrontal areas (mPFC). While in most areas listeners' brain activity lagged behind that of speakers', intriguingly, listeners' neural responses in mPFC preceded speakers' responses, suggesting ongoing prediction. Furthermore, the strength of coupling in general, and the strength of predictive coupling in particular was associated with how well listeners comprehended the story (see also Dikker, Silbert, Hasson and Zevin, 2014 about predictions and ISC). Importantly, a series of further studies have showed that brain-to-brain coupling in higher-level areas is not merely an effect of the shared auditory environment, but signals a shared understanding of the narrative content (Honey, Thompson, Lerner and Hasson, 2012; Lerner, Honey, Silbert and Hasson, 2011; Yeshurun et al., 2017), suggesting that such coupling could be the neurophysiological correlate of conceptual alignment (Pickering and Garrod, 2004).

### 4.1.2. Brain-to-brain coupling in two-way communication

While the results from one-way communication studies are elucidating, most of our interactions are two-way, involving back-and-forth information exchange. To capture the conceptual difference between the two types of communication - as in the previous chapter - we use the term *interactivity*. Unlike one-way, two-way communication is interactive, meaning that the agents continuously influence each other's contributions while engaged in turn-taking. Interactivity does not only take precedence in everyday life, but – according to the influential interactive alignment model (IAM) - is also construed as central for developing a shared understanding during communication, making it a desirable target of study (Pickering and Garrod, 2004, 2013; Garrod and Pickering, 2009).

However, extending the research on neural coupling to two-way interactions comes with a price, namely the need for simultaneous recordings from multiple scanners (termed hyperscanning, for reviews see Babiloni and Astolfi, 2014; Konvallinka and Roepstorff, 2012; Schoot, Hagoort and Segaert, 2016). While the first fMRI hyperscanning study was published more than 15 years ago (Montague et al., 2002), such setups are still considered unique and pose technical challenges (Bilek et al., 2015). Due to these difficulties, there have only been two fMRI hyperscanning studies published so far directly relevant to two-way communication (we explore the implications of relevant research from other imaging modalities - mainly from fNIRS - in the discussion).

The first one is by Spiegelhalder et al. (2014), and - to our knowledge - is the only study that has investigated live verbal communication with dual-fMRI. In their experiment, pairs of participants were shown descriptions of life events (e.g. “being lied to”) and either one of them was asked to describe such an event while the other listened (speaker-listener trials) or they were both instructed to imagine such an event on their own (imagery trials). In speaker-listener trials participants repeatedly switched roles. By using speakers' averaged motor and premotor activity as regressors for listeners' responses, Spiegelhalder et al. (2014) found evidence of neural coupling during speaker-listener trials. Speakers' motor-related activity was predictive of listeners' activity in auditory and medial parietal areas. These results go beyond locally specific alignment of neural activity as they describe coupling across speaker and listener but also across different regions. However, while Spiegelhalder et al. (2014) recorded from pairs of participants simultaneously, their study consisted of a series of one-way communication steps and did not explore interactivity that would have necessitated hyperscanning.

Another study by Stolk et al. (2014) focused on the development of mutual understanding in an interactive, non-verbal communication game. In their task, a communicator sent location and direction information to an addressee by moving a token on a grid. Importantly, participants had to develop the mapping between movements and potential meanings over the course of multiple trials, with task success yielding a measure for the development of conceptual alignment. They found that cerebral coherence across the communicator and the addressee in the right superior temporal gyrus was linked with mutual understanding and could not be explained by stimulus properties or rhythms inherent to the task. Strictly speaking, the communication game employed by Stolk et al. (2014) did not require interactivity, as task-relevant information traveled in only one-way, from the communicator to the addressee. However, in practice, communicators also had to interpret addressees' actions to establish a successful convention, as evidenced by detailed behavioral analysis of the task (see de Ruiter et al., 2010).

While we consider both works groundbreaking, they are only partially relevant to interactive verbal communication. The study by Spiegelhalder et al. (2014) is in many ways similar to those by Stephens et al. (2010) and Silbert et al. (2014), as their task focused on short segments with one-way communication. Stolk et al. (2014) did investigate an interactive situation, however, their task was abstracted away from naturalistic communication. Our aim with the present study was to complement earlier approaches by both (1) focusing on verbal communication and (2) manipulating interactivity, allowing us to compare one-way and two-way interactions in a brain-to-brain coupling framework.

### 4.1.3. Testing the role of interactivity

Several models of communication argue that mutual adaptation and prediction in bilateral, interactive situations elicit stronger and faster coupling of neural systems yielding shared understanding and successful communication (Friston and Frith, 2015; Gallotti et al., 2017; Pickering and Garrod, 2013). Hence we expect stronger and more predictive coupling in interactive (two-way) than in non-interactive (one-way) verbal interactions. This hypothesis is supported by the results of Stephens et al., 2010) who found that - at least in the context of story comprehension, - successful communication was linked with stronger brain-to-brain coupling in an area exhibiting predictive activity in listeners (Stephens et al. (2010). Further support is provided by joint action studies showing that interaction elicits stronger motor predictions regarding the partner's actions relative to observation (Kourtis et al., 2010, 2013; Menoret et al., 2014; see also Wilson and Knoblich, 2005).

To test our hypothesis, we employ a version of the joint storytelling paradigm introduced in the previous chapter, while simultaneously recording neural responses from both pair members using fMRI scanners. We manipulate interactivity again by asking pairs to invent stories either jointly or individually.

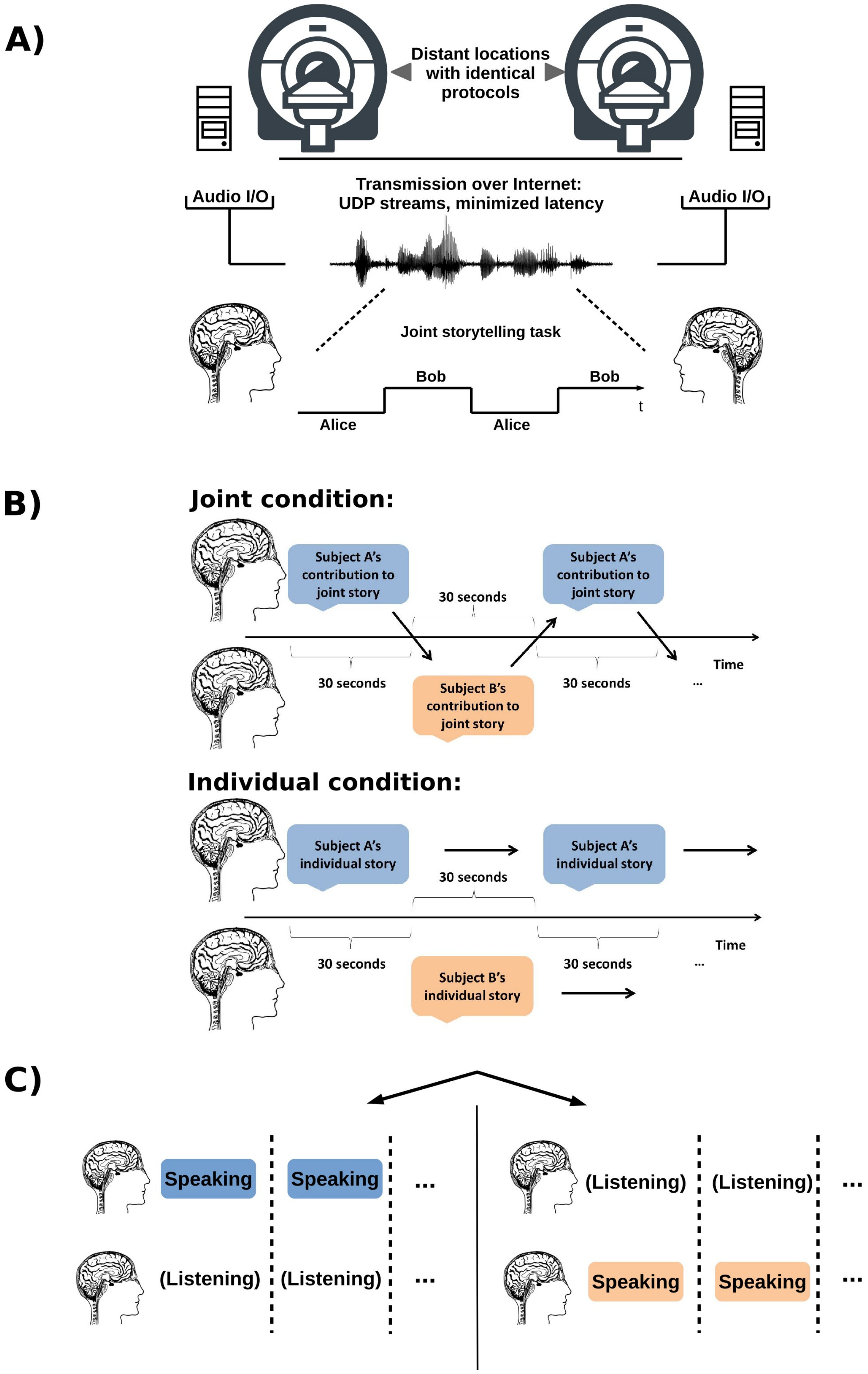

A)
Distant locations with identical protocols
Audio I/O
Transmission over Internet: UDP streams, minimized latency
Audio I/O
Joint storytelling task
Bob
Bob
Alice
Alice
t
B)
Joint condition:
Subject A's contribution to joint story
30 seconds
Subject A's contribution to joint story
Time
30 seconds
30 seconds
...
Subject B's contribution to joint story
Individual condition:
Subject A's individual story
30 seconds
Subject A's individual story
Time
30 seconds
30 seconds
...
Subject B's individual story
C)
Speaking
Speaking
...
(Listening)
(Listening)
...
(Listening)
(Listening)
...
Speaking
Speaking
...

*Figure 4.1.* Outline of the hyperscanning setup and the task. A) The two fMRI suites were connected over the internet, using a dual UDP stream of audio. B) Pairs of participants were asked to improvise stories together. We manipulated the interactivity of storytelling. In the Joint condition, participants told one story together, by taking turns. In the Individual condition, they both told a story individually, while keeping the same turn-taking structure. C) For ISC analysis, the storytelling data was segmented according to speaker-listener roles. Each story was divided into two speaker-listener streams.

## 4.2. Methods

### 4.2.1. Participants

18 subjects (in 9 pairs) took part in the study. We excluded 1 pair due to excessive movement, resulting in 16 subjects (8 pairs) as the final sample (11 females, mean age = 26.80, *SD* = 5.43, two left-handed). Most of our participants were recruited from the graduate student population of the Department of Psychological and Brain Sciences. They all had normal or corrected-to-normal vision and hearing. All subjects gave written informed consent. Study procedures were approved by the Committee for the Protection of Human Subjects at Dartmouth College. Participants were reimbursed financially after participation (40 USD).

### 4.2.2. Apparatus

Figure 4.1A provides a schematic overview of the setup. An fMRI hyperscanning study of verbal communication brings about unique challenges (see e.g. Bilek et al., 2015, Baecke et al., 2015). We were primarily concerned with the following four aspects: hardware and settings at different research sites, recording speech in the scanner, movement while speaking and controlling timing using a network connection. We address each of these issues below.

The study was conducted at two distant sites ~130 miles apart, at Dartmouth Brain Imaging Center (DBIC), Dartmouth College, NH and at the Center for Brain Science (CBS), Harvard University, MA. We relied on identical scanner hardware and protocols at both sites (see image acquisition details below). The same type of microphones and earphones (see below) were used, and the two experimental control computers had identical software environment, and nearly-identical hardware. To our best knowledge, no major equipment or protocol detail was different across sites.

For audio recording we used FOMRI-III (Optoacoustics, www.optoacoustics.com) MRI compatible optical microphones with active noise-cancellation (up to 40 dB). The noise-cancellation feature is achieved by the use of two orthogonal microphones built into the device, sampling both the sound of interest and the background noise simultaneously. Noise-cancellation is implemented online, in a near-immediate fashion making it possible to record intelligible speech from both participants in a pair. Sound was displayed via the same type of devices at both sites, two sets of S14 MRI compatible earphones by Sensimetrics (www.sens.com).

In order to mitigate the effects of speech in terms of movement, participants wore personalized headcases (foam helmet stabilizers) from CaseForge (www.caseforge.co). Our participants all had previous experience with fMRI studies, and movement for 16 of the original 18 participants stayed below 2 mm maximum displacements. The subjects with excessive movement were excluded. Subjective reports indicated that the headcases were comfortable to wear, even for longer periods.

Synchronization across research sites was achieved by a simple mechanism. Task events were synchronized to a common timeline across control computers but not to scanner backticks. At the beginning of each trial, the control software negotiated a common start time across the sites, and all events were timestamped relative to the common timeline. In order for this approach to work, we had to fulfill three assumptions: (1) high degree of synchrony across control computer clocks to start with, (2) subsequent monitoring of network transmission time and computer clock drift, and (3) subsequent interpolation of images to align them temporally across the sites (see below at temporal alignment). Regarding synchrony across computer clocks, the two control computers were aligned to the same Network Time Protocol stratum 2 servers at Dartmouth College using ntpd (http://doc.ntp.org/4.1.0/ntpd.htm), providing a synchrony of ~3-5 ms (as tested with timestamped packet exchanges). Clock drift and network transmission time were both estimated from the timestamps attached to each audio packet sent over the network. Assuming a constant clock difference for a given trial and that network transmission time on average is equal in both directions, by aggregating timestamp data from both packet stream we could estimate both clock drift and network transmission time. For each pair and trial, clock drift estimates were in the range of 4 - 10 ms, while network transmission time was in the range of 6-11 ms. Occasional jumps in network transmission time (> 30 ms) affected less than 0.5% of the packets. As we employed a continuous buffer of 512 ms, transmission jitter proved to be inconsequential.

Experimental control software was written in Python. Audio connection was implemented as dual UDP streams of uncompressed 16bit audio with fixed buffer size, relying on the PyAudio package (v0.2.10). Visual displays were controlled via PsychoPy functions (v1.85, Pierce et al., 2009).

### 4.2.3. Stimuli and procedure

Participants were randomly assigned to pairs and performed the study together while undergoing scanning at DBIC and CBS. We asked participants to perform the storytelling task described in the previous chapter (30 speech turns, 30 s each; visual display of time left, turn number and speaker-listener roles). However, unlike in the behavior study, we did not provide any instructions regarding emotional valence trajectories and, thus, were manipulating only the factor Interactivity (Joint vs. Individual, see Fig. 4.1B). Due to this change the participants performed only two trials, one Joint and one Individual, the order of which was counterbalanced across pairs. For these two trials the topics were “Alien encounter” and “Family with child in trouble,” counterbalanced across trial types. Following the two storytelling trials and their re-tell phases, participants performed two additional tasks in the scanner individually: an audio listening and a reading task. The results of the latter two runs are not discussed in the present paper. At the end of the experiment, participants filled out a short questionnaire about their experience, were debriefed and reimbursed.

### 4.2.4. fMRI image acquisition

We used a 3.0 Tesla Siemens Magnetom Prisma whole-body MRI system (Siemens Medical Solutions, Erlangen, Germany) with a 32-channel head coil at both sites. The same scanning protocol was employed at the two sites.

We used gradient-echo echo-planar imaging for the four task runs with SMS (simultaneous multi-slice) of 4. The parameters were as follows: TE, 32 ms; TR, 727 ms; flip angle, 53°; resolution, 3 × 3 mm; matrix size, 80 × 80; FOV, 240 × 240 mm; 40 transverse slices with full brain coverage, interleaved order; slice thickness, 3 mm. At the beginning of the scanning session, we also acquired a high-resolution T1-weighted (T1w) anatomical scan (TE, 2.32 ms; TR, 2300 ms), with a voxel resolution of 1 × 1 × 1 mm, FOV 240 × 240 × 220 mm. B0 distortions were computed from gradient-recalled images with different echo-times (Jezzard and Balaban, 1995; Hutton, Bork, Josephs et al., 2002). For fieldmap acquisition we used echo times of 4.92 and 7.38 ms, flip angle 60° and voxel resolution of 3 × 3 × 3 mm.

### 4.2.5. fMRI data preprocessing

For most preprocessing steps we used the FMRIPREP pipeline (v1.04, Esteban et al., 2018). Many internal operations of FMRIPREP use Nilearn (Abraham et al., 2014), principally within the BOLD-processing workflow. For more details of the pipeline see http://fmriprep.readthedocs.io/en/stable/. Parts of the following three paragraphs were generated using the citation tool of FMRIPREP at https://fmriprep.readthedocs.io/en/stable/citing.html.

Each T1w volume was corrected for intensity non-uniformity using N4BiasFieldCorrection (from ANTs, Tustison et al., 2010, v2.1.0 for all ANTs tools) and skull-stripped using antsBrainExtraction (using the OASIS template). Brain surfaces were reconstructed using recon-all from FreeSurfer (Dale, Fischl & Sereno, 1999, v6.0.0 for all Freesurfer tools), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov, Evans, McKinstry, Almli & Collins, 2011) was performed through nonlinear registration with the antsRegistration tool (Avants, Epstein, Gorssman & Gee, 2008), using brain-extracted versions of the T1w volume and the template. Brain tissue segmentation of cerebrospinal fluid, white-matter and gray-matter was performed on the brain-extracted T1w using FAST (from FSL, Zhang, Brady & Smith, 2001, v5.0.9 for all FSL tools).

Functional data was slice time corrected using 3dTshift (from AFNI v16.2.07, Cox, 1996) and motion corrected with MCFLIRT (Jenkinson, Bannister, Brady & Smith, 2002). Distortion correction was performed using fieldmaps processed with FUGUE (from FSL, Jenkinson, 2003). This was followed by co-registration to the corresponding T1w using boundary-based registration with 9 degrees of freedom (with bbregister from FreeSurfer, Greve & Fischl, 2009). Motion correcting transformations, field distortion correcting warp, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using Lanczos interpolation (with antsApplyTransforms from ANTs).

Physiological noise regressors were extracted by applying CompCor (Behzadi, Restom, Liau & Liu, 2007). Principal components were estimated for anatomical CompCor variants (aCompCor). Six components were calculated within the intersection of the subcortical mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run. Frame-wise displacement (Power et al., 2014) was calculated for each functional run using the implementation of Nipype (Gorgolewski et al., 2011, 2017). Nuisance regression was performed with the six motion parameters, framewise displacement and the first five aCompCor components using Nilearn (v0.4.1). Voxelwise time series were extracted from the nuisance-regressed data by applying the brain mask of the template using Nilearn. Time series were standardized.

### 4.2.6. Temporal alignment and data parsing

Following nuisance regression and time series extraction, we temporally aligned the recordings from the two sites. For this step, we estimated average network transmission time and computer clock drift from the timestamps of the audio packets separately for each pair. Then in one interpolation step (using shape-preserving piecewise cubic interpolation) we adjusted for network transmission time, computer clock differences, audio buffer length and also adjusted our TRs so that their spacing was the same for each speech turn relative to speech turn start. The last step was required as we parsed our data along speech turn boundaries and speech turn length (30 s) was not a multiple of TR (0.727 s). The interpolation and all further analyses steps were carried out with custom scripts written in MATLAB.

To characterize coupling between pair members, we utilized the brain-to-brain coupling framework relying on ISC (Stephens et al., 2010, Silbert et al., 2014). Since participants in our task switched roles in every turn, we first segmented our data into continuous speaker-listener streams by separately concatenating fMRI data from odd and even turns (shown in Fig. 4.1C). In other words, all speech turns where participant “Alice” was the speaker and participant “Bob” was the listener were concatenated into one 4D image, and all turns with opposite roles were aggregated into another 4D image. Thus, our dataset was parsed into two separate speaker-listener time series in each condition for each pair. By doing so, we obtained a speaker-listener time series for each participant as a speaker, resulting in 16 – 16 data sets for the Joint and Individual trials.

### 4.2.7. Speaker-listener coupling

Next, for each speaker-listener pairing and for each voxel we formed a linear model of listener’s time course using a linear combination of speaker’s time-shifted voxel time series as the model (Fig. 4.2). Formally, we solved the equation below for the beta weights in the least-squares sense:

$$v_{listener}^{model}(t) = \sum_{\tau=-\tau_{max}}^{\tau=\tau_{max}} \beta_i v_{speaker}(t+\tau)$$

Here $v_{listener}(t)$ is the time course of the voxel in listener’s brain, $\beta_i$ are the linear coefficients corresponding to each time shift and $v_{speaker}(t+\tau)$ is the time course of speaker’s voxel shifted by τ*TR in time. The maximum time shift, $\tau_{max}$ was set to 6 TRs (4.36 s), as a trade-off between the number of model parameters and the time window covered by the time shifts. We have not explored the behavior of the model with different maximum time lags, but note that Stephens et al. (2010) found their results to be stable with regards to changes in the maximum shift. For each speaker-listener pairing the model was characterized by the map of model fits ($R^2$ values). Please note that while our approach did not consider the potential effects of temporal autocorrelation in the BOLD signal, Stephens et al. (2010)

showed that for very similar time shifts, the autocorrelation is not substantial and the results are virtually unaffected when the predictor variables are decorrelated.

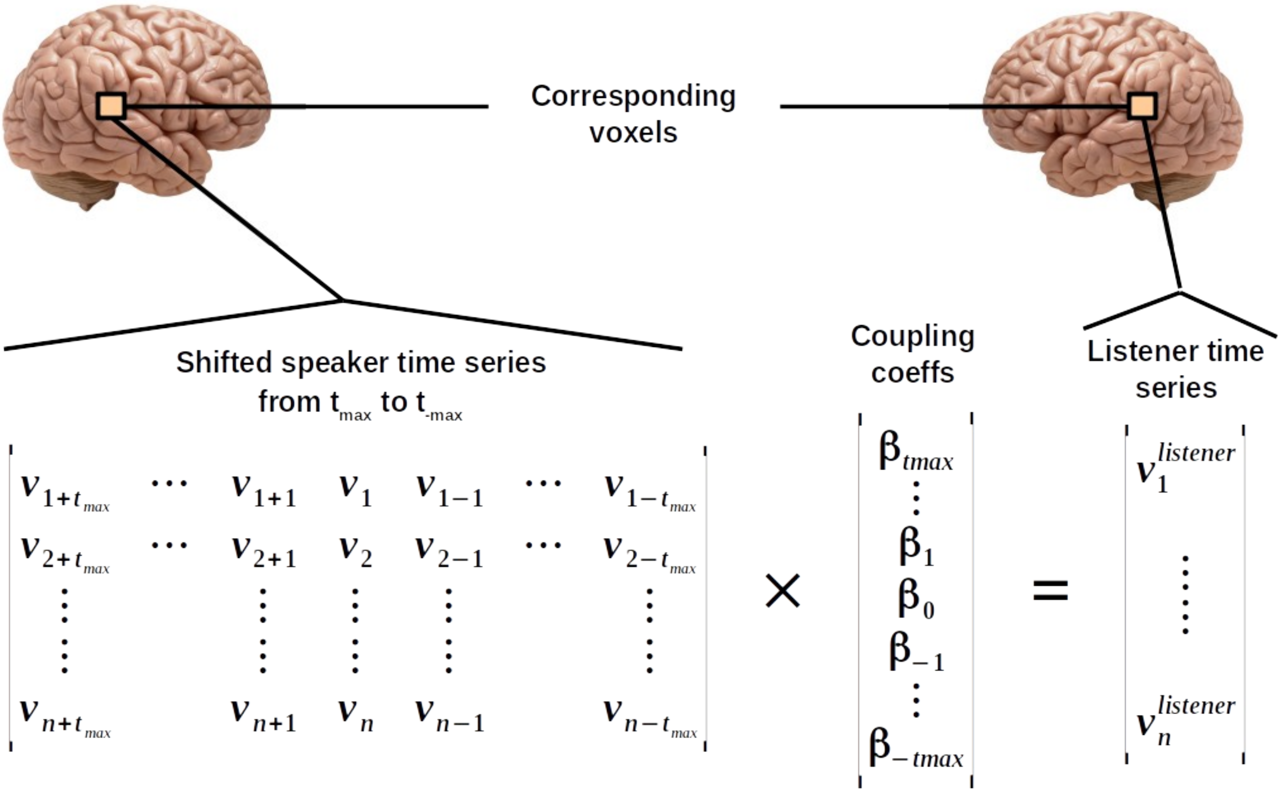


*Figure 4.2.* Core ISC analysis. For each voxel, time-shifted voxel timeseries from the speaker are used to predict listener voxel time series in a simple linear model sense.

### 4.2.8. Average coupling

For a group-level coupling map, we wanted to be able to test against the null hypothesis that the observed coupling on the pair-level is not due to pair-specific processes but due to more general, situation-specific factors (specifically, movement due to speaker-listener role changes at the speech turn boundaries, periodically repeated preparations for speech-turn changes or the shared visual environment—that is, the timer and the instructions on the screen). To this end, we first calculated coupling in pseudo-pairs using data from speakers and listeners from different pairs. We only used pseudo pairs where the two participants were from different measurement sites, thus we modeled speaker-listener coupling in 56 pseudo-pairs, resulting in 112 – 112 Joint and Individual trial pairings. This data set corresponded to the null hypothesis outlined above. Then we performed a random permutation test with 10,000 permutations for each voxel using the real- and the pseudo-pair data with the mean model fit as the test statistic. The estimated voxelwise *p* values were corrected with the FDR procedure at *q* = 0.05 (Benjamini & Hochberg, 1995, Benjamini & Yekutieli, 2001).

We performed the above analysis with all data aggregated across conditions, but also separately for the Joint and Individual trials.

### 4.2.9. Joint vs. Individual contrast

To identify areas where speaker-listener coupling was different across Joint and Individual trials, we compared the differences between Joint and Individual model fit maps in real pairs to the same differences in pseudo pairs. A random permutation test with 10,000 permutations estimated the significance of any difference between real and pseudo pairs.The resulting map of *p* values was then corrected with FDR (*q* = 0.05).

### 4.2.10. Temporal lag in coupling

To analyze the time shift characterizing the speaker-listener coupling best, for each voxel that predicted the listeners reliably we performed a contrast analysis on the linear weights of speakers’ shifted voxel time series. We defined three categories of time shifts:

delayed (speaker precedes, with time lags from -1.5 to - 4.4 s), synchronous (with time lags from -1.5 to +1.5 s) or advanced (listener precedes, with time lags from +1.5 to +4.4 s). For each category, a contrast was used to evaluate the weights corresponding to that particular time shift relative to the weights outside the category. The sum of contrast weights always added up to zero. For example, for the delayed (speaker precedes) category the contrast was defined as

$$\left[0.25, 0.25, 0.25, 0.25, -0.125, -0.125, -0.125, -0.125, -0.125, -0.125, -0.125, -0.125, -0.125\right]$$

. After applying the contrast, we assigned *p* values by performing a random permutation test against pseudo pair contrasts with the mean contrast value as the test statistic. We applied FDR ($q$ = 0.05) correction to the results.

## 4.3. Results

### 4.3.1. Average speaker-listener coupling

To create a general speaker-listener coupling map we averaged over all stories and pairs and tested the results against pseudo-pair data. After correcting for multiple comparisons we found a number of areas coupled across speakers and listeners (see Figure 4.3): the bilateral superior temporal gyri (STG), including both the primary auditory cortex (A1) and associative areas involved in speech processing; the bilateral precentral gyri and its neighboring areas, including mainly the ventrolateral part of primary motor cortex (MC), with overlap into premotor (PMC) and - to some degree - primary somatosensory cortices (SC); and the bilateral supplementary motor area (SMA). We observed a slight lateralization with larger clusters in the left hemisphere, with the largest difference in the SMA.

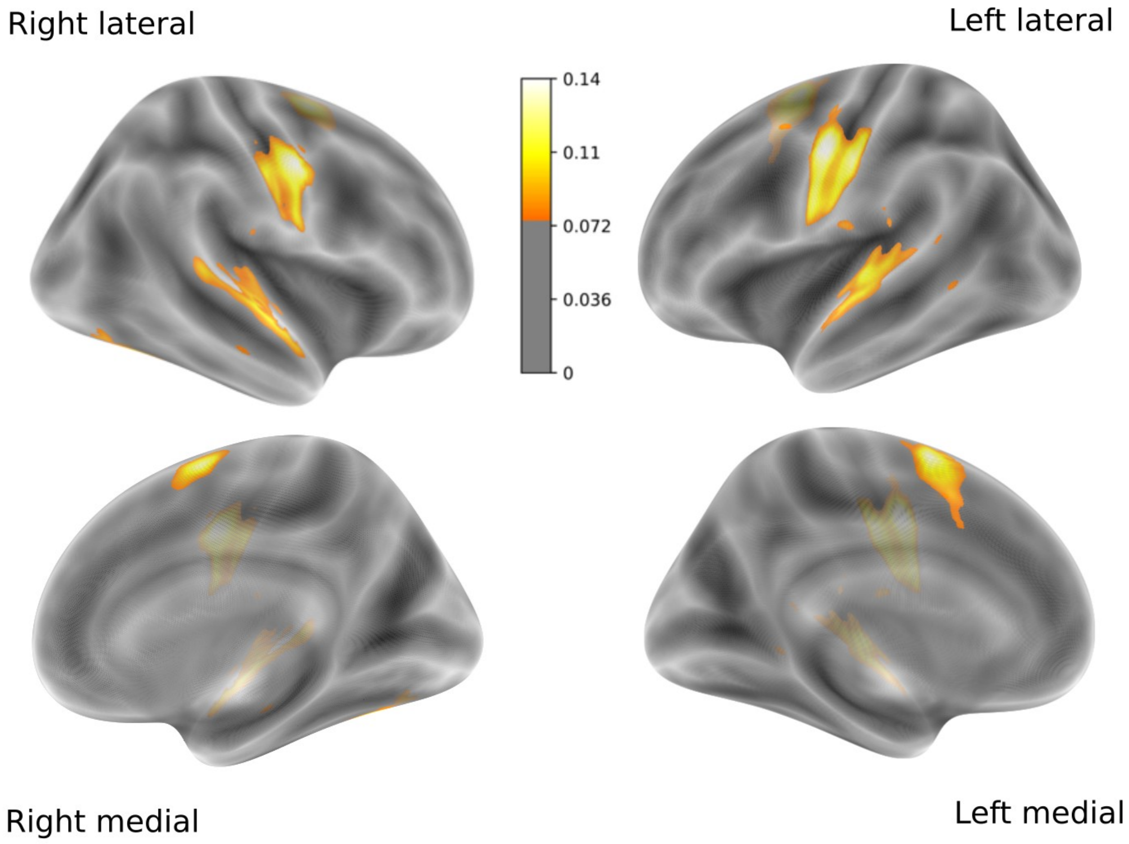


*Figure 4.3.* Average speaker-listener coupling map, lateral and medial views of both hemispheres. Voxelwise statistical significance was derived from the model fit map averaged over Joint and Individual trials and tested against pseudo-pair data with random permutations (FDR corrected). The color bar depicts coupling model fit strength (averaged $R^2$ values from the coupling models of significant voxels). Coupling map is projected on the fsaverage surface template.

### 4.3.2. Joint vs. Individual trials

Figure 4.4 and Figure 4.5 depict the averaged speaker-listener coupling maps as calculated separately for the Joint and Individual conditions. The contrast was significant in only one cluster (see Figure 4.6), the anterior portion of the left SMA.

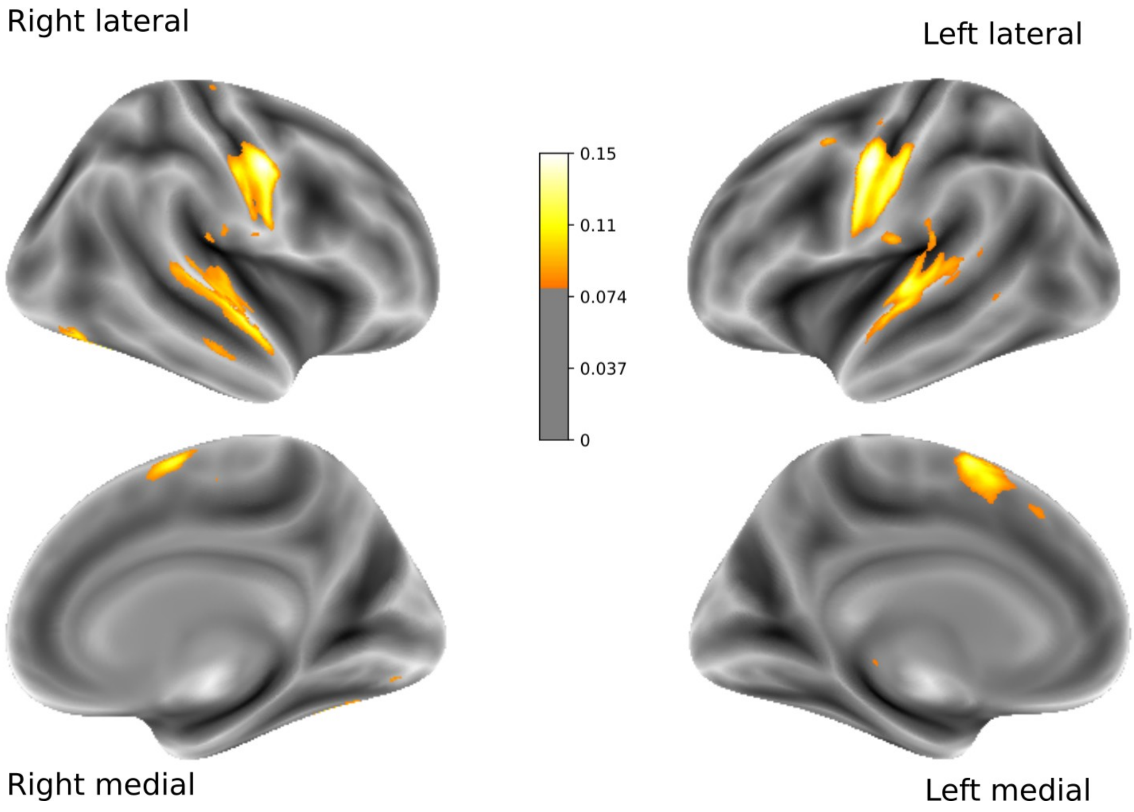


*Figure 4.4.* Average speaker-listener coupling in the Joint condition, lateral and medial views of both hemispheres. Voxelwise statistical significance was derived from the model fit map averaged over Joint trials and tested against pseudo-pair data with random permutations (FDR corrected). The color bar depicts coupling model fit strength (averaged $R^2$ values from the coupling models of significant voxels). Coupling map is projected on the fsaverage anatomical surface.

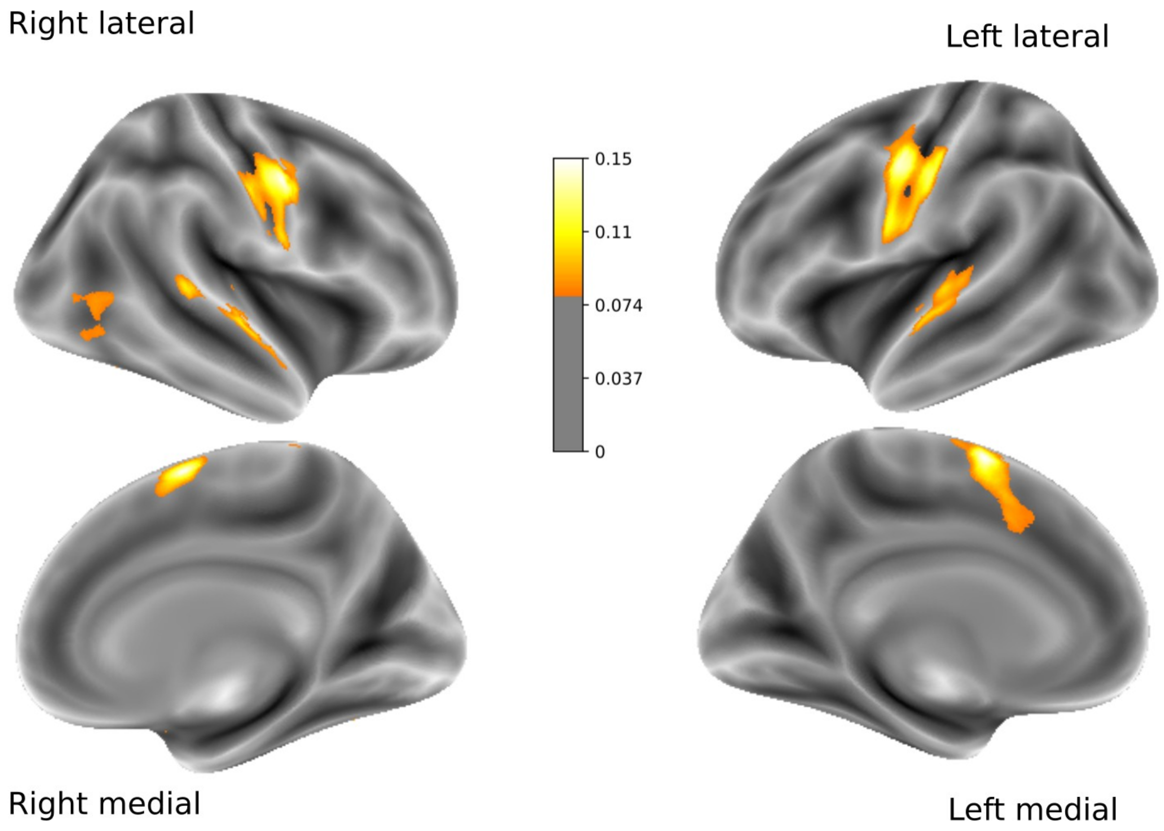


*Figure 4.5.* Average speaker-listener coupling in the Individual condition, lateral and medial views of both hemispheres. Voxelwise statistical significance was derived from the model fit map averaged over Individual trials and tested against pseudo-pair data with random permutations (FDR corrected). The color bar depicts coupling model fit strength (averaged $R^2$ values from the coupling models of significant voxels). Coupling map is projected on the fsaverage anatomical surface.

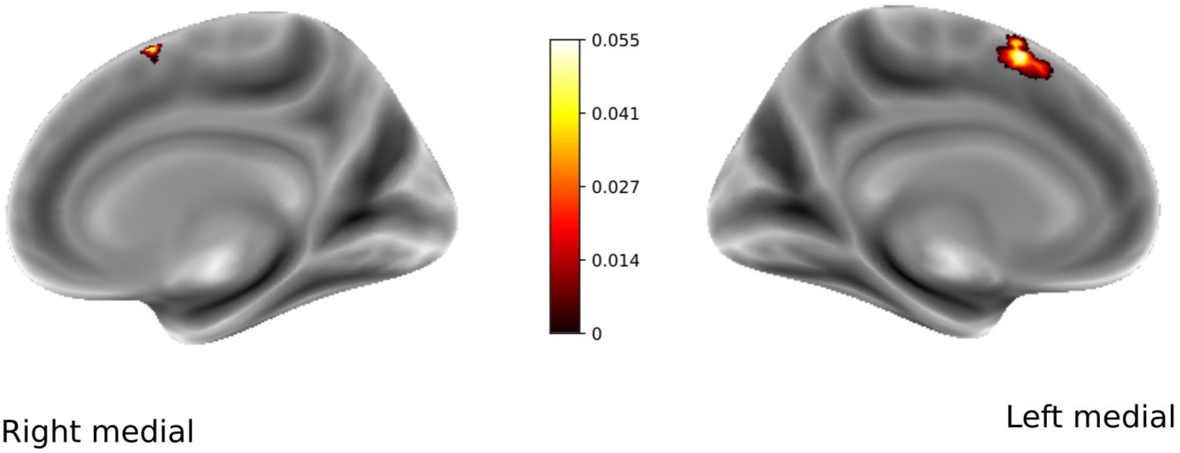


*Figure 4.6.* Joint vs. Individual condition contrast. To compare coupling across conditions, we compared model fit differences between Joint and Individual conditions across real and pseudo pairs (voxelwise random permutation test with FDR correction). Colors show the difference in average $R^2$ values between the conditions, with positive values indicating that coupling was stronger in the Joint than in the Individual condition. The difference can be seen in a small cluster in the left SMA (centered at -3, 9, 55 in MNI coordinates). Coupling contrast map is projected on the fsaverage anatomical surface.

### 4.3.3. Time shifts in speaker-listener coupling

We applied a contrast analysis to the linear weights of the shifted voxel time series model to detect the temporal lags captured by the coupling model. The results are depicted in Figure 4.7.

We found synchronous coupling (time shift between -1.5 and +1.5 s) in the bilateral MC and PMC. Interestingly, the anterior part of the SMA, where we found stronger coupling in the Joint than in the Individual condition, fell into the “listener precedes” category, suggesting that the anterior SMA is involved in predicting speakers’ behavior primarily in the Joint condition.

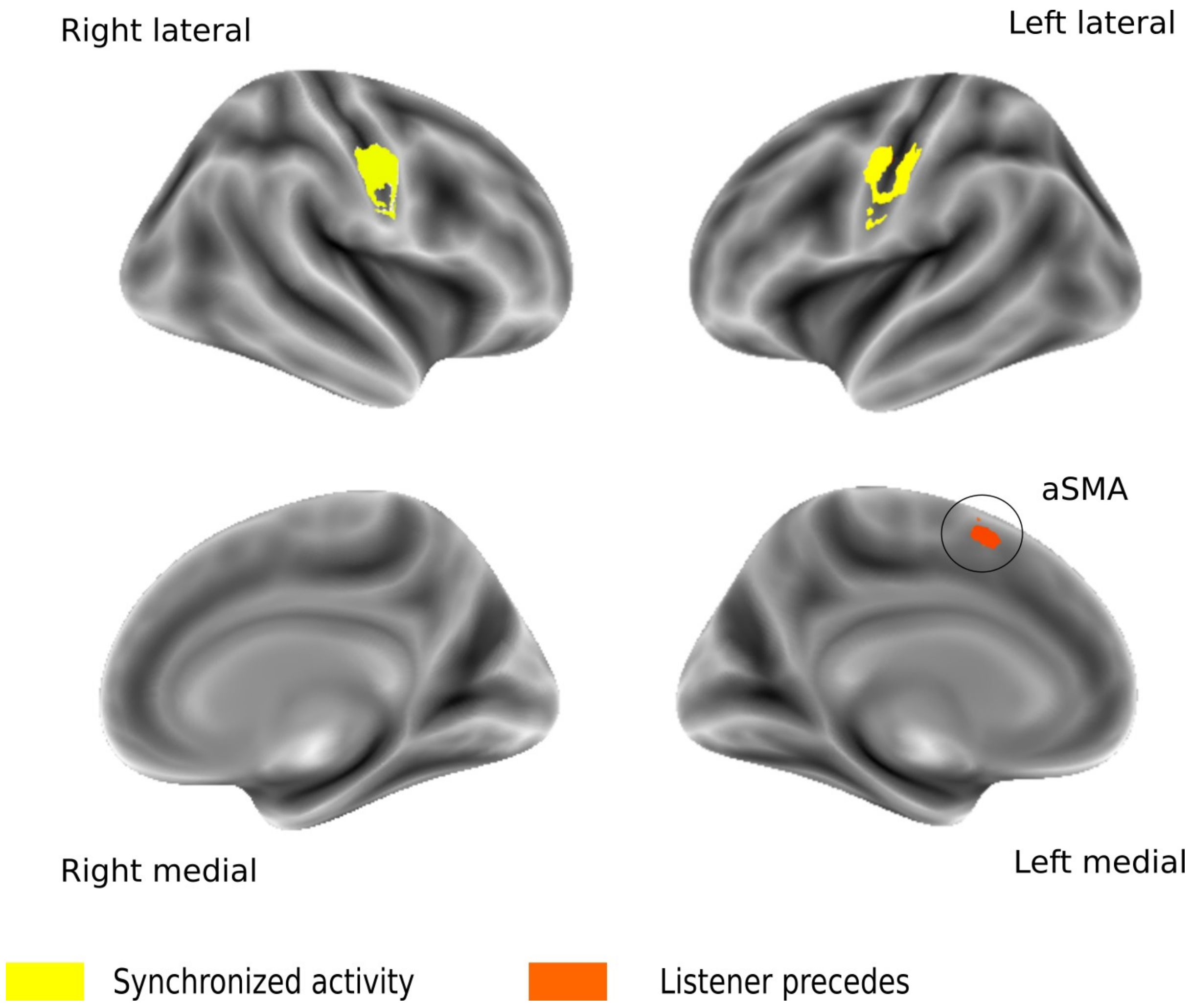


*Figure 4.7.* Time lags in speaker-listener coupling. Yellow color indicates that listeners' activity on average is synchronous with that of speakers' (time lags from -1.5 to +1.5 s). Red color indicates that listeners' activity on average precedes that of speakers' (time lags from +1.5 to +4.4 s). There was synchronized activity in bilateral MC (overlap with PMC and SC), while listeners' neural activity preceded speakers' activity in the left anterior SMA. No coupling with a reliable 'speaker precedes' profile was found. Time lag map is projected on the fsaverage anatomical surface. aSMA: anterior supplementary motor area.

## 4.4. Discussion

In our study we extended the brain-to-brain coupling framework to two-way communication and tested the effects of interactivity on the network of coupled brain areas. In summary we found that two-way communication results in (1) greater coupling in the SMA and (2) in the same area we find that listeners' activity precedes speakers', suggesting more pronounced prediction processes in the real interactions.

### 4.4.1. Overall coupling across conditions

While the two tasks were different, it is informative to compare our results to previous findings from Stephens et al. (2010) and Silbert et al. (2014). The difference between the coupling maps of the Joint and Individual conditions is not relevant for this comparison (see more regarding the difference below), hence we focus on the averaged coupling map (Figure 4.4). There is overlap with previous results. Namely, we found coupling in auditory areas (bilateral A1 and STG), in bilateral MC and PMC, and in bilateral SMA. With the exception of MC, these regions were all reported as part of the overlapping production-comprehension network in Silbert et al. (2014). Importantly, we see the reproduction of coupling results both in areas traditionally more associated with speech comprehension (A1 and STG, see e.g. Saur et al., 2010) and in areas classically involved in speech production (bilateral MC, PMC and SMA, see e.g. Meister, Wilson, Deblieck, Wu & Iacoboni, 2007). Besides the overlap though, there are also important differences. First, we see widespread coupling in bilateral MC and PMC, affecting larger areas than reported by Stephens et al (2010) or Silbert et al (2014), even over-spilling into primary somatosensory areas. This is a novel coupling result not reported before. Second, we did not find coupling in a wide range of areas relative to Silbert et al. (2014): bilateral TPJ,  temporal poles, middle temporal gyrus, PC, PCC and mPFC. While these areas were not all reported by Stephens et al. (2010), the bilateral TPJ, PC, PCC and the mPFC are identified as part of the coupled speaker-listener network in their findings too.

The finding that coupling in bilateral motor areas is more widespread than in previous experiments is not necessarily surprising. There are both empirical and theoretical considerations that support the role of MC, PMC and SMA both in speech production and comprehension, and as such, these areas are natural candidates to exhibit coupling across interlocutors.

Regarding the primary motor cortices, there is ample evidence that MC are involved in speech processing (e.g. Correia, Jansma and Bonte, 2015; Pulvermüller and Fadiga, 2010, 2016; for a review see Hickok and Poeppel, 2016), even if the necessity of this activity for comprehension is a contentious issue (e.g. Hickok, Houde and Rong, 2011). Furthermore, MC activity in speech processing is often accompanied with somatosensory activations (Schomers and Pulvermüller, 2016).

PMC and SMA are involved in speech production, more specifically in speech act planning and feedback monitoring (e.g. Parkinson, Flagmeier, Mansen et al., 2012; Behroozmand, Shebek & Hansen et al., 2016). Importantly, PMC and SMA are also thought to play an important role in comprehension. In a series of fMRI studies, Wilson et al. (2004, 2007, see also Iacoboni, 2008) found bilateral premotor activations in speech comprehension, comparable to that in production, and a TMS study (Meister et al., 2007) further supported the causal role of these regions in speech perception. Frameworks that try to account for the overlap between comprehension and production systems (e.g. Friederici, 2012; Hickok et al., 2011; Rauschecker & Scott, 2009) postulate that PMC generates predictions in the form of articulatory motor plans, performing a very similar role in both speech production and speech comprehension. Similarly, SMA is speculated to perform the same action regulatory role in both production and comprehension (Hertrich, Dietrich & Ackermann, 2016).

Relative to the Stephens et al. (2010) and Silbert et al. (2010) experiments, our task involved turn-taking in both the Joint and the Individual trials, and this aspect might be the reason behind stronger coupling in motor areas in our experiment. Pair members needed to switch roles repeatedly, meaning that they had to monitor speech turn boundaries and prepare for either to start or to stop talking. We parsed our data in a manner that avoided actual turn-taking for each speaker-listener pairing we modeled. Nevertheless, turn-taking related processes were still involved in the task situation. We would like to point out though that—in our opinion—the speech-turn based structure of our data could not have induced the couplings we observe in PMC and SMA. First, our coupling results reflect pair-specific processes as they were tested against pseudo-pairs. Second, we would expect pair members to show opposite patterns at turn boundaries (preparation to stop vs. preparation to talk), and not simply mirror each other.

Instead, one specific way for turn-taking to yield strong coupling in PMC and SMA is through generating temporal predictions. Turn-taking is known to exhibit universal laws (Stivers et al., 2009), one of which is the minimization of silence between speech segments. It has also been shown that turn-taking is “too fast” for cognition and can only be achieved by predictive processes engaged during the other’s turn, the neural correlates of which can be detected (e.g. Bögels, Magyari & Levinson, 2015; for an overview see Levinson, 2016). In recent EEG and MEG studies on turn-taking (Bögels et al., 2015; Mandel, Bourguignon, Parkkonen & Hari, 2016) turn-taking - related effects were localized mainly to premotor / motor areas. In their review on this topic, Scott, McGettigan & Eisner (2009) have argued that the role of motor activity in speech comprehension is primarily to extract the underlying rhythms in speech, in order to generate temporal predictions—both for parsing information and to accurately react to speech segment endings. This view is in line with a more general understanding of the role that motor plans and motor activity play both in perception and joint actions—namely, that it predicts the timing of perceptual outcomes, enabling fast and accurate reactions (Wilson & Knoblich, 2005).

The above argument for the role of motor area-related predictive processes induced by turn-taking might explain the strong bilateral coupling in MC, PMC and in SMA, but it would not predict our negative findings—namely, the lack of coupling in the network of TPJ, the precuneus, the insula and the mPFC. A null result is hardly conclusive, so we can only speculate about the absence of these regions in the averaged coupling map. One possibility though is that the stories the participants invented in our task were less consistent than those in previous studies, presenting a challenge in terms of long timescale narrative integration. The turn-taking structure might have led to relative disruptions (large changes) in the narrative. It has been shown that theó regions in question have long temporal receptive windows, integrating information over many seconds, even minutes, and thus play an important role in tracking the content of a narrative story (Hasson et al., 2015, Lerner et al., 2011). It could be the case that the stories in our task are less coherent, more fragmented than the ones used in previous studies. Another possibility is that the data segmentation procedure we applied disrupted long integration windows more severely than short ones.

### 4.4.2. Joint vs. Individual trials

We hypothesized that in the Joint condition participants were engaged in constant coordination and adaptation to each other’s contributions, resulting in stronger predictive coupling relative to the Individual condition. The Joint - Individual coupling map contrast and the time lag analysis supported our hypothesis, revealing a cluster in the anterior part of the left SMA that (1) showed stronger coupling in the Joint condition and (2) was engaged in predictions as its activity preceded that of speakers’. These results suggest that coupling in the left SMA is linked to the unique requirements of two-way communication.

What is the role of SMA in verbal interactions? There is evidence for the involvement of SMA in auditory processing and imagination, including stimuli as emotional nonverbal vocalizations (e.g. laughter; for a recent review see Lima, Krishnan & Scott, 2016). However, SMA has also been linked to speech and language as an area involved in temporal predictions (Kotz & Schwartze, 2010; Hertrich et al., 2016). Both in production and comprehension, it has been argued that the anterior portion of SMA acts as a temporal organizer, providing structure for motor acts, and temporal predictions for parsing incoming information (Kotz & Schwartze, 2010). Thus, coupling in the area of the SMA could reflect coupling in terms of temporal structure. This role of the SMA in temporal predictions is further supported by the results of the time lag analysis: activity of left SMA in listeners preceded speakers' corresponding SMA activity. This finding suggests that SMA in listeners was involved in predictions and these predictive processes were weaker in the Individual trials. We assume that Joint trials require a closer, more attentive tracking of speaker's speech, as a response to current content is required from listener in the subsequent turn. However, predictions could be generated regarding different aspects of speakers' activity. One possibility is that such closer tracking initiates articulatory motor predictions starting from the SMA (as described for action observation by Schubotz, 2007), similar to an inner "shadowing" of the ongoing dialogue. Another possibility is that predictions are purely concerned with timing information, similar to our discussion of PMC coupling elicited by turn-taking. In either case, we find it a very interesting open question if such predictive coupling could be generalized to other situations and serve as a marker of a higher degree of engagement. We hope to learn more about SMA coupling in the future by by systematically analyzing the content and nonverbal features of the stories generated during the task, and relating them to SMA activity.

Regarding the temporal lags detected by the coupling model in other areas, we found synchronous coupling in bilateral MC and PMC. As our lag categories are quite long time windows, the synchronous coupling is consistent with the idea that PMC is generating timing predictions in listeners, as such predictions could happen very close in time to the actual stimuli. Interestingly, the contrasts revealed no significant lag category in temporal areas. A potential reason for this could be that the lag is at the boundary (around 1.5 s) of current categories.

### 4.4.3. Motor coupling and fNIRS results

While there are not many relevant fMRI hyperscanning studies, there has been a rise in fNIRS hyperscanning experiments in recent years. Several papers utilized the flexibility fNIRS offers for investigating interactions (Jiang et al., 2012; Jiang et al., 2015; Liu et al., 2016; Nozawa et al., 2016; Osaka et al., 2015). In general, all of these experiments report brain-to-brain synchrony in frontal areas but the exact results vary depending on the task and manipulation employed. For example, Jiang et al. (2012) studied face-to-face versus back-to-back dialogue and monologue and found that brain-to-brain wavelet coherence (in the left inferior frontal regions) was only present in the face-to-face condition. Liu et al. (2016) employed a joint Jenga game with cooperative versus obstructive conditions. During the game, participants were encouraged to freely discuss their strategy. Interestingly, their results showed increased coherence in right prefrontal cortex both during cooperation and obstructive interaction, relative to rest, suggesting that synchronization may not depend on shared goals. The considerable variability in fNIRS results is due, in part, to a small number of available channels and to the selection of varying regions for those channels. As such, different experiments cast light (literally and metaphorically) on different areas, resulting in an incomplete picture that continues to develop. Nevertheless, it could be the case that the marked involvement of premotor areas in our results is related to the repeated findings of frontal coherence in fNIRS studies. With the exception of the study by Osaka et al. (2015), the fNIRS hyperscanning research cited above employed tasks with some sort of turn-taking, potentially involving temporal predictions. While this argument is pure speculation at this point, the results of Liu et al. (2017) might provide some credibility to it, as they demonstrated convergent brain-to-brain synchrony results in fNIRS and fMRI during the one-way storytelling-listening paradigm.

### 4.4.4. Limitations

One limitation of our study is the design of the itself. We used the Joint trials as a model of minimally-constrained, two-way verbal communication and the Individual trials as a model of repeated instances of one-way communications. It is unclear though if the Individual condition is fulfilling its role. In the previous chapter on behavioral results from the same paradigm we saw some evidence indicating that participants in the Individual trials are also more similar to each other lexically and semantically than pseudo-pairs, suggesting that some priming is introduced into the task. Another limitation is that the task given was improvisational in nature, as arguably many conversations have some other goal than to invent interesting stories together. Our concern is that jointly improvised stories are a less predictable form of interactions than everyday conversations, engaging slightly different processes. Additionally (and perhaps more obviously), while fMRI is currently the most effective neuroimaging technique at capturing precisely localized neural activity, the very nature of requiring participants to speak to a remote partner through an open audio channel within the confines of an MRI bore (where participants need to lay down) is a strong deviation from how humans naturally interact in the real-world. Human conversations typically occur face-to-face and are perceived through both visual and auditory systems. It is well known that watching a speaker’s face improves speech intelligibility during verbal communication (Sumby & Pollack, 1954), and this increased speech intelligibility may strengthen the neural coupling between interlocutors. As such, in our present study, and in all previous fMRI hyperscanning studies, the lack of visual facial stimuli during the experimental task may inherently reduce the strength of neural coupling that would otherwise be achieved in real-world scenarios.

Lastly, it has recently been evidenced that humans are exceptionally similar to friends in how they perceive and respond to the world around them (Parkinson, Kleinbaum, & Wheatley, 2018). Previous personal knowledge about a conversational partner may therefore be associated with increased similarity of real-time mental responding and may enhance the strength of neural coupling during verbal interaction, relative to interlocutors who are complete strangers (Parkinson, Kleinbaum, & Wheatley, 2018). In our study, we did not assess whether any given pair of participants knew each before the start of the experiment. Since our participants were selected from the undergraduate and graduate student populations at Dartmouth College and Harvard University, it is possible that the members of any given pair were familiar with each other before performing the storytelling task. Thus, at the present moment we do not know whether friendship—or more generally, social network proximity—may have influenced the strength or extent of the neural couplings found across participant pairs.

# Chapter Five. General discussion

The goal of my thesis was to explore human communication from different view points: as a coordination tool in joint action, as multilevel alignment and as neural coupling. I presented three empirical studies covering the topics of efficiency in helping, alignment as an aspect of verbal interactions and brain-to-brain coupling as the neural model of dialogues. In the present chapter I summarize the findings and highlight promising future directions.

## 5.1. Overachiever helpers and underachiever helpees: Helping as a case of joint action

In Chapter 2, we construed helping as a case of joint action where the helper assumes the helpee's goal as her own. This approach enabled us to focus on efficiency in helping and testing (1) whether helpers optimize their communication for increasing helpees' performance and (2) whether helpees maximize their performance by utilizing helpers' communication to its fullest extent. We designed four experiments testing these questions, each employing a different version of a novel two-person cueing paradigm. In two experiments the helpee performed a reaction time task and the helper could select cues appearing for the helpee before he reacted to an imperative stimulus. Importantly, the helper was aware of the helpee's task and the upcoming stimuli, and, thus, she could select potentially helpful cues for the helpee. We could determine helper's efficiency on the basis of her cueing strategy (that is, whether the cues mapped responses or other aspects of the task) and the effort spent in the process, while the helpee's efficiency was characterized by the cueing gain. Two further experiments served as non-social controls where the cues were generated by the computer. Across experiments we manipulated the task difficulty, the saliency of cue-response mappings and the source of the cues (human helper or computer).

Overall, we found that helpers' behavior was aimed at improving helpees' performance, in line with our definition of helping. Helpers preferred to send maximally helpful cues and did so at the expense of their own effort. However, helpees were less efficient than helpers as they did not show the cueing gain that would have been expected if they made optimal use of the cues. Overall, helpees did not fully rely on helpers' cues but preferred to perform the task at least partially themselves. Moreover, helpees' apparent inefficiency was not a result of occasional helper errors or slow build-up of trust, as helpees displayed similar tendencies in computer cueing versions of the experiment. Analyses of RT time series revealed that cues from human helpers were utilized even faster than those from a computer. We conclude that helpees' behavior is best understood as a general preference for partial reliance on cues, in line with a split-effort model.

Our findings uncovered the dynamics of an interesting tension between helper and helpee, potentially present in many everyday helping situations too. According to our experiments, helpers who have committed themselves would not spare any effort in accomplishing helpee's goal, while helpees were reluctant to rely on the helpers. This latter result relates our work to earlier results in both behavioral economics and social psychology experiments on helping. For example, Hennig-Schmidt et al. (2008) found an avoidance of unequal or unfair offers by receivers even when the offers were unfairly advantageous for them. In social psychology, receiving help has been recognized as potentially threatening to one's self-esteem and competence (Nadler and Fisher, 1986), or as a debt that will need to be repaid (Greenberg and Shapiro, 1971) in the future. Here we extended earlier work by quantifying both helpees' reluctance and helpers' over-eagerness in terms of cognitive effort and showed its relevance to everyday coordination problems. Furthermore, we provided a task setting that connects the social aspects of helping to the cognitive and motor processes involved in coordination and simple forms of communication.

Importantly, we focused on helping as a communicative scenario, presenting an additional coordination problem for helper and helpee. Helpers' signals did not only need to be helpful but also easily recognizable by helpees. Similar to teaching, helpers needed to find the right way of communication (salient cue mappings) with the right content (helpful cue-response mappings). However, by using cues as signals, our study was also conceptually similar to studies on sensorimotor communication (SMC, Pezzulo et al., 2018). As in the case of SMC, in principle, helpers' signals were not always needed to be recognized as intentional communication but could have operated as changes in the environment making it easier for the partner to perform the task.

## 5.2. Alignment in verbal interaction as a function of interactivity and individual goals

In Chapter 3, we investigated alignment in verbal interactions. Alignment is ubiquitous not only as behavioral mimicry but also as paralinguistic and linguistic similarity, convergence, and synchrony present at multiple levels. The processes behind behavioral mimicry (automatic imitation and action simulation, Heyes, 2011; Wilson and Knoblich, 2005) support predictions and, hence, are important for interpersonal coordination. Similarly, the processes behind alignment in dialogues are assumed to support coordination in communication, that is, the development of shared understanding.

Given the importance of alignment in verbal interactions, we tested if important aspects of everyday interactions, namely interactivity and goal overlap modulated alignment strength. We contrasted predictions of the interactive alignment model (Pickering and Garrod, 2004) with alternatives based on research on automatic imitation. More specifically, we tested (1) if back-and-forth, two-way interactions led to stronger alignment than interactions consisting of one-way speaker-listener segments; and (2) if sharing the same goal or having different goals in an interaction affect alignment directly. In our paradigm, pairs of participants took turns in inventing stories. In interactive trials, the pair invented one storyline together, while in non-interactive trials they told independent stories that ran in parallel. Pair members were further instructed to follow either the same or different valence trajectories, so that they had the same goal or different goals regarding the intended direction of storylines.

We found that interactivity and the match between individual goals affected alignment at different levels. Shared goals led to stronger convergence and weaker correlations mainly in prosodic features (speech rate and pause length), irrespective of interactivity, while interactivity resulted in larger similarity syntactically, lexically and semantically, irrespective of matching goals. The findings were in line with the predictions of both the interactive alignment model and the automatic imitation account, but for different speech features.

While not conclusive, our findings suggest a dissociation of the processes behind alignment in prosodic and purely linguistic features. A potential reason is that prosody can serve several functions at once. On the one hand, prosody contributes to language comprehension by providing important cues about the speaker’s meaning (e.g. Hellbernd and Sammler,  2016; Hirschberg, 2004). For example, prosodic cues help disambiguating syntax (e.g. Allbritton, McKoon and Ratcliff, 1996) and understanding irony (e.g. Bryant and Fox Tree, 2002, 2005). On the other hand, prosody comprehension is partially independent from linguistic comprehension (Wildgruber et al., 2009; Cutler, Dahan and van Donselaar, 1997) and might be better characterized as part of a multi-modal emotion recognition network. Similarly, prosody production is intimately linked with affective display, including affiliation/disaffiliation with the partner in an interaction (e.g. Banse and Sherer, 1996). To the degree that prosody reflects emotional states, prosodic alignment might be more susceptible to purely imitative mechanisms (e.g. emotional contagion, Nummenmaa, Hirvonen, Parkkola and Hietanen, 2008) rather than linguistic features.

While we found different forms of alignment across a number of speech features, alignment was not as widespread or strong as expected on the basis of current theories. For example, convergence was missing at several levels and we even found overall divergence over speech turns across conditions in syntax and lexicon. Taken together with other experiments yielding similar results (e.g. Levitan and Hirschberg, 2011; Fusaroli et al., 2017; De Looze et al., 2014), our findings speak to the necessity of better measures of alignment (e.g. cross-recurrence analysis, Louwerse et al., 2012; Fusaroli and Tylén, 2016; Duran and Fusaroli, 2017) or a revision to the supposed central role of alignment in discourse (e.g. Fusaroli et al., 2014). A promising avenue for future research consists of validating both alignment-based measures and dialogue structures derived from discourse analysis (e.g. structure of compound contributions, Howes, Purver, Healey, Mills and Gregoromichelaki, 2011) or topic modeling (e.g. Eizenstein and Barzilay, 2008) against task success in well-defined communicative scenarios (e.g. Reitter and Moore, 2014).

## 5.3. Neural coupling in verbal interactions

In Chapter 4, we investigated brain-to-brain coupling in verbal interactions using an fMRI hyperscanning setup. As in Chapter 3, we employed a joint storytelling task but without the Goal manipulation, focusing solely on comparing Joint and Individual conditions. We relied on inter-subject correlations (ISC) for detecting coupling, capturing linear dependencies between brains over the course of the interaction. For the present study, we focused on speaker-listener segmentations of the data yielding consistent roles for participants for each analysis. This approach enabled us to compare our results to previous studies on one-way (non-interactive) communication. While built on earlier studies, our experiment was unique for using an fMRI hyperscanning setup with realistic verbal interactions.

Across conditions and relative to pseudo pairs, we found widespread coupling in a bilateral network of auditory and motor areas. Interestingly, this coupling did not fully replicate earlier results. In particular, there was no coupling in areas associated with mentalizing and narrative integration (temporo-parietal junction, posterior cingulate, precuneus and medial prefrontral cortex). A potential explanation for this discrepancy is that our task involved relatively short turns that lasted 30 s. This may have disrupted coupling in areas with long temporal receptive windows (Lerner et al., 2011). Furthermore, our coupling results regarding motor areas were stronger than in previous experiments. Again, a potential explanation for this result is that the turn-taking structure elicited predictions involving the listeners' production system (Scott, McGettigan and Eisner, 2009; Mandel et al., 2016).

The contrast between Joint vs. Individual condition showed that there was stronger coupling in the Joint condition in one area, the anterior portion of the left SMA. Intriguingly, the same area showed predictive properties in listeners, preceding speakers' neural activity. The SMA is often theorized to be involved in temporal predictions of one's own or other's action sequences, including speech production and comprehension (Schwartze, Rothermich and Kotz, 2011). Our results suggest that interactivity elicits stronger predictive SMA activity, presumably reflecting temporal predictions about the structure of a speaker's speech stream present both in speakers and listeners.

Here we provided a first analysis of the data set focusing on the Joint – Individual contrast using ISC. In the near future, we will try to replicate these results on an independent data set, extending the coupling measures to include non-linear methods, as well. Another promising future direction is to identify the neural correlates of the semantic and topic-level structure of the storylines participants invented.

## 5.4. Conclusion

In the current thesis we studied communication from a joint action perspective, that is, as a coordination problem. This view argues that human collaborations (joint actions) in general require the coordination of actions and underlying representations and that communication is both a type of collaboration in itself and a coordination device enabling further joint actions. Joint action research has identified a range of mechanisms supporting different types of coordination, suggesting that the same mechanisms play a role in communication as well. This is an integrative view: human communication - including language use – is assumed to lie on a continuum of joint actions and to be supported by an interplay of different mechanisms not necessarily specific to intentional communication. Given the wide variety of theories and approaches in communication research, we think that such an integrative framework is invaluable.

In our research we focused on different aspects of communication, employing distinct methods in each empirical chapter. We relied on cognitive psychology to study communication from an efficiency point of view, We used natural language processing tools to capture alignment in verbal interactions, and applied neuroimaging to characterize interactions in terms of coupled neural dynamics. Communication research is an interdisciplinary field and we believe that future advances will require ever-closer integration of traditionally separate approaches, including discourse analysis, behavioral experiments, corpus linguistics, big-data analysis and neuroscience. Technological advances in recent years have made such integration easier and will likely continue to do so. For example, experimental results from psycholinguistics have been validated on large internet-based corpora collected from thousands of people (e.g. Doyle et al., 2016), and researchers could identify the neural correlates of the development of the putative shared understanding in a communication game using fMRI hyperscanning (Stolk et al., 2014). An especially promising avenue for future research is the application of imaging methods to hypotheses derived from pragmatics and discourse analysis (cf. Bögels and Levinson, 2017).

# References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, *143*(6), 2304.

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., ... & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, *8*, 14.

Abu-Shikhah, N., & Deriche, M. (2001). A robust technique for harmonic analysis of speech. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on* (Vol. 2, pp. 877-880). IEEE.

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 714.

Ames, D. L., Honey, C. J., Chow, M. A., Todorov, A., & Hasson, U. (2015). Contextual alignment of cognitive and neural dynamics. *Journal of cognitive neuroscience*, *27*(4), 655-664.

Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., & Haynes, J. D. (2011). Flow of affective information between communicating brains. *Neuroimage*, *54*(1), 439-446.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., ... & Sotillo, C. (1991). The HCRC map task corpus. *Language and speech*, *34*(4), 351-366.

Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., Isabella, R., Fallani, F. D. V., ... & Caltagirone, C. (2011, August). Study of the functional hyperconnectivity between couples of pilots during flight simulation: an EEG hyperscanning study. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (pp. 2338-2341). IEEE.

Astolfi, L., Cincotti, F., Mattia, D., Fallani, F. D. V., Salinari, S., Vecchiato, G., ... & He, B. (2010, August). Imaging the social brain: multi-subjects EEG recordings during the “Chicken's game”. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE* (pp. 1734-1737). IEEE.

Astolfi, L., De Vico Fallani, F., Toppi, J., Cincotti, F., Salinari, S., & Vecchiato, G. (2011). Imaging the social brain by simultaneous hyperscanning of different subjects during their mutual interactions. *IEEE Intell. Syst*, *26*, 38-45.

Atlas, L., Clark, P. & Schimmel, S. (2010 September). Modulation Toolbox Version 2.1 for MATLAB, https://sites.google.com/a/uw.edu/isdl/projects/modulation-toolbox, University of Washington.

Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action co-representation: the joint SNARC effect. *Social neuroscience*, *3*(3-4), 410-420.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, *12*(1), 26-41.

Babiloni, C., Vecchio, F., Infarinato, F., Buffo, P., Marzano, N., Spada, D., ... & Perani, D. (2011). Simultaneous recording of electroencephalographic data in musicians playing in ensemble. *Cortex*, *47*(9), 1082-1090.

Babiloni, F., & Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neuroscience & Biobehavioral Reviews*, *44*, 76-93.

Baecke, S., Lützkendorf, R., Mallow, J., Luchtmann, M., Tempelmann, C., Stadler, J., & Bernarding, J. (2015). A proof-of-principle study of multi-site real-time functional imaging at 3T and 7T: implementation and validation. *Scientific reports*, *5*, 8413.

Bahrami, B., Olsen, K., Latham, P.E., Roepstorff, A., Rees, G. & Frith, C.D. (2010). Optimally interacting minds. *Science, 329*, 1081-1085.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G. & Frith, C.D. (2012a). Together, Slowly but Surely: The Role of Social Interaction and Feedback on the Build-Up of Benefit in Collective Decision-Making. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 3-8.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G. & Frith, C.D. (2012b). What failure in collective decision-making tells us about metacognition. *Philosophical Transactions of the Royal Society B, 367*, 1350-1365.

Bahry, D.L. & Wilson, R.K. (2006). Confusion or fairness in the field? Rejections in the ultimatum game under the strategy method. *Journal of Economic Behavior & Organization, 60*, 37-54.

Balcetis, E. E., & Dale, R. (2005). An exploration of social modulation of syntactic priming. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 27(27), 184-189.

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*, *54*(1), 39-57.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, *70*(3), 614.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *The Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Barragan, R. C., & Dweck, C. S. (2014). Rethinking natural altruism: Simple reciprocal interactions trigger children's benevolence. *Proceedings of the National Academy of Sciences*, *111*(48), 17071-17074.

Bašnáková, J., van Berkum, J., Weber, K., & Hagoort, P. (2015). A job interview in the MRI scanner: How does indirectness affect addressees and overhearers?. *Neuropsychologia*, *76*, 79-91.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59-78.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of personality and social psychology*, *79*(6), 941.

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, 557–564.

Becchio, C., Sartori, L., & Castiello, U. (2010). Toward you: The social side of actions. *Current Directions in Psychological Science*, *19*(3), 183-188.

Behnel, M., Cummins, C., Sichelschmidt, L., & De Ruiter, J. (2013). Priming and conceptual pacts in overhearers' adoption of referring expressions. *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*(35), 1869-1874.

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard III, M. A., & Greenlee, J. D. (2015). Sensory–motor networks involved in speech production and motor control: An fMRI study. *Neuroimage*, *109*, 418-428.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage*, *37*(1), 90-101

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, *29*, 1165-1188.

Bellemare, C., Kröger, S., & Van Soest, A. (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, *76*(4), 815-839.

Ben-Yakov, A., Honey, C. J., Lerner, Y., & Hasson, U. (2012). Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *Neuroimage*, *63*(1), 501-506.

Bestmann, S., Harrison, L.M., Blankenburg, F., Mars, R.B., Haggard, P., Friston, K.J. & Rothwell, J.C. (2008). Influence of Uncertainty and Surprise on Human Corticospinal Excitability during Preparation for Action. *Current Biology, 18(*10), 775-780.

Bickhard, M. H. (2007). Language as an interaction system. *New Ideas in Psychology*, *25*(2), 171-187.

Bilek, E., Ruf, M., Schäfer, A., Akdeniz, C., Calhoun, V. D., Schmahl, C., ... & Meyer-Lindenberg, A. (2015). Information flow between interacting human brains: Identification, validation, and relationship to social expertise. *Proceedings of the National Academy of Sciences*, *112*(16), 5207-5212.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. Sebastopol, CA, US: O'Reilly Media, Inc.

Bishop, L., & Goebl, W. (2018). Beating time: How ensemble musicians' cueing gestures communicate beat position and tempo. *Psychology of music*, *46*(1), 84-106.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific reports*, *5*, 12881.

Bögels, S., & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction*, *50*(1), 71-89.

Boghrati, R., Hoover, J., Johnson, K. M., Garten, J., & Dehghani, M. (2018). Conversation level syntax similarity metric. *Behavior research methods*, *50*(3), 1055-1073.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American economic review*, *90*(1), 166-193.

Bonin, F., De Looze, C., Ghosh, S., Gilmartin, E., Vogel, C., Polychroniou, A., ... & Campbell, N. Investigating fine temporal dynamics of prosodic and lexical accommodation. *INTERSPEECH 2013*,

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13-B25.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, *104*(2), 163-197.

Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem?. *Trends in cognitive sciences*, *9*(10), 489-495.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: inferential processes versus action simulation. *Current biology*, *17*(24), 2117-2121.

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.

Brennan, S. E. (2002). Visual co-presence, coordination signals, and partner effects in spontaneous spoken discourse. *Cognitive Studies*, *9*(1), 7-25.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., and Zelinsky, G. J. (2008). Coordinating cognition: the costs and benefits of shared gaze during collaborative search. *Cognition,* 106, 1465–1477.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274-291.

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. In *Psychology of learning and motivation* (Vol. 53, pp. 301-344). Academic Press.

Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and symbol*, *17*(2), 99-119.

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice?. *Language and speech*, *48*(3), 257-277.

Burks, A. W. (1949). Icon, index, and symbol. *Philosophy and phenomenological research*, *9*(4), 673-689.

Burle, B., van den Wildenberg, W., & Ridderinkhof, K. R. (2005). Dynamics of facilitation and interference in cue-priming and Simon tasks. *European Journal of Cognitive Psychology*, *17*(5), 619-641.

Cacioppo, J. T., & Decety, J. (2011). Social neuroscience: challenges and opportunities in the study of complex behavior. *Annals of the New York Academy of Sciences*, *1224*(1), 162-173.

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.

Chang, F., Dell, G., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–272.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579-1601.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. Journal of Personality and Social Psychology, 76(6), 893–910.

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual review of psychology*, *64*, 285-308.

Chen, D., & Manning, C. (2014). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 740-750). Association for Computational Linguistics*.*

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior*, *22*(2), 245-258.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259-294.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, *13*(1991), 127-149.

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: University of Chicago Press.

Clark, H. H. (1994). Discourse in production. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA, US: Academic Press.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H.H. & Krych, M.A. (2004). Speaking while monitoring addresses for understanding. *Journal of Memory and Language, 50*, 62-81.

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse studies*, *7*(4-5), 507-525.

Clarke, S., Francová, A., Székely, M., Butterfill, S., & Michael, J. (2018, December 8). Joint action goals reduce visuomotor interference from a partner's incongruent actions. https://doi.org/10.31234/osf.io/tdbne

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*(2), 214-230.

Constable, M. D., Bayliss, A. P., Tipper, S. P., Spaniol, A. P., Pratt, J., & Welsh, T. N. (2016). Ownership status influences the degree of joint facilitatory behavior. *Psychological science*, *27*(10), 1371-1378.

Controzzi, M., Singh, H., Cini, F., Cecchini, T., Wing, A., & Cipriani, C. (2018). Humans adjust their grip force when passing an object according to the observed speed of the partner's reaching out movement. *Experimental brain research*, *236*(12), 3363-3377.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, *29*(3), 162-173.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, *23*(3), 371-414.

Critcher, C.R., Inbar, Y. & Pizarro, D.A. (2013). How Quick Decisions Illuminate Moral Character. *Social Psychological and Personality Science, 4*(3), 308-315.

Csibra, G. (2008). Action mirroring and action understanding: An alternative account. In Haggard, P., Rossetti, Y. & Kawato, M. (Eds.), *Sensorimotor foundations of higher cognition. Attention and performance XXII*, (pp. 435-459). Oxford, UK: Oxford University Press.

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage*, *59*(3), 2430-2437.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, *40*(2), 141-201.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179-194.

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The self-organization of human interaction. In *Psychology of learning and motivation* (Vol. 59, pp. 43-95). New York, NY: Academic Press.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*(5), 381-385.

Deelstra, J. T., Peeters, M. C., Schaufeli, W. B., Stroebe, W., Zijlstra, F. R., & van Doornen, L. P. (2003). Receiving instrumental support at work: when help is not welcome. *Journal of applied psychology*, *88*(2), 324-331.

De Looze, C., Scherer, S., Vaughan, B., & Campbell, N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, *58*, 11-34.

del Prado Martín, F. M. & Bois, J. W. D. (2015). Syntactic Alignment is an Index of Affective Alignment: An Information-Theoretical Study of Natural Dialogue. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings & P. P. Maglio (eds.), *CogSci*, : cognitivesciencesociety.org. ISBN: 978-0-9911967-2-2

De Ruiter, J. P., Noordzij, M. L., Newman-Norlund, S., Newman-Norlund, R., Hagoort, P., Levinson, S. C., & Toni, I. (2010). Exploring the cognitive infrastructure of communication. *Interaction Studies*, *11*(1), 51-77.

De Smedt, T., Daelemans, W. (2012). Pattern for Python. *Journal of Machine Learning Research, 13*, 2031–2035.

De Weerd, H., Verbrugge, R., & Verheij, B. (2015). Higher-order theory of mind in the tacit communication game. *Biologically Inspired Cognitive Architectures*, *11*, 10-21.

Dikker, S., Silbert, L. J., Hasson, U., & Zevin, J. D. (2014). On the same wavelength: predictable language enhances speaker–listener brain-to-brain synchrony in posterior superior temporal gyrus. *Journal of Neuroscience*, *34*(18), 6267-6272.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., ... & Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, *27*(9), 1375-1380.

Doyle, G., Yurovsky, D., & Frank, M. C. (2016, April). A robust framework for estimating linguistic alignment in Twitter conversations. In *Proceedings of the 25th international conference on world wide web* (pp. 637-648). International World Wide Web Conferences Steering Committee. Retrieved from http://gdac.uqam.ca/WWW2016-Proceedings/proceedings/p637.pdf.

Dötsch, D., & Schubö, A. (2015). Social categorization and cooperation in motor joint action: evidence for a joint end-state comfort. *Experimental brain research*, *233*(8), 2323-2334.

Dumas, G. (2011). Towards a two-body neuroscience. *Communicative & integrative biology*, *4*(3), 349-352.

Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement. *PloS one*, *12*(6), e0178140.

Durso, F. T., Rea, C. B., & Dayton, T. (1994). Graph-theoretic confirmation of restructuring during insight. *Psychological Science*, *5*(2), 94-98.

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E. J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*(2), 454-465.

Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. *INTERSPEECH 2009*, 2779-2782.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51*, 667-695.

Eisenstein, J., & Barzilay, R. (2008). Bayesian unsupervised topic segmentation. In Lapata, M. & Tou Ng, H. (Eds), *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 334-343). Association for Computational Linguistics.

Elsner, C., D'ausilio, A., Gredebäck, G., Falck-Ytter, T., & Fadiga, L. (2013). The motor cortex is causally related to predictive eye movements during action observation. *Neuropsychologia*, *51*(3), 488-492.

Endo, S., Pegman, G., Burgin, M., Toumi, T., & Wing, A. M. (2012, June). Haptics in between-person object transfer. In *Haptics: Perception, Devices, Mobility, and Communication: 8th International Conference, EuroHaptics 2012, Tampere, Finland, June 13-15, 2012 Proceedings* (Vol. 7282, p. 103). Berlin: Springer.

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Aliaga, A. E., ... & Oya, H. (2018). FMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 16, 111-116.

Evans, A.M. & Krueger, J.I. (2009). The Psychology (and Economics) of Trust. *Social and Personality Psychology Compass, 3,* 1003-1017.

Fehr, E. & Schmidt, K.M. (1999). A Theory of Fairness, Competition and Cooperation. *The Quarterly Journal of Economics, 114*(3), 817-868.

Fehr, E. & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*, 785-791.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2011). Unbiased nonlinear average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313-327.

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse processes*, *27*(1), 35-53.

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in cognitive sciences*, *16*(5), 262-268.

Friesen, C.K., Ristic, J. & Kingstone, A. (2004). Attentional Effects of Counterpredictive Gaze and Arrow Cues. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 319-329.

Friston, K., & Frith, C. (2015). A duet for one. *Consciousness and cognition*, *36*, 390-405.

Funane, T., Kiguchi, M., Atsumori, H., Sato, H., Kubota, K., & Koizumi, H. (2011). Synchronous activity of two people's prefrontal cortices during a cooperative task measured by simultaneous near-infrared spectroscopy. *Journal of Biomedical Optics*, *16*(7), 077011.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science*, *23*(8), 931-939.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147-157.

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive science*, *40*(1), 145-171.

Fusaroli, R., Tylén, K., Garly, K., Steensig, J., Christiansen, M. H., & Dingemanse, M. (2017). Measures and mechanisms of common ground: backchannels, conversational repair, and interactive alignment in free and task-oriented social interactions. *Proceedings of the Annual Meeting of the Cognitive Science Society, 39*(39), 2055-2060.

Gambi, C., & Pickering, M. J. (2017). Models linking production and comprehension. *Handbook of Psycholinguistics*, 157-182.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737-767.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, *13*(3), 361-377.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in human neuroscience*, *5*, 11.

Gallagher, H.L., Jack, A.I., Roepstorff, A. & Frith, C.D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage, 16*, 814-821.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in cognitive sciences, 8*(9), 396-403.

Gallotti, M., Fairhurst, M. T., & Frith, C. D. (2017). Alignment in social interactions. *Consciousness and Cognition*, *48*, 253-261.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*(3), 181-215.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy?. *Trends in cognitive sciences*, *8*(1), 8-11.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292-304.

Gergle, D., Kraut, R.E. & Fussell, S.R. (2004). Language Efficiency and Visual Technology: Minimizing Collaborative Effort with Visual Information. *Journal of Language and Social Psychology, 23*(4), 1-27.

Gerhard, D. (2003). *Pitch extraction and fundamental frequency: History and current techniques*. Technical Report. Retrieved from http://sapyc.espe.edu.ec/evcarrera/DSP/pitch.pdf.

Gerven, S. V., & Xie, F. (1997). A comparative study of speech detection methods. In *Fifth European Conference on Speech Communication and Technology*. Retrieved from https://www.isca-speech.org/archive/archive_papers/eurospeech_1997/e97_1095.pdf.

Giese, M. A., & Rizzolatti, G. (2015). Neural and computational mechanisms of action processing: interaction between visual and motor representations. *Neuron*, *88*(1), 167-180.

Giles, H., Coupland, N. & Coupland, J. (1992) Accommodation theory: Communication, context and consequences. In Giles, H., Coupland, J. & Coupland (Eds.), *Contexts of accommodation* (pp. 1–68). Cambridge, UK: Cambridge University Press.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905-922.

Gonzalez D.A., Studenka B.E., Glazebrook C.M. & Lyons J.L. (2011). Extending end-state comfort effect: do we consider the beginning state comfort of another? *Acta Psychologica, 136*(3), 347-353.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in neuroinformatics*, *5*, 13.

Gorgolewski, Krzysztof J., Esteban, Oscar, Ellis, David Gage, Notter, Michael Philipp, Ziegler, Erik, Johnson, Hans, … Ghosh, Satrajit. (2017, May 21). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13.1 (Version 0.13.1). Zenodo. https://zenodo.org/record/581704

Gough, P. M., Nobre, A. C., & Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *Journal of Neuroscience*, *25*(35), 8010-8016.

Gräfenhain, M., Carpenter, M. & Tomasello, M. (2013). Three-Year-Olds' Understanding of the Consequences of Joint Commitments. PLoS ONE 8*(9),* e73039.

Greenberg, M. S., & Shapiro, S. P. (1971). Indebtedness: An adverse aspect of asking for and receiving help. *Sociometry*, 290-301.

Greenberg, M. S., & Westcott, D. R. (1983). Indebtedness as a mediator of reactions to aid. *New directions in helping*, *1*, 85-112.

Gregory, S. W. (1990). Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior*, *14*(4), 237-251.

Gregory Jr, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of personality and social psychology*, *70*(6), 1231.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*(1), 63-72.

Grice, H.P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hadley, L. V., Novembre, G., Keller, P. E., & Pickering, M. J. (2015). Causal role of motor simulation in turn-taking behavior. *Journal of Neuroscience*, *35*(50), 16516-16520.

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive psychology*, *30*(3), 304-337.

Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al.(1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(2), 392– 406.

Hari, R., & Kujala, M. V. (2009). Brain basis of human social interaction: from concepts to brain imaging. *Physiological reviews*, *89*(2), 453-479.

Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of social interaction in human brain function. *Neuron*, *88*(1), 181-193.

Haruno, M., Wolpert, D. M., & Kawato, M. (2003). Hierarchical MOSAIC for movement generation. *International congress series, 1250*, 575-590.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in cognitive sciences*, *16*(2), 114-121.

Hasson, U., & Honey, C. J. (2012). Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, *62*(2), 1272-1278.

Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B* , *371*(1693), 20150366.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. S*cience*, *303*(5664), 1634-1640.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, *14*(1), 40-48.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends in cognitive sciences*, 19(6), 304-313.

Healey, P. G., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PloS one*, *9*(6), e98598.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic bulletin & review*, *7*(2), 185-207.

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70-86.

Hennig-Schmidt, H., Li, Z. Y., & Yang, C. (2008). Why people reject advantageous offers—Non-monotonic strategies in ultimatum bargaining: Evaluating a video experiment run in PR China. *Journal of Economic Behavior & Organization*, *65*(2), 373-384.

Hertrich, I., Dietrich, S., & Ackermann, H. (2016). The role of the supplementary motor area for speech and language processing. *Neuroscience & Biobehavioral Reviews*, *68*, 602-610.

Heyes, C. (2011). Automatic imitation. *Psychological bulletin*, *137*(3), 463-483.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407-422.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, *29*(1), 2-20.

Hickok, G. & Poeppel, D. (2015). Neural basis of speech perception. In Hickok, G. & Small, S. (Eds), *Neurobiology of Language* (pp. 299-310). San Diego, CA, US: Academic Press.

Hill, K. (2002). Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature*, *13*(1), 105-128.

Hirschberg, J. (2004). Pragmatics and intonation.  In L. R. Horn, L.R. & Ward, G. (Eds), *The handbook of pragmatics* (pp. 515-537). Malden, MA, USA: Blackwell Publishing Ltd.

Hommel, B. (1993). The relationship between stimulus processing and response selection in the Simon task: Evidence for a temporal overlap. *Psychological research*, *55*(4), 280-290.

Hommel, B., Pratt, J., Colzato, L. & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science, 12*(5), 360-365.

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, 849–78.

Hommel, B., Proctor, R. W., & Vu, K. P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological research*, *68*(1), 1-17.

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica, 136(2)*, 189-202.

Honey, C. J., Thompson, C. R., Lerner, Y., & Hasson, U. (2012). Not lost in translation: neural responses shared across languages. *Journal of Neuroscience*, *32*(44), 15277-15283.

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*(6), 949-960.

Howes, C., Healey, P. G., & Purver, M. (2010). Tracking lexical and syntactic alignment in conversation. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 32(32), 2004-2009.

Howes, C., Purver, M., Healey, P. G., Mills, G., & Gregoromichelaki, E. (2011). On incrementality in dialogue: Evidence from compound contributions. *Dialogue & Discourse*, *2*(1), 279-311.

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., & Turner, R. (2002). Image distortion correction in fMRI: a quantitative evaluation. *Neuroimage*, *16*(1), 217-240.

Iacoboni, M. (2008). The role of premotor cortex in speech perception: Evidence from fmri and rtms. *Journal of Physiology-Paris*, *102*(1-3), 31-34.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime’s prediction error given both prior and recent experience. *Cognition*, *127*(1), 57-83.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825-841.

Jenkinson, M. (2003). Fast, automated, *N*-dimensional phase-unwrapping algorithm. *Magnetic resonance in medicine*, *49*(1), 193-197.

Jezzard, P., & Balaban, R. S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic resonance in medicine*, *34*(1), 65-73.

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience*, *32*(45), 16064-16069.

Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L., & Lu, C. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences*, *112*(14), 4274-4279.

Jordan, J. J., Hoffman, M., Nowak, M. A., & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences*, *113*(31), 8658-8663.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current opinion in neurobiology*, *9*(6), 718-727.

Keysers, C., & Gazzola, V. (2014). Hebbian learning and predictive mirror neurons for actions, sensations and emotions. *Philosophical Transactions of the Royal Society B*, *369*(1644), 20130175.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in cognitive sciences*, *15*(8), 352-357.

Kauppi, J. P., Jääskeläinen, I. P., Sams, M., & Tohka, J. (2010). Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency. *Frontiers in neuroinformatics*, *4*, 5.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., ... & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLoS computational biology*, *13*(2), e1005350.

Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Quarterly journal of experimental psychology*, *63*(11), 2220-2230.

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction?. *Frontiers in human neuroscience*, *6*, 215.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology-Learning Memory and Cognition*, *25*(6), 1534-1555.

Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological science*, *12*(6), 467-472.

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1499), 2021-2031.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: theory and data. In *Psychology of learning and motivation, (*Vol. 54, pp. 59-101). Academic Press.

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in cognitive sciences*, *14*(9), 392-399.

Kourtis, D., Sebanz, N., & Knoblich, G. (2010). Favouritism in the motor system: social interaction modulates action simulation. *Biology Letters*, *6*, 758-761,

Kourtis, D., Sebanz, N., & Knoblich, G. (2013a). Predictive representation of other people's actions in joint action planning: An EEG study. *Social neuroscience*, *8*(1), 31-42.

Kourtis, D., Knoblich, G., & Sebanz, N. (2013b). History of interaction and task distribution modulate action simulation. *Neuropsychologia*, *51*(7), 1240-1247.

Kourtis, D., Knoblich, G., Woźniak, M., & Sebanz, N. (2014). Attention allocation and task representation during joint action planning. *Journal of Cognitive Neuroscience*, *26*(10), 2275-2286.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of personality and social psychology*, *4*(3), 343.

Kraut, R. E., Lewis, S. H., & Swezey, L. W. (1982). Listener responsiveness and the coordination of conversation. *Journal of personality and social psychology*, *43*(4), 718.

Kusner, M., Sun, Y., Kolkin, N., & Weinberger, K. (2015, June). From word embeddings to document distances. *International Conference on Machine Learning,* 957-966.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).

Lankinen, K., Saari, J., Hari, R., & Koskinen, M. (2014). Intersubject consistency of cortical MEG signals during movie viewing. *NeuroImage*, *92*, 217-224.

Lee, C. C., Black, M., Katsamanis, A., Lammert, A. C., Baucom, B. R., Christensen, A., ... & Narayanan, S. S. (2010). Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. *INTERSPEECH 2010*, 793-796.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, *31*(8), 2906-2915.

Levinson, S. C. (1983). *Pragmatics*. Cambridge, UK:  Cambridge University Press.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences*, *20*(1), 6-14.

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *INTERSPEECH 2011*, 3081-3084.

Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in neurosciences*, *39*(8), 527-542.

Lindenberger, U., Li, S. C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC neuroscience*, *10*(1), 22.

Liu, N., Mok, C., Witt, E. E., Pradhan, A. H., Chen, J. E., & Reiss, A. L. (2016). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga game with face-to-face communication. *Frontiers in human neuroscience*, *10*, 82.

Liu, Y., Piazza, E. A., Simony, E., Shewokis, P. A., Onaral, B., Hasson, U., & Ayaz, H. (2017). Measuring speaker–listener neural coupling with functional near infrared spectroscopy. *Scientific Reports*, *7*, srep43293.

Loehr, J. D., & Vesper, C. (2016). The sound of you and me: novices represent shared goals in joint action. *Quarterly Journal of Experimental Psychology*, *69*(3), 535-547.

Loftus, G.E & Masson, M.E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*(4), 476-490.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive science*, *36*(8), 1404-1426.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494-1502.

Magyari, L., & de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in psychology*, *3*, 376.

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, *91*, 5-27.

Mandel, A., Bourguignon, M., Parkkonen, L., & Hari, R. (2016). Sensorimotor activation related to speaker vs. listener role during natural conversation. *Neuroscience letters*, *614*, 99-104.

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.

Mason, A. H., & MacKenzie, C. L. (2005). Grip forces when passing an object to a partner. *Experimental brain research*, *163*(2), 173-187.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315.

McEllin, L., Sebanz, N., & Knoblich, G. (2018). Identifying others’ informative intentions from movement kinematics. *Cognition*, *180*, 246-258.

McGarva, A. R., & Warner, R. M. (2003). Attraction and social coordination: Mutual entrainment of vocal activity rhythms. *Journal of Psycholinguistic Research*, *32*(3), 335-354.

Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, *84*(3), 658-673.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*(19), 1692-1696.

Menenti, L., Gierhan, S. M., Segaert, K., & Hagoort, P. (2011). Shared language: overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological science*, *22*(9), 1173-1182.

Menenti, L., Garrod, S. C., & Pickering, M. J. (2012). Toward a neural basis of interactive alignment in conversation. *Frontiers in human neuroscience*, *6*, 185.

Ménoret, M., Varnet, L., Fargier, R., Cheylus, A., Curie, A., Des Portes, V., ... & Paulignan, Y. (2014). Neural correlates of non-verbal social interactions: a dual-EEG study. *Neuropsychologia*, *55*, 85-97.

Ménoret, M., Bourguignon, M., & Hari, R. (2015). Modulation of Rolandic Beta-Band Oscillations during Motor Simulation of Joint Actions. *PloS one*, *10*(7), e0131655.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. London, UK: Routledge.

Meyer, M., van der Wel, R. P., & Hunnius, S. (2013). Higher-order action planning for individual and joint object manipulations. *Experimental brain research*, *225*(4), 579-588.

Michael, J., Sebanz, N., & Knoblich, G. (2016). The sense of commitment: A minimal approach. *Frontiers in psychology*, *6*, 1968.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

Mills, G. J. (2014). Dialogue in joint activity: complementarity, convergence and conventionalization. *New ideas in psychology*, *32*, 158-173.

Misyak, J., Noguchi, T., & Chater, N. (2016). Instantaneous conventions: The emergence of flexible communicative signals. *Psychological science, 27(12)*, 1550-1561.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., ... & Fisher, R. E. (2002). Hyperscanning: simultaneous fMRI during linked social interactions.

Nadler, A., & Fisher, J.D. (1986). The role of threat to self-esteem and perceived control in recipient reaction to help: Theory development and empirical validation. *Advances in experimental social psychology, 19*, 81-122.

Nadler, A. (2015). The other side of helping: Seeking and receiving help. In Schroeder, D. A., & Graziano, W. G. (Eds.), *The Oxford handbook of prosocial behavior* (307-328). New York, NY, US: Oxford University Press

Nakahara, K., Hayashi, T., Konishi, S., & Miyashita, Y. (2002). Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science*, *295*(5559), 1532-1536.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum

Newman-Norlund, S. E., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A., De Ruiter, J. P., Hagoort, P., & Toni, I. (2009). Recipient design in tacit communication. *Cognition*, *111*(1), 46-54.

Ninio, A., & Bruner, J. (1978). The achievement and antecedents of labelling. *Journal of child language*, *5*(1), 1-15.

Noordzij, M. L., Newman-Norlund, S. E., De Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2009). Brain mechanisms underlying human communication. *Frontiers in Human Neuroscience*, *3*, 14.

Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2012). Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cerebral Cortex*, *22*(12), 2894-2903.

Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2013). Motor simulation and the coordination of self and other in real-time joint action. *Social cognitive and affective neuroscience*, *9*(8), 1062-1068.

Noy, L., Dekel, E., & Alon, U. (2011). The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proceedings of the National Academy of Sciences*, *108*(52), 20947-20952.

Nummenmaa, L., Hirvonen, J., Parkkola, R., & Hietanen, J. K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage*, *43*(3), 571-580.

Nummenmaa, L., Lahnakoski, J. M., & Glerean, E. (2018). Sharing the social world via intersubject neural synchronisation. *Current opinion in psychology*, *24*, 7.

Ondobaka, S., de Lange, F. P., Newman-Norlund, R. D., Wiemers, M., & Bekkering, H. (2012). Interplay between action and movement intentions during social interaction. *Psychological Science*, *23*(1), 30-35.

Osaka, N., Minamoto, T., Yaoi, K., Azuma, M., Shimada, Y. M., & Osaka, M. (2015). How two brains make one synchronized mind in the inferior frontal cortex: fNIRS-based hyperscanning during cooperative singing. *Frontiers in psychology*, *6*, 1811.

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *Neuroimage*, *61*(1), 314-322.

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature communications*, *9*(1), 332.Paulus, M. (2014). The emergence of prosocial behavior: why do infants and toddlers help, comfort, and share?. *Child Development Perspectives*, *8*(2), 77-81.

Paxton, A., & Dale, R. (2013). Argument disrupts interpersonal synchrony. *Quarterly journal of experimental psychology*, *66*(11), 2092-2102.

Paxton, A., & Dale, R. (2017). Interpersonal movement synchrony responds to high-and low-level conversational constraints. *Frontiers in psychology*, *8*, 1135.

Pearce, W.B. (1973). Trust in Interpersonal Communication. Paper presented at the Annual Meeting of the International Communication Association, Montreal, 1973. https://files.eric.ed.gov/fulltext/ED087069.pdf

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in neuroinformatics*, *2*, 10.

Pele, O., & Werman, M. (2008, October). A linear time histogram metric for improved sift matching. In *European conference on computer vision* (pp. 495-508). Springer, Berlin, Heidelberg.

Pele, O., & Werman, M. (2009). Fast and robust earth mover's distances. In *Computer vision, 2009 IEEE 12th international conference on* (pp. 460-467). IEEE.

Penner, L.A., Dovidio, J.F., Piliavin, J.A. & Schroeder, P.A. (2005). Prosocial Behavior - Multilevel Perspectives. *Annual Review of Psychology, 56*(14), 1-28.

Pesquita, A., Whitwell, R. L., & Enns, J. T. (2017). Predictive joint-action model: A hierarchical predictive approach to human cooperation. *Psychonomic bulletin & review*, 1-19.

Pezzulo, G., Donnarumma, F., and Dindo, H. (2013). Human sensorimotor communication: a theory of signaling in online social interactions. PLoS ONE 8:e79876. doi: 10.1371/journal.pone.0079876

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.

Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., & Castelfranchi, C. (2018). The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of life reviews*.

Pfitzinger, H. R. (1998). Local speech rate as a combination of syllable and phone rate. In *Fifth International Conference on Spoken Language Processing,* paper 0523.

Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2018). Infant and adult brains are coupled to the dynamics of natural communication. *bioRxiv*, 359810.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in cognitive sciences*, *3*(4), 136-141.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347.

Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in human neuroscience*, *8*, 132.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: a theory and review. *Psychological Bulletin, 144*(10) , 1002-1044.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological bulletin*, *134*(3), 427-459.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, *84*, 320-341.

Prinz, W. (1997). Perception and action planning. *European journal of cognitive psychology*, *9*(2), 129-154.

Pulvermüller, F., Moseley, R. L., Egorova, N., Shebani, Z., & Boulenger, V. (2014). Motor cognition–motor semantics: action perception theory of cognition and communication. *Neuropsychologia*, *55*, 71-84.

Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of experimental psychology*, *71*(2), 264-272.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience*, *12*(6), 718.

Ray, M. & Welsh, T.N. (2011). Response selection during a joint action task. *Journal of Motor Behavior, 43*(4), 329-332.

Reitter, D., Moore, J. D. & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. *Proceedings of the Annual Meeting of the Cognitive Science Society, 28*(28), 685-690.

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive science*, *35*(4), 587-637.

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, *76*, 29-46.

Riordan, M. A., Kreuz, R. J., & Olney, A. M. (2014). Alignment is a function of conversational dynamics. *Journal of Language and Social Psychology*, *33*(5), 465-481.

Roberts, G. (1998). Competitive altruism: from reciprocity to the handicap principle. *Proceedings of the Royal Society of London B: Biological Sciences*, *265*(1394), 427-431.

Robinson, J. S., Page-Gould, E., & Plaks, J. E. (2017). I appreciate your effort: Asymmetric effects of actors' exertion on observers' consequentialist versus deontological judgments. *Journal of Experimental Social Psychology*, *73*, 50-64.

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: a review. *Frontiers in psychology*, *2*, 307.

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.

Rosenbaum, D. A. (1980). Human movement initiation: specification of arm, direction, and extent. *Journal of Experimental Psychology: General*, *109*(4), 444-474.

Rosenbaum, D. A. (1983). The movement precuing technique: Assumptions, applications, and extensions. *Advances in psychology, 12*, 231-274.

Rubin, N., Nakayama, K., & Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Current Biology*, *7*(7), 461-467.

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.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In Schenkein, J. (Ed), *Studies in the organization of conversational interaction* (pp. 7-55). Academic Press.

Scharoun, S. M., Scanlan, K. A., & Bryden, P. J. (2016). Hand and grasp selection in a preferential reaching task: the effects of object location, orientation, and task intention. *Frontiers in psychology*, *7*, 360.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and brain sciences*, *36*(4), 393-414.

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., & Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences*, *107*(20), 9388-9393.

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In *Coordination: Neural, behavioral and social dynamics* (pp. 281-308). Berlin, Germany: Springer,

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2018). When height carries weight: Communicating hidden object properties for joint action. *Cognitive science*, *42*(6), 2021-2059.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive psychology*, *21*(2), 211-232.

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in cognitive sciences*, *11*(5), 211-218.

Schweitzer, A., & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. *INTERSPEECH 2013,* 525-529.

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*(2), 226-233.

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

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, *10*(2), 70-76.

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353-367.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences*, *104*(18), 7361-7366.

Shany-Ur, T., Poorzand, P., Grossman, S.N., Growdon, M.E., Jang, J.Y., Ketelle, R.S., Miller, B.L. & Rankin, K.P. (2012). Comprehension of insincere communication in neurodegenerative disease: lies, sarcasm, and theory of mind. *Cortex, 48*, 1329–1341.

Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326.

Schoot, L., Hagoort, P., & Segaert, K. (2016). What can we learn from a two-brain approach to verbal interaction?. *Neuroscience & Biobehavioral Reviews*, *68*, 454-459.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43), E4687-E4696.

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature communications*, *7*, 121

Sorkin, R.D., Hays, C.J. & West, R. (2001). Signal-Detection Analysis of Group Decision Making. *Psychological Review, 108*(1), 183-203.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and cognition.* Oxford, UK: Basil Blackwell.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (2010). Epistemic Vigilance. *Mind and Language, 25*, 359-393.

Spiegelhalder, K., Ohlendorf, S., Regen, W., Feige, B., van Elst, L. T., Weiller, C., ... & Tüscher, O. (2014). Interindividual synchronization of brain activity during live verbal communication. *Behavioural brain research*, *258*, 75-79.

Springer, A., Parkinson, J., & Prinz, W. (2013). Action simulation: time course and representational mechanisms. *Frontiers in psychology*, *4*, 387.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425-14430.

Stevens, J.R. & Hauser, M.D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences, 8*(2), 60-65.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587-10592.

Stolk, A., Noordzij, M. L., Verhagen, L., Volman, I., Schoffelen, J. M., Oostenveld, R., ... & Toni, I. (2014). Cerebral coherence between communicators marks the emergence of meaning. *Proceedings of the National Academy of Sciences*, *111*(51), 18183-18188.

Stolk, A., Verhagen, L., & Toni, I. (2016). Conceptual alignment: How brains achieve mutual understanding. *Trends in Cognitive Sciences*, *20*(3), 180-191.

Street Jr, R. L., Brady, R. M., & Putman, W. B. (1983). The influence of speech rate stereotypes and rate similarity or listeners' evaluations of speakers. *Journal of Language and Social Psychology*, *2*(1), 37-56.

Street Jr, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, *11*(2), 139-169.

Sugden, R. (1995). A theory of focal points. *The Economic Journal*, *105*, 533-550.

Sugden, R., & Zamarrón, I. E. (2006). Finding the key: the riddle of focal points. *Journal of Economic Psychology*, *27*(5), 609-621.

Tilk, O., & Alumäe, T. (2016). Bidirectional Recurrent Neural Network with Attention Mechanism for Punctuation Restoration. *INTERSPEECH 2016,* 3047-3051. Available at http://bark.phon.ioc.ee/punctuator#

Tomasello, M. (1988). The role of joint attentional processes in early language development. *Language sciences*, *10*(1), 69-88.

Torrey, C., Fussell, S. sR., & Kiesler, S. (2008). Trying to be helpful: Social challenges for smart robots. In *Workshop-Proceedings of ACM/IEEE Human-Robot Interaction Conference (HRI2008), Amsterdam* (pp. 23-26).

Tricomi, E., Rangel, A., Camerer, C.F. & O'Doherty, J.P. (2010). Neural evidence for inequality-averse social preferences. *Nature, 463*, 1089-1091.

Tsoi, L., Dungan, J., Waytz, A., & Young, L. (2016). Distinct neural patterns of social cognition for cooperation versus competition. *NeuroImage*, *137*, 86-96.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*, *29*(6), 1310-1320.

Twenge J.M., Baumeister, R.F., DeWall, C.M., Ciarocco, N.J. & Bartels, J.M. (2007). Social exclusion decreases prosocial behavior. *Journal of Personality and Social Psychology, 92*(1), 56-66.

Tylén, K., Weed, E., Wallentin, M., Roepstorff, A., & Frith, C. D. (2010). Language as a tool for interacting minds. *Mind & Language*, *25*(1), 3-29.

Van de Calseyde, P. P., Keren, G., & Zeelenberg, M. (2014). Decision time as information in judgment and choice. *Organizational Behavior and Human Decision Processes*, *125*(2), 113-122.

van Schie, H. T., Mars, R. B., Coles, M. G., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature neuroscience*, *7*(5), 549.

Van Swol, L. M., & Drury-Grogan, M. L. (2017). The effects of shared opinions on nonverbal mimicry. *SAGE Open*, *7*(2), 2158244017707243.

Verhoef, T., Kirby, S., & De Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, *43*, 57-68.

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8-9), 998-1003.

Vesper, C., van der Wel, R., Knoblich, G., and Sebanz, N. (2011). Making oneself predictable: reduced temporal variability facilitates joint action coordination. Exp. Brain Res. 211, 517–530. doi: 10.1007/s00221-011-2706-z

Vesper, C., van der Wel, R. P., Knoblich, G., & Sebanz, N. (2013). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 48-61.

Vesper, C., and Richardson, M. (2014). Strategic communication and behavioral coupling in asymmetric joint action. Exp. Brain Res. 232, 2945–2956. doi: 10.1007/s00221-014-3982-1

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.

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., ... & Schmitz, L. (2017a). Joint action: mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in psychology*, *7*, 2039.

Vesper, C., Schmitz, L., & Knoblich, G. (2017b). Modulating action duration to establish nonconventional communication. *Journal of Experimental Psychology: General*, *146*(12), 1722

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. S*cience*, *311*(5765), 1301-1303.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child development*, *77*(3), 640-663.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, *11*(3), 271-294.

Watson, M. E., Pickering, M. J., & Branigan, H. P. (2004, January). Alignment of reference frames in dialogue. *Proceedings of the Annual Meeting of the Cognitive Science Society, 26*(26), 1434-1439.

Webb, J. T. (1969). Subject speech rates as a function of interviewer behaviour. *Language and Speech*, *12*(1), 54-67.

Weiner, E. J., & Labov, W. (1983). Constraints on the agentless passive. *Journal of linguistics*, *19*(1), 29-58.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, *11*(4), 277-281.

Wilkes-Gibbs, D. & Clark, H.H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31*, 183-194.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological bulletin*, *131*(3), 460.

Winner, E., Brownell, H., Happé, F., Blum, A. & Pincus, D. (1998). Distinguishing lies from jokes: theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language, 62*, 89–106.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature neuroscience*, *3*(11s), 1212.

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 B: Biological Sciences*, *358*(1431), 593-602.

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same story, different story: the neural representation of interpretive frameworks. *Psychological science*, *28*(3), 307-319.

Yu, R., & Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *Neuroreport*, *17*(16), 1747-1751.

Zaki, J. (2013). Cue Integration: A Common Framework for Social Cognition and Physical Perception. *Perspectives on Psychological Science, 8(3)*, 296-312.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, *20*(1), 45-57.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162.

# Appendix A

## 7.1. Experiment 1/A

**Compatibility effect across helpers and helpees.** To understand more about the compatibility effect displayed by helpees in the joint condition, we compared the compatibility effect both in terms of RT and ER in Exp. 1/A across helpers and helpees. With RT, a 2x2 repeated measures ANOVA with the factors Role (Helper vs. Helpee) and Condition (Joint vs. Baseline) yielded no significant interaction or main effects (all *F*s < 1). With ER, the same ANOVA showed a nearly significant main effect of Condition (*F*(1, 42) = 3.75, *p* = .06, $\eta^2_p$ = .08), with a smaller compatibility effect on ER in the Joint (*M* =2.03%, *SD* = 2.92) than in the Baseline condition (*M* = 3.98%, *SD* = 6.65). The main effect of Role and the interaction were not significant (*F*s < 1).

## 7.2. Experiment 1/B

**RT distribution analysis across the joint/cueing conditions of Exp. 1/A and 1/B.** In Exp. 1/B we found that helpers showed a larger compatibility effect in the cueing condition than in the baseline. We hypothesized that this effect was due to helpees' RT being less spread out in this condition, maximizing the spatial incompatibility effect, but only in Exp. 1/B. To test this idea, we compared helpees' RT distribution across the joint condition of Exp. 1/A and the cueing condition of Exp.1/B using quintiles. We performed a 2x5 mixed design ANOVA with between-subjects factor Error Source (Human vs. Computer) and within-subject factor Quintile (from 1st to 5th). Mauchly's test indicated that the sphericity assumption was violated ($\chi^2(9) = 205.17$, $p < .001$), thus we corrected degrees of freedom using the Greenhouse – Geisser estimate of sphericity ($\varepsilon = .337$). The trivial main effect of Quintile was significant ($F(1, 160) = 179.25$, $p < .001$, $\eta^2_p = .82$), whereas the main effect of Error Source was not significant ($F(1, 40) < .01$ ). As can be seen in Figure A1, the interaction was in the predicted direction, with helpees in Exp. 1/B showing larger mean RT in the first quintiles, but smaller in the last quintiles. However, the interaction was not significant: $F(1, 160) = 3.43$, $p = .057$, $\eta^2_p = .08$.

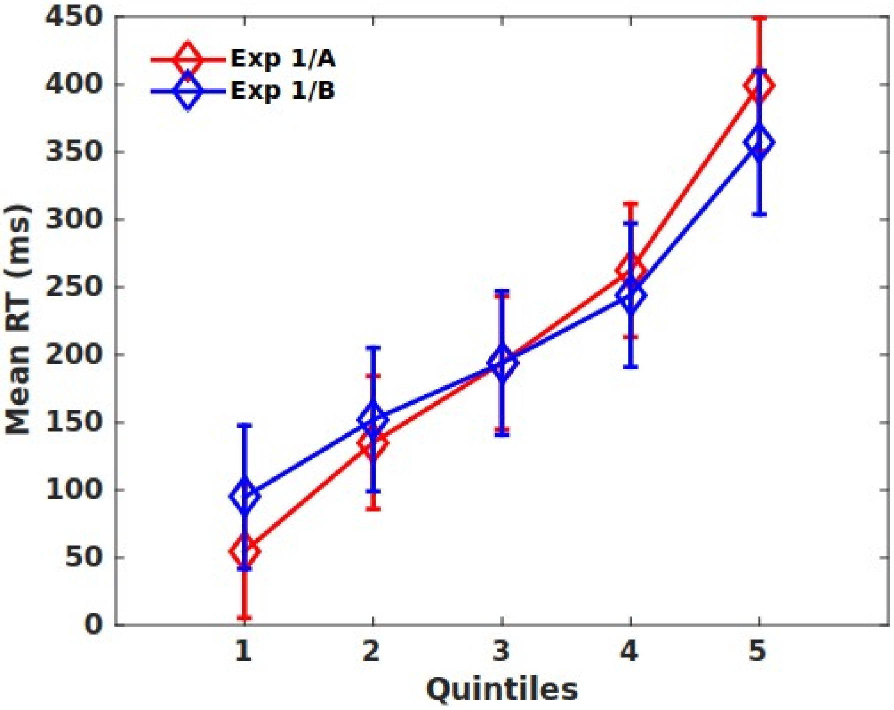


*Figure A1*. Helpees' RT distribution in the joint condition of Exp. 1/A and the cueing condition of Exp. 1/B. The red line corresponds to the quintiles in Exp. 1/A, the blue one to the quintiles in Exp. 1/B. Error bars depict *SEM*. Note that the data in Exp. 1/B is overall less variable across quintiles.

## 7.3. Experiment 2/A

**ANOVA on helpers' RT data using the whole Other group (*n* = 15).** The weakness of this analysis – relative to the version reported in the main text using a subset of the Other-cueing group (response-change and incompatibility cueing) - is that there is substantial variability in the number of RTs among helpers in the full Other cueing group, as some helpers used the cue in almost all trials while others used it hardly at all. The strength though is that we rely on the predefined categories and not on post hoc selections. We can only report RT results here for the whole group, as ER are undefined for some of the helpers in the whole Other-cueing group.

The same 2x2 repeated measures ANOVA reported on helpers' RT data in Exp. 2/A (factors Condition: Joint vs. Baseline; Cueing: Response vs. Other) was also performed with the whole Other-cueing group ($n = 15$). The results are shown in Figure A2. The main effect of Condition was significant ($F(1, 33) = 34.57$, $p < .001$, $\eta^2_p = .51$), with larger RT in the Joint condition ($M = 1191$ ms, $SD = 430$) than in the Baseline condition ($M = 796$ ms, $SD = 163$). We found no significant main effect of Cueing ($F(1, 33) = 2.55$, $p = .12$, $\eta^2_p = .07$), but the interaction was significant: $F(1, 33) = 4.83$, $p = .035$, $\eta^2_p = .13$. Pairwise comparisons showed that the difference between the Response-cueing group and the Other-cueing group just failed to reach significance ($t(33) = 2.02$, $p = .052$, Cohen's $d = .69$) in the Joint condition, but was clearly not significant in the Baseline condition ($t(33) = .50$, $p = .62$, Cohen's $d = .04$).

The results are very similar to the ones obtained with the restricted Other-cueing group in the main text. Using the whole group we found though an interaction between Condition and Cueing that was missing in the restricted group. However, the interaction would not affect the interpretation of the data.

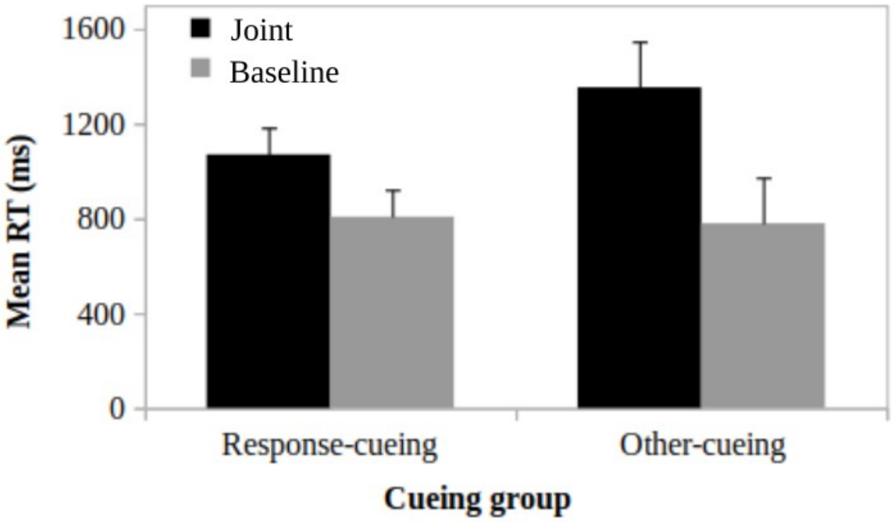


*Figure A2*. Helpers' mean response times in Experiment 2/A. Black bars depict the mean RT in the joint while gray bars depict mean RT in the baseline condition. Response-cueing refers to the group of helpers signaling the required responses for helpees (*n* = 20) and Other-cueing refers to all helpers falling into the Other category (*n* = 15). Error bars show within-subject confidence intervals, calculated separately for the two groups of helpers (95% CI, based on Loftus & Masson, 1994).

**ANOVA on helpees' RT and ER data using the whole Other group (*n* = 15).** As with helpers' data, we calculated the 2x2 repeated measures ANOVAs on helpees' RT and ER data with the full Other-cueing group (factors Condition: Joint vs. Baseline; Cueing: Response vs. Other). Figure A3 depicts the results.

The ANOVA on helpees' RT data resulted in significant main effects both for Condition (*F*(1, 33) = 9.98, *p* = .003, $\eta^2_p$ = .23) and Cueing (*F*(1, 33) = 5.68, *p* = .023, $\eta^2_p$ = .15). Mean RT was larger in the Baseline (*M* = 648 ms, *SD* = 300) than in the Joint condition (*M* = 805 ms, *SD* = 147), and was also larger in the Other-cueing group (*M* = 808 ms, *SD* = 147) of helpees than in the the Response-cueing group (*M* = 665 ms, *SD* = 195).

There was also a significant interaction: *F*(1, 33) = 9.37, *p* = .004, $\eta^2_p$ = .22. Pairwise comparisons revealed that in the Joint condition helpees in the Response-cueing group had smaller RT (*M* = 529 ms, *SD* = 316) than helpees from the Other-cueing group (*M* = 806 ms, *SD* = 188, *t*(31) = 3.23, *p* = .003, Cohen's *d* = 1.03, using Welch's t-test because of unequal variances), while there was no such difference in the Baseline condition: *t*(33) = .19, *p* = .85, Cohen's *d* = .07.

The same ANOVA on ER yielded a significant main effect of Condition (*F*(1, 33) = 6.76, *p* = .014, $\eta^2_p$ = .17), with larger ER in the Baseline (*M* = 5.10%, *SD* = .86) than in the Joint condition (*M* = 3.64%, *SD* = .62). We found no main effect of Cueing (*F*(1, 33) = .52, *p* = .48, $\eta^2_p$ = .02). There was a significant interaction: *F*(1, 33) = 4.26, *p* = .047, $\eta^2_p$ = .11. Pairwise comparisons showed that in the Joint condition there was a significant difference (*t*(33) = 2.37, *p* = .024, Cohen's *d* = .81) between ER of helpees from the Response-cueing group (*M* = 4.41%, *SD* = 3.29) and of helpees from the Other-cueing group (*M* = 7.17%, *SD* = 3.59). There was no such difference in the Baseline condition: *t*(33) = .62, *p* = .54, Cohen's *d* = .21.

There is one meaningful difference relative to the results obtained using the restricted Other-cueing group. In the main text, for ER, the only effect was the main effect of Condition. The difference suggests that with the whole Other-cueing group the error-reducing effect of Cueing is more narrow, present mainly in the Joint condition of the Response-cueing group.

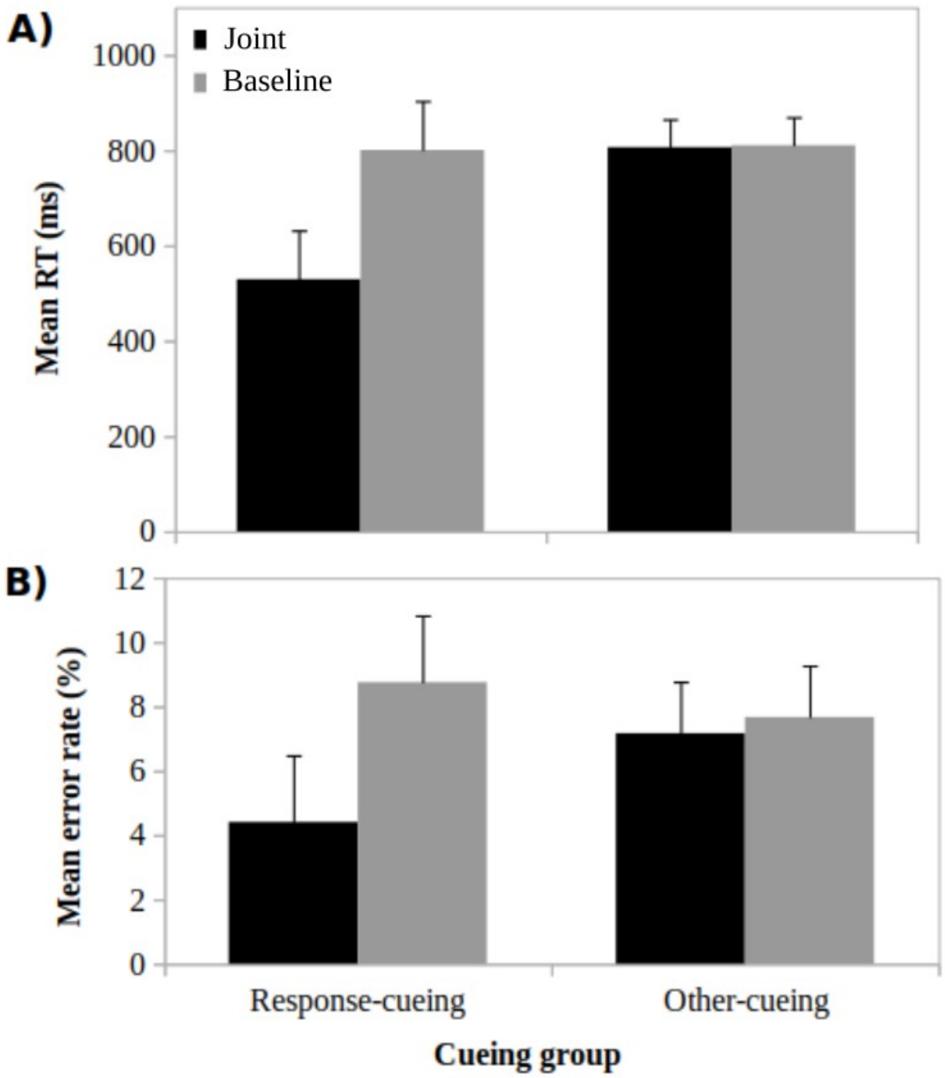


*Figure A3.* (A) Helpees' mean response times in Experiment 2/A. (B) Helpees' mean ER in Experiment 2/A. In both panels, black bars depict the means from the joint condition while gray bars depict the means from the baseline conditions. Response-cueing refers to the group where helpers signaled the required responses for helpees (*n* = 20) and Other-cueing refers to all pairs falling into the Other category (*n* = 15). Error bars show within-subject confidence intervals, calculated separately for the two groups of helpees (95% CI, based on Loftus & Masson, 1994).

## 7.4. Cue utilization

**Linear slope differences as a function of trials included.** In the main text we reported the results for linear slope coefficients calculated on the first ten trials. To show that the main effect of Cue source is not dependent on including a specific number of trials in the analysis, we computed the mean coefficient values against the number of trials included from the beginning of the experiment, both for the human cueing and computer cueing experiments in Figure A4. The difference is robust for the first ~13 trials.

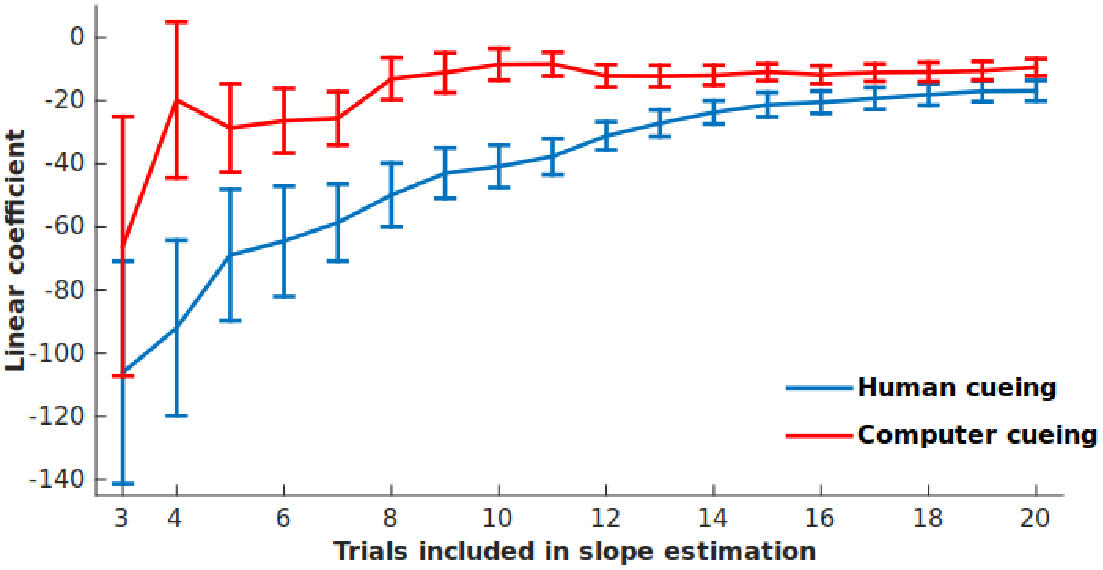


*Figure A4.* Average linear coefficients as a function of the number of trials included from the beginning of helpees' RT data in the joint/cueing conditions of all four experiments. The blue line shows the linear coefficients for the group of helpees from experiments with human cueing (collapsed data from Exp. 1/A and 2/A), while the red line shows the same for helpees from experiments with computer cueing (collapsed data from Exp. 1/B and 2/B). Error bars depict *SEM*.

# Appendix B

## ANOVA-type output tables for LMMs

All tables below show results from Type I SS ANOVAs, with Satterthwaite's method for correcting degrees of freedom.

Table B1

LMM results on median pitch difference in Real pairs.

**Dependent variable: median pitch difference (Hz)**

| | Sum of sq | Mean sq | Numerator DF | Denumerator DF | F value | P value | |
|---|---|---|---|---|---|---|---|
| Interactivity | 11.462 | 11.462 | 1.0 | 44.0 | 0.165 | 0.6864 | |
| Goal | 90.067 | 90.067 | 1.0 | 44.0 | 1.298 | 0.2608 | |
| Turn number | 1656.778 | 1656.778 | 1.0 | 5036.0 | 23.870 | < 0.000 | *** |
| Interactivity:Goal | 83.498 | 83.498 | 1.0 | 44.0 | 1.203 | 0.2787 | |
| Interactivity:Turn number | 301.915 | 301.915 | 1.0 | 5036.0 | 4.350 | 0.0371 | * |
| Goal:Turn number | 70.032 | 70.032 | 1.0 | 5036.0 | 1.009 | 0.3152 | |
| Interactivity:Goal:Turn number | 204.695 | 204.695 | 1.0 | 5036.0 | 2.949 | 0.0860 | |

Notes: *** *p* < .001; ** *p* < .01; * *p* < .05;

Table B2

LMM results on median pitch correlation coefficients in Real pairs.

**Dependent variable: pitch correlation coeffs (Fisher-Z transformed)**

| | Sum of sq | Mean sq | Numerator DF | Denumerator DF | F value | P value |
|---|---|---|---|---|---|---|
| Interactivity | 4.70E-02 | 4.70E-02 | 1.0 | 132.0 | 0.499 | 0.4811 |
| Goal | 1.18E-03 | 1.18E-03 | 1.0 | 132.0 | 0.013 | 0.9109 |
| Interactivity:Goal | 1.03E-02 | 1.03E-02 | 1.0 | 132.0 | 0.109 | 0.7414 |

Notes: *** *p* < .001; ** *p* < .01; * *p* < .05;

Table B3

LMM results on median pause length difference in Real pairs.

**Dependent variable: median pause length difference (ms)**

| | Sum of sq | Mean sq | Numerator DF | Denumerator DF | F value | P value | |
|---|---|---|---|---|---|---|---|
| Interactivity | 843.382 | 843.382 | 1.0 | 44.0 | 0.249 | 0.6205 | |
| Goal | 8164.363 | 8164.363 | 1.0 | 44.0 | 2.407 | 0.1279 | |
| Turn number | 2497.789 | 2497.789 | 1.0 | 5036.0 | 0.736 | 0.3908 | |
| Interactivity:Goal | 1239.093 | 1239.093 | 1.0 | 44.0 | 0.365 | 0.5487 | |
| Interactivity:Turn number | 23.352 | 23.352 | 1.0 | 5036.0 | 0.007 | 0.9339 | |
| Goal:Turn number | 20492.883 | 20492.883 | 1.0 | 5036.0 | 6.042 | 0.0140 | * |
| Interactivity:Goal:Turn number | 115.418 | 115.418 | 1.0 | 5036.0 | 0.034 | 0.8537 | |

Notes: *** *p* < .001; ** *p* < .01; * *p* < .05;

Table B4

LMM results on median pause length correlation coefficients in Real pairs.

**Dependent variable: median pause length correlation coeffs (Fisher-Z transformed)**

| | Sum of sq | Mean sq | Numerator DF | Denumerator DF | F value | P value | |
|---|---|---|---|---|---|---|---|
| Interactivity | 0.150 | 0.150 | 1.0 | 176.0 | 2.044 | 0.1545 | |
| Goal | 0.376 | 0.376 | 1.0 | 176.0 | 5.120 | 0.0249 | * |
| Interactivity:Goal | 0.098 | 0.098 | 1.0 | 176.0 | 1.327 | 0.2509 | |

Notes: *** *p* < .001; ** *p* < .01; * *p* < .05;

Table B5

LMM results on speech rate difference in Real pairs.

**Dependent variable: speech rate difference (syl/sec)**

| | Sum of sq | Mean sq | Numerator DF | Denumerator DF | F value | P value | |
|---|---|---|---|---|---|---|---|
| Interactivity | 0.0907 | 0.0907 | 1.0 | 44.0 | 0.490 | 0.4876 | |
| Goal | 0.0121 | 0.0121 | 1.0 | 44.0 | 0.066 | 0.7991 | |
| Turn number | 0.1610 | 0.1610 | 1.0 | 5036.0 | 0.870 | 0.3510 | |
| Interactivity:Goal | 0.0094 | 0.0094 | 1.0 | 44.0 | 0.051 | 0.8224 | |
| Interactivity:Turn number | 0.1737 | 0.1737 | 1.0 | 5036.0 | 0.939 | 0.3327 | |
| Goal:Turn number | 4.8313 | 4.8313 | 1.0 | 5036.0 | 26.108 | < 0.00 | *** |
| Interactivity:Goal:Turn number | 0.0794 | 0.0794 | 1.0 | 5036.0 | 0.429 | 0.5125 | |

Notes: *** *p* < .001; ** *p* < .01; * *p* < .05;

Table B6

LMM results on speech rate correlation coefficients in Real pairs.

**Dependent variable: speech rate correlation coeffs (Fisher-Z transformed)**

| | Sum of sq | Mean sq | Numerator DF | Denumerator DF | F value | P value | |
|---|---|---|---|---|---|---|---|
| Interactivity | 0.030 | 0.030 | 1.0 | 176.0 | 0.563 | 0.4542 | |
| Goal | 0.376 | 0.376 | 1.0 | 176.0 | 6.996 | 0.0089 | ** |
| Interactivity:Goal | 0.007 | 0.007 | 1.0 | 176.0 | 0.129 | 0.7202 | |

Notes: *** *p* < .001; ** *p* < .01; * *p* < .05;

Table B7

LMM results on syntactic similarity in Real pairs.

**Dependent variable: syntactic similarity (std scale)**

| | Sum of sq | Mean sq | Numerator DF | Denumerator DF | F value | P value | |
|---|---|---|---|---|---|---|---|
| Interactivity | 2.95E-03 | 2.95E-03 | 1.0 | 44.0 | 4.874 | 0.0325 | * |
| Goal | 1.72E-04 | 1.72E-04 | 1.0 | 43.8 | 0.284 | 0.5968 | |
| Turn number | 2.67E-03 | 2.67E-03 | 1.0 | 5031.9 | 4.420 | 0.0356 | * |
| Interactivity:Goal | 6.60E-05 | 6.60E-05 | 1.0 | 44.0 | 0.109 | 0.7428 | |
| Interactivity:Turn number | 3.61E-03 | 3.61E-03 | 1.0 | 5031.8 | 5.973 | 0.0146 | * |
| Goal:Turn number | 7.40E-04 | 7.40E-04 | 1.0 | 5031.8 | 1.223 | 0.2689 | |
| Interactivity:Goal:Turn number | 5.74E-06 | 5.74E-06 | 1.0 | 5031.8 | 0.009 | 0.9224 | |

Notes: *** *p* < .001; ** *p* < .01; * *p* < .05;

Table B8

LMM results on lexical recurrence in Real pairs.

**Dependent variable: lexical recurrence rate (%)**

| | Sum of sq | Mean sq | Numerator DF | Denumerator DF | F value | P value | |
|---|---|---|---|---|---|---|---|
| Interactivity | 17536.357 | 17536.357 | 1.0 | 43.7 | 400.008 | < 0.000 | *** |
| Goal | 1.186 | 1.186 | 1.0 | 45.6 | 0.027 | 0.8701 | |
| Turn number | 1348.738 | 1348.738 | 1.0 | 5036.0 | 30.765 | < 0.000 | *** |
| Interactivity:Goal | 26.982 | 26.982 | 1.0 | 43.6 | 0.615 | 0.4370 | |
| Interactivity:Turn number | 80.511 | 80.511 | 1.0 | 5036.0 | 1.836 | 0.1754 | |
| Goal:Turn number | 81.556 | 81.556 | 1.0 | 5036.0 | 1.860 | 0.1726 | |
| Interactivity:Goal:Turn number | 0.244 | 0.244 | 1.0 | 5036.0 | 0.006 | 0.9405 | |

Notes: *** *p* < .001; ** *p* < .01; * *p* < .05;

Table B9

LMM results on semantic similarity in Real pairs.

**Dependent variable: semantic similarity (std scale)**

| | Sum of sq | Mean sq | Numerator DF | Denumerator DF | F value | P value | |
|---|---|---|---|---|---|---|---|
| Interactivity | 5.00E-03 | 5.00E-03 | 1.0 | 42.5 | 10.740 | 0.0021 | ** |
| Goal | 5.19E-04 | 5.19E-04 | 1.0 | 43.2 | 1.1139 | 0.2971 | |
| Turn number | 2.73E-03 | 2.73E-03 | 1.0 | 5036.0 | 5.861 | 0.0155 | * |
| Interactivity:Goal | 1.46E-03 | 1.46E-03 | 1.0 | 41.1 | 3.1324 | 0.0842 | |
| Interactivity:Turn number | 3.80E-05 | 3.80E-05 | 1.0 | 5036.0 | 0.082 | 0.7752 | |
| Goal:Turn number | 2.81E-03 | 2.81E-03 | 1.0 | 5036.0 | 6.0285 | 0.0141 | |
| Interactivity:Goal:Turn number | 3.99E-05 | 3.99E-05 | 1.0 | 5036.0 | 0.0857 | 0.7698 | |

Notes: *** *p* < .001; ** *p* < .01; * *p* < .05;

# ACKNOWLEDGEMENT

## TO EXTERNAL FUNDING AGENCIES
## CONTRIBUTING TO PHD DISSERTATIONS

Name of doctoral candidate: *Adam Boncz*

Title of dissertation: *Communication as Joint Action: The role of cognitive alignment and coupling*

Name of supervisor(s): Gunther Knoblich, Thalia Wheatley

External funding agency: **European Research Council**

Acknowledgement:

This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) under **grant agreement No 609819 – SOMICS.**