INVESTIGATING THE PRODUCTION AND PERCEPTION OF MOVEMENT CUES IN JOINT ACTION

Luke McEllin

Central European University

Department of Cognitive Science

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

Primary Supervisor: Natalie Sebanz

Secondary Supervisor: Günther Knoblich

Declaration of Authorship

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

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

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

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

McEllin, L., Knoblich, G., & Sebanz, N. (in prep). Comparing the perception of velocity based and interval based synchrony in joint action

Luke McEllin

Abstract

The movements of those engaged in social interactions are laden with meaning, and reflect a whole host of mental states, including intentions and attitudes towards a co-actor. The aim of this thesis was to investigate how the movements of actors engaged in joint actions provide us with information about their informative intentions, and the interpersonal relations of those interacting with each other. Our first study investigated how actors modulate the kinematics of their actions in order to provide informative cues to co-actors, and demonstrated that actions that are identical instrumentally can have different kinematic signatures depending on the informative intentions of the actor (i.e. the intention to coordinate, or the intention to teach). Our second study set out to investigate whether or not observers are able to use kinematic cues to understand an actor's informative intentions, and demonstrated that not only can observers detect the presence of informative intentions on the basis of movement cues, but they can also discriminate between different informative intentions. Our third study aimed to investigate how different types of interpersonal synchrony affect third person perception of the relations between two actors, and found that the movement cues reflecting different types of synchrony have a direct effect on our perception of a performance in terms of the affiliation between the performers, and how aesthetically pleasing we find these performances. In the final section of this thesis, our findings are discussed with respect to their implications for theories of direct perception of mental states, as well as their applications to our understanding of teaching and learning, and human robot interaction.

Acknowledgements

First and foremost, I would like to thank my supervisors Nati and Guenther, who have been nothing short of fantastic in their supervision during my time as a PhD student. Not only am I grateful for their massive contribution to the work in this thesis, but I also thankful for the wealth of knowledge that I feel that I have acquired whilst working with them.

I would like to thank both current and previous members of the CEU department of Cognitive Science, in particular the SOMBY lab who have provided many types of input with regards to my work, from novel theoretical insights, to discussions about the nuances of different statistical tests. SOMBY lab managers, Annamaria, David, Fruszi and Vanda, also deserve plaudits for their help with data collection. Thanks are also in order for Reka, Andi and Eszter, without whom, I'm sure the whole department would descend into chaos!

It hasn't all been work though; I have also met so many fun and interesting people. Martin, Nazli, Otavio, Arianna, Bozsi, Darran, James, Julien and Mateusz….. and many more. You guys have always been there to provide guidance when needed, whether it be guidance in the lab, or guidance to the pub.

I would like to thank my parents. Apart from the obvious, I would also like to express my gratitude for the unwavering support and encouragement that they have shown me throughout the years.

Finally, I would like to thank my fiancé, Judit, who has been there to celebrate with me during the high points, and keep me motivated when I felt like giving up. Thank you for keeping my head firmly attached to my shoulders, even when I felt like it was about to fall off (which would no doubt be problematic with regards to writing a PhD thesis). Because I can't add her as a co-author on any of my papers, I will dedicate this thesis to her!

Table of Contents

Chapter 1. Introduction

Research has shown that movements can reflect a whole host of mental states, from low level instrumental intentions such as whether they will pour from or drink from a bottle (Cavallo et al. 2016), to how confident they are when betting in a game of poker (Slepian, Young, Rutchick & Ambady, 2013). Movements also reflect whether or not one is engaged in a social interaction (Sartori, Becchio, Bara & Castiello, 2009), and even whether or not one is trying to inform a co-actor about their intentions (Pezzulo, Donnarumma & Dindo, 2013). People can also derive the interpersonal relations of multiple actors, on the basis of how the actors' movements relate to each other, for example how synchronized two people are when walking reflects their level of rapport and affiliation (Miles, Nind & Macrae, 2009).

During the last decade researchers have begun to focus on what these movement cues can tell us about people engaged in various social interactions such as joint actions in which people intend to coordinate their movements in space and time, as well contexts such as teaching, in which people transmit information to others through their movements (Sebanz, Bekkering & Knoblich, 2006; Pezzulo et al. 2013). Investigating the production and perception of movement cues in these types of social interactions yields interesting questions with regards to the informative purpose that movement cues serve for those engaged in these social interactions, as well as what these movement cues can tell third-party observers about the relations of those engaged in these social interactions.

Considering the above, I identified three open questions which we aimed to answer using a series of experiments. The first question concerns the production of informative action modulations produced in social interactions. Specifically, are action modulations produced in coordination and teaching serving a generic informative purpose (such as

ostention), or are people modulating their actions in more fine grained ways that optimize the efficacy of the social interaction (chapter two)? The second questions focused on how sensitive people are to the informative intentions underlying these action modulations. Do people only detect that they are being communicated to, or can they derive specific informative intentions underlying these movements (chapter three)? Our final question shifted the focus from movement cues produced by individuals interacting with each other to movement cues produced by groups. Specifically, we aimed to use different modes of synchrony (i.e. synchrony of movement intervals and synchrony of velocity profiles) as a cue produced by multiple individuals engaged in social interaction, and what these cues convey about the interpersonal relations of these individuals. We also aimed to investigate whether these movement cues can go as far as affecting our aesthetic experience of these interactions (chapter four). This chapter will give a review of the relevant literature for these questions, with later chapters reporting how we addressed these questions experimentally.

1.1 Producing movement cues to facilitate joint action and teaching

In social interactions actors share a large amount of information and need to be on an informational 'common ground', which can be described as a mutual understanding that what has been communicated has been understood well enough for the purposes of the current interaction (Clark & Brennan, 1991; Clark, 1996). This is essential in order to form shared task representations which support the prediction of our partner's actions, as well as an understanding of how each actor will contribute to the joint goal (Sebanz; Bekkering & Knoblich, 2006; Vesper, Sebanz & Knoblich, 2010). Communication systems such as language and gesture are commonly used to form a common ground, with information sharing in these modalities allowing for the alignment of task representations (Clark, 1996). Communication through these modalities is supported by strategies which allow for a

communicator to make the message clearer, thus more understandable for an interaction partner. For example, people exaggerate the prosody of their speech in order to maintain the attention of an addressee (Fernald & Simon, 1983), to emphasize particularly relevant information (Wilson & Wharton, 2006). Likewise, people also exaggerate kinematic parameters of their gestures such as amplitude and speed in order to emphasize relevance (Trujillo, Simanova, Bekkering & Ozyurek, 2018). Thus, the above demonstrates that people modulate their communication channel in ways that increase the likelihood of their message being understood, thus supporting the formation a common ground.

Importantly, we also use the movements of our co-actors in order to establish a common ground, with information contained in the execution of ones actions influencing the interpretation of ones actions as well as the wider social interaction (Clark, 2005). Indeed, this is afforded by a close link between perception and action, which have been proposed to share a common representational code which allows for the use of one’s own motor system for the understanding of others actions, as well as predicting how they will unfold (Prinz, 1997; Knoblich & Flach, 2001). In a similar way that those engaged in a dialogue exaggerate the prosody of their speech or the amplitude of their gestures in order to make their message less ambiguous, people engaged in social interactions also modulate their actions in order to make them less ambiguous and easier to predict. This has been labelled 'sensorimotor communication' by Pezzulo Donarumma and Dindo (2013), who suggested that people modulate their actions by deviating from the optimal action trajectory, thus compromising their own action efficiency, in order to produce informative cues for their co-actors. Here, the instrumental action can double up as a channel through which an actor can inform a co-actor with regards to understanding the action being executed, and predicting how it will unfold.

In joint actions, in which two or more agents are required to coordinate their actions in space and time, the ability to understand and predict the spatial and temporal aspects of a co-actors actions is of utmost importance (Sebanz, Bekkering & Knoblich, 2006). This ability relies on the mapping of observed actions onto our own motor system, which is possible due to the close links between perception and action (Wolpert, Doya & Kawato, 2003; Prinz, 1997). There is evidence that people perform informative modulations of their instrumental actions in order to support prediction in joint actions. One study by Sacheli et al. (2013) investigated this phenomenon in dyads who were required to grasp a bottle in synchrony, in either an imitative or complimentary manner. Dyads were assigned to the role of leader and follower, with the leader being instructed to either grasp the top of the bottle with a precision grip, or grasp the bottom of the bottle with a power grip, and the follower being instructed to either an imitative or complimentary movement to the leader. They found that compared to followers, leaders moved with a more exaggerated spatial profile in terms of movement height and grip aperture and reached peak velocity earlier. These modulations served an informative purpose, disambiguating the leaders' movements, allowing the follower to predict the target location (top or bottom of bottle) of the leaders' movement earlier, thus having more time to prepare an appropriate response and coordinate with the leader.

The ability to understand the actions of an interaction partner is also crucial for teaching through demonstration, in which experts show novices how to perform a particular action by using their movements as a model of how the action should be performed (Csibra & Gergely, 2009; Gergely & Csibra, 2005). The ability to understand and encode a demonstrated action sequence and it's components is important in order to learn from demonstration, with this ability, like joint action, also depends on the mapping of an observed movement onto one's own motor system, and is made possible by the close link between perception and action (Wohlsclager & Gattis & Bekkering, 2003; Rizzolati & Craighero,

2004). Developmental psychologists have observed a sensorimotor communication like phenomenon labelled 'motionese' in which adults exaggerate their movements when teaching through demonstration in order to produce cues that support the infants learning of what is being demonstrated. Brand, Baldwin and Ashburn (2002) carried out a study in which adults demonstrated how to use objects to either children or to other adults. They found that compared to adult directed demonstrations, child directed demonstrations where characterized by more exaggerated and punctuated movements, with these action modulations serving to inform novices in a teaching context. These informative action modulations support infants learning of demonstrated actions by guiding their attention to the most learning relevant parts of the action, and allowing them to more effectively parse an action sequence in order to understand and encode the overall structure of the action as well as its components (Brand & Shallcross, 2008; Nagai & Rohlfing, 2008). Although the motionese literature primarily focused on informative cues produced by adults to inform infants, these phenomena are commonly observed in many expert-novice interactions, with parent-infant interactions being the most natural form of expert-novice interaction (Csibra & Gergley, 2006).

Joint action and teaching are two different social contexts in which people modulate their actions, for different informative purposes. Those engaged in joint actions produce informative movement cues in order to enhance a co-actors prediction of the timing and target location of ones actions, whilst those who are teaching through demonstration produce movement cues that guide a learners attention to learning relevant information such as the structure of an action and its components. This leads us to the question of whether the action modulations produced in joint action and teaching serve a general informative purpose when engaged in a social interaction, e.g. ostensively communicating an informative intent

(Sperber & Wilson, 2004), or are the action modulations conveying specific information that is useful for completing the task successfully?

Comparing informative movement modulations in joint action and teaching also yields interesting questions with regards to how people learn when engaged in joint action. Institutionalized pedagogy is just one of the many ways in which we learn, and being engaged in coordinated social interactions is one of them (Rogoff, 2003), yet the ways in which teaching and learning occurs through joint action is yet to be explored. One possibility is that informative action modulations produced in order to support spatial and temporal prediction in joint action double up as learning relevant cues, due to their resemblance to the cues produced in order support the understanding of the structure of an action sequence in teaching.

In order to answer the above questions, we compared the informative kinematic modulations produced in joint actions and in teaching in order to investigate whether the same or different cues are produced in order to enhance spatial and temporal prediction in joint action, and to guide a learner's attention to learning relevant information in teaching contexts.

1.2 Perceiving informative intent from movement cues

In their critique of the 'motor theory of social cognition', Jacob and Jeannerod (2005) proposed that it is highly unlikely that people can derive the social and communicative intention of an actor from simulating observed actions, with this ability yielding the ability only to derive motor intentions. They used a thought experiment that involved imagining that Jill wants to inform John that she wants to leave the very loud party that they are attending, and is pointing to her wristwatch in order to convey this desire. However, Jill's watch is also

broken, meaning that John does not know whether Jill is pointing to her watch because it is broken or because she wants to leave, or she is pointing to express her frustration at her broken watch. In a grave underestimation of John's abilities, the authors suggested that he could not derive anything beyond motor intentions from Jill's movements, and concluded that he would not be able distinguish between these two communicative intentions on the basis of the pointing action alone. This conclusion assumes that the kinematic signatures of an action can only differ on the basis of instrumental intentions, and does not leave open the possibility that instrumentally identical actions may have different kinematic signatures on the basis of informative intentions. However, it is plausible that Jill pointing to convey her frustration at her broken watch, and pointing to convey that she is sick of the party and wants to leave would elicit qualitatively different movement patterns, even though her instrumental goal is the same. Moreover, it is likely that John has motor experience with pointing due to frustration, and pointing to communicate that he wants to leave a place, meaning that he could use his motor system in order to understand whether or not Jill is frustrated at her watch, or if she wants to leave. Indeed, there is evidence that joint action and teaching intentions can be reflected in an actor’s movements (Pezzulo et al. 2013; Brand et al. 2002), suggesting that it is possible that instrumentally identical actions can have different kinematic signatures based on the informative intentions held by the actor. Considering this, we aimed to investigate whether or not people also detect and discriminate between an actor’s informative intentions on the basis of movement cues. We also aimed to investigate whether or not people can discriminate between different types of informative intentions (i.e. coordination or teaching) on the basis of these cues. This section will discuss some of the ways in which people can use movement cues in order to derive various mental states, particularly social and non-social intentions.

How people move reflects a lot about different facets of their mental state, with observers being able to derive these by watching these movements. For example, people can derive how confident people are in their responses to a stimulus using the length of their movement onset time (Patel, Fleming & Kilner, 2012), and how hesitant one is to bet in a poker game using the level of jerk in the players movements when putting their chips into the pot (Slepian, Young, & Ambady, 2013). Interestingly, those observing dance performances can accurately identify the emotions of the performers on the basis of their movements (Dittrich, Troscianko, Lea & Morgan, 1996). With regards to an actor's intentions, there is evidence that people can derive instrumental intentions, and even social intentions on the basis of their movements (Cavallo et al., 2016; Manera et al. 2010), but whether or not people can derive informative intentions from observing movements is still an open question.

A recent study by Cavallo et al. (2016) aimed to identify the specific kinematic signatures of different types of instrumental actions, and whether people could identify the instrumental intentions underlying these intentions of the basis of these kinematic signatures. To this end, they recorded people reaching from a bottle with the intention to either pour from the bottle, or drink from the bottle and identified the kinematic parameters such as wrist height and grip aperture, which best discriminated between these two actions. They found that people could accurately predict whether the actor was going to drink or pour from the bottle on the basis of these reach to grasp movements. Moreover, they found out that they could manipulate the 'visibility' of these intentions by showing movements that contained more or less of the discriminant kinematic parameters.

As well as using movement cues in order to derive instrumental intentions, there is also evidence that people can discriminate between different social intentions on the basis of movement cues. A study by Georgiou, Becchio, Glover and Castellio (2007) found that

cooperative and competitive actions when placing objects on a table had different patterns of kinematics, with competitive movements being faster with less of an exaggerated spatial profile, compared to cooperative movements which were slower and more exaggerated. Moreover, Manera et al. (2010) found that people could reliably discriminate between actions performed with a cooperative intention and actions performed with a competitive intention on the basis of these movement cues.

An fMRI study by Becchio et al. (2012) revealed differential activation in mirroring and mentalizing areas when viewing social movements, and when viewing non-social movements. Specifically, there was stronger activation when viewing social movements compared to non-social movements combined with increased activity in mirroring networks when viewing social movements, compared to non-social movements. This finding highlights the role of motor simulation and its relationship between mentalizing in decoding an actor's social intentions from low level kinematics in the absence of contextual information.

The above research demonstrates that people can reliably identify the instrumental and social intentions of an actor through observing their movements, using their own motor system to derive these intentions on the basis of the observed motor parameters. However, whether or not observers can derive the informative intention of an actor engaged in a social interaction has yet to be explored. Movement cues produced in social contexts such as competitive scenarios do not necessarily reflect that an actor has an intention to inform a co-actor, and discriminating between these social contexts does not necessarily mean that people perceive informative cues produced by an actor. Considering this, and considering the fact that people produce informative movement cues in joint actions and teaching, we aimed to investigate how people perceive these informative cues, and whether or not people can actually derive more than motor intentions from observing peoples movements (Jacob &

Jeannerod, 2005). Specifically, we wanted to know whether people can discriminate between actions performed with an informative intention and non-informative individual actions on the basis of kinematic cues. We also investigated whether people could discriminate between actions performed with different informative intentions on the basis of kinematic cues. Again, we focused on actions performed with coordination intentions and teaching intentions as they have been shown to lead to different kinematic signatures, even when the action is instrumentally similar. Moreover, this comparison has the potential to yield interesting insights with regards to the relationship between cues to joint action coordination and cues to teaching through demonstration.

1.3 Perceiving interpersonal relations from movement cues

Early evidence that movement cues can reflect various facets of the interpersonal relations between two or more actors comes from a classic study by Heider and Simmel (1944), who investigated whether or not people attributed animacy to two triangles and a circle moving in relation to one another around a box, at different speeds and in different configurations. They found that people consistently described these objects as if they were humans engaged in a social interaction, attributing human like traits to the individual shapes, and humanlike social relations between the objects, on the basis of their 'roles' in the animation. This study demonstrates our remarkable sensitivity to movement cues in social interactions, with people being able to make rich inferences and attributions about social interactions and the relations between those engaged in social interactions, based on very simple movement cues.

More recently, people have begun to investigate the relationship between interpersonal synchrony and the perception of the interpersonal relations between actors. Research on interpersonal coordination has demonstrated that being synchronized with

another person can foster a whole host of prosocial effects. For example, tapping in synchrony with another person can lead to higher levels of affiliation and liking with that person (Hove & Riesen, 2009), and singing and waving in synchrony with another person can lead to increased cooperation with that person in an economic game, even if this cooperation comes at a cost (Wilthermuth & Heath, 2009). The link between synchrony and prosociality leads to the question of how synchrony affects our perception of affiliation between people, as well as how cues in coordination affect our perception of interpersonal relations more generally. Aiming to investigate how third person perception of synchrony affects how we perceive the affiliation between actors, Miles (2009) found that participants rated point light displays of people walking synchronously with a stable phase relationship as having more rapport than those walking asynchronously with a unstable phase relationship.

Interestingly, there is also evidence demonstrating that relational movement cues can influence an observers' aesthetic experience of a performance. Although, aesthetics research has traditionally emphasized the role of the visual system in aesthetic experience (Zeki, 2001), more recent work has demonstrated the influence that the motor system has on aesthetic experience (Calvo-Merino, Jola, Glaser & Haggard, 2008; Cross, Kirsch, Ticini & Schutz-Bosbach, 2011). With regards to group performances, a study by Vicary, Sperling, von Zimmerman, Richardson and Orgs (2017) found that the levels of synchrony between dancers in a performance predicted spectators level of arousal (as indexed by heart rate), as well as their subjective enjoyment and aesthetic experience of the performance, demonstrating that relational movement cues play a role in aesthetic experience, as well as individual movement cues.

The above studies provide us with evidence that synchrony can be used as a cue in order to understand the interpersonal relations of those engaged in a social interaction.

However, these studies have focused on what we call interval based synchrony, in which actors are only required to synchronize the end points of their movements. However, as well as synchronizing the intervals of their movements, actors can also synchronize by aligning the velocity profiles of their movements in order to achieve frame to frame synchrony. Compared to interval based synchrony, this velocity based synchrony is more sophisticated and requires continuous alignment, rather than just holding a phase relationship. Using a task in which actors were required to move a slider from side to side and synchronize with their co-actors' Noy, Dekel & Alon (2011) found that although both experts and novices could achieve interval based synchrony by coordinating the turning points of their movements, only experts could successfully achieve velocity based synchrony. Experts' velocity based synchrony was characterized by their smooth and aligned velocity profiles, and novices' failure to reach velocity based synchrony was reflected by the jitter of the participants' movements, with followers' velocity profile jittering around leaders movements. Using the same paradigm, Hart, Noy, Feinger-Schall and Alon (2014) found that when synchronizing, experts' were able to modulate the shape of their velocity profile in order to make their movements more predictable and easier to align with, whereas novices did not break away from their idiosyncratic manner of moving when attempting to synchronize. These studies demonstrate that interval based synchrony and velocity based synchrony can be distinguished on the basis of how the timing of the two actors' movements relate to each other, with velocity based synchrony reflecting a deeper and more sophisticated mode of synchrony than interval based synchrony. Considering this, our aim was to investigate whether cues to interval based synchrony and velocity based synchrony differ with regards to their contributions to our perception of how coordinated and affiliated two actors are. We also aimed to investigate how interval based and velocity based synchrony differ with regards to how they influence people's aesthetic experiences when observing a social interaction.

1.4 Movement cues in joint action: Open questions

The next three chapters will report how we addressed the questions raised in this chapter. Chapter 2: *Producing cues for informative intentions in joint action and teaching* will report a series of experiments used a virtual xylophone paradigm which allowed us to record the movements of individuals engaged in joint action coordination and teaching interactions, in order to investigate the differences between the informative action modulations that people produce in these different types of social interaction. Chapter 3: *Perceiving informative intentions in joint action and teaching using movement cues* will focus on how people can use these movement cues in order to recognize social intentions of people engaged in social actions. We will report a series of experiments which employed a visual categorization task in which we investigated whether participants could discriminate between actions performed with an informative intention, and non-informative individual actions on the basis of movement cues produced in these actions. This study also investigated whether people could use movement cues in order to discriminate between action produced with the intention to coordinate in a joint action, and actions produced with the intention to teach. In Chapter 4: *Perceiving joint action using relational movement cues*, we present a study in which we presented participants with dyads engaged in a slider based coordination task in which individuals were required to synchronize with each other. Here, we aimed to investigate how observers judge the level of coordination and level of affiliation of the individuals engaged in the social interaction, on the basis of synchrony cues from the dyads movements. We also aimed to investigate to what extent these cues influenced an observers aesthetic experience when watching the social interaction.

Chapter 2. Kinematic Markers of Demonstration and Coordination

2.1 Introduction

In social interactions, people often modulate their instrumental actions to carry additional communicative signals (Sartori, Becchio, Bara & Castiello, 2009; Pezzulo, Donnarumma & Dindo, 2013). For example, when reaching for a bottle one can modulate the trajectory and speed in order to communicate whether one is intending to pour from the bottle or to take it away. This in turn allows an interaction partner to prepare an appropriate motor response. Likewise, an expert sommelier demonstrating to a novice how to pour wine from a bottle into a glass may slow down and exaggerate her movements to highlight particular aspects of the action. This type of communicative action modulation has been labelled sensorimotor communication and can be defined as communication through the same channel as the executed action (Pezzulo et al., 2013).

Evidence for sensorimotor communication comes from two separate domains. On the one hand, systematic action modulations have been observed in pedagogical contexts where parents modulate the kinematics of their actions in order to highlight the significance of particular actions and in order to communicate knowledge of the structure of actions to their infants (e.g., Brand, Baldwin & Ashburn, 2002). On the other hand, research on joint action has found that actors modulate the kinematics of their actions in order to enhance spatial and temporal prediction, making these actions easier for a partner to coordinate with (Pezzulo et al. 2013; Vesper & Richardson, 2014). The observed action modulations in pedagogical contexts and in joint action contexts have received different explanations in terms of the underlying mechanisms. At the same time, it has proven difficult to draw conclusions about

similarities and differences with regard to mechanisms involved in teaching and joint action coordination because the studies from these two different domains differ in many ways.

The current study aimed to directly compare sensorimotor communication in teaching and in joint action contexts, in order to better understand what drives the emergence of particular kinematic modulations. In a series of three experiments, we measured how trained individuals played melodies on a virtual xylophone in order to demonstrate these sequences to a learner (Experiment 1), to play in synchrony with a naive partner (Experiment 2), or to play in synchrony with a partner who had been trained on the same sequences (Experiment 3). To motivate the specific questions and hypotheses for our study, we review prior research on sensorimotor communication in pedagogical and joint action contexts in the next two sections.

2.1.1 Modulating actions to demonstrate

There is evidence that when demonstrating actions to children or naive observers, human adults modulate their instrumental actions in order to make them more informative; this is known as motionese (Brand, Baldwin & Ashburn, 2002). When interacting with infants, caregivers modulate their kinematics in order to put more emphasis on the meaning and significance of the actions, and to elucidate the structure of these actions (Brand et al., 2002). For example, in a study by Brand, Baldwin and Ashburn (2002), mothers were asked to demonstrate features of novel objects to infants or adults. They found that demonstrations directed towards infants were not only more engaging and simple than demonstrations directed at adults, but were also more punctuated and exaggerated.

Infants show more attention to, and are more likely to imitate actions containing motionese (Brand & Shallcross, 2008; Koterba & Iverson, 2009; Nagai & Rohlfing, 2009).

Koterba and Iverson (2009) carried out a study in which they had caregivers demonstrate features of objects which the infant would then interact with. They found that infants looked longer at demonstrations containing motionese (actions performed with a higher amplitude, and more repetitions) and interacted with objects longer when the caregiver's demonstration had contained motionese.

It has been proposed that the main function of motionese is to enhance attention, and to highlight boundaries between action units (Brand et al., 2002). In line with this view, evidence from robotics research has demonstrated that motionese can influence visual attention and pattern recognition. In a study using a bottom up attention model based on saliency, Nagai and Rohlfing (2009) found that motionese increased the saliency of particular actions, resulting in increased visual attention to important parts of the action sequence. They concluded that motionese helps to guide attention to end states of an action, and facilitates pattern recognition by allowing observers/ robots to extract primitives from the observed action sequence.

More generally, it has been proposed that humans have a unique sensitivity to communicative intentions underlying observed actions (Csibra & Gergely, 2009). Infants selectively attend to and generalize information from actions containing ostensive cues that signal that an action is intended to be communicative. Features of motionese, such as exaggerated movement amplitudes, may convey that an action is intended to be communicative and that learning-relevant information is being provided, thus drawing attention to the action and facilitating imitation. Although much of the evidence informing this account stems from infant studies, ostensive communication is thought to guide teaching and learning in adults as well (Gergely & Csibra, 2013; Sperber & Wilson, 2002). Investigating how instrumental actions are performed in a demonstration context, the present

study contributes to answering the question of whether adults modulate actions for other adults (for modulations of speech and gestures, see Campisi & Özyürek, 2013).

2.1.2 Modulating actions to coordinate

Studies on action kinematics in the context of joint action coordination suggest that the need to coordinate also generates communicative modulations of action. In joint action, these modulations may allow one to make coordination smoother and more efficient, by making actions more informative (Pezzulo et al. 2013). A study by Sacheli et al. (2013) used a task in which two participants were instructed to grasp a bottle synchronously, with either a power grip or a precision grip. One of the participants had advance information about the action to be performed while the other had to rely on the actions of the informed task partner to select the appropriate grip. Informed participants modulated kinematic parameters, such as wrist height and grip aperture, as well as reducing the velocity of the reaching movement. The authors suggested that these kinematic modulations made participants' actions more predictable, allowing their co-actors to understand the goal of the action earlier.

Converging evidence comes from a study by Vesper and Richardson (2014) in which participants were instructed to synchronize a sequence of taps. Crucially, only the "leader" had knowledge of the target locations. The maximum height of the leader's actions was significantly higher when followers had full vision of their actions, compared to a condition in which followers could only see the start and end of the leader's actions, and compared to an individual baseline that did not involve coordination. The authors suggested that increasing the movement amplitude allowed the leaders to make their actions easier for the followers to predict.

But how exactly could such subtle modulations of kinematic parameters facilitate action prediction? The key idea here is that action prediction relies on internal forward models that guide not only predictions about the sensory outcomes of our own actions (Wolpert & Flanagan, 2001), but are also involved in making predictions about others' actions (Wolpert, Doya & Kawato, 2003). Sensorimotor communication could improve the predictive efficacy of these internal models, with less ambiguous actions allowing one to select the most appropriate model to predict another's actions either earlier or more accurately (Vesper & Richardson, 2014; Sacheli et al. 2013). A related proposal is based on the observation that actions performed with communicative intentions tend to deviate from the most typical, most efficient trajectory. Deviations from the most efficient path can be seen as a cost invested in helping a task partner to disambiguate one particular action from other possibilities, which ultimately benefits both interaction partners in a coordination context (Pezzulo et al., 2013). Such action modulations can be helpful even if the partner does not recognize the actor's communicative intent. However, more inefficient actions (such as actions performed with high amplitudes) can also serve as cues to an actor's communicative intention (ibid).

So far, joint action research has investigated how modulating kinematics can support coordination of discrete actions which require knowledge of an upcoming target (Vesper & Richardson 2014). It has not been investigated whether and how kinematics are modulated in a joint task that requires coordination of a whole sequence of actions, which is typical of many real life joint actions, such as playing a piece of music together. Furthermore, previous tasks investigated situations where the less knowledgeable partner needed to make both spatial and temporal predictions based on the movements of the more knowledgeable individual. It is still unknown whether actions are modulated differentially depending on whether the interaction partner needs to predict only the timing or also the target location of

the actions. Dissociating these two possibilities can inform the debate on whether distinct kinematic modulations occur in different action contexts (Becchio et al. 2014; Jacob & Jeannerod, 2005), and can reveal whether people make unique adjustments depending on the knowledge state of their joint action partner (Sartori et al. 2009).

2.1.3 Coordination vs Demonstration

The literature on motionese and joint action coordination demonstrates that people communicate through the channel of instrumental actions for various social purposes. In joint actions, kinematics are modulated in order to effectively achieve coordination, by enhancing spatial and temporal motor prediction. In teaching contexts kinematics are modulated to enhance learning, by guiding an observer's attention, facilitating parsing of the structure of the action, and conveying the relevance of particular aspects of the observed actions. So far, it has not been investigated whether sensorimotor communication is employed to a similar extent in teaching and joint action contexts, and whether the particular ways in which actions are modulated in these contexts imply specific kinematic signatures (Cavallo et al. 2016) or whether the same kinematic modulations can serve different functions in different interpersonal contexts.

The current study used a virtual xylophone paradigm in order to investigate sensorimotor communication in three interactive contexts. In the "demonstration interaction" (Experiment 1), participants were required to teach xylophone melodies that they had learnt to play to a student watching their actions. There was no requirement to coordinate their actions with the student, implying that any modulations of the demonstrator's actions compared to an individual baseline would be related to the need to transmit knowledge of the action sequences to the student. In the "unequal knowledge coordination interaction" (Experiment 2) participants played xylophone melodies that they had previously learnt to

play individually, together with a follower who did not know the melodies. This implies that any systematic action modulations relative to the individual baseline likely served to facilitate both spatial and temporal action prediction. Finally, in the "equal knowledge coordination interaction" (Experiment 3), participants played melodies together with a task partner who had received equal practice with the melodies. In this type of interaction, action modulations reflect the need to facilitate temporal coordination.

If sensorimotor communication occurs not only in parent-infant interactions in the form of motionese, but also in teaching interactions between an adult teacher and student, we should find that teachers in our study modulate their actions to facilitate learning, such as moving more slowly and with higher movement amplitudes. Furthermore, if the informed individual in a joint action modulates actions not only when performing single actions (as investigated in previous studies), but also when performing a sequence of actions, we should observe sensorimotor communication in the unequal knowledge coordination interaction. Finally, if we find similar kinematic modulations regardless of whether or not the joint action partner has knowledge about the action sequences to be performed, we can conclude that sensorimotor communication plays a role for temporal action coordination in the absence of any knowledge asymmetries concerning action goals (Vesper et al., 2016).

2.2 Experiment 1: Demonstration

Our first experiment aimed to investigate kinematic cues produced in a teaching context, in which a demonstrator was required to teach xylophone melodies to a student. We tested whether, compared to an individual baseline, demonstrators would modulate their actions to help learners encode sequences of demonstrated actions, facilitating parsing them in terms of sub-goals. We expected similar action modulations for demonstrating an action sequence to another adult as for demonstrating action sequences to infants and small children,

including slower movements and higher movement amplitudes compared to individual performance of the demonstrator.

2.2.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 24 participants (15 males, 9 females), with a mean age of 26.6 (SD = 3.41). Participants were required not to have received any musical training, and to be proficient English speakers. Participants worked together in dyads (12 dyads) and were randomly assigned to the role of demonstrator or student. All participants gave informed consent, and were given 1500 Forint (approximately 5 Euros) worth of vouchers for their participation. In order to justify our sample size we carried out a g*power analysis. Because our study used a novel paradigm, we decided to justify our sample size based on a large effect size ($\eta p2$ = .14). For our design, the analysis determined that 12 participants would give us sufficient statistical power in this experiment[1].

Apparatus and Stimuli

A computer program which emulated a xylophone was created in MATLAB. Using an EPSON projector, 10 keys were projected onto a table that was covered by a white sheet of paper (see Figure 1). The size of the projection was 120 cm x 79 cm (resolution of 1800x1000 pixels). Each key was 5 cm wide (75 pixels), and 24 cm long (304 pixels) with a gap of 4 cm between keys. The keys were blue on a grey background. Each key had a tone label at the top and bottom corresponding to the musical labels for tones on the pentatonic scale (c1, d1, e1, g1, a1, c2, d2, e2, g2, a2). Two areas of 3 cm x 3 cm (45 x 38 pixels), one 5

cm (64 pixels) above and one 5 cm (64 pixels) below the horizontal center of the xylophone, were colored in red and served as starting points for the two participants in each trial.

Drumsticks (length: 39 cm, width: 1.5 cm) served as xylophone mallets with a motion sensor attached to the head of the sticks. The heads of the drumsticks were wrapped in soft material, in order to protect the motion sensor, and to minimize any sound caused by the sticks hitting the table.

A Polhemus G4 motion tracker recorded three dimensional cartesian coordinates of mallet heads. Sampling frequency was 120 HZ, resulting in a frame of data approximately every 8ms. The MATLAB program used 10 fixed areas around the center of each projected key to determine whether the motion sensor on the mallet head 'touched' the key. The tone corresponding to the key was played when the mallet was less than 2.5 cm from the center of the key on the left-right axis, less than 12 cm on the depth axis, and less than 1.5 cm away on the height axis. We used a 1.5 cm threshold on the height axis in order to compensate for the delay in producing the tone that was caused by the program.

We constructed short xylophone melodies consisting of a sequence of four tones each. Tones were drawn pseudo-randomly from the two octave pentatonic scale that the 10 keys comprised, in such a way that any tone (key) occurred not more than once in a sequence. Because of the way the sequences were generated, the distance between successive keys varied randomly between 1 and 9 within the sequences.

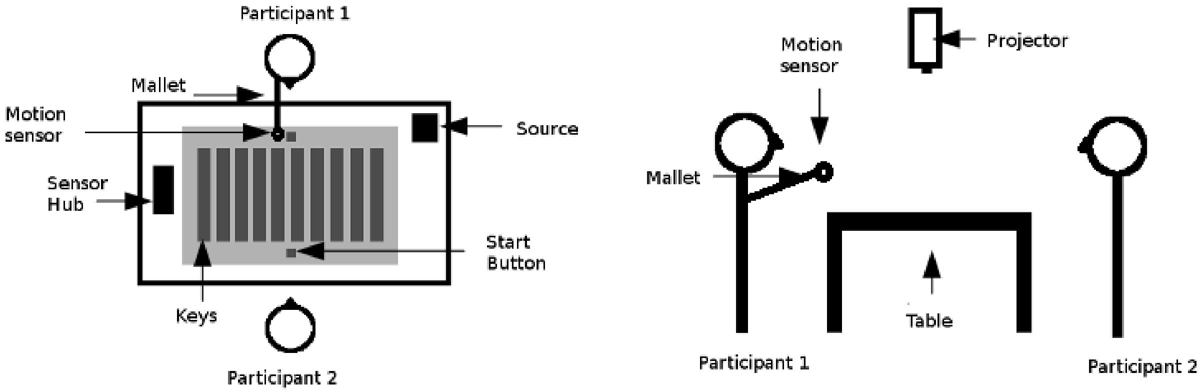

Figure 1: Sketch of the experimental setup, from a bird's eye view (left) and a lateral view (right). See Method section of Experiment 1 for detailed description of apparatus and procedure.

Procedure

For each of four melodies, demonstrators first completed a training phase. Second, they played the melody alone (individual condition). Third, they demonstrated the melody to the student (joint condition), who then attempted to reproduce the melody. This procedure was repeated three times with the same four melodies so that the demonstrator trained, performed, and demonstrated each melody across three consecutive blocks with increasing practice for each of the four melodies. The order of melodies was randomized within each block.

Prior to the experiment, we explained the task to the participant and familiarized them with the equipment. They were asked to remove anything containing metal from their body or pockets, in order to minimize any electromagnetic interference with the motion tracker. Importantly, participants were told that verbal communication was not permitted, other than when the demonstrator instructed the student to turn around and face the xylophone. They were also instructed to keep the mallet within the projection of the xylophone, and to make

sure that they always hit the table when striking a key. They also completed a practice block, in order to ensure that they were clear on the procedure, and knew how to use the equipment.

Training phase. Each training phase started with screen instructions that informed demonstrators that they were about to practice a melody. Once they pressed the space bar on a computer keyboard, they were presented with the virtual xylophone. A white circular cue (with a diameter of 5cm or 75 pixels) was projected consecutively on top of each of the four keys that needed to be played to produce the melody. The interval between consecutive cues was 1000ms. Demonstrators memorized the order and position of each key, and then touched the start area on their side of the xylophone with their mallet (see Figure 1). Once they touched the start area, the cues disappeared from the keys, and they attempted to reproduce the melody. After playing the melody, demonstrators were informed on a computer screen whether they had played the melody correctly. This training consisted of six repetitions in order to ensure that demonstrators remembered the melody. During training students faced the wall so that they could not see the xylophone. They also could not hear the melodies the demonstrators were producing because demonstrators heard the tones they were producing through closed headphones.

Individual condition. Demonstrators performed the melody without cues. They first touched the start area and then played the melody learned during the training phase, as accurately as possible from memory. Demonstrators were not given feedback about whether they played the melody correctly. This procedure was repeated twice. Demonstrators heard the tones they produced through closed headphones. As in the training phase, the student could not see or hear the sequence.

Joint condition. Demonstrators remained at the same side of the xylophone they had occupied during the individual condition. They unplugged their headphones and instructed

students to turn around so that they faced the xylophone from the opposite side. The demonstrator played the melody to the student twice. Then the student attempted to play the melody once. The tones produced by the xylophone were now audible to both through loudspeakers. Participants were not given feedback about whether they played the melody correctly.

Data Analysis

The raw data consisted of arrays of three-dimensional coordinates at a sampling rate of 120 frames per second for each performance of a melody. From these arrays we identified single key strokes (four per trial). The first stroke of each melody was excluded. This movement was incomparable to the remaining strokes because the movement originated from the start area that was outside of the array of xylophone keys. Thus, only strokes two, three and four were analyzed from each trial. Because the experiment aimed to find out how individuals modulate kinematic parameters for demonstration our focus was on comparing the 72 strokes of the demonstrator in the individual condition and the 72 strokes of the demonstrator in the joint condition. The 36 strokes from the student were only used to compute error rates.

The kinematic data were standardized so that for every stroke, all of the coordinates and time started at zero. Once transformed, the data were passed through a low pass Butterworth filter in order to reduce noise in the recordings. Incorrect strokes (landing on the wrong key) were excluded from the analysis (see Table 1). Furthermore, for each key to key distance (number of keys crossed by a stroke), values above or below 3 SD of the mean for that distance were treated as outliers and also excluded from the analysis.

	Demonstration		Coordination		Equal Coordination	
Block	Individual	Joint	Individual	Joint	Individual	Joint
1	1.74	1.74	2.77	1.39	5.56	2.74
2	2.78	1.39	2.88	8.33	2.56	4.48
3	5.56	1.92	4.86	4.17	4.86	6.59

Table 1: Percentage errors for demonstrators/leaders, from all three experiments, for joint and individual conditions of every block.

From the remaining data we derived three kinematic parameters: Maximum height, average ascent velocity, and average descent velocity. Maximum height was computed as the maximum value on the height axis for each stroke. Ascent velocity was computed as the average speed of a movement from the previous key to the maximum height on the way to the next key. Accordingly, descent velocity was computed as the average speed of a movement from the maximum height to the next key.

All dependent variables were analyzed with a repeated measure ANOVAs with the factors Block (1, 2, and 3) and Condition (individual and joint). In order to control for the variability in kinematic parameters associated with the key to key distance of different strokes, we added this variable as a covariate for all of our analyses.

Incorrect strokes for all participants were excluded from the analysis (see Table 1). For each key distance, values above or below 3 SD of the mean for that key distance were treated as outliers and excluded from the analysis.

2.2.2 Results

Maximum Height

The results for maximum height are displayed in Figure 2. It was significantly larger in the joint (demonstration) condition than in the individual condition as revealed by a significant main effect of Condition, $F(1,11) = 12.75$, $p = .004$, ηp2 = .54. There was no significant main effect of Block, $F(2,11) = .275$, $p = .76$, ηp2 = .02 and no significant interaction ,$F(2,11) = 2.231$, $p = .13$, ηp2 = .17.

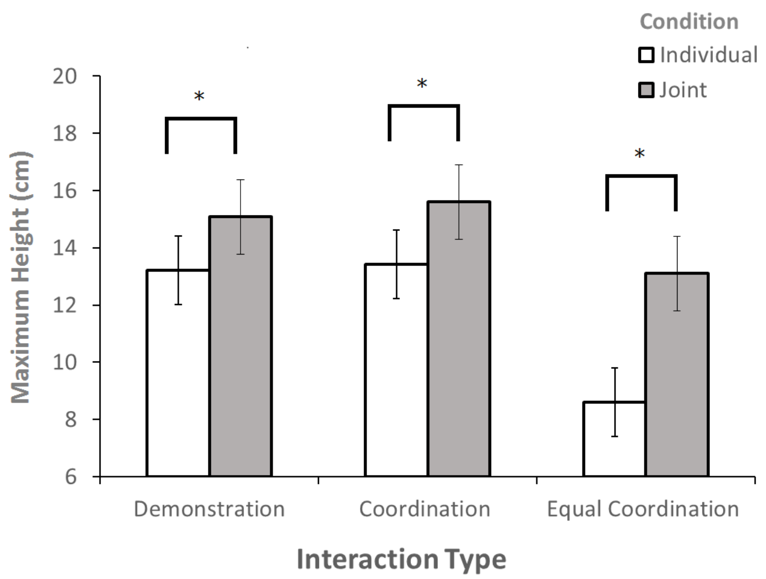

Figure 2: Mean maximum height of demonstrators'/leaders' movements, for Experiment 1, 2 and 3 (from left to right). Black lines indicate significant within group effects. Error bars represent +/- 1 standard error of the mean.

Ascent velocity

Figure 3 displays the results for ascent velocity. There was no significant main effect for Condition, $F(1,11) = 2.322$, $p = .156$, ηp2 = .174, but there was a significant main effect of Block, $F(2,11) = 3.64$, $p = .043$, ηp2 = .249. Ascent velocity in block 3 ($M = 9.61$, $SD = 2.01$) was higher than in block two ($M = 9.17$, $SD = 1.84$), and ascent velocity in block two was also higher than in block one ($M = 8.70$, $SD = 1.36$). There was no significant interaction, $F(2,11) = .15$, $p = .266$, ηp2 = .013.

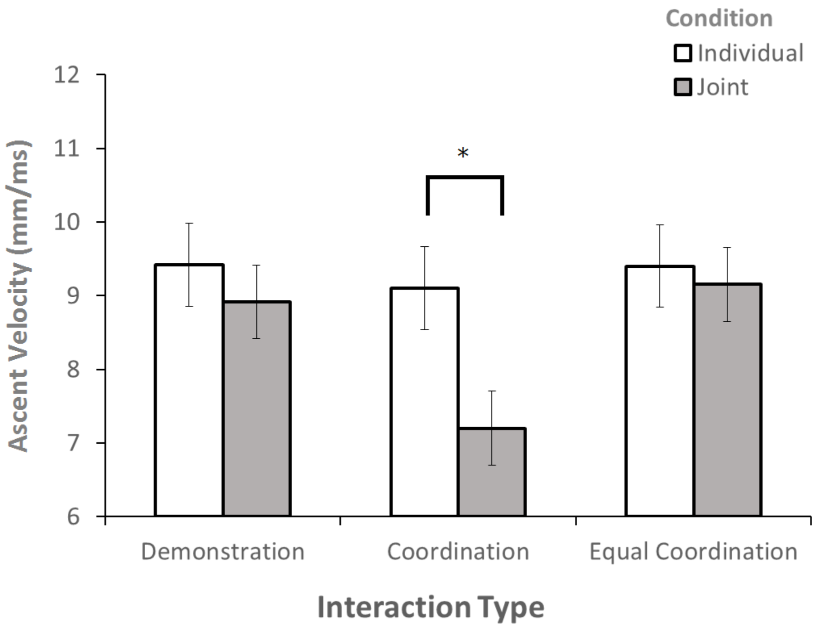

Figure 3: Mean ascent velocity of demonstrators'/leaders' movements, for Experiment 1, 2 and 3 (from left to right). Black lines indicate significant within group effects. Error bars represent +/- 1 standard error of the mean.

Descent velocity

The results for descent velocity are displayed in Figure 4. The ANOVA revealed no significant main effect of Block, $F(2,11) = 1.407$, $p = .266$, ηp2 = .113, and the main effect of Condition fell short of significance, $F(1,11) = 4.436$, $p = .06$, ηp2 = .287. There was no significant interaction between Block and Condition $F(1,11) = 2.153$, $p = .14$, ηp2 = .164.

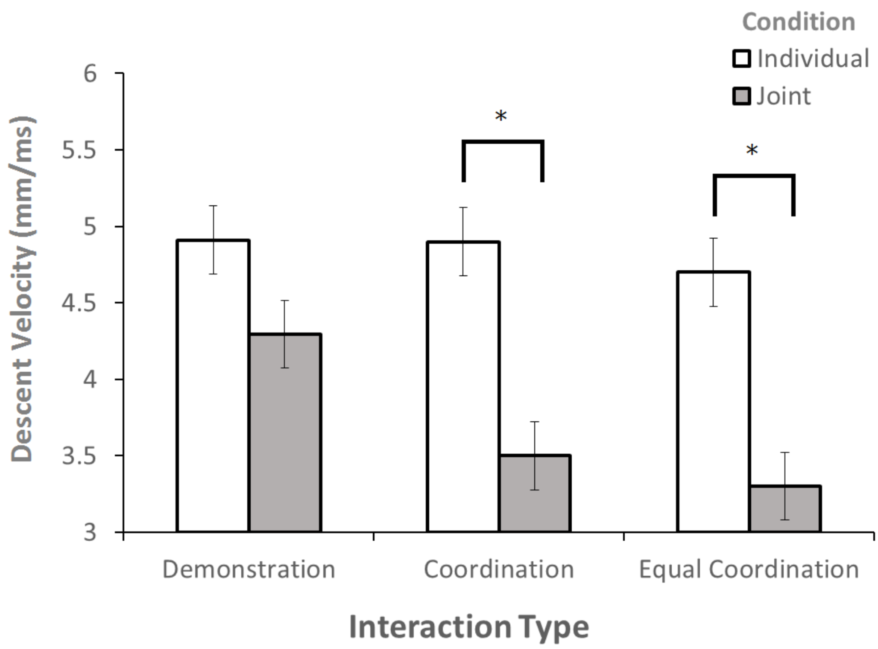

Figure 4: Mean descent velocity of demonstrators'/leaders' movements, for Experiment 1, 2 and 3 (from left to right). Black lines indicate significant within group effects. Error bars represent +/- 1 standard error of the mean.

Error Data

The error data for demonstrators are displayed in the two leftmost columns of Table 1. There were no significant main effects of Block, $F(2,11) = .6$, $p = .75$, $\eta p2 = .05$, or Condition, $F(1,11) = 2.02$, $p = .18$, $\eta p2 = .16$, and there was no significant interaction, $F(2,11) = 1.6$, $p = .25$, $\eta p2 = .13$. The error data for students is displayed in the left column of Table 2. There was also no significant main effect of Block for the students' errors, $F(2,11) = .5$, $p = .62$, $\eta p2 = .04$.

Block	Demonstration	Coordination	Equal Coordination
1	11.43	12.5	10.1
2	9.42	14.28	4.49
3	8.31	4.51	7.64

Table 2: Percentage errors for students/followers, from all three experiments.

2.2.3 Discussion

The results of Experiment 1 show that demonstrating a melody to a student systematically affected a demonstrator's kinematics. Compared to an individual baseline participants moved their mallet substantially higher when an uninformed partner was instructed to repeat a melody following their demonstration. This could reflect an attempt to help the student encode a particular action/tone sequence (Brand et al., 2002). The within participant comparisons for ascent velocity and descent velocity fell short of significance, despite showing large effect sizes. This could indicate that our study was underpowered for within-subject comparisons, and suggests that the results on velocity need to be interpreted with caution. However, we chose our sample size based on calculations from a mixed ANOVA (conducted in Experiment 2 and 3), because the between subject comparisons of Demonstration and Unequal Coordination, and Unequal Coordination and Equal Coordination are the most crucial comparisons for this study. Future studies with larger sample sizes may serve to determine whether velocity modulations reliably occur in the context of demonstrating action sequences.

2.3 Experiment 2: Coordination

This experiment investigated kinematic modulations in a joint action context where temporal coordination between two partners with unequal knowledge is required. In particular, we investigated whether participants modulate the same or different kinematic parameters as in demonstration when they are required to communicate information to an uninformed partner to facilitate spatial and temporal predictions about the next step in a sequence (tone in a melody). Based on earlier research using a different coordination task (Vesper & Richardson, 2014), we predicted that movement height would be increased. We also predicted that systematic modulations of velocity might play a greater role in the coordination context due to the real-time nature of the predictions involved (Sacheli et al., 2013).

2.3.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 26 participants (10 males, 15 females), with a mean age of 25.7 (SD = 3.35). Participants were required not to have received any musical training, and to be proficient English speakers. The leader in one dyad had a very high error rate in both the individual and joint condition (> 3SD from mean) and this dyad was therefore excluded. A g*power analysis based on a large effect size ($\eta p2 = .14$) determined that we needed 12 participants for sufficient power for the comparison between this experiment and Experiment 1.

Apparatus and Stimuli

These were the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 1 with the following exceptions. Participants were randomly assigned to 'leader' (knowing the melody to be played together) and 'follower' (not knowing the melody played together). Similarly to the demonstrators in Experiment 1, leaders performed the training phase and individual condition while the 'follower' stood facing the wall and could not hear their partner playing. In the joint condition leader and follower twice played the same melody (only known by the leader) together. Leader and follower were instructed to play the correct keys as synchronously as possible. Both participants' mallets triggered tones corresponding to the key touched so that they received feedback about their asynchrony in playing the same tone.

Data analysis

For the leaders, we used the same design and computed the same dependent variables as in Experiment 1. To assess differences between kinematic cues for demonstration and joint action coordination we directly compared demonstrators' and leaders' performance in Experiment 1 and 2 using a 3x2x2 mixed ANOVA with the within factors Block and Condition and the between factor Experiment (demonstration, coordination).

To assess coordination between leader and follower, we computed mean absolute asynchrony between leaders and followers for the joint condition. In order to determine whether coordination between leaders and follower during joint action was better than chance, we generated surrogate dyads, randomly pairing leaders and followers from different dyads. This resulted in 12 surrogate pairs. The mean values of the surrogate pairs were used as baseline for the asynchrony observed in the joint condition.

2.3.2 Results

Maximum Height

There was a significant effect of Condition, $F(1,11) = 6.06$, $p = .032$, ηp2 = .355, with maximum height in the joint condition being significantly larger than in the individual condition (see middle of Figure 2). There was no significant main effect of Block, $F(2,11) = .232$, $p = .79$, ηp2 = .021 and there was no significant interaction between Block and Condition, $F(2,11) = .124$, $p = .884$, ηp2 = .011. The between experiment comparison revealed a main effect of Condition, $F(1,22) = 15.45$, $p = .001$, ηp2 = .412, but no significant effect of the factor Experiment $F(1,22) = .049$, $p = .83$, ηp2 = .002. There was also no interaction between Condition and Experiment, $F(2,22) = .1$, $p = .76$, ηp2 = .005, or Block and Experiment, $F(2,22) = .021$, $p = .979$, ηp2 = .021 and no three way interaction between Condition, Block and Experiment, $F(2,22) = .838$, $p = .439$, ηp2 = .037.

Ascent velocity

There was a significant main effect of Condition, $F(1,11) = 29.89$, $p < .001$, ηp2 = .732. Participants were significantly slower in the joint condition, compared to the individual condition (see middle of Figure 3). The main effect of Block was not significant, $F(2,11) = .964$, $p = .397$, ηp2 = .081 nor was the interaction between Block and Condition, $F(1,11) = 1.06$, $p = .364$, ηp2 = .088. The between experiment comparison showed no significant main effect of Experiment $F(1,22) = 2.31$, $p = .14$, ηp2 = .1. However, there was a significant main effect of condition, $F(1,22) = 24.35$, $p < .001$, ηp2 = .525, and a significant interaction between Condition and Experiment, $F(1, 22) = 7.65$, $p = .01$, ηp2 = .25. Ascent velocity between the individual and joint condition differed only in Experiment 2, where leaders were required to synchronize their actions with those of their followers (see Figure 3). There was no interaction between Block and Experiment, $F(2,22) = .549$, $p = .582$, ηp2 = .024, and no

three way interaction between Condition, Block and Experiment, $F(1,22) = .353$, $p = .71$, ηp2 = .016.

Descent velocity

There was a significant main effect of condition, $F(1,11) = 57.76$, $p < .001$, ηp2 = .84, with leaders being slower in the joint condition, compared to the individual condition (see Figure 4). There was no significant main effect of block, $F(2,11) = .97$, $p = .39$. ηp2 = .08, and no interaction between block and condition, $F(2,11) = .31$, $p = .761$, ηp2 = .03. In the comparison of Experiments there was a significant main effect of condition, $F(1,22) = 34.19$, $p < .001$, ηp2 = .598. Moreover there was an interaction between Condition and Experiment, $F(1, 22) = 5.24$, $p = .032$, ηp2 = .19, with descent velocity being slower only in the joint condition of Experiment 2 where leaders synchronized their actions with followers than in the individual condition (Figure 4). There was no significant main effect of Experiment, $F(1,22) = 2.35$, $p = .14$, ηp2 = .1, no interaction between Block and Experiment, $F(2,22) = .89$, $p = .42$, ηp2 = .04 and no three way interaction between Condition, Block and Experiment $F(2,22) = 2.29$, $p = .11$, ηp2 = .09.

Error data

The error data for leaders are displayed in the two middle columns of Table 1. There was a significant effect of Condition, $F(1,11) = 6.053$, $p = .03$, ηp2 = .34, with significantly more errors in the joint condition than in the individual condition. There was no significant main effect of Block, $F(2,11) = .87$, $p = .43$, ηp2 = .07, and no significant interaction between Block and Condition, $F(2,11) = .63$, $p = .54$, ηp2 = .07. The between experiment comparison revealed no significant effect of the factor Experiment, $F(1,22) = 3.76$, $p = .07$, ηp2 = .13, and no main effect of Condition, $F(1,22) = 1.04$, $p = .32$, ηp2 = .04, but an interaction

between Condition and Experiment, $F(1,22) = 7.718$, $p = .011$, ηp2 = .25. In Experiment 2, leaders made significantly more errors in the joint condition compared to the individual condition. There was no significant interaction between Block and Condition, $F(2,22) = 1.53$, $p = .23$, ηp2 = .06, or between Block and Interaction, F,(2,22) = 1,45, p = ,24, ηp2 = .06. There was also no three way interaction between Condition, Experiment, and Block, $F(2,22) = .28$, $p = .74$, ηp2 = .01.

The followers' error data is displayed in the middle column of Table 2. For the followers' errors there was a significant main effect of Block, $F(2,11) = 4.417$, $p = .024$, ηp2 = .29, with followers' errors being significantly lower in Block 3 ($M = 4.5$, $SD = 7.2$), than in Blocks 1 ($M = 12.5$, $SD = 11.1$) and 2 ($M = 14.2$, $SD = 11.1$), as revealed by pairwise comparisons, $p = .035$ and $p = .042$ respectively, bonferroni corrected). The between experiment comparison (errors of students in Experiment 1 vs. errors of followers in Experiment 2) revealed a significant effect of Block, $F(2,22) = 3.61$, $p = .035$, ηp2 = .14. There was no significant effect of Experiment $F(1,22) = .059$, $p = .810$, ηp2 = .003, or no interaction between Block and Experiment, $F(2,22) = 1.68$, $p = .2$, ηp2 = .07.

Asynchrony

In order to create a baseline to compare to the performance of our dyads, we generated surrogate dyads by randomly pairing leaders and followers across different dyads, resulting in 12 surrogate pairs. We iterated this process 10 times (resulting in 120 surrogate pairs), and then used the mean values of the surrogate pairs as our baseline.[2] For both real and surrogate dyads, we set the time of every key hit to zero, by subtracting the onset time from the overall time for one hit. This made the duration for each key stroke more comparable, allowing us to compute asynchrony even when participants had different onset times.

For the asynchronies we carried out a 3 x 2 ANOVA with block (one, two three) and Dyad Type (real, surrogate) as within subjects factors. The ANOVA revealed a main effect of Dyad Type (real, surrogate), $F(1,11) = 198.81$, $p < .001$, $\eta p2 = .95$. Asynchrony was significantly lower in real dyads compared to the surrogate dyads (see Figure 5). There was also a main effect of Block, $F(2,11) = 9.74$, $p = .001$ $\eta p2 = .47$, and a significant interaction between Block and Dyad Type, $F(2,11) = 9.4$, $p = .001$, $\eta p2 = .46$ (see Figure 6). Only real pairs showed an initial drop in asynchrony from Block 1 ($M = .27$, $SD = .14$) to Block 2 ($M = .13$, $SD = .06$). However, a pairwise comparison fell short of reaching significance, $p = .05$ (bonferroni corrected). Surprisingly, surrogate pairs showed a drop in asynchrony from Block 2 (M = .59, SD = .12) to Block 3 (M = .43, SD = .07).

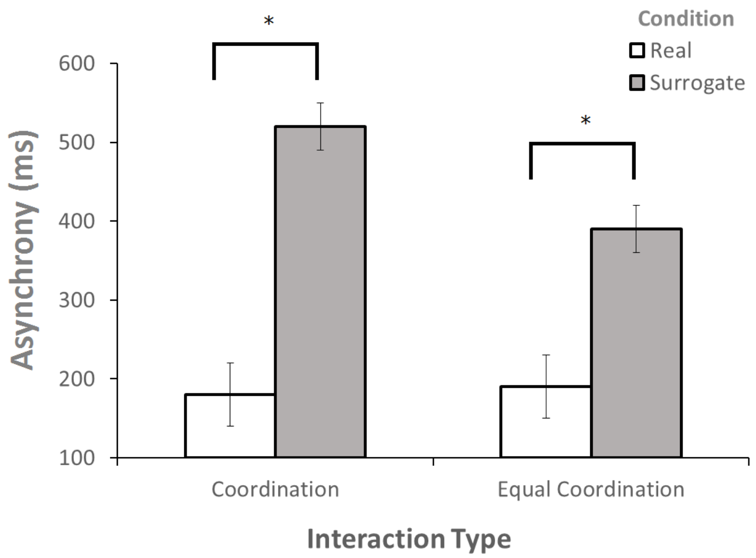

Figure 5: Asynchronies for real and surrogate pairs for Experiment 2 and 3 (from left to right). Black lines indicate significant within group effects. Error bars represent +/- 1 standard error of the mean.

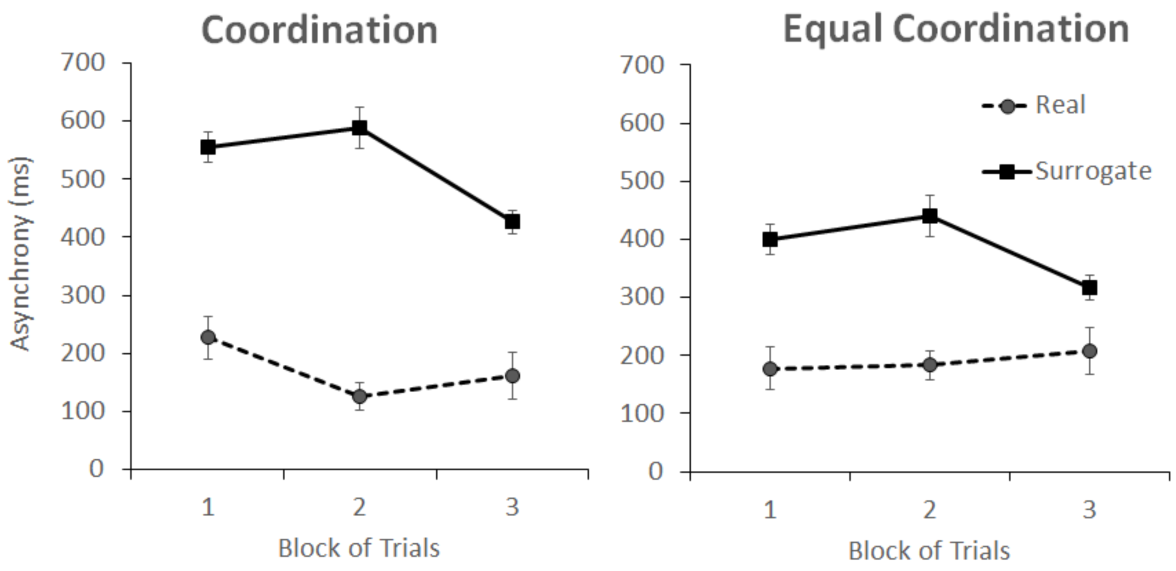

Figure 6: Asynchronies for real and surrogate pairs across block one, two and three (from left to right) for Experiment 2 and 3. Error bars represent +/- 1 standard error of the mean.

2.3.3 Discussion

Participants moved the mallet higher when their task was to synchronize with an uninformed follower than when they performed action sequences on their own. The fact that similar increases in movement height were observed in Experiment 1 and 2 indicates that movement height cannot only serve as a kinematic marker of demonstration but that it can also serve to facilitate joint action coordination. Thus, exaggerating movement and systematically deviating from the most efficient trajectory may highlight task relevant knowledge in different task contexts. Another possibility is that the same kinematic marker may have different functions in the context of demonstration and joint action coordination. Exaggeration of movement height in Experiment 2 may have resulted from leaders trying to support a follower's online prediction of the time and location of their ongoing actions rather than from leaders trying to support learning of a goal directed action sequence.

Both ascent and descent velocity showed context-specific modulations. Leaders in a joint action moved more slowly, both during the ascent and descent phase, when synchronizing with followers who did not know the melodies to be performed jointly, compared to demonstrators who performed the melodies for a student watching their actions. This indicates that slowing down is used as a means to support unknowing followers who need to predict the location and timing of the next step in an action sequence. Although we cannot completely rule out the role of ascent and descent velocity for demonstration (due to the possible lack of statistical power for this comparison) there is clear evidence for the role of these kinematic parameters in joint action coordination.

Overall, the increase in movement height and the slowing down in Experiment 2 are in line with earlier results on action modulations in the context of joint action coordination (Sacheli et al. 2013; Vesper & Richardson, 2014). They demonstrate that modulations of instrumental action that were observed for the coordination of discrete actions occur also when two people need to coordinate a whole sequence of instrumental actions of the same kind.

The higher number of errors in the joint condition compared to the individual condition could reflect the increased task difficulty, due to the leader having to coordinate with the follower. The relatively high error rates of the followers may have also led leaders to make more errors. Followers made significantly fewer errors in Block 3 than in Blocks 2 and 1, suggesting that they learned to remember the melody while performing joint actions with the leader.

Asynchronies were lower in real pairs than in surrogate pairs showing that leaders and followers successfully synchronized their actions, improving from the initial part of the experiment. The drop of asynchrony in surrogate pairs towards the end of the experiment is

likely due to a decrease in temporal variability during individual performance. Indeed, we found that the variability of participants' movement duration was numerically lower in Block 3, for both leaders and followers, which could explain the decrease in asynchrony for surrogate pairs in Block 3.

2.4 Experiment 3: Coordinating with equal knowledge

The previous two experiments leave open two questions. First, it is not clear whether exaggerating movement height serves to overcome knowledge asymmetries regardless of the particular context or whether it can serve different functions in different task contexts, such as also supporting prediction during online coordination. To address this issue, Experiment 3 investigated performance in a joint action context where both leader and follower knew the melody to be played together (equal coordination). If the exaggeration of movement height persists, then this would indicate that it is not a specific marker for knowledge transfer.

Second, it is not clear whether the modulation of velocity parameters in Experiment 2 reflects an attempt to provide information that supports prediction of the spatial target of an action or of the timing of the action or both. The reason is that in Experiment 2 followers needed to both predict which key the leader intended to hit and when they were going to hit it. In the equal coordination of Experiment 3 leaders needed to only support the followers' temporal predictions because the followers also knew the melodies (and the leaders knew this). If reductions in velocity during ascent and descent mainly serve to support temporal predictions then they should also be observed in Experiment 3.

2.4.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 26 participants (12 males, 14 females), with a mean age of 27.21 (SD = 5.18). Participants were required not to have received any musical training, and to be proficient English speakers. One dyad had to be excluded from the analysis due to equipment malfunction. A g*power analysis based on a large effect size ($\eta p2 = .14$) determined that 12 participants would give us sufficient power for the comparison between this experiment and Experiment 2.

Apparatus and Stimuli

These were the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 2 with the following exceptions. Both leader and follower completed the training phase and the individual condition (followers first and leaders second). The order of follower/leader was fixed to keep the procedure of the leader (from whose performance the main kinematic parameters are derived) as close as possible to Experiment 2 where the joint condition immediately followed the leader's training and individual performance. Again, we ensured that participants could not hear or see the other participant practice (using headphones and making the inactive participant face the wall).

Data analysis

This was the same as in Experiment 2. To assess differences between kinematic cues for a joint action where the follower knew (equal coordination, Experiment 3) or did not know the melody (coordination, Experiment 2) we directly compared leaders' performance

and asynchronies between the two experiments using a 3x2x2 mixed ANOVA with the within factor Block and Condition and the between factor Experiment (coordination, equal coordination).

2.4.2 Results

Maximum Height

The ANOVA revealed a significant main effect of Condition, $F(1,11) = 43.86$, $p < .001$, ηp2 = .799. Maximum height was significantly larger in the joint condition than in the individual condition (see Figure 2). There was no main effect of Block, $F(2,11) = 2.12$, $p = .144$, ηp2 = .16, or no interaction between Block and Condition, $F(2,11) = .375$, $p = .692$, ηp2 = .033. The between experiment comparison showed a significant effect of Condition, $F(1,22) = 36.59$, $p < .001$, ηp2 = .623 and a significant effect of Experiment, $F(1,22) = 5.13$, $p = .034$, ηp2 = .19. Moreover there was an interaction between Condition and Experiment, $F(1,22) = 4.682$, $p = .042$, ηp2 = .18. Maximum height was significantly lower in the Equal coordination interaction compared to the Unequal coordination interaction, but only for the individual trials. There was no interaction Between Block and Experiment, $F(2,22) = 1.371$, $p = .264$, ηp2 = .06, or no three way interaction between Condition, Block and Experiment, $F(2,22) = .46$, $p = .64$, ηp2 = .02.

Ascent velocity

There was no significant main effect of Condition, $F(1,11) = .26$, $p = .62$, ηp2 = .02 or Block, $F(2,11) = .71$ $p = .48$, ηp2 = .07. Also, we did not find an interaction between Condition and Block, $F(2,11) = .13$, $p =$,88, ηp2 = .01. Accordingly, the between experiment comparison revealed a significant main effect of Condition, $F(1,22) = 15.27$, $p = .001$, ηp2 = .41, and an interaction between Condition and Experiment, $F(1,22) = 9.75$, $p = .005$, ηp2 =

.31. Ascent velocity was only lower in the joint condition compared to the individual condition in Experiment 2 where followers did not know the melody (see Figure 3). There was no significant main effect of Experiment, $F(1,22) = 2.54$, $p = .13$. ηp2 = .1, and no interaction between Experiment and Block, $F(2,22) = .48$, $p = .62$, ηp2 = .2. There was no three way interaction between Condition, Block and Experiment, $F(2,22) = .28$, $p = .76$, ηp2 = .013.

Descent velocity

There was a significant main effect of condition, $F(1,11) = 60.44$, $p < .001$, ηp2 = .85, with descent velocity being significantly lower in the joint condition, compared to the individual condition (see Figure 4). In the between experiment comparison there was a significant main effect of Condition, $F(1,22) = 118.18$, $p < .001$, ηp2 = .84, but no significant main effect of Experiment, $F(1,22) = .83$, $p = .37$, ηp2 = .04. There was no significant interaction between Condition and Experiment, $F(1,22) = .009$, $p = .92$, ηp2 = .00, or no interaction between Block and Experiment, $F(1,22) = .62$, $p =.55$, ηp2 = .03. There was no three way interaction between Condition, Block and Experiment, $F(2,22) = 1.3$, $p = .28$, ηp2 = .06.

Error data

The two right columns of Table 1 show error rates for leaders in Experiment 3. There was no significant main effect of Condition, $F(1,11) = .37$, $p = .56$, ηp2 = .03, or Block, $F(2,11) = .26$, $p = .77$, ηp2 = .02. There was no interaction between Condition and Block, $F(2,11) = .31$, $p = .73$, ηp2 = .03. The between experiment comparison revealed a main effect of Condition, $F(1,22) = 5.31$, $p = .03$, ηp2 = .18, but no main effect of Experiment, $F(1,22) = .04$, $p =$,85, ηp2 = .002, and no significant interaction between Condition and Experiment,

$F(1,22) = 2.51$, $p = .17$, ηp2 = .1, or Block and Experiment, $F(2,22) = .77$, $p = .47$, ηp2 = .03. There was also no three way interaction between Condition, Block and Experiment, $F(2,22) = .89$, $p = .42$, ηp2 = .04 . The right column of Table 1 shows error rates for followers in Experiment 3. There was no significant effect of Block for the followers' errors, $F(2,11) = 1.41$, $p = .27$, ηp2 = .11. There was also no main effect of the between subject comparison of Experiment, $F(1,22) = .811$, $p = .38$, ηp2 = .04 . However, there was an interaction between Block and Experiment, $F(1,22) = 3.58$, $p = .036$, ηp2 = .14. Errors in Block 3 were significantly lower than errors for Block 1 and 2, but only in Experiment 2 (unequal knowledge coordination).

Asynchrony

There was a significant main effect of Dyad Type, $F(1,11) = 33.55$, $p < .001$, ηp2 = .42. Asynchrony was significantly lower for the real dyads, compared to the surrogate dyads (Figure 5). There was also a significant interaction between Block and Dyad Type, $F(2,11) = 3.91$, $p = .035$, ηp2 = .26 with asynchrony for surrogate pairs significantly larger in Block 2 ($M = .441$, $SD = .12$) than in Block 3 ($M = .31$, $SD = .07$). The between experiment comparison revealed a significant main effect of Dyad Type, $F(1,22) = 169.95$, $p < .001$, ηp2 = .86. There was also a significant main effect of Experiment $F(1,22) = 6.1$, $p = .022$, ηp2 = .22, with asynchrony lower in the equal knowledge coordination interaction ($M = .29$, $SD = .11$) than in the unequal knowledge coordination interaction ($M = .35$, $SD = .15$). The analysis also revealed an interaction between Dyad Type and Experiment, $F(1,22) = 13.19$, $p = .001$, ηp2 = .39 (see Figure 5). The difference in asynchrony between real and surrogate pairs was smaller in Experiment 3 than in Experiment 2. The analysis also revealed a main effect of Block, $F(2,22) = 6.45$, $p = .004$, ηp2 = .23 and a significant interaction between Block and Dyad Type, $F(1,22) = 11.56$, $p < .001$, ηp2 = .55. In both Experiment 2 and Experiment 3,

asynchrony was lower in Block 3, compared to Blocks 2 and 1, but only for surrogate pairs. There was no significant three-way interaction between Dyad Type, Block and Experiment $F(2,22) = .83$, $p = .44$, $\eta p2 = .04$.

2.4.3 Discussion

The results of Experiment 3 showed that leaders exaggerated movement height even when the followers they were synchronizing with knew the melodies to be played together. This indicates that exaggerated movement height is not always a marker of knowledge transfer. Rather, it can serve different purposes in different task contexts including temporal prediction in the present experiment.

The analysis of velocity modulations revealed an important difference between the ascent and descent phase of movements directed at a particular location. Slow down during the early part of the movement while raising the mallet occurred only in Experiment 2 but not in Experiment 3. Thus the slowing of early parts of the movement seems to serve spatial prediction in the context of joint action coordination (Vesper & Richardson, 2014). Slow down during the later part of the movement, bringing the mallet down to the key, was equally pronounced in Experiment 2 and 3 indicating that it serves to facilitate temporal coordination during joint action.

Again, we found that asynchronies were lower in real pairs than in surrogate pairs, showing that leaders and followers successfully synchronized their actions. Surprisingly, there was no difference between the size of the asynchronies in Experiments 2 and 3 for the real pairs, implying that participants performed the joint task equally well, regardless of whether or not the follower knew the melody. This is interesting because the fact that participants can perform the task just as well when only one participant knows the melody

suggests that leaders can successfully overcome asymmetries in knowledge using cues produced from their movements. As in Experiment 2, asynchronies for surrogate pairs were lower in Block 3, compared to Block 1 and 2. Like in Experiment 2, a small numerical decrease in variability for followers' movement duration may explain this finding.

2.5 General Discussion

Instrumental actions are sometimes not just instrumental actions. In the context of teaching and joint action, they tend to be modulated in order to communicate relevant information to a learner or to a joint action partner. But how sensitive is our motor system to the needs of communication and coordination? Are kinematic modulations unspecific and general in nature, so that across different social interactions the same raising of an arm or the slowing down of a hand serve different informative functions? Or are there distinct kinematic markers of demonstration and joint action coordination, uniquely facilitating the acquisition of task-relevant knowledge and the spatial and temporal predictions required for interpersonal coordination? To answer these questions, across three experiments we investigated sensorimotor communication in interactive contexts which required participants to modulate their kinematics in order to either inform a student of the structure of an action sequence (Experiment 1), to provide spatial and temporal information to a coordination partner (Experiment 2), or to provide only temporal information to a coordination partner (Experiment 3).

There were two key findings, reflected in modulations of movement height and velocity. Firstly, we found that across all three interaction contexts participants exaggerated the spatial trajectory of their movements by increasing the maximum height, compared to when they performed the same action sequence individually. This suggests that increases in movement

height play a role both in communicating knowledge of the structure of an action, and in communicating information to facilitate spatial and temporal prediction.

The finding that teachers increased movement height when demonstrating an action sequence to a student is in line with the motionese literature (Brand et al., 2002). The increase in maximum height could have served to draw students' attention and to allow them to more effectively parse the observed action sequence, in order to understand the structure of the sequence in terms of sub-goals. The fact that similar increases in height were observed across teaching and joint action coordination implies that deviations from the most efficient movement path in terms of height can serve different purposes, and speaks against the possibility that height (or more generally, a less efficient way of acting) serves as a distinct marker of pedagogical intentions (Csibra & Gergely, 2013).

The observed increase in maximum height in the coordination interactions is in line with findings by Vesper and Richardson (2014) who also found that participants modulated their movement height in order to facilitate coordination with an uninformed partner. The increase in height may serve to create a steeper slope when descending, which can facilitate predictions of movement direction and help disambiguate the target location of the movement. However, one critical extension to earlier studies is that we found increases in movement height even in the absence of the need to transfer spatial information. Experiment 3 showed that leaders increased movement height when coordinating with a partner who was informed about the sequence of movement targets. This demonstrates that maximum height serves temporal prediction in the context of joint action.

But how can increasing movement height support temporal predictions? It has been suggested that people predict others' actions based on simulations of motor commands and associated action outcomes (internal models; Wolpert, 2003), and that sensorimotor

communication enhances a co-actor's selection of the most appropriate internal model in order to coordinate with their partner (Vesper et al. 2014; Wolpert, 2003). This may apply not only to the spatial prediction of an action's trajectory. Selecting the correct internal model earlier may allow for more effective prediction of the timing of a co-actor's action. Potentially, exaggerating the spatial trajectory of an action increases not only the spatial accuracy of predictions by allowing one to more clearly distinguish between action alternatives and predict the correct trajectory, but also increases the temporal accuracy of predictions by allowing one to anticipate earlier when a co-actor will reach the target location of the action. More generally, this suggests that exaggerating the spatial trajectory of one's action to enhance a coordination partner's internal model selection plays a role in both spatial and temporal aspects of interpersonal coordination.

Another possibility is that increasing the movement height served to keep the co-actor's attention, more generally. As well as drawing attention to the structure of an action in a pedagogical context, keeping a co-actor attentive and engaged is also important in joint action so that the co-actor can predict and adapt to the actor's movements effectively. Thus, exaggerating the spatial trajectory of one's actions could be a kinematic modulation that spans across a wide range of social actions, with the purpose of maintaining a co-actor's attention (Sartori, Becchio, Bara, & Castiello, 2009). This explanation is not incompatible with the idea that sensorimotor communication can enhance the selection of internal models; we believe that sensorimotor communication may serve both these purposes.

The second key finding was that velocity during the ascent and descent phase of the performed movements was differentially modulated depending on the interaction context and thus provides a distinct marker of the interactive context. Although we cannot draw any firm conclusions as to whether or not participants modulated the speed of their movements during

demonstration, we provide clear evidence that participants modulated their ascent and descent velocity differently depending on the knowledge of their co-actor.

Decreases in descent velocity were observed whenever participants had to achieve temporal coordination, regardless of whether their partner was informed or uninformed about the action sequence to be performed. This extends earlier findings from tasks where one individual was always uninformed (Sacheli et al., 2013; Vesper & Richardson, 2014), and suggests that slowing down movements before approaching a target serves to facilitate temporal prediction. Slowing down could have enhanced temporal prediction simply by providing the follower with more time to predict when the leader would hit the key. Also, moving slower towards the key provided the follower more time to achieve synchrony with the leader, ensuring that they were coordinated when hitting the key.

Participants decreased the velocity of the ascent phase of their movement specifically when coordinating their actions with an uninformed partner. This indicates that reducing ascent velocity may specifically serve to facilitate spatial predictions. Together with an exaggerated spatial trajectory, slowing down the ascent velocity may have provided the follower with more information and provided them with more time to predict the trajectory of the leader's movement. This was not necessary in the equal knowledge coordination interaction where followers never needed to make predictions about the final destination of the leader's movements. The fact that coordination was equally successful regardless of whether only one or both participants knew the action sequence suggests that the cues provided in terms of increased movement height and slower ascent and descent velocity were effective.

The finding that people differentially modulate their ascent and descent velocity depending on their co-actor's knowledge can provide us with new insights into joint action

planning. When engaging in joint actions, people form representations of a co-actor's task (Sebanz, Knoblich & Prinz, 2005, and of the properties of objects constraining a co-actor’s actions (Schmitz, Vesper, Sebanz & Knoblich, in press), which can help them adjust their own actions in the service of coordination. We believe that our findings provide a critical extension to this research. The fact that people modulate the kinematics of their actions differently depending on whether or not their partner is informed is evidence that actors represent their co-actor’s knowledge state, and adjust their action plans accordingly. At the planning stage of their action, people use a representation of the other’s knowledge in order to determine whether their movements need to be predictable spatially or temporally (or both) and execute their actions accordingly. Indeed, earlier studies already demonstrate that for sensorimotor communication, people adjust their action plans depending on the perceptual access of a co-actor (Sartori et al., 2009; Vesper & Richardson, 2014; Vesper et al. 2016). We take this one step further by showing that people also consider the epistemic state of their co-actor in the absence of any perceptual cues reflecting the other’s epistemic state.

As well as informing us about how people achieve joint action coordination, the present research may also allow us to better understand how one can learn from participating in coordinated joint actions (Rogoff et al., 2003). The fact that there is overlap between sensorimotor communication in demonstration and joint action coordination points to the fact that kinematic cues produced in interpersonal coordination contexts may also provide effective learning cues. One possibility is that the cues produced when coordinating can also enhance learning by elucidating the structure of an action. Indeed, research has demonstrated that people segment action sequences into meaningful segments based on low level movement features of the actions (Zacks, Kumar, Abrams & Mehta, 2009). Moreover, Nagai and Rohlfing (2009) showed that motionese provides low level perceptual cues which can enhance learning by guiding an observer’s attention to learning relevant information. Given

that people segment action sequences, and that motionese guides low level attention, kinematic cues produced in joint action coordination could serve as learning relevant cues, by modulating a co-actor's attention in such a way that he or she can effectively extract meaningful segments and sub-goals of an action.

A limitation of the current study is that it did not address the performance of the student/follower in the interaction in any detail. Because our design focused on action modulations of the demonstrators/leaders we could not examine the effects of these kinematic modulations on the follower in a controlled and systematic way. Furthermore, because our task used simple key strokes from one position to another position, we ended up with simple action sequences that had very homogeneous component actions. Specifically, the structure of a key stroke is the same, regardless of the start and end position. Indeed, there is evidence that people can use low level movement features in order to parse complex and hierarchical action sequences into meaningful segments (Zacks et al. 2009; Buchsbaum et al. 2015). Given this, whether kinematic cues as observed in the present study also support parsing of more complex, less homogeneous action sequences is an interesting question left open by the present research. Future studies could address these issues, by investigating how kinematic cues affect people's performance on a variety of tasks. This could be done by using a task in which participants are shown a pre-recorded action sequence with artificially modulated kinematics. This would give experimenters control over the kinematic cues produced, allowing them to systematically investigate their effects on participants' performance in learning and coordination tasks.

In sum, people flexibly modulate their kinematics for sensorimotor communication given different task constraints, and given the epistemic state of their partner. Sensorimotor communication plays a role in both communicating the structure of an action, and facilitating

spatial and temporal prediction. Increases in movement height occurred across different interaction contexts, indicating that increases in height and deviations from the most efficient movement path are not solely a mark of teaching. However, demonstration, joint action coordination with an informed partner, and joint action coordination with an uninformed partner were associated with distinct velocity profiles. These findings demonstrate that people adjust their actions in fine-grained ways depending on a co-actor's knowledge state and the need for spatial and temporal prediction. Wider implications of our findings are: i) they indicate the possibility that we incorporate higher level representations of the knowledge state of a co-actor into our joint action plans; ii) kinematic cues produced in a joint action context can also convey learning relevant information. Future research is needed to investigate the effects of these action modulations on individual learning and on joint action performance in more detail.

Chapter 3. Identifying Informative Intentions from Movement Kinematics

3.1 Introduction

People derive mental states such as intentions and expectations from observing the movements of others (Cavallo et al. 2016; Grezes, Frith & Passingham, 2004). Using early movement kinematics of perceived actions, observers can discriminate between different instrumental intentions (Cavallo et al. 2016; Manera et al. 2011). In addition, informative intentions can also be reflected in kinematics. On the one hand, people acting together produce informative action modulations in order to support interpersonal coordination by facilitating spatial and temporal prediction (Pezzulo, Donnarumma & Dindo, 2013; Vesper & Richardson, 2014; Vesper et al., 2016). On the other hand, parents and teachers modify their movements to support learning through demonstration by highlighting the structure of an action (Brand, Baldwin & Ashburn, 2002). These findings suggest that the same action can be modulated in different ways to convey different informative intentions to an observer.

But can observers actually identify informative intentions based on movement kinematics? The first aim of the present study was to investigate whether people can discriminate actions with informative intentions from actions without informative intentions using kinematic cues. The second aim was to investigate whether people are able to distinguish different interactive intentions based on kinematic cues. Specifically, we asked whether observers can tell whether perceived agents are intending to teach a co-actor or whether they intend to perform a coordinated joint action with a co-actor.

3.1.1 Perceiving intentions from actions

Much of the research on perception of individuals' intentions has focused on perception of instrumental actions. This research has demonstrated that humans have the ability to derive

different mental states of an actor by observing the kinematics of their actions. For instance, people can recognize whether an actor intends to cooperate or compete (Manera et al. 2011), whether or not an actor has a false belief (Grezes, Frith & Passingham, 2004) or even whether or not an actor has a deceptive intention (Runeson & Frykholm, 1983). Even though these actions are not intended to inform, people can still read mental states from them.

A recent study by Cavallo and colleagues (2016) demonstrated that people can discriminate observed actors' instrumental intentions based on early kinematic features of the action. In their study participants observed reach to grasp movements of actors intending to grasp a bottle in order to pour from it, or in order to drink. They found that kinematic features such as wrist height and grip aperture predicted how well an observer could discriminate between the two different underlying intentions. Moreover, the accuracy of participants' discrimination between the two underlying intentions could be modulated by modifying kinematic parameters that predicted classification accuracy. In contrast to many earlier studies, Cavallo et al. (2016) were not only able to show that different intentions can be discriminated, but they could also quantify the contribution of different kinematic parameters to the accuracy of identifying a particular intention.

There is also evidence that movement kinematics carry information about social intentions. Becchio, Sartori, Bulgheroni and Castiello (2008) carried out a study in which participants were required to grasp an object to build a tower together with a co-actor, either with a cooperative intention (build the tower together) or a competitive intention (place the object at the bottom of the tower before the other participant). They showed that compared to competitive actions, cooperative actions had a larger trajectory, were slower, and displayed a smaller grip aperture. Another study by Manera et al. (2011) demonstrated that people could discriminate between cooperative and competitive intentions when perceiving reach to grasp

movements. Moreover, participants could still discriminate between competitive and cooperative intentions when viewing point light displays of reach to grasp movements, demonstrating that dynamic kinematic cues were used to discriminate between different intentions.

Evidence obtained in sports experts indicates that identifying intentions from action kinematics taps into motor simulation. Aglioti, Cesari, Romani and Urgesi (2008) demonstrated that expert basketball players could predict the accuracy of a free throw on the basis of the player's kinematics, whereas expert watchers and novices could not. Similarly, Sebanz and Shiffrar (2009) found that expert basketball players could distinguish real passes from fake passes by observing another player’s actions, both when the actions were shown in videos and when they were shown as point-light displays. In contrast, novice basketballers were not able to discriminate real and fake passes. These results imply that motor expertise can be a pre-condition for identifying intentions from an observed agent’s kinematics.

In sum, previous research shows that movement kinematics provide a rich source of information that observers can use to make predictions about observed agents’ intentions. Even when instrumental actions are not intended to inform the observer, they are nonetheless a rich source of information due to dedicated perceptual processing of kinematic cues (Becchio et al. 2012) and people’s ability to map observed actions onto their own motor repertoire (Rizzolati & Sinigaglia, 2010; Ansuini, Cavallo, Bertone & Becchio, 2015).

3.1.2 Sensorimotor communication in joint action coordination and teaching

The kinematics of an action do not only provide cues to intention as a side effect of an actor’s performance, but they can also reflect an actor’s intention to inform another agent (Sperber & Wilson, 2004). Thus, action kinematics can be actively used as a channel of information for

joint action coordination and communication. Pezzulo and colleagues (2013) coined the term 'sensorimotor communication' for this active use of kinematics to inform. Sensorimotor communication is special compared to other forms of communication in that communication is superimposed on performed instrumental actions. Specifically, actors make instrumental actions informative by modulating kinematic parameters so that the actions become more predictable and less ambiguous (Pezzulo, Donnarumma & Dindo, 2013).

Sensorimotor communication is often observed in joint actions, where co-actors make their actions more informative in order to effectively achieve interpersonal coordination. In a study by Sacheli and colleagues (2013), two participants were instructed to grasp a bottle synchronously with either a power or a precision grip. Crucially, only the 'leader' knew which part of the bottle to grasp, while the 'follower' relied on the leader's actions to select the appropriate grip. Compared to followers, leaders reduced the velocity of their movements, and modulated wrist height and grip aperture. This made their movements more informative, communicating task relevant information to their joint action partner. It is also important to note that sensorimotor communication is only produced when informative cues are required, which is evidenced by findings demonstrating that actors no longer produce kinematic cues when their co-actor already has access to the information necessary to complete the joint task (Pezzulo & Dindo, 2011; Leibfried, Grau-Moya, & Braun, 2015).

Developmental research on imitation shows that sensorimotor communication also occurs in teaching contexts, with teachers adjusting their actions to make them more informative for the learner. Brand, Baldwin and Ashburn (2002) found that when mothers demonstrated actions to their children, their movements were more punctuated and pronounced, with a larger range of motion. This was labelled 'motionese' and has been shown to facilitate imitation of observed actions. Infants are more likely to imitate actions

containing motionese, compared to actions without motionese (Koterba & Iverson, 2009). It has been proposed that motionese enhances understanding of the goal structure of the action by guiding attention to important parts of an action sequence (Nagai and Rohlfing, 2009). These studies can be taken as evidence that sensorimotor communication is important for teaching through demonstration.

Using a virtual xylophone playing task, McEllin, Knoblich and Sebanz (2017) directly compared sensorimotor communication in joint action and in teaching through demonstration. Participants who had been trained to play melodies on a virtual xylophone produced different kinematic cues when trying to play the melodies in synchrony with a novice, compared to when they were demonstrating melodies to a novice. Specifically, modulations of movement height were used to support both teaching and coordination, modulations of the acceleration phase (ascent) of a movement were used to support spatial prediction in joint action coordination, and modulations of the deceleration phase (descent) of a movement were used to support temporal prediction in joint action coordination. This indicates that different kinematic cues are produced to support different informative intentions. In joint action kinematic cues are optimized to make the communicator's action more predictable, whereas in teaching kinematic cues are optimized to orient the learner's attention.

3.1.3 Reading informative intentions from actions

The finding that communicators modulate the kinematics of their actions differentially in joint action and teaching contexts raises the question of whether the recipients of the communication can identify communicators' informative intentions from observing their movements. We first aimed to investigate whether the recipients of sensorimotor communication can distinguish instrumental actions that have an informative intention

superimposed from regular instrumental actions. Given that actors differentially modulate kinematics for different informative intentions (coordination vs teaching), we further aimed to investigate whether people can distinguish different informative intentions based on the kinematics of observed actions. Finally, we aimed to investigate which types of kinematic cues make people perceive that an actor has a coordination intention or a teaching intention.

We used a task in which participants were presented with a point light-display of a mallet movement that corresponded to an actor playing simple melodies on a virtual xylophone. Participants were asked to categorize the displays as reflecting individual action, demonstration for teaching, or part of a coordinated joint action. The observed movements were synthesized so that they corresponded to fundamental movement laws. Maximum height and velocity profile of the movements were systematically varied because they had been identified as the main cues communicators used in coordination and teaching contexts in our previous study (McEllin et al., 2017). Artificially modulating kinematic parameters rather than using natural kinematics gave us full experimental control over the kinematic cues in the display.

Assuming that intentions can be rendered 'visible' based on the kinematic signatures of actions, we made the following predictions. First, we predicted that participants would be able to discriminate between actions without informative intention (individual) and actions with an informative intention (teaching and joint), on the basis of kinematic cues. More specifically, given that actions that are intended to improve joint action coordination have been shown to be slower with a larger maximum height (Vesper and Richardson, 2014; McEllin et al. 2017), we predicted that participants would use exaggerated movement height and duration in order to categorize actions as joint, compared to individual actions. Based on the finding that teaching actions are characterized by more exaggerated movements (larger

maximum height or larger range of motion) (Brand et al. 2002; McEllin et al. 2017) and by a slower pace (Dunst, Gorman & Hamby, 2012), we predicted that participants would also use exaggerated height and duration in order to categorize actions as teaching actions, compared to individual actions. Second, we predicted that participants would be able to discriminate between different informative intentions (joint action versus teaching) on the basis of different kinematic cues.

3.2 Experiment 1a: Discriminating Individual and Joint Actions

The aim of Experiment 1 was to investigate whether people can discriminate actions that are performed with an informative intent in the context of joint action coordination from regular instrumental actions. Based on previous findings on kinematics produced during joint action (Vesper & Richardson, 2014), we predicted that participants would infer from exaggerated movement height and duration that the observed movement reflected an intent to inform a task partner about movement goals in order to achieve joint action coordination.

3.2.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 20 participants (13 males, 7 females), with a mean age of 25.4 (SD = 4.3). All participants gave informed consent and were given 1500 Forint (approximately 5 Euros) worth of vouchers for their participation. This study was approved by the United Ethical Review Committee for Research in Psychology (EPKEB). Informed consent was obtained from all participants, and they were fully briefed and debriefed before and after the experiment.

Apparatus and Stimuli

Using data from previous experiments (McEllin et al., 2017), we synthesized point light displays of sequences of mallet movements reflecting the playing of melodies on a xylophone (see Figure 1). The virtual xylophone had ten projected keys, each 5cm wide and 24cm long, separated by a 4cm gap. Participants were required to learn simple action sequences, by moving the xylophone mallet from key to key in order to play a melody. To derive realistic parameters for our synthesized movements we computed, from the participants' individual performances without informative intent, the mean trajectories for movements of one, two and three key distances, for left and right movements. This resulted in six movement primitives, movements of one, two, and three keys to the left and right, which could be configured to synthesize, with appropriate resting times on the keys (100 ms delay between movements), action sequences reflecting the playing of melodies. While synthesizing the action sequences we used a pseudo-random sequence of the movement primitives with the added constraint that there had to be a direction change at least every two movements. This served to ensure that the mallet did not move off the xylophone displayed. Twenty unique six element action sequences were synthesized (see Appendix 1).

We artificially modulated the kinematic parameters of movement height and duration, by increasing (exaggeration) or decreasing (suppression) these parameters by 25% relative to the movement height and duration of the mean trajectory (the transformation was applied to all samples of the trajectory). For each kinematic parameter this resulted in three levels of modulation, (suppressed, original, exaggerated). Every action sequence was subject to each level of both height modulation and duration modulation, resulting in nine different height-duration combinations, for every action sequence.

The data were animated using MATLAB psychophysics toolbox. A lateral view of the xylophone was represented by ten blue rectangles (96 x 15 dimensions), arranged horizontally, and separated with a (72 pixel) gap. These dimensions were proportionate to the dimensions of the original xylophone. The xylophone mallet was represented by a green circle, which moved in accordance with the motion data. Please see Figure 1 for a sketch of an example of one trial. The vertical and horizontal motion data were transformed into pixels and scaled down to fit within the dimensions of the animated xylophone. Data were presented at a rate of 60HZ, with a frame of data being sampled and presented every 16ms. Responses were recorded using a custom designed button box.

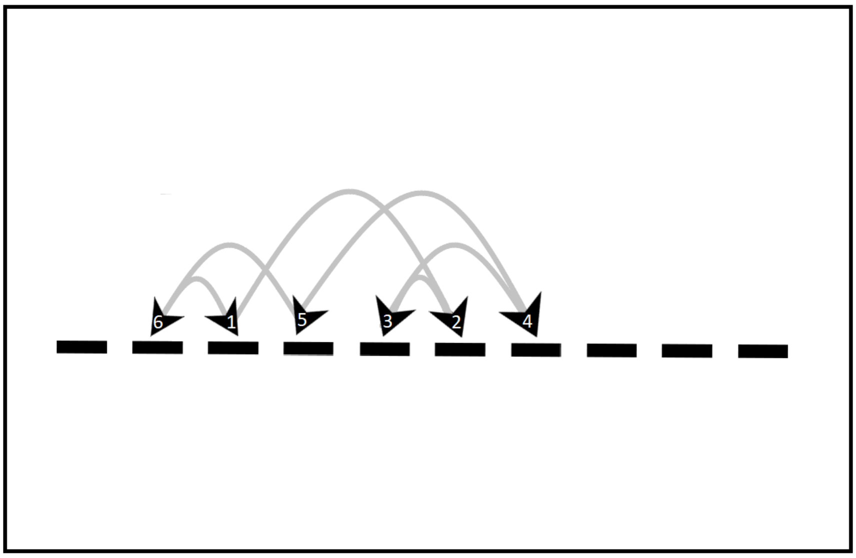

Figure 1: Graphical depiction of the stimulus for one trial. Numbers represent the movement order for the action sequence, and arrows depict movement direction and end position of each movement.

Procedure

Participants were told that they would complete a task in which they would have to decide whether a xylophone sequence played showed a participant playing alone (individual) or a

participant playing together with another participant (joint). Participants were then provided with information about the individual condition and the joint condition from the previous set of experiments (McEllin et al., 2017). We described the individual condition as a task in which the observed participant played a xylophone sequence alone. We described the joint condition as a task in which the observed participant played the action sequence together in synchrony with an unknown participant who did not know the sequence.

Participants were told that half of the action sequences they were about to observe were from the individual condition, and half of the action sequences were from the joint condition. They were also told that participants played the exact same action sequences in both conditions. Participants were then familiarized with the current stimuli, being shown a frame depicting the xylophone and the mallet. They were told that for each trial the data from one of the two conditions would be reanimated, with the green circle representing the mallet head. We then had participants complete two practice trials, in order to further familiarize them with the kinematic displays and the decision they were asked to make.

In each trial of the main experiment participants were presented with a 500ms fixation cross, followed by an animation of one of the action sequences. The duration of the action sequences ranged from 2460-4100 ms. Then participants were presented with a prompt screen which instructed them to indicate whether the action sequence they just watched had been played individually or as part of a coordinated joint action, by responding on a button box.

Each participant completed 180 trials judging 20 different action sequences for each height-duration modulation. The order of action sequences with different height-duration modulations was fully randomized. Whether participants categorized an action as individual or joint with a left or right button press was counterbalanced across participants.

Design

This experiment had a 3 x 3 within participant design, with the factors height modulation (suppressed, baseline, exaggerated) and duration modulation (suppressed, baseline, exaggerated). Our dependent variables were percentage of trials judged as joint (% Joint).

3.2.2 Results

A 3 x 3 within-participants ANOVA with the factors height modulation and duration modulation revealed a significant main effect of height modulation, $F(2,19) = 72.89$, $p < .001$, $\eta p2 = .79$ (see left panel of Figure 2). The percentage of joint choices was significantly larger for exaggerated height than for original height and for suppressed height. Moreover, the percentage of joint choices for original height was significantly larger than for suppressed height. There was no significant main effect of duration, and no interaction between height and duration (all $p > .05$).

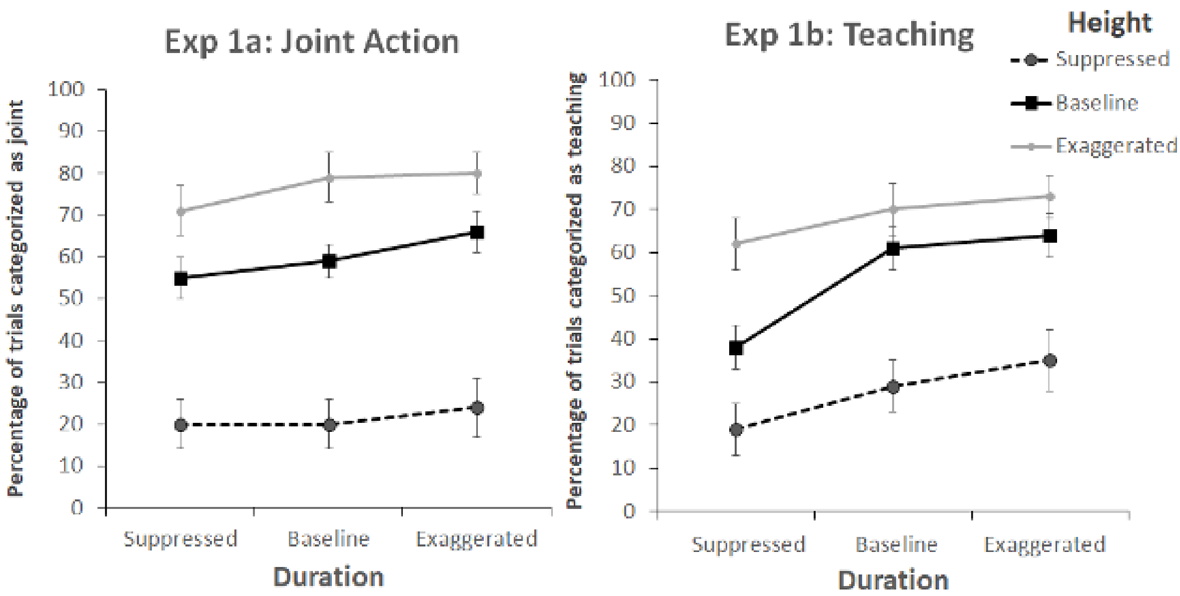

Figure 2: Interaction between Height and Duration for Experiments 1a and 1b. Error bars represent +/- 1 SEM.

3.3 Experiment 1b: Discriminating Individual and Teaching Actions

Experiment 1a demonstrated that participants use movement height as a cue to discriminate individual actions from actions performed with an informative intent in the context of joint action. This provides evidence that people can use kinematic cues to distinguish actions performed with an informative intent from actions performed without an informative intent. Another type of social interaction where actors modulate the kinematics of their movements to inform their co-actors is teaching. Here, the modulations serve to enhance attention to learning relevant information (Brand et al. 2002; McEllin et al., 2017). Experiment 1b asked whether people can discriminate between actions performed with the intention to teach and non-informative instrumental actions on the basis of kinematic cues. We predicted that participants would use movement height to discriminate actions performed with teaching intentions from regular instrumental actions, given the evidence for exaggeration of spatial parameters in teaching (Brand et al. 2002; McEllin et al. 2017). It is also possible that participants would use longer movement duration as an indication of a teaching intention, given that demonstrations for novice learners tend to be slower paced (Dunst, Gorman & Hamby, 2012).

3.3.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 20 participants (12 males, 8 females), with a mean age of 23.7 (*SD* = 3.5). All participants gave informed consent and were given 1500 Forint (approximately 5 Euros) worth of vouchers for their participation.

Apparatus and Stimuli

Like in Experiment 1a, participants were presented with point-light displays of artificially generated six-element xylophone sequences, in which height and duration were modulated.

Procedure and Design

The procedure was like Experiment 1a, except that participants were asked to decide whether the animated action sequence showed an individual playing alone, or an individual teaching a learner. They were provided with information about the individual condition and the teaching condition from the previous set of experiments (McEllin et al., 2017). They were told that half of the action sequences were from an individual playing alone and half of the action sequences were from an individual teaching. The teaching condition was described as a task in which the observed participant was required to demonstrate the action sequence to an unknown student who was required to watch and reproduce what was observed. Like in Experiment 1a, whether participants categorized individual and teaching actions with a left or right button press was counterbalanced across participants. The design was the same as Experiment 1a but with the percentage of trials judged as teaching (% Teaching) as the dependent variable.

3.3.2 Results

The 3 x 3 within-subjects ANOVA with height modulation and duration modulation as factors (see right panel of Figure 2) revealed a main effect of height, $F(2,19) = 19.92$, $p < .001$, $\eta p2 = .51$, with the percentage of teaching choices being significantly larger for exaggerated height than for original height and for suppressed height, and percentage of teaching choices for original height significantly larger than for suppressed height (all pairwise comparisons < .05). We also found a main effect of duration, $F(2,19) = 5.05$, $p =$

.011, $\eta p2$ = .21, with the percentage of teaching choices being significantly larger for original duration than suppressed duration. Moreover, there was an interaction between height and duration, $F(4, 19) = 3.19$, $p = .018$, $\eta p2$ = .144. There was a lower percentage of teaching choices for movements with a lower movement height and shorter duration.

3.3.3 Discussion Experiment 1a and 1b

Taken together, the results from Experiment 1a and 1b demonstrate that people are sensitive to sensorimotor communication and can infer informative intentions using low level kinematic cues. Exaggerated movement height made participants more likely to judge actions as joint rather than individual (Experiment 1) and as reflecting the intention to teach (Experiment 2). Longer movement duration did not increase judgments of actions as joint rather than individual. This was unexpected given our earlier findings where participants acting in a joint coordination context moved more slowly than when acting alone (Mc Ellin et al., 2017). It could be that movement height was a dominant cue in the present task, leaving open the question whether in the absence of height modulations people would use action duration to discriminate between actions performed with the intention to engage in coordinated joint action and individual actions. Movement duration had some effect on judgments of teaching intentions, with faster actions being judged unlikely to reflect a teaching intention.

Although participants were informed that half of the trials were individual trials and half of the trials were joint/teaching, participants seemed to be slightly biased towards categorizing trials as joint or teaching actions. One possibility could be that this reflects a more general bias towards perceiving social relations given minimal cues to interaction (Heider & Simmel, 1944). However, this bias cannot explain the observed results, as it does not imply a systematic effect of particular movement cues on judgments.

3.4 Experiment 2

Experiment 2 aimed to investigate whether people can discriminate between different types of informative intentions based on the different types of kinematic cues produced in these contexts. Specifically, we asked whether people can discriminate between actions performed with the intention to coordinate in a joint action and actions performed with the intention to teach. We did not make specific predictions for how participants would use height cues, given that an exaggerated movement height is observed in both joint action coordination and teaching (Vesper & Richardson, 2014; McEllin et al.2017) and that participants used maximum height to identify both the intention to teach and the intention to coordinate in Experiment 1a and 1b.

For duration, prior findings motivate two opposing predictions. On the one hand, longer durations may increase judgments of a teaching intention, given that demonstration often entails slower movements (Dunst, Gorman & Hamby, 2012), and given the findings of Experiment 1b where longer duration served as a cue towards teaching. On the other hand, we found in an earlier study measuring the kinematics involved in producing the xylophone melodies (McEllin et al., 2017) that performing the actions in a joint context with a partner resulted in slower movements while demonstrating the actions to an observer did not reliably lead to a slowing down. This predicts that exaggerated movement duration would serve as a cue to joint action.

3.4.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 20 participants (10 males, 10 females), with a mean age of 24.7 (*SD* = 4.7). All

participants gave informed consent and were given 1500 Forint (approximately 5 Euros) worth of vouchers for their participation.

Apparatus and Stimuli

Like in Experiment 1a and 1b, participants were presented with point-light displays of artificially generated six-element xylophone sequences, in which height and duration were modulated.

Procedure

The procedure was like Experiment 1a and 1b, but participants were asked to discriminate between actions performed with the intention to coordinate in a joint action, and actions performed with the intention to teach. They were provided with information about the joint condition and the teaching condition before the experiment started and were told that half of what they observed were joint actions and half were teaching actions. Like in Experiment 1a and 1b, whether participants categorized joint and teaching actions with a left or right button press was counterbalanced across participants. The design was the same as in Experiment 1a, that is, the dependent variable was the percentage of trials judged as joint (% Joint).

3.4.2 Results

The 3 x 3 within-subjects ANOVA with height modulation and duration modulation as factors (see Figure 3) showed a significant main effect of height, $F(2,19) = 3.49$, $p = .041$, $\eta p2 = .155$, with percentage of joint choice increasing as a function of height modulation. There was no significant main effect of duration and no significant interaction between height and duration.

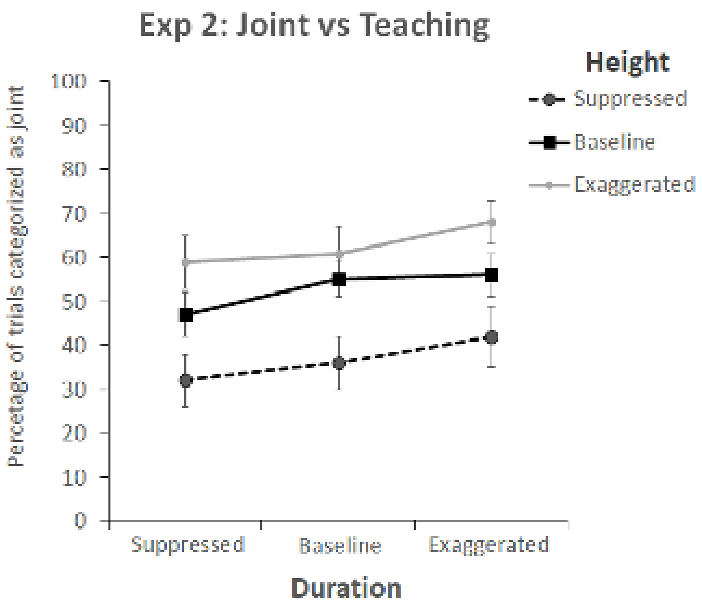

Figure 3: Interaction between Height and Duration for Experiment 2. Error bars represent +/- 1 SEM.

3.4.3 Discussion

The results of Experiment 2 demonstrate that people can use kinematic cues in order to discriminate between actions performed with different informative intentions. Unexpectedly however, we found that participants used movement height, but not duration to discriminate between actions performed with the intention to coordinate in a joint action and actions performed with a teaching intention. These findings imply that exaggerated movement height is more likely interpreted as an attempt to achieve interpersonal coordination during joint action than to serve teaching purposes. Note, however, that the effect of height in Experiment 2 is considerably smaller than in the previous two experiments, implying higher uncertainty in discriminating the underlying intentions of the movement.

3.5 Experiment 3a

The results of the previous experiments could be taken to suggest that participants hardly use timing cues to discriminate between actions performed with different intentions. However, the role of action duration may have been underestimated because the height modulation may have dominated participants' judgments. Furthermore, McEllin et al. (2017) demonstrated that participants differentially modulate their movement velocity during different movement phases depending on what they intended to inform a joint action partner about. Specifically, we found that ascent velocity of the mallet (the movement speed from the xylophone up to the maximum height) was modulated when participants were informing a partner about spatial locations. Descent velocity (the movement speed from the maximum height to the target key) was modulated when participants were informing a partner about the timing of their movements. Thus, subtle changes in velocity parameters may also be used to discriminate between different informative intentions.

In the ensuing four experiments (3a-4b), we aimed to further investigate whether timing cues can be used to discriminate between informative intentions and instrumental intentions, and between different informative intentions. In Experiment 3a, we asked whether in the absence of height modulations people use action duration to discriminate between actions performed with the intention to engage in coordinated joint action, and individual instrumental actions. We manipulated the duration of the up stroke and down stroke of the mallet resulting in different ratios of ascent and descent velocity. Because ascent velocity has been shown to support spatial prediction, we hypothesized that exaggerated duration of the upstroke (a slow-down in ascent velocity relative to descent velocity) would be used to identify the intention to engage in a coordinated joint action. Because descent velocity has been shown to support temporal prediction, we predicted that exaggerated duration of the down stroke (a slow-down in descent velocity relative to ascent velocity) would also be used in order to identify an intention to engage in coordinated joint action.

3.5.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 20 participants (14 males, 6 females), with a mean age of 24.6 (*SD* = 5.6). All participants gave informed consent and were given 1500 Forint (approximately 5 Euros) worth of vouchers for their participation.

Apparatus and Stimuli

Apparatus and Stimuli were the same as in Experiment 1a and 1b, except for how the kinematic parameters were modulated. We used the same action sequences as in Experiment 1a and 1b. Ascent duration was increased by a factor of 30%, 60% and 90% (see Appendix 2). Descent duration was kept constant. We then combined the ascent and descent durations and normalized them so that the duration matched the original overall duration, thus increasing the ascent duration relative to the descent duration. We did the same for the duration of the descent phase of the movements, increasing the descent duration by 30%, 60% and 90% (see Appendix 2), and then normalizing the overall duration to increase the proportion of the descent duration, relative to the ascent duration. We also had an individual baseline in which we never modulated the ascent or descent duration. We created each of these seven ascent-to-descent ratios for each of the twenty action sequences that we used in the previous experiments. Again, these action sequences were animated as point-light displays.

To dissociate effects of overall duration from effects of specific ascent and descent modulations, we also manipulated overall action duration. Every action sequence had a

randomly modulated duration, which ranged from the original duration (3280ms) to double the original duration (6560ms).

Procedure and Design

The procedure was the same as experiment 1a, but with 140 trials instead of 180 trials. The baseline duration and each ascent and descent duration modulation of the twenty action sequences were presented in a random order. Like in Experiment 1a, participants were provided with information about the individual condition and the joint condition, before being instructed to decide whether each of the observed action sequences was an individual action or performed with the intention to coordinate in a joint action.

We added the baseline as a level of both the ascent and descent factors, in order to compare each of the modulated actions to the unmodulated actions. We also created a factor of speed, by performing a median split based on movement duration in order to split the stimuli into slow and fast actions. This resulted in a design with a 2 x 4 within-participant comparison for ascent exaggeration (baseline, 30%, 60% and 90%) and speed (slow, fast) and a 2 x 4 within-participant comparison for descent exaggeration (baseline, 30%, 60%, 90%) and speed (slow, fast). Percentage of trials judged as joint (% joint) was the dependent variable.

3.5.2 Results

Ascent Exaggeration

We carried out a 2 x 4 ANOVA for ascent exaggeration (left panel of Figure 4) with the factors speed (fast, slow) and exaggeration (baseline, 30%, 60%, 90%). This analysis yielded a main effect of speed, $F(1,19) = 128.5$, $p < .001$, $\eta p2 = .87$ (Figure 6), and a main

effect of exaggeration, *F*(3,19) = 4.7, *p* = .005, *ηp2* = .2 (Figure 7). The percentage of trials judged joint was higher for slow movements than for fast movements, and percentage of trials judged as joint increased as a function of ascent exaggeration. However, there was no interaction between speed and exaggeration, *F*(3,19) = 2.1, *p* = .11, *ηp2* = .1.

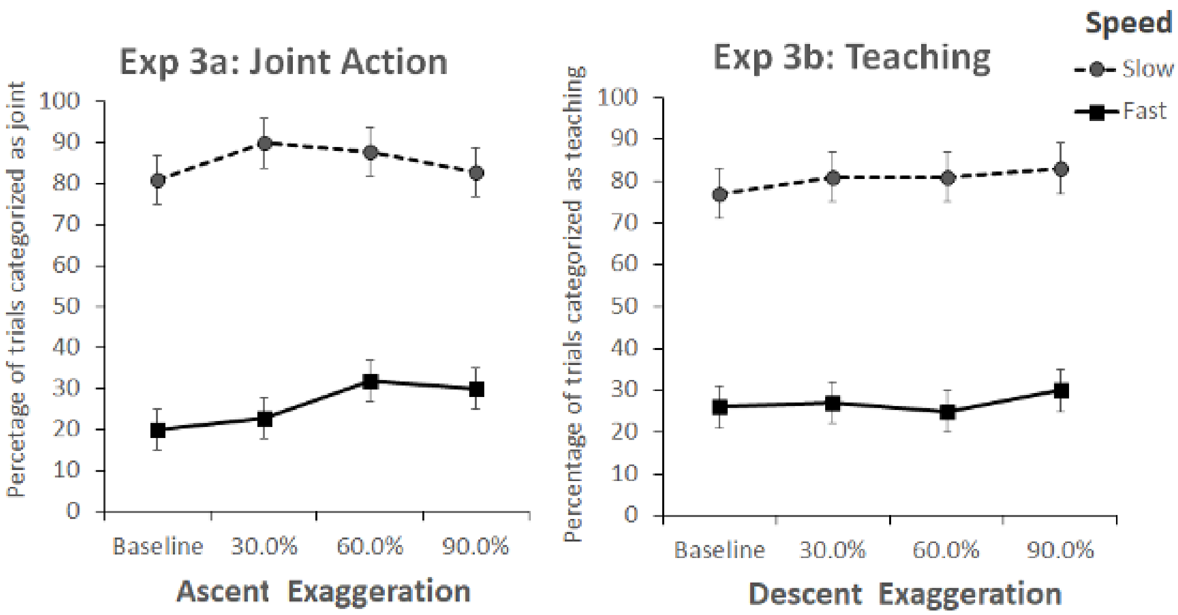

Figure 4: Interaction between Ascent Exaggeration and Speed for Experiment 3a and 3b. Error bars represent +/- SEM.

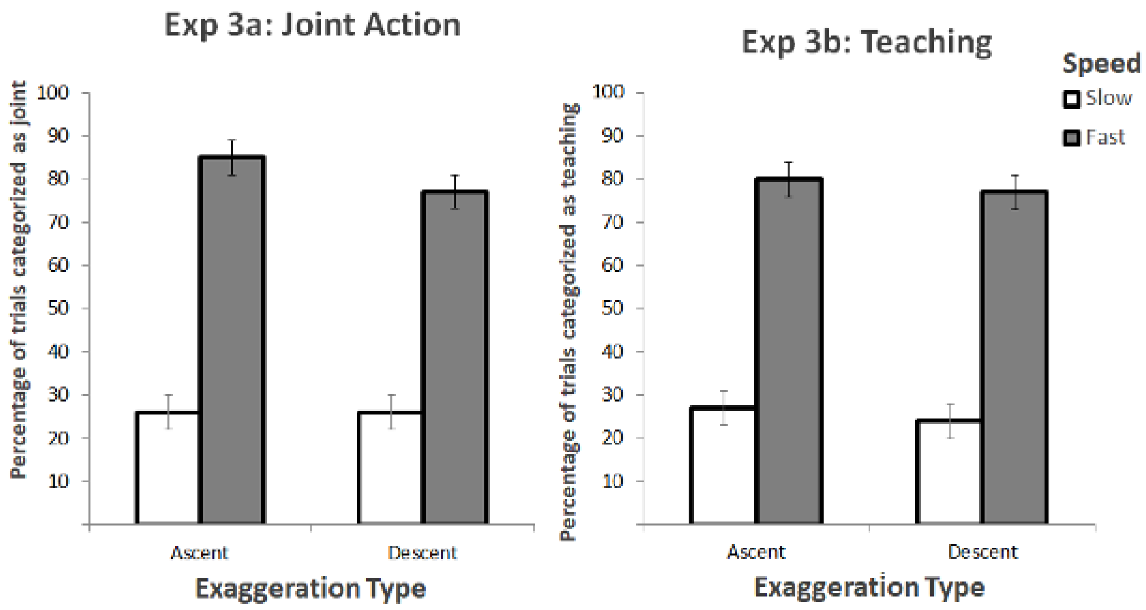

Figure 6: Main effect of Speed for both Ascent and Descent modulations. Error bars represent +/- 1 SEM.

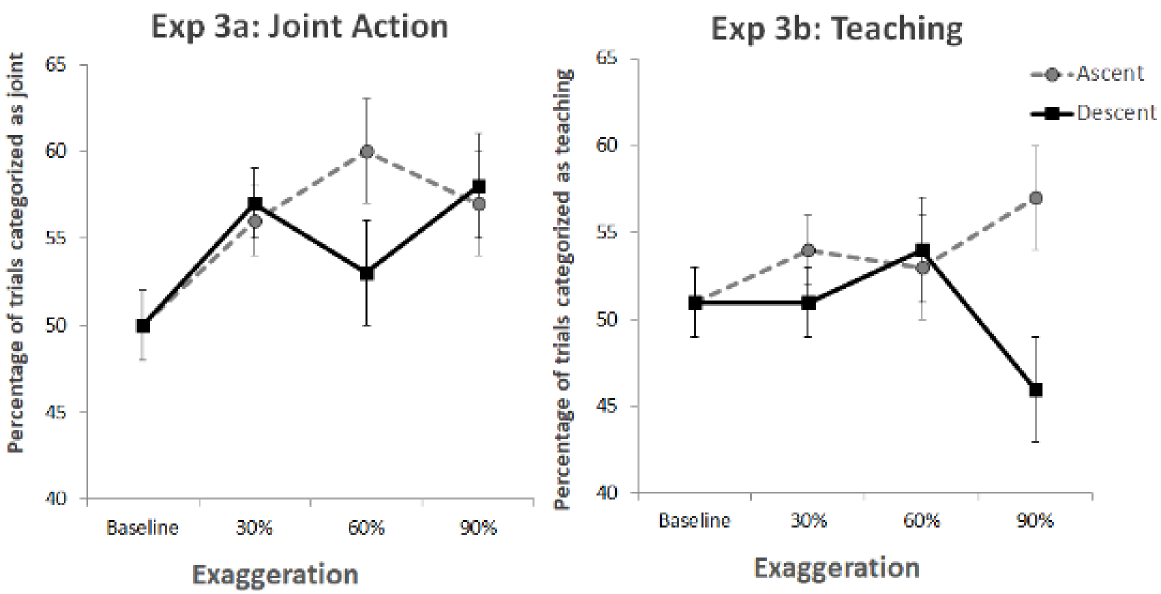

Figure 7: Main effects of Exaggeration for both Ascent and Descent Modulations. Error bars represent +/- 1 SEM

Descent Exaggeration

We carried out a 2 x 4 ANOVA for descent exaggeration (left panel of Figure 5) with the factors of speed (slow, fast) and exaggeration (baseline, 30%, 60%, 90%). The ANOVA revealed a main effect of speed, $F(1,19) = 120.09$, $p < .001$, $\eta p2 = .86$ (Figure 6), and exaggeration, $F(3,19) = 2.98$, $p = .039$, $\eta p2 = .136$. The percentage of trials judged joint was higher for slow movements than for fast movements, and the percentage of trials judged as joint increased as a function of descent exaggeration (Figure 7). There was no interaction between speed and exaggeration, $F(3,19) = .25$, $p = .86$, $\eta p2 = .01$.

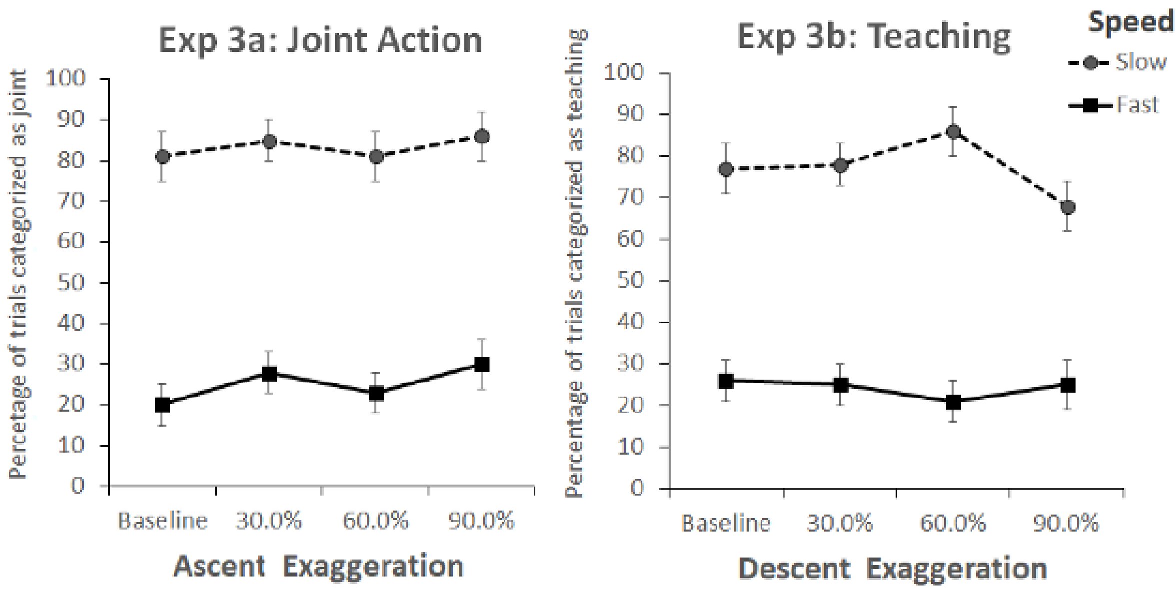

Figure 5: Interaction between Descent Exaggeration and Speed for Experiment 3a and 3b. Error bars represent +/- 1 SEM.

3.6 Experiment 3b

This experiment aimed to investigate whether people use information from the velocity profile of an observed action in order to infer teaching intentions. In particular, we

investigated whether participants can decide whether an observed action was performed alone or whether it was performed with the intention to teach a learner, on the basis of the speed and ascent and descent ratio of that action. Because the results from Experiment 1b indicate that action duration serves as a cue to teaching, we predicted that participants would mostly rely on overall movement speed in order to discriminate between individual movements and movements performed with the intention to teach.

3.6.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 20 participants (11 males, 9 females), with a mean age of 22.8 (*SD* = 2.7). All participants gave informed consent and received 1500 Forint (approximately 5 Euros) worth of vouchers for their participation.

Apparatus and Stimuli

Like in Experiment 3a, participants were presented with point-light displays of artificially generated six-element xylophone sequences, in which ascent and descent duration and overall duration were modulated.

Procedure

This was the same as in Experiment 3a, except that participants were familiarized with the individual condition and the teaching condition, and then instructed to decide whether the observed action was an individual action or a teaching action.

Design

This was the same as in Experiment 3a, except that percentage of teaching choices was the dependent variable.

3.6.2 Results

Ascent Exaggeration

Like in Experiment 3a, we carried out a 2 x 4 ANOVA for ascent exaggeration (right panel of Figure 4) with the factors of speed and exaggeration. It revealed a main effect of speed, $F(1,19) = 65.85$, $p < .001$, $\eta p2 = .78$ (Figure 6), but no main effect of exaggeration, $F(3,19) = .82$, $p = .49$ (Figure 7), $\eta p2 = .04$. The percentage of trials judged as teaching was higher for slow movements compared to fast movements. There was no interaction between speed and exaggeration, $F(3,19) = .33$, $p = .81$, $\eta p2 = .02$.

Descent Exaggeration

For descent exaggeration, we carried out a 2 x 4 ANOVA with speed and exaggeration as within-participant factors (right panel of Figure 5). It revealed a main effect of speed, $F(1,19) = 60.23$, $p < .001$, $\eta p2 = .76$ (Figure 6), but no effect of exaggeration, $F(3,19) = 1.65$, $p = .188$, $\eta p2 = .08$ (Figure 7). The percentage of trials judged as teaching was higher for slow movements compared to fast movements. However, there was a significant interaction between speed and exaggeration, $F(3,19) = 3.17$, $p = .031$, $\eta p2 = .14$, with the percentage of teaching choices being lower for 90% descent exaggeration during slow speed.

3.6.3 Discussion Experiment 3a and 3b

The results from Experiment 3a and 3b show that people can use temporal cues to detect an actor's informative intentions (both coordination and teaching) and suggest that in our first two experiments, the duration modulation had not been salient enough for the participants.

Interestingly, we also found that participants used the relative length of the ascent phase and descent phase of the movements in order to discriminate between individual actions and joint actions (Experiment 3a), while they did not use this information to discriminate between individual actions and teaching actions (Experiment 3b). This may provide some indication that people are more sensitive to temporal cues in actions performed with an intention to coordinate, compared to actions performed with an intention to teach. However, an experiment comparing these two types of intention would be needed in order to provide conclusive evidence.

3.7 Experiment 4a

Experiment 4a aimed to investigate whether people can use temporal cues in order to discriminate between actions performed with different informative intentions (joint action and teaching). Considering that people have been shown to modulate the ratio of ascent to descent velocity in order to enhance spatial and temporal prediction in joint action (Sacheli et al. 2013; McEllin et al.), we predicted that participants would categorize movements with larger ratios of ascent to descent velocity as joint actions, as they understand the role this informative modulation plays in spatial and temporal prediction.

3.7.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 20 participants (13 males, 7 females), with a mean age of 23.4 (SD = 5.2). All participants gave informed consent and received 1500 Forint (approximately 5 Euros) worth of vouchers for their participation.

Apparatus and Stimuli

Like in Experiment 3a and 3b, participants were presented with point-light displays of artificially generated six-element xylophone sequences, in which ascent and descent duration and overall duration were modulated.

Procedure

This was the same as experiment 3a and 3b, except that participants were provided with information about the teaching condition and the joint condition, and then instructed to categorize the observed actions as being performed with an intention to coordinate or as being performed with the intention to teach.

Design

This was the same as Experiment 3a; percentage of trials judged as joint was the dependent variable.

3.7.2 Results

Ascent Exaggeration

We carried out a 2 x 4 within-participant ANOVA and found neither a significant main effect of speed, $F(1,19) = .18$, $p = .67$, $\eta p2 = .01$ (Figure 8), nor a significant main effect of

exaggeration, *F*(3,19) = .77, *p* = .52, *ηp2* = .04. The interaction between exaggeration and speed was also not significant, *F*(3,19) = 1.5, *p* = .22, *ηp2* = .08.

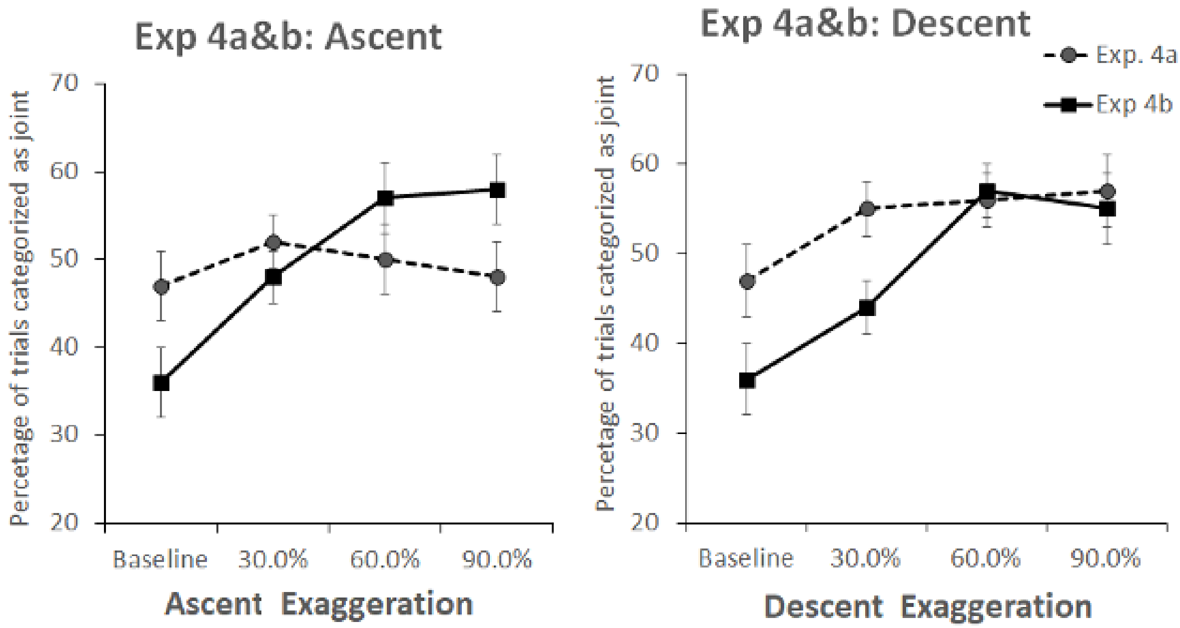

Figure 8: Interaction between Exaggeration (Ascent: Left, Descent: Right) and Experiment (4a: Speed factor present. 4b: Speed factor absent). Error bars represent +/-1 SEM.

Descent Exaggeration

A 2 x 4 within-participant ANOVA revealed a significant main effect of exaggeration, *F*(3,19) = 3.73, *p* = .016, *ηp2* = .16 (Figure 8), but no main effect of speed, *F*(1,19) = .34, *p* = .56, *ηp2* = .02. Percentage of trials judged as joint increased as a function of descent exaggeration. There was no significant interaction between exaggeration and speed, *F*(3,19) = 1.61, *p* = .2, *ηp2* = .08.

3.8 Experiment 4b

Experiment 4a provided first evidence that participants can use the ratio between the ascent and descent duration to discriminate between actions performed with an intention to

coordinate and actions performed with a teaching intention. Experiment 4b served to replicate this finding and to determine whether discrimination becomes more reliable when overall speed does not vary.

3.8.1 Method

Participants

Using an online participant database (Sona systems, www.sona-systems.com), we recruited 20 participants (9 males, 11 females), with a mean age of 21.6 (SD = 1.8). All participants gave informed consent and received 1500 Forint (approximately 5 Euros) worth of vouchers for their participation.

Apparatus and Stimuli

Like in Experiment 4a, participants were presented with point-light displays of artificially generated six-element xylophone sequences, in which ascent and descent duration modulated. However overall duration was not modulated.

Procedure

This was the same as Experiment 4a.

Design

This was the same as Experiment 4a.

3.8.2 Results

Ascent Exaggeration

We carried out a one-way ANOVA with the factor exaggeration. We found a significant main effect of exaggeration, $F(3,19) = 7.01$, $p < .001$, $\eta p2 = .27$. The percentage of trials judged as joint increased as a function of ascent exaggeration (Figure 8).

Descent Exaggeration

We carried out a one-way ANOVA that revealed a significant main effect of exaggeration, $F(3,19) = 6.69$, $p = .001$, $\eta p2 = .26$. The percentage of trials judged as joint increased as a function of descent exaggeration (Figure 8).

3.8.3 Discussion Experiment 4a and 4b

Experiment 4a did not provide any evidence that participants use movement speed in order to discriminate between actions performed with the intention to coordinate and actions performed with the intention to teach. However, Experiment 4b provided evidence that people use exaggerated ascent and descent durations in order to discriminate between joint actions and teaching actions. These findings demonstrate that people can use specific information about the velocity profile in order to discriminate between different types of informative intentions.

Interestingly, we found that participants only used ascent exaggeration in the absence of any overall speed cues, whereas participants can use descent exaggeration regardless of whether speed cues are present or not. This could indicate that with regards to coordinated joint actions, people have stronger expectations about the descent phase of a movement, compared to the ascent phase. Considering that ascent velocity is typically used to inform a task partner about spatial movement parameters and descent velocity is used to inform a task partner about movement timing, it may be the case that the temporal requirements of joint action were more salient for the participants than the spatial requirements.

3.9 General Discussion

We aimed to investigate whether people can discriminate between actions performed with informative intentions and purely instrumental individual actions and whether they can discriminate between actions with different informative purposes such as the intention to perform a coordinated joint action coordination and the intention to teach through demonstration.

Regarding the first aim, previous research has demonstrated that people can detect the instrumental and social intentions of an actor on the basis of kinematic signatures (Cavallo et al. 2016; Manera et al. 2011). We extend this research by demonstrating that people can also detect an actor's informative intentions as expressed through sensorimotor communication. Our findings demonstrate that people use different movement cues in order to distinguish instrumental actions performed with an informative intention from individual instrumental actions without informative intention. Actions that systematically deviate from the easiest way of individually performing an effective instrumental action are understood as fulfilling some informative purpose (Pezzulo et al. 2013; 2018). Our findings challenge theories of social cognition suggesting that movement cues alone are not sufficient for detecting intentions beyond motor intentions (Jacob & Jeannerod, 2005). Minimally, the findings demonstrate that there are some instances where informative intentions are derived from the kinematics of an observed movement.

An important goal for future research is to quantify the accuracy with which informative intentions can be identified. In the present study, we exaggerated natural movement kinematics to be able to specify and dissociate the contribution of different movement parameters. This approach allowed us to measure participants' tendency to attribute particular informative intentions as a function of exaggeration of particular

movement cues, while their judgments were not right or wrong. Exposing participants to actual kinematics from teaching and joint action coordination contexts can contribute to the understanding of the efficiency of sensorimotor communication in these contexts.

Theories of communication assert that in order for communication to succeed, one needs to explicitly recognize an interaction partner's 'communicative intention' (Sperber & Wilson, 2004). Our results could be taken to suggest that sensorimotor communication can be sufficient for making communicative intentions explicit. However, it is possible that participants merely derived informative intentions, which specify the kind of information to be transmitted rather than making the actor's communicative intention explicit. As understanding informative intentions seems to be sufficient in many joint action and teaching contexts, it may be the case that only very large deviations from optimal performance elicit explicit attributions of communicative intentions (such as when the observer sees the other waving the mallet to grab attention). Further research could aim to identify kinematic parameters that discriminate between actions produced in a context requiring the detection of communicative intentions and actions produced in a context requiring only the identification of informative intentions.

The second aim of the present study was to investigate whether people can discriminate between actions performed with the intention to inform a co-actor in a joint action, and actions performed with the intention to inform a student in a teaching context. We found that participants reliably used movement height in order to discriminate between actions performed with an intention to coordinate in a joint action and actions performed with the intention to teach. This is somewhat surprising given that modulations of the spatial trajectory serve not only to enhance spatial and temporal prediction in joint action, but also to highlight the structure of an action in teaching (McEllin et al. 2017). It could be that people

expect modulations of the spatial trajectory of one’s movement to be more crucial for spatial and temporal prediction than for highlighting the structure of an action because they have more experience with exaggerating the spatial trajectory of their movements when being engaged in coordinated joint actions compared to when teaching through demonstration.

Furthermore, we found that people reliably categorize movements with elongated ascent or descent ratios as joint actions rather than teaching actions. This suggests that participants perceive an elongated ascent or descent phase of a movement as an informative action modulation when trying to coordinate in a joint action, but not when teaching. Indeed, there is evidence that people elongate the ascent phase of their movement in order to make themselves spatially predictable, that they and elongate the descent phase of their movement in order to make themselves temporally predictable (McEllin et al. 2017; Sacheli et al. 2013). Considering this, it is likely that participants perceived the longer ascent phase as an action modulation which provides actors with more time to make spatial predictions about their co-actor’s targets. Likewise, longer relative descent phases could be perceived as modulations which provide actors with more time to make temporal predictions about their co-actor’s movements.

It is possible that the ability to detect informative intentions arises through participants' simulation of performing the observed actions. Indeed, Becchio, Sartori & Castiello (2010) proposed that the ability to understand an actor’s intentions through observing their early kinematics relies on motor simulation, mapping the kinematics onto their own motor repertoire in order to predict how the action will unfold. This is supported by evidence showing stronger activation of mirroring networks when observing cooperative and competitive actions, compared to individual actions (Becchio et al. 2012). The same mechanism could be employed in order to detect sensorimotor communication. Given that

people deviate from optimality in order to send informative signals in social contexts (Pezzulo et al. 2013), they may also understand that an action is informative when it systematically deviates from the easiest way of performing the action in ways that they themselves would use to signal an informative intent to an observer.

Given that we used a musical task, one may wonder whether the observed actions were actually in our participants' motor repertoire. Although participants may not have direct experience with playing a xylophone, it is likely that they have experience with instrumentally similar actions (e.g., moving to a sequence of locations in a particular order), so that they could understand the actions of the xylophone players using motor simulation. Moreover, we found in our earlier study using the xylophone task that non-musicians reliably produced informative movement cues (McEllin et al. 2017), suggesting that the observed actions were indeed within our observers' repertoire. One discrepancy between the kinematics produced in performing the task and the kinematics used to infer informative intentions is that participants relied less on timing cues when judging observed actions. It is possible that limited expertise with the observed actions made it easier to detect deviations in movement height than subtle deviations in timing.

A way to further test the role of motor simulation would be to investigate performance on the present task in a population of people who lack motor experience of producing informative cues in social interactions. A recent study demonstrated that those with autistic spectrum conditions (ASC) are less likely to produce sensorimotor communication in coordination contexts (Curioni, Minio-Paluello, Sacheli, Candidi & Aglioti, 2017). Thus, for the current task one could predict that individuals with ASC would not be able to reliably discriminate between actions produced with an informative intention and non-informative actions on the basis of kinematic cues, due to the lack of experience in producing

sensorimotor communication. In a similar vein, experts should be more sensitive to detect kinematic cues signalling different informative intentions in their domains of expertise. If expectations of sensorimotor communication are driven by experience in producing kinematic cues, increasing expertise should increase sensitivity to these cues.

A further issue that would be interesting to address is to determine when kinematic modulations stop to act as informative cues and become noise. In Experiment 3b, we unexpectedly found that participants were less likely to categorize slow movements with very long descent phases as teaching actions. One explanation for this finding could be that participants perceive very long descent phases as reflecting hesitance or uncertainty, rather than perceiving them as being informative in a teaching context. Alternatively, it could be that very large kinematic exaggerations are interpreted as mistakes or bad performance rather than signalling an informative intent in a teaching context. This may have occurred specifically for teaching actions but not joint actions because participants expect modulating ascent and descent ratio to be useful to achieve interpersonal coordination, but not for teaching (as evidenced by Experiment 3a, 4a and 4b). More generally, this finding suggests that although deviating from individual efficiency by exaggerating kinematic parameters allows one to provide useful and informative cues in social interactions such as coordination and teaching, there may be a threshold at which kinematic exaggeration is no longer informative, and actually makes the performed action more ambiguous and harder to predict. Further research should investigate at what point deviation from optimality actually begins to violate the process of mapping observed actions onto our own motor system rather than facilitating it.

Chapter 4: Perceiving joint action using relational movement cues

4.1 Introduction

Humans are remarkably sensitive to the interpersonal relations between multiple actors engaged in social interactions, with the ability to attribute a whole host of goals and intentions to interacting actors on the basis of minimal movement cues (Heider & Simmel, 1944). Most of the previous research on this issue has focused on how relational cues affect causality and intention attribution (e.g. Scholl & Tremoulet, 2000). Less is known about how relational movement cues affect our perception of the interpersonal coordination of multiple agents engaged in joint actions such as when watching two dancers performing a piece together.

Observers seem to be particularly awed by joint performances that are well coordinated despite relying on a significant amount of improvisation. Recent studies on joint improvisation have demonstrated that the actions of expert improvisers have a kinematic profile that is distinct from the actions of novice improvisers (Noy, Dekel & Alon, 2011). But is not known whether observers perceive joint movements characterized by these distinct profiles as being particularly well coordinated. Thus, the first aim of the present study was to investigate whether relational movement parameters that characterize the joint movements of improvisation experts make observers perceive enhanced interpersonal coordination.

Studies on interpersonal coordination have demonstrated that synchrony creates affiliation between those moving in synchrony (Hove & Riesen, 2009; Wilthermuth & Heath, 2009). Seeing others performing synchronous movements also leads observers to judge the observed actors as having a high level of rapport (Miles, Nind & Macrae, 2009). However, these studies have been restricted to repetitive or choreographed movement such as walking side by side. Less is known about how affiliation is perceived in more open-ended joint actions, such as when two people improvise together. Thus, the second aim of the present

research was to investigate whether relational movement parameters that characterize the joint movement of improvisation experts make people perceive more affiliation between actors.

Even though performances that consist of experts improvising with each other, such as dance and acting are so commonly enjoyed by spectators, the extent to which movement cues produced by actors in these contexts affect an observers aesthetic experience has yet to be investigated. Although the visual system plays a central role in the perception of beauty with regards to many types of art (Zeki, 2001), there is increasing evidence that the sensorimotor system is involved in the aesthetic experience of dynamic performances, with movement cues driving how beautiful we find an observed action sequence (Calvo-Merino, Jola, Glaser & Haggard, 2008; Cross, Kirsch, Ticini & Schutz-Bosbach, 2011). Moving beyond individual movements, a study by Vicary, Sperling, von Zimmerman, Richardson and Orgs (2017) found that the levels of synchrony between dancers in a choreographed performance predicted spectators' level of enjoyment and aesthetic experience of the performance. However, less is known about the extent to which relational movement features of expert's movements in an open ended and improvisational performance have an effect on our aesthetic experience. Considering this, our third aim was to investigate whether relational movement parameters that characterize the joint movement of improvisation experts are perceived as especially aesthetically pleasing.

4.1.1 Interval-based and velocity-based synchronization

In order to make specific predictions concerning the observation of joint improvisation, it is important to distinguish between interval-based and velocity-based synchronization. Tasks such as finger tapping or walking side by side that have traditionally been used to study interpersonal synchrony (e.g. Kovanalinka et al. 2011; Schmidt & Richardson, 2008) can be seen as involving interval-based synchronization, with actors

aligning the timing of their movements in such a way that they reach their end point at the same time. Interval based synchrony is supported by anticipatory mechanisms and adaptation mechanisms such as phase and period correction, which help actors to ensure that their movements repeatedly end at the same time as their co-actors' movements (Konvalinka et al. 2011; Repp & Keller, 2008).

Synchronization in which actors' movements are continuously aligned throughout the duration of the movement can be described as velocity based because this type of synchronization requires the whole velocity profile of movements to be overlapping. While little research has investigated people engaging in velocity based synchrony, recent work suggests that this type of synchronization may be a mark of coordination expertise in joint improvisation. Noy, Dekel, and Alon (2011) set out to investigate how novices and experts coordinate open-ended and continuous movements. Based on the improvisational theatre exercise known as the mirror game in which people are required to continuously mirror each other, they designed a one-dimensional version, in which pairs of individuals facing each other were required to move sliders from side to side with the instruction to 'synchronize and imitate each other, create interesting patterns, and enjoy playing'. They found that novice improvisers could achieve interval-based synchrony, successfully coordinating their movements at the end points (where direction changes occurred). However, they could not achieve velocity-based synchrony, as they relied on a leader-follower strategy, with one participant (the "leader") moving with a smooth trajectory, and the other participant (the "follower") 'jittering' around this smooth trajectory at a rate of 2-3 Hz. However, experts could successfully achieve both interval-based and velocity-based synchrony, aligning their velocity profiles and synchronizing their movements seamlessly without any jitter, as if they were both leaders who were improvising together; the authors labelled this 'co-confident motion'. In a later study, Hart, Noy, Feinger-Schall and Alon (2014) investigated the

skewness (symmetry) and kurtosis (sharpness) of the velocity profiles of those playing the mirror game. They found that experts' movements could be characterized by a sine-wave like velocity profile, with low skewness (symmetrical acceleration and deceleration phase) and low kurtosis (relatively linear acceleration and deceleration phase). Importantly, this occurred regardless of the participants' individual velocity characteristics when not engaged in co-confident motion, suggesting that experts could achieve velocity-based synchronization by converging or agreeing on a particular movement style that is easy to align with.

4.1.2 Present study

The present study aimed to investigate the difference between interval based synchrony and velocity based synchrony in terms of how these types of synchrony inform judgements of a joint performance. Based on recordings of single movements, we created displays of artificial dyads showing two moving dots. Participants were told that each dot represented the movements of one partner in a pair improvising together. We investigated the effects of interval based synchrony by creating high asynchrony and low asynchrony dyads. We manipulated different cues to velocity based synchrony in two separate experiments in order to provide converging evidence that interval based synchrony and velocity based synchrony differentially affect judgements of coordination, affiliation, and aesthetics. In Experiment 1, we investigated the effects of jitter as a cue to velocity based synchrony by comparing judgments for dyads where the velocity profiles of both actors were smooth with dyads in which one actor jittered around the other actor's movement trajectory. In Experiment 2, we investigated the shape of actors' velocity profiles as a cue to velocity based synchrony by comparing dyads where both actors' velocity profiles were the same shape in terms of kurtosis with dyads where the actors had different velocity profiles in terms of kurtosis.

The advantages of this approach over using natural movements was that it gave us complete experimental control over the kinematic cues that participants were presented with, and we could ensure that participants were not presented with other cues that may influence their responses. Moreover, we could manipulate jitter and kurtosis separately, in order to investigate the relative contributions that these velocity-based cues have on the perception of jointly improvised performances.

As in previous studies, we expected that interval based synchrony would have an effect on participants' judgements of coordination, affiliation and aesthetics. Also, given that expert performance in joint improvisation is characterized by velocity based synchrony, we expected this type of synchrony to play a particularly important role in driving participants' judgements, with regards to our three questions. Moreover, we predicted that interval based synchrony would act as a precursor for velocity based synchrony, with the effects of velocity based synchrony on judgements of coordination, affiliation and aesthetics would be stronger when performers' could synchronize their movement intervals. This was because it is likely that in a real life interaction, at least when synchronizing the same movements, performers would not be able to align their velocity profiles if their movement intervals were misaligned. Moreover, we expected that the degree of alignment between two velocity profiles would be more visible when the performers' movement intervals were also aligned.

4.2 Experiment 1: Jitter

4.2.1 Method

Apparatus and Stimuli

In order to gather movement data to generate our animations, we carried out a pilot in which we recorded individual participants moving the sliders of an adapted version of the mirror game (Noy et al. 2011) from side to side. We used a wooden box (80cm x 40cm x 10

cm), with two sliders (8cm x 4cm) that moved smoothly across horizontal tracks (60cm x 1cm), which were spaced 20cm apart. These dimensions were as similar as possible to the original mirror game apparatus used by Noy and colleagues. We recorded movement data from six participants using a Polhemus G4 motion tracker, by attaching the sensors of this system to the top of the sliders (see Figure 1 for a sketch of the setup). The participants were instructed to make either short, medium or long movements to the left and right, at a speed comfortable for them. We aggregated the movement trajectories of these movements across all participants, and took the mean length, duration and peak velocity for each of these aggregated movements. We then used these aggregated movements in order to generate our experimental stimuli. We generated twelve unique twelve-element sequences out of short, medium and long strokes to the left and to the right. Each of these sequences was composed of four short, medium and long movements in a random order, alternating between left and right movements.

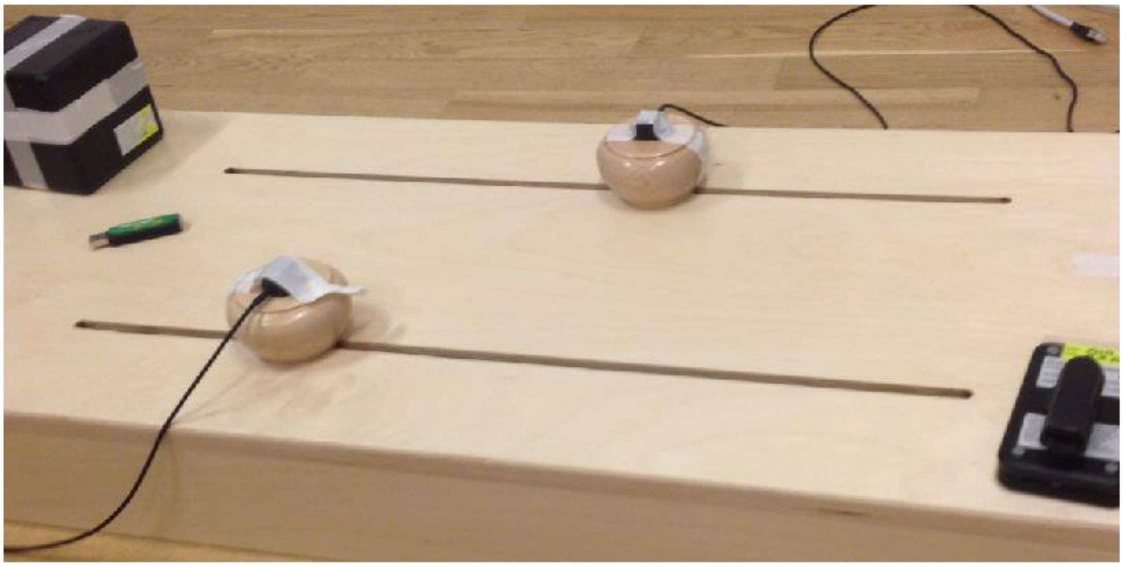

Figure 1: A photo showing the version of the mirror game that we used in order to record movement data for our stimuli.

From these sequences, we created artificial dyads who moved together with varying levels of interval based synchrony. To create the impression that the perceived dyads intended to align the end state of their movements, we used the same movement sequence for both members of the virtual dyad. For each of these dyads we created high asynchrony (low interval based synchrony) and low asynchrony (high interval based asynchrony) interactions, with high jitter (low velocity based synchrony) and low jitter (high velocity based synchrony) interactions for each synchrony type (see Figure 2 for an example of two movements for each of these types of interaction). This resulted in 48 unique interactions.

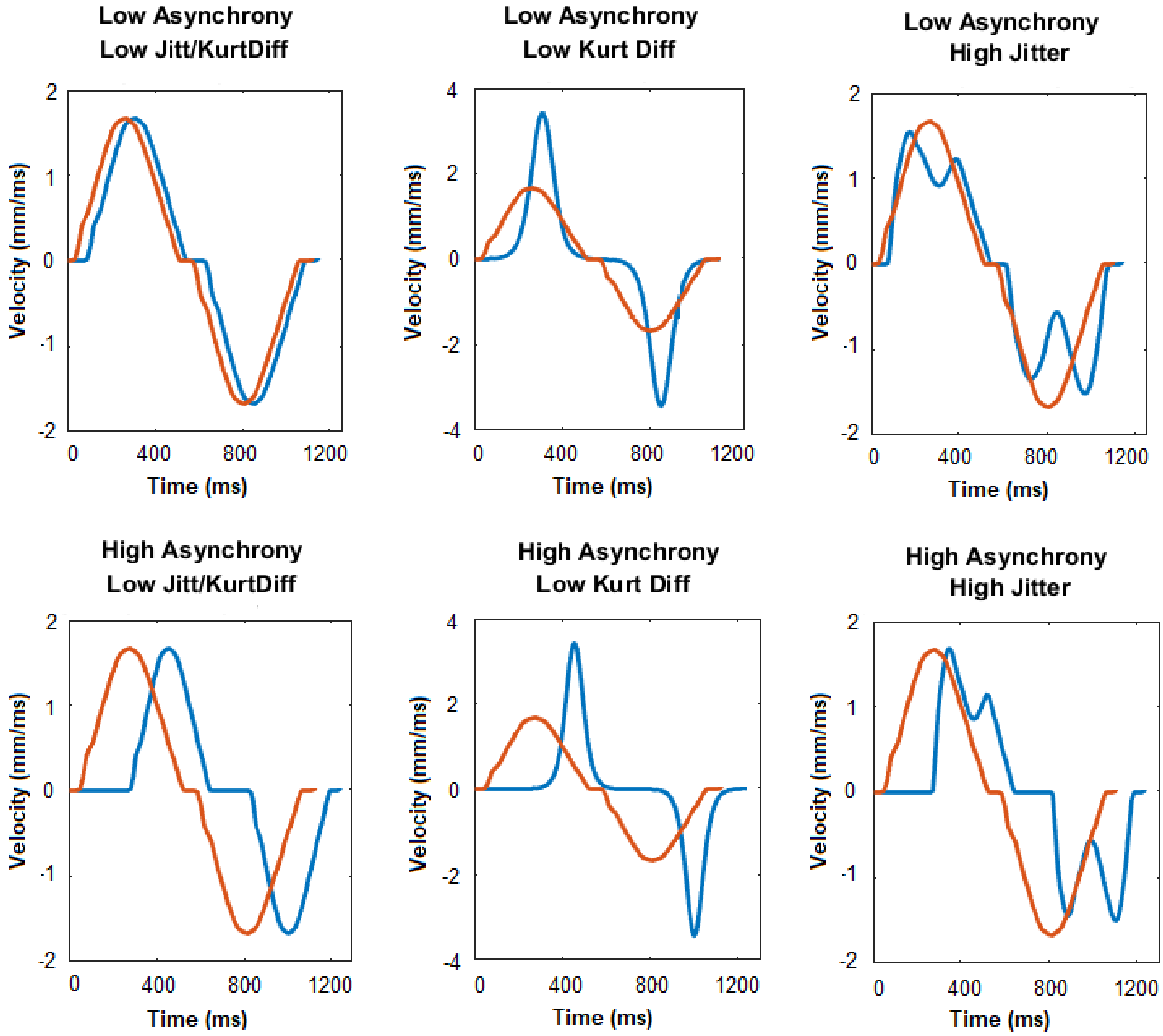

Figure 2: Example of the velocity profiles for two strokes, for each of our types of interaction.

We created high asynchrony sequences with a mean absolute asynchrony of 200ms by increasing the onset of one of the virtual performer's movements by a value between 100-300ms, and low asynchrony movements with an absolute asynchrony of 40ms by increasing the onset of one of the player's movements by a value between 20-60ms. In order to minimize any perception of leader-follower dynamics due to asynchrony, we ensured that both performers lagged approximately equally for each of the interactions by increasing each of the members of the dyads movement onset times an equal amount of times within each interaction. We also constrained our asynchrony generation by ensuring that the mean signed asynchrony for each interaction was below 20ms.

We created low jitter interactions by having both performers move with the same smooth velocity profile, and created high jitter interaction by making one performers velocity profiles 'jitter' around the other performer's velocity profile. Starting from the same smooth velocity profile for both performers, we added a 2-3hz sine wave with a random amplitude over one of the 'followers' (the virtual performer who was lagging at this stroke) velocity profile, which created jitter by causing temporary increases of decreases in the performer's acceleration. This created the appearance of a follower's movement jittering around a leader's movement. We chose to use a random amplitude rather than a fixed amplitude as the constant amplitude of the sine wave would have created a predictable increase and decrease of acceleration of the same magnitude.

We animated the interaction with a red dot and a green dot (with a diameter of 50 pixels) representing the sliders, and two tracks (1200 pixels x 20 pixels) representing the mirror game tracks. The red dot and green dot were superimposed over the tracks, and moved in one dimension, from side to side on the basis of the transformed dyad data. All the stimuli were animated in MATLAB psychophysics toolbox on a 1920x1080 pixel display, at a frame rate of 60Hz.

Procedure

Participants were first familiarized with the original mirror game apparatus, and given a detailed explanation with regards to how the apparatus works. We then explained the mirror game to them, explaining that we had two individuals sitting face to face to move one slider each from side to side, in order to synchronize, create interesting patterns together, and to enjoy playing. Participants were then familiarized with our experimental stimuli, and given the chance to ask any questions before completing practice trials. Participants completed two practice trials, and then were given another opportunity to ask questions before the full experiment began.

In each trial of the main experiment participants were first presented with the prompt screen, which would provide them with advance information about the judgement that they would be required to make. They would be randomly presented with one of three prompts, which were; 'How much do the two individuals LIKE each other'; 'How COORDINATED are the two individuals'; 'How INTERESTING AND BEAUTIFUL do you find the interaction between the two individuals'.

They were then presented with a fixation cross for 500ms, and then viewed a performance consisting of a virtual dyad engaging in one of the four interaction types (lasting around 8 seconds). They were then presented with the question screen, which instructed them to rate the dyad on a 1-6 likert scale, with respect to the prompt that they had received before viewing the performance. There was also a reminder of the question above the likert scale. Responses were made using left and right button presses and then pressing enter to choose the desired response (see Figure 3 for an example of one trial).

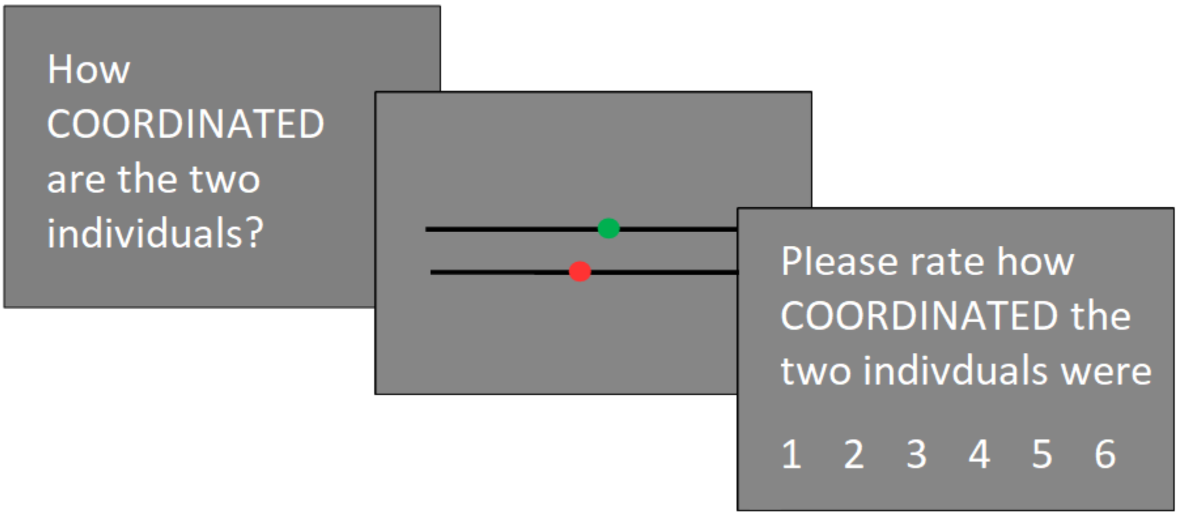

Figure 3: An illustration of one trial. The first frame is the prompt screen, the second screen is an illustration of the interaction, and the third screen is the question screen.

Participants completed this procedure for each of the 48 generated interactions, for the three questions, resulting in 144 trials. The order of interactions and question type was fully randomized.

4.2.2 Results

We wanted to investigate how participants' ratings of liking, coordination, and aesthetics depend on our two factors (asynchrony and jitter). To this end, we carried out three separate 2 x 2 ANOVAs with asynchrony (low, high) and jitter (absent, present) as within-subjects factors, for each of the three questions.

Coordination

The 2 x 2 ANOVA for the coordination ratings (see Figure 4) revealed a significant main effect of asynchrony $F(1,23) = 150.26$, $p < .001$, $\eta p2 = .87$, with coordination ratings being higher for low asynchrony trials than for high asynchrony trials (all pairwise, $p < .001$). There was a significant main effect of jitter, $F(1,23) = 82.13$, $p < .001$, $\eta p2 = .78$, with

coordination ratings being higher for jitter absent trials than for jitter present trials (all pairwise, $p < .001$). We also found an interaction between asynchrony and jitter, $F(1,23) = 27.01$, $p < .001$, $\eta p2 = .54$, with the effects of jitter being larger for low asynchrony trials than for high asynchrony trials.

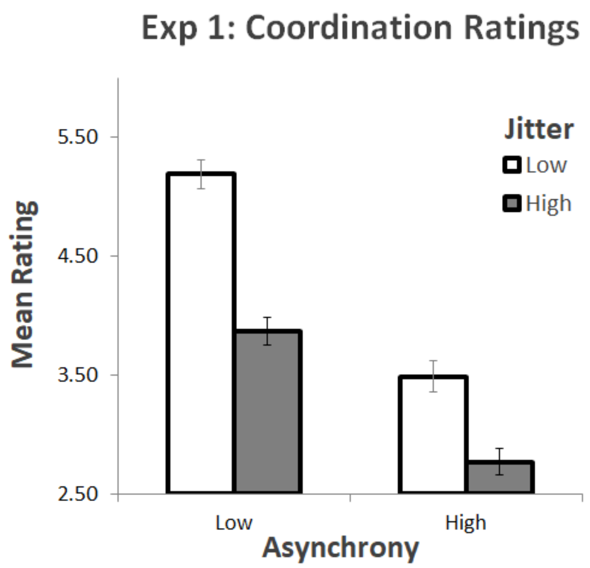

Figure 4: Interaction between jitter and asynchrony for coordination ratings. Error bars represent +/- 1 SEM.

Affiliation

Our 2 x 2 ANOVA for the liking ratings (see Figure 5) revealed a significant main effect of asynchrony, $F(1,23) = 33.15$, $p < .001$, $\eta p2 = .59$, with liking ratings being higher for low asynchrony trials than for high asynchrony trials (all bonferroni corrected pairwise comparisons, $p < .001$). There was also a main effect of jitter, $F(1,23) = 51.92$, $p < .001$, $\eta p2 = .69$, with liking ratings being significantly higher for jitter absent trials than for jitter present trials (all bonferroni corrected pairwise comparisons, $p < .001$). There was also an interaction between asynchrony and jitter, $F(1,23) = 16.61$, $p < .001$, $\eta p2 = .42$, with the effects of jitter being larger for low asynchrony trials than for high asynchrony trials.

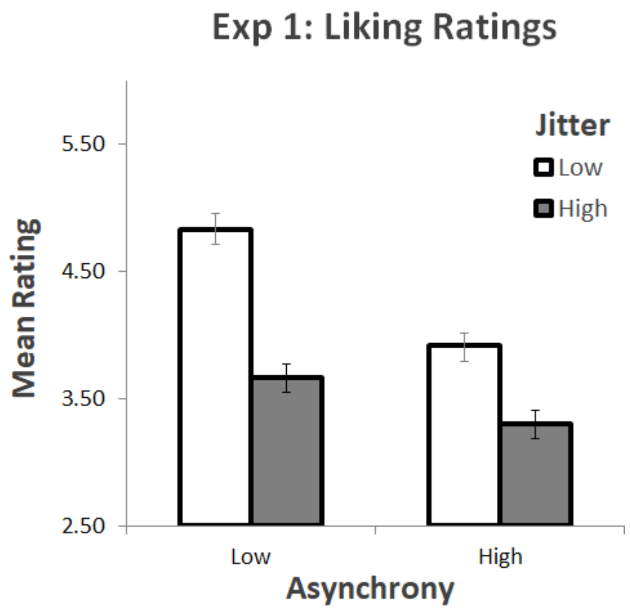

Figure 5: Interaction between jitter and asynchrony for ratings of liking. Error bars represent +/- 1 SEM.

Aesthetics

The 2 x 2 ANOVA for aesthetics ratings (see Figure 6) yielded a significant main effect of asynchrony, $F(1,23) = 10.96$, $p = .003$, $\eta p2 = .32$, with aesthetic ratings being higher for low asynchrony trials than for high asynchrony trials (pairwise comparisons, $p = .003$). We also found a significant main effect of jitter, $F(1,23) = 27.65$, $p < .001$, $\eta p2 = .55$, with aesthetic ratings being higher for jitter absent trials than for jitter present trials (pairwise comparisons, $p < .001$). There was also an interaction between asynchrony and jitter, $F(1,23) = 8.81$, $p = .007$, $\eta p2 = .28$, with the effects of jitter being larger for low asynchrony trials than for high asynchrony trials.

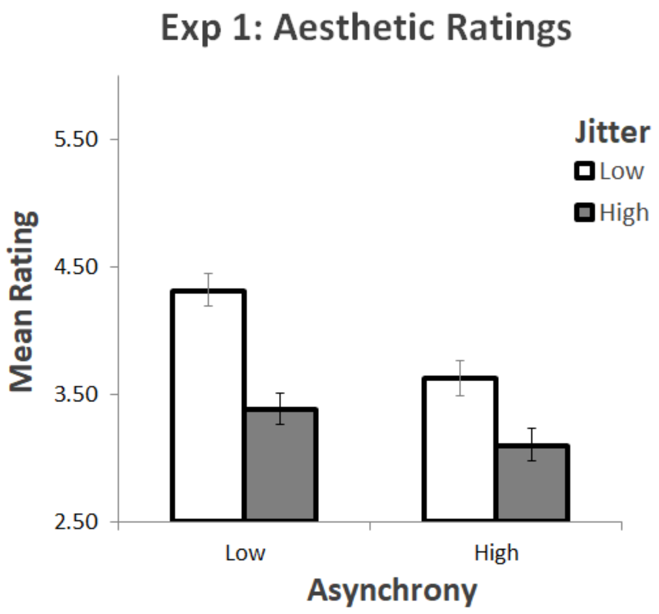

Figure 6: Interaction between jitter and asynchrony for aesthetics ratings. Error bars represent +/- 1 SEM.

4.2.3 Experiment 1: Discussion

Our first aim was to compare the effects of interval based synchrony and velocity based synchrony on participants' judgements of coordination between two actors. We found that participants used both interval based synchrony (asynchrony) and velocity based synchrony (jitter) in order to make judgements about the level of coordination between two actors. Moreover, the interaction between synchrony and jitter demonstrates that perception of interval based synchrony is a precondition to perceiving velocity based synchrony, with performers being rated as most coordinated when they can successfully coordinate both on the level of the timing of their movement intervals, and on the level of their velocity profiles.

Our second aim was to compare the effects that interval based synchrony and velocity based synchrony have on perceived affiliation between two actors. Although both interval based synchrony and velocity based synchrony informed participants' judgements of affiliation between two performers, ratings of affiliation were highest when performers displayed both interval based and velocity based synchrony. This finding indicates that the

combination of interval and velocity based synchrony leads to a perception of a deeper or more profound affiliation, compared to interval based synchrony alone.

Our third aim was to compare the effects of interval based synchrony and velocity based synchrony on an observer's aesthetic experience of a performance. The interaction between jitter and asynchrony indicates that people's aesthetic experience of a performance is strongest when performers are aligned both in terms of the intervals of their movements and the velocity profiles of their movements. This demonstrates that relational movement cues that affect aesthetic experience go beyond stable phase relationship, with alignment in a fine grained fashion leading to a more profound aesthetic experience.

4.3 Experiment 2: Kurtosis

4.3.1 Method

Apparatus and Stimuli

The apparatus and stimuli were exactly the same as in Experiment 1, but instead of manipulating jitter in order to create low and high velocity based synchrony we manipulated the kurtosis of the velocity profiles of the movements in order to give the two virtual players matching or mismatching velocity profile shapes.

We manipulated the kurtosis of the movements by generating bell curves with different kurtosis (how flat or sharp the peak of the distribution is), and transforming the velocity profiles of the original movements on the basis of these curves. For our low kurtosis difference (high velocity based synchrony) condition, we had both virtual players move with a velocity profile with a steady 'sine wave' like low kurtosis (2.2), which was based on the values from Hart et al. (2014). This created a relatively linear acceleration and deceleration phase of the movement, with a relatively constant acceleration. For our high kurtosis difference condition (low velocity based synchrony) we had one virtual player move with a sharp 'bell shaped' like velocity profile (kurtosis of 3), with a more exponential and less

constant acceleration and deceleration phase, and the other virtual player moving with a low kurtosis (2.2), leading to a steady acceleration and deceleration phase. For each of these dyads we created high asynchrony and low asynchrony interactions, with high kurtosis difference and low kurtosis difference (see Figure 2 for an example of two movements for each of these types of interaction). This resulted in 48 unique interactions.

Procedure

The procedure was the same as in Experiment 1.

4.3.2 Results

Like in Experiment 1, we carried out a 2 x 2 ANOVA with asynchrony (high, low) and kurtosis difference (low, high) as within-subjects factors, for each of the three questions.

Coordination

Our 2 x 2 ANOVA for the coordination ratings (see Figure 7) revealed a significant main effect of asynchrony $F(1,23) = 83.82$, $p < .001$, $\eta p2 = .79$, with coordination ratings being higher for low asynchrony trials than for high asynchrony trials (all pairwise, $p < .001$). There was a significant main effect of kurtosis, $F(1,23) = 41.19$, $p < .001$, $\eta p2 = .64$, with coordination ratings being higher for kurtosis absent trials than for kurtosis present trials (all pairwise, $p < .001$). We also found an interaction between asynchrony and kurtosis, $F(1,23) = 17.71$, $p < .001$, $\eta p2 = .44$ with the effects of kurtosis being higher for low asynchrony trials than for high asynchrony trials.

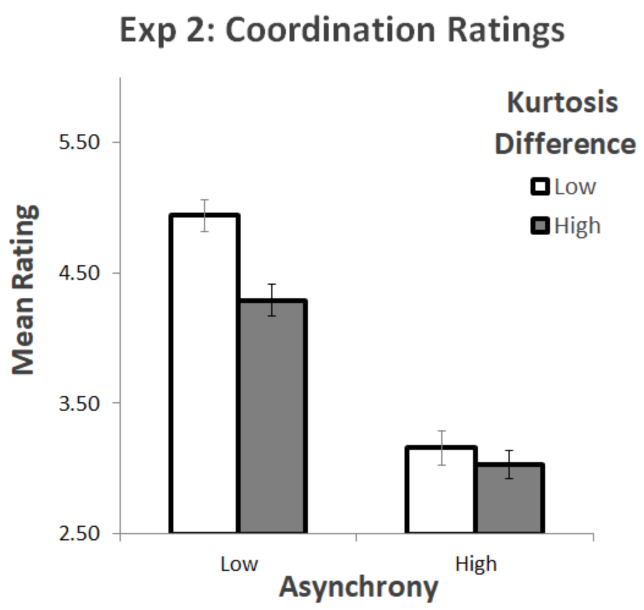

Figure 7: Interaction between kurtosis difference and asynchrony for ratings of coordination. Error bars represent +/- 1 SEM.

Affiliation

The 2 x 2 ANOVA for the liking ratings (see Figure 8) revealed a significant main effect of asynchrony, $F(1,23) = 46.13$, $p < .001$, $\eta p2 = .67$, with liking ratings being higher for low asynchrony trial than for high asynchrony trials (all pairwise, $p < .001$). There was also a main effect of kurtosis, $F(1,23) = 22.73$, $p < .001$,$\eta p2 = .5$, with liking ratings being significantly higher for kurtosis absent trials than for kurtosis present trials (all pairwise, $p < .001$. There was also an interaction between asynchrony and kurtosis, $F(1,23) = 8.45$, $p = .008$, $\eta p2 = .27$, with the effects of kurtosis being higher for low asynchrony trials than for high asynchrony trials.

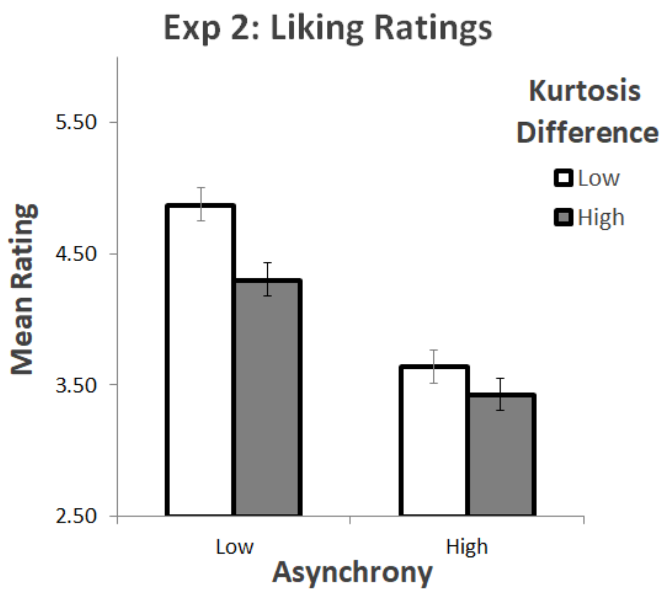

Figure 8: Interaction between kurtosis difference and asynchrony for ratings of liking. Error bars represent +/- 1 SEM.

Aesthetics

The 2 x 2 ANOVA for aesthetics ratings (see Figure 9) yielded a significant main effect of asynchrony, $F(1,23) = 10.32$, $p = .004$, $\eta p2 = .31$, with aesthetic ratings being higher for low asynchrony trials than for high asynchrony trials (pairwise comparisons, $p = .004$). We also found a significant main effect of kurtosis, $F(1,23) = 20.16$, $p < .001$, $\eta p2 = .47$, with aesthetic ratings being higher for kurtosis absent trials than for kurtosis present trials (all pairwise, $p < .001$). There was no interaction between asynchrony and kurtosis, $F(1,23) = 1.32$, $p = .26$, $\eta p2 = .05$.

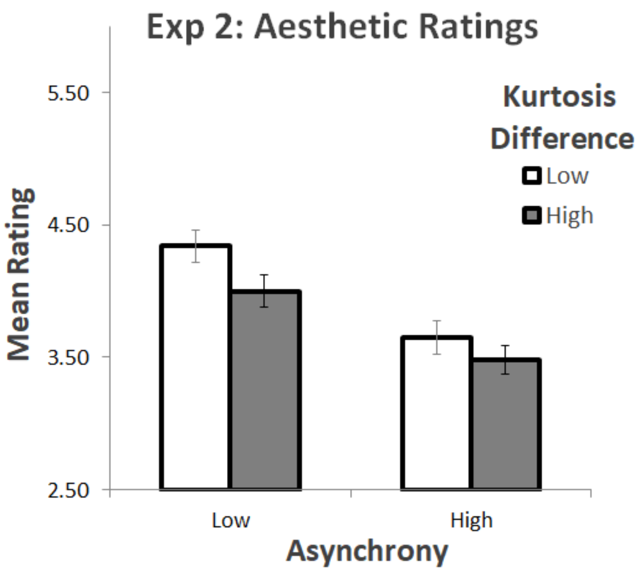

Figure 9: Interaction between kurtosis difference and asynchrony for ratings of aesthetics. Error bars represent +/- 1 SEM.

4.3.3 Experiment 2: Discussion

Experiment 2 aimed to replicate and generalize the findings of Experiment 1. Rather than jitter, we used the difference in the kurtosis of the performers' velocity profile as a cue to velocity based synchrony. We found that participants used both synchrony and kurtosis in order to inform their judgements about how coordinated two actors were, as well as how much the two actors like each other. Moreover, the interaction between synchrony and kurtosis suggests that interval based synchrony is a precondition for velocity based synchrony, with performances that are synchronized both in terms of the timing of the end points of their movements and in terms of their velocity profiles yielding highest coordination ratings, and the perception of strongest affiliation. With regards to aesthetic experience there was no interaction between asynchrony and kurtosis for ratings of aesthetic experience. Although this does not provide evidence that perception of asynchrony may not act as a precondition for the perception of kurtosis in this context, aesthetic experience was still most profound when both interval based and velocity based synchronization are present.

4.4 General Discussion

The current study aimed to investigate the effects of cues to interval based synchrony and velocity based synchrony on third party judgements of a joint performance. To recap, interval based synchrony consists of actors aligning their movements in order to match the timing of the end state of their movements (e.g. tapping), whereas velocity based synchrony consists of actors aligning their movements in order to match their whole velocity profile (e.g. coordinating dance moves together).

Our first question aimed to investigate whether velocity based synchrony leads to observers judging performers as more coordinated, compared to performers who are only engaging in interval based synchrony. In line with our predictions, we found that performers engaged in velocity based synchrony as well as interval based synchrony appear more coordinated than performers engaged in interval based synchrony only. It is likely that observers perceive velocity based synchrony as being more coordinated than interval based synchrony because the former constitutes a more fine grained mode of coordination, which occurs on a frame by frame temporal scale, compared to interval based synchrony which occurs on a movement by movement temporal scale. Indeed, Noy et al. (2011) proposed that smooth interactions in which velocity profiles are synchronized and jittery interactions in which velocity profiles are not synchronized can be characterized differently with regards to the mechanisms by which the dyad are aligning their movements. Jittery interactions can be described in terms of one predictive-reactive controller which aims to track and adjust to the leaders' movements, whereas smooth interaction can be characterized as mirrored predictive-reactive controllers which are aligned with a continuous bidirectional information flow. Because velocity based synchrony reflects a more sophisticated and fine grained mode of coupling between two performers, it leads these performers to be perceived as more coordinated than those engaged in interval based synchrony.

Our second question concerned the effects that interval based and velocity based synchrony have on an observer’s perception of the level of affiliation between two performers. As we predicted, we found that performers who were engaged in velocity based synchrony appeared as liking each other more than those who were engaged in interval based synchrony. The effect of interpersonal synchrony on the perception of affiliation has been proposed to reflect the appearance of 'social connectedness' that synchrony creates, with actors appearing to be moving as a cohesive unit (Marsh, Schmidt, Baron & Richardson, 2006; Miles et al. 2009).It is possible that velocity based synchronization yields the strongest judgements of liking because the finer grained synchrony produced by performers able to align their velocity profiles makes these performers appear as even more connected than those who are only able to synchronize their movements at the end points.

Alternatively, velocity based synchrony may lead to an increased perception of affiliation due to the fact that the performers are able to reach a more intricate form of synchrony, compared to simply holding a phase relationship. Velocity based synchrony requires the ability to continuously align with a co-performer in order to achieve coordination (Noy et al. 2011). The ability that dyad has to stay continuously aligned and synchronized throughout a movement may reflect a deeper understanding of each other’s movements, thus a deeper rapport because the performers appear as 'being on the same page'. Our third question aimed to investigate the role that interval based synchrony and velocity based synchrony play in an observer’s aesthetic experience of a performance. We found that observers found performances containing velocity based synchrony more aesthetically pleasing than performances containing only interval based synchrony.

Although research into aesthetic experience has traditionally focused on static visual cues (McManus et al. 1981; Jacobsen et al. 2005), and dynamic movement cues from individual performances (Calvo-Merino et al. 2008), a recent study by Vicary et al. (2017)

demonstrated that relational movement cues in the form of interval based synchrony play a role in aesthetic experience. Our findings extend this research by demonstrating that in addition to interval based synchrony velocity based synchrony also plays a role in aesthetic experience. Vicary and colleagues explained the relationship between interval based synchrony and aesthetic experience by suggesting that synchronized behaviour signals coalitional strength to an audience. In this view, synchrony is a cue to formidability and group strength (Fessler & Holbrook, 2016), with this show of group strength making a performance aesthetically pleasing. Considering that synchronizing on a frame by frame timescale requires more skill and reflects a greater level of expertise than synchronizing on an interval timescale (Noy et al. 2011; Hart et al. 2014), and that velocity based synchrony projects a deeper level of affiliation compared to interval based synchrony, it is possible that velocity based synchrony signals greater coalitional strength than interval based synchrony. This greater coalitional strength associated with velocity based synchrony could then lead to a more profound aesthetic experience compared to the relatively weaker coalitional strength signalled by interval based synchrony.

An alternative account of how aesthetic experience arises from movement cues suggests that aesthetic experiences when observing dance performances can be explained by a willingness to integrate these spectacular and impressive movements into our own motor system (Cross et al. 2011; Kirsch, Urgesi & Cross, 2016). This is supported by evidence which suggests that perceived difficulty of an action predicts the aesthetic experience when watching dance (Cross, Kirsch, Ticini & Schutz-Bosbach, 2011). Moreover, it has also been found that watching actions that an observer cannot perform lead to increased activity in mirror neuron areas, compared to actions that can be performed. Taken together, these studies can be taken as evidence that aesthetic experience of dance can be explained at least in part by greater motor activity, due to an attempt to assimilate the observed movements onto one’s

own motor repertoire. This account of the relationship between movement cues and aesthetic experience may also explain our findings that velocity based synchrony yields a more profound aesthetic experience than mere interval based synchrony. As mentioned previously, velocity based synchrony reflects a higher level of expertise with regards to dance and improvisation than interval based synchrony, and can be seen as a more sophisticated and intricate mode of coordination (Noy et al. 2011). Considering this, the fact that velocity based synchrony leads to a stronger aesthetic experience than interval based synchrony could be due to observers' desire to assimilate this expert like manner of coordinating onto their own motor repertoires.

With regards to explaining our findings, we have provided both perceptual and sensorimotor explanations. On one hand, velocity-based synchrony could lead to an increased perception of affiliation, because this more fine grained level of synchrony leads to a greater appearance of 'social connectedness' (Marsh et al. 2006; Miles et al. 2009), and signals greater coalitional strength, thus providing a rewarding and aesthetically pleasing display (Fessler & Holbrook, 2016; Vicary et al. 2017). Attributions of causality and animacy on the basis of relational cues are assumed to be the result of specialized perceptual systems which contain innate assumptions about the properties of the motion (Scholl & Tremoulet, 2000). Perhaps attributions of affiliation and beauty are the result of perceptual systems which are specialized for making inferences about interpersonal dynamics on the basis of relational movement cues which specify how connected and cohesive multiple actors are.

On the other hand, velocity based synchrony could lead to a greater level of perceived affiliation because two performers who are able to align their movements on a fine grained temporal scale are engaging in a more intricate mode of coordination, thus sharing more information with each other and being on the same page. This is arousing and aesthetically pleasing for observers due to their attempts to assimilate actions associated with this more

sophisticated mode of coordination onto their own motor repertoire. This explanation supposes that relational movement cues are processed by the motor system, in order to allow for observers to understand the observed actions through motor simulation, by matching (or attempting to match) these actions onto their own motor repertoire.

Future research should aim to investigate the relative contributions that specialized perceptual systems, and motor simulation have with regards to processing relational cues. Investigating the extent to which this ability depends on either perceptual or motor experience of synchrony, particularly in infants, may yield interesting insights with regards to the role that the perceptual system and the motor system play with regards to how relational cues influence the perception of affiliation and aesthetic experience.

One thing to consider is the role that intentionality plays in how interval and velocity based synchrony differ in their contributions to aesthetic experience. A study by Eskenazi and colleagues (2015) demonstrated that observing actors coordinate their actions when they have a shared intention to coordinate reflects activation in the human reward system, compared to when two actors coordinate incidentally, suggesting that observers find observing those intentionally synchronize as rewarding. It is possible that interval based synchrony and velocity based synchrony can have different levels of intentionality, with velocity based synchrony requiring those to intentionally synchronize their movements, whereas interval based synchrony can be achieved incidentally. Although in our study, the virtual performers were instructed to intentionally synchronize, future research should investigate how people attribute shared intentionality to those either aligning their movement intervals or their velocity profiles. One could then investigate how shared intentionality attribution mediates the relationship between interval based and velocity based synchrony and the perception of affiliation and aesthetic experience.

Another interesting avenue for future research concerns how the ability to enter and exit from synchrony affects the perception of a joint performance. As well as being able to maintain synchrony, the ability to enter and exit synchrony at will also reflects affiliation between two performers, as it reflects control over the interaction as well as the ability to effectively repair ruptures in an interaction (Dahan, Noy, Hart, Mayo & Alon, 2016). With regards to both liking and aesthetics, one could make the prediction that the ability to enter and exit synchrony is more important than being continuously synchronized. This could also differ depending on what type of judgement participants are required to make. For example, judgements of coordination could be strongest for continuously aligned performances, whereas performances which continuously enter and exit synchrony could yield stronger ratings of liking due to signalling the ability to effectively repair ruptures in an interaction. Moreover, judgements of aesthetics could be stronger for interactions that enter and exit synchrony because rather than being continuously aligned as the performers are displaying more skill by being able to control exactly when they align with each other.

Chapter 5: General Discussion

This thesis aimed to investigate the production and perception of movement cues produced in social interactions, and identified three questions which were addressed empirically in three studies.

Our first study set out to investigate informative action modulations produced in coordination and teaching contexts. Specifically, we wanted to investigate whether kinematic modulations in coordination and teaching are purpose general for communication (e.g. ostension), or whether people modulate their actions differently in these contexts, depending on the information required by a co-actor to achieve the joint goal (to predict a partner's actions to coordinate, or to understand the structure of an action to learn). The first key finding was that informative action modulations in order to enhance spatial and/or temporal prediction in coordinated joint actions resulted in different kinematic signatures than informative action modulations in order to highlight the structure of an action sequence in teaching. Moreover, we also found that particular action modulations were used both to support teaching through demonstration, and to support coordinated joint action, suggesting that informative action modulations in coordinated joint actions may also support learning in joint action contexts.

Our second study aimed to investigate whether or not people could detect an actor's informative intentions associated with joint action and teaching, on the basis of movement cues. We found that people could reliably discriminate between actions performed with only instrumental intentions, and actions performed with informative intentions on the basis of movement cues such as movement height and the timing of the action. Moreover,

participants could also use movement cues such as movement height and timing in order to discriminate between actions performed with the intention to teach through demonstration, and actions performed with the intention to coordinate in a joint action. This study demonstrated that observers can understand high level informative intentions on the basis of low level movement cues.

Our final study moved beyond movement cues produced by individuals and investigated third person perception of movement cues produced by dyads. We aimed to investigate how movement cues that reflect interval based and velocity based modes of synchrony affect perception of the relations between a dyad. We also aimed to investigate how these movement cues can drive our aesthetic experience when watching a performance. Because cues to velocity based synchrony reflect a more sophisticated and intricate mode of coordination, performances containing these cues yielded greater judgements of coordination and affiliation, and had a more profound effect on the observers' aesthetic experience.

The current chapter will aim to discuss some of the wider implications of the studies included in this thesis, as well as some of the questions that our research has raised. We will first discuss how our findings can inform the debates around the extent to which mental states can be directly perceived from an actor's movements. Secondly, we will discuss how our findings with regards to informative and relational movement cues can advance our understanding of teaching and learning, both on the interpersonal level and on the cultural level. Finally, we will discuss some of the ways in which our research can contribute to the field of human-robot interaction.

5.1 Direct perception: Can people directly perceive mental states from informative and relational movement cues?

The direct social perception theory posits that mental states can be directly perceived from an actor's behaviour (Gallagher, 2008). This theory suggests that inferential processes are not required to derive mental states from behaviour. Rather, the perceptual system is considered 'smart' and sophisticated enough to directly see the mental states of an actor, considering the rich meaning reflected in human behaviour. This is considered possible due to the perceptual system being endowed with various tools such as an automatic motor resonance system (Gallagher, 2008; Gallese, 2005) and the ability to automatically perceive 'affordances', which are features of an object that allow for it to be interacted with or manipulated (Gallagher, 2008; Gibson, 1979).

Recent work by Becchio, Koul, Ansuini, Bertone and Cavallo (2017) proposed that the ability to directly perceive mental states depends on the degree to which an action contains movement features that predict a given mental state. With regards to instrumental intentions, whether an actor intends to grasp a bottle in order to pour or to drink can be directly perceived due to the movement features associated with these two instrumental intentions. Considering the findings from our current set of studies, can observers directly perceive informative intentions in the same way that they can directly perceive instrumental intentions, or does understanding these higher-level intentions require additional inferential processes? Likewise, with regards to relations between actors, can people also directly perceive the relations between multiple agents on the basis of relational movement cues, or does this also require additional inferential processing?

Contrary to suggestions that people cannot detect anything more than instrumental intentions on the basis of movement features (Jacob & Jeannerod, 2005), we demonstrate that people can successfully discriminate between different informative intentions on the

basis of movement cues. Does this mean that observers can directly perceive informative intentions on the basis of these cues? Perhaps the same processes involved in detecting instrumental intentions are also involved in detecting informative intentions. Because the movements of an actor reflect their instrumental intentions in the form of movement cues, it is possible that movement cues associated with coordination and teaching also render these informative intentions observable, meaning that they can be directly perceived. For example, a slow-down in the acceleration phase of a movement could be directly perceived as an attempt to coordinate a joint action, because this movement feature predicts being in a coordinated interaction and a slow-down in acceleration affords coordination.

Another possibility is that observing a sub-optimally executed action leads observers to recruit additional inferential processes in order to derive the social intention underlying the action, with observers inferring whether the action was executed with the intention to coordinate, or with the intention to teach on the basis of their knowledge of what is required in order to teach or to coordinate.

Future research should aim to investigate whether an informative intention can be directly perceived or whether additional inference is needed in order to discriminate between different informative intentions. Perhaps investigating patterns of neural activity evoked by observing informative and instrumental intentions may help us understand the extent to which understanding informative intentions from kinematic cues requires the recruitment of additional inferential processes. For example, it has been demonstrated that in addition to the mirror neuron areas evoked by instrumental reach to grasp movements (specifically the IFL and the IFG), reach to grasp movements which contain social intentions to cooperate or compete evoke activity in the mentalizing network (i.e. the TPJ and the mPFC), suggesting that the understanding of these intentions requires more than just motor resonance, and

recruits additional mentalizing processes (Becchio et al. 2012). Could it be possible that observing actions that contain informative intentions also lead to activation in the mentalizing network? Perhaps this finding could be taken as evidence that motor resonance alone is not sufficient for detecting an informative intention from an observed action, with the understanding that an actor has the intention to inform requiring one to make inferences about the actor's mental state.

With regards to the perception of relational cues, whether they can also be directly perceived in the same manner as individual movement cues is also an open question. Can attitudes such as how much two actors like each other be directly perceived from the kinematics of their interaction? Judgements of rapport and affiliation are said to be the result of coordination creating the appearance of 'social connectedness' between actors (Miles et al. 2009). This can be taken to suggest that affiliation effectively has a kinematic signature that characterizes the relations between two actors, which could allow affiliation to be detected on the basis of this kinematic signature. Considering this, it is possible that affiliation and perhaps other relations between actors can be directly perceived on the basis of relational cues.

However, there is also evidence that speaks against the idea that co-actor's attitudes towards each other can be directly perceived. A study by Lakens and Stel (2011) found that when actors synchronized spontaneously, ratings of rapport and entitativity increased as a function of synchronization. However, when actors were instructed to synchronize, entitativity ratings increased as a function of synchrony but rapport ratings did not, suggesting that the relationship between synchrony and rapport is not purely perceptual, but was also inferred using contextual information. This finding points to the possibility that relations between actors cannot be directly perceived and require additional inferential

processes. Other sources of information that are non-perceptual such as knowledge of the joint task, or knowledge of the wider context (e.g. whether they have been instructed to synchronize) may also be required in order to make attributions about two actors on the basis of the relational cues produced.

Considering whether or not the intention to synchronize is directly observable could also yield interesting questions. Coordination can either be emergent in which actors automatically fall into synchrony, or it can be planned in which people intentionally aim to synchronize their movements with one another (Knoblich, Butterfill & Sebanz, 2011). To our knowledge, there are no studies that have investigated the different kinematic signatures of emergent and planned coordination. So it is currently unknown whether people can use relational movement cues in order to directly perceive whether or not two actors have an intention to synchronize. A study which first characterizes the kinematics of planned and emergent synchrony, and then investigates whether participants can discriminate between these types of synchrony on the basis of these kinematics could yield interesting results. One possibility is that falling into synchrony simply leads to interval based synchrony with timing of the end states of a movement synchronizing, but in order to achieve velocity based synchrony, actors are required to intentionally synchronize, perhaps rendering the intention to synchronize visible through the dyads' kinematics.

In sum, we believe that one challenge for the direct social perception theory concerns explaining how higher-level mental states can be derived from kinematic cues. Beyond instrumental intentions such as the ones investigated by Becchio and colleagues, movements carry information that reflect a whole host of higher level social intentions and mental states. Whether these movement cues render these types of mental states observable, or whether

additional inferential processes are needed in order to derive mental states from these movement cues is still open for debate.

5.2 Teaching and Learning in Joint Action: From individuals to societies

It goes without saying that teaching and learning play a central role in human social life, with many of the practices that we carry out with each other on a daily basis being learned from what has been communicated to us. Considering this, there has been little research investigating the role that sensorimotor communication plays in teaching and learning, both in small scale and large scale joint actions. This section aims to discuss some of the ways in which investigating sensorimotor communication in the context of teaching and learning can advance our understanding of how information is transmitted between individuals and societies. We will first discuss how sensorimotor communication can be used to scaffold learning, before exploring how sensorimotor communication can support learning through the haptic modality. We will then discuss how sensorimotor communication can also support teaching and learning both on a group, and on a societal level.

5.2.1 Scaffolding

Sensorimotor communication entails actions deviating from optimality in order to produce movement cues in order to support interpersonal coordination, depending on the knowledge state of co-actors (Pezzulo et al. 2013). When a co-actor does not need to be informed because they have all the information necessary to complete the task, movement cues are not produced (Candidi et al. 2015). Likewise, it has also been demonstrated that caregivers no longer produce motionese when their children can successfully execute the demonstrated action sequence (Fukuyama et al. 2015). This raises questions with regards to how actors use sensorimotor communication in order to scaffold a co-actor's learning.

When scaffolding a teacher is required to identify the learning requirements of a learner, in order to understand what needs to be taught, and what the learner can do competently (Vygostky, 1978; Wood, Bruner & Ross, 1976). The fact that sensorimotor communication can be used flexibly depending on the knowledge state of a co-actor suggests that it could be a useful tool with regards to scaffolding a co-actor's learning, either when teaching through demonstration, or when trying to coordinate in a repeated interaction as a leader and a follower. Future research could aim to investigate whether actors can use sensorimotor communication flexibly even within action sequences, in order to support learning of parts of an action sequence that an actor struggles with.

5.2.2 Haptic information sharing

Our research, as well as research into sensorimotor communication more generally has focused on movement cues which provide observers or co-actors with visual information. However, those engaged in joint action often share information haptically. Whether it be dancing or carrying a table, people are often haptically coupled when trying to coordinate a joint action. Although it has been suggested that sensorimotor communication can come through haptic channels (Pezzulo et al. 2018), there is little research investigating how movement cues reflecting different informative intentions are transmitted through the haptic channel. One study by Van der Wel, Knoblich and Sebanz (2011) demonstrated that when instructed to move a rod together at different speeds and frequencies, participants shared information haptically by increasing the force of their movements in order to create a force overlap between their movements and their co-actor's movements. This study demonstrates that haptic cues can be produced in order to stabilize joint action coordination, pointing to the possibility that informative intentions can be reflected in the haptics of an actor's movements. With regards to producing informative movement cues in order to enhance joint action coordination, further research should investigate how haptic

communication is used by actors who possess task knowledge to inform co-actors who do not possess task knowledge (e.g. when carrying a sofa, an actor with visual access to the room using haptic cues to signal direction to a co-actor who does not have visual access to the room). Such a study should consist of a leader who possesses task knowledge and a follower who does not possess task knowledge trying to coordinate their actions whilst being haptically coupled, and not possessing any other channel of communication.

In the same way that the intention to inform a co-actor visually yields differing kinematic signatures depending on one's informative intentions, is it possible that the intention to inform a co-actor haptically also leads to different kinematic signatures, depending on the type of informative intention the actor has (e.g. to coordinate or to teach)? For instance, when considering teaching and coordination, would teaching a co-actor a dance sequence by haptically guiding them through the key steps yield a different pattern of kinematics than trying to coordinate the same dance sequence with a co-actor haptically? Moreover, can an actor tell whether a co-actor is trying to teach or trying to coordinate from feeling these haptic cues? This could be investigated by using similar methods employed in the current research, but constraining the task in such a way that participants can only receive information haptically.

The comparison of haptic and visual cues with regards to detecting informative intentions also raises the question of whether or not these cues can be used by modalities other than the one that they were communicated through. With regards to using haptic cues to discriminate between teaching and coordination intentions, it could be interesting to investigate whether people can only discriminate between these different informative intentions when these cues are presented haptically, or whether they could also discriminate

between these different informative intentions visually, on the basis of the different kinematic signatures associated with these haptic cues.

With regards to teaching and learning through joint action, considering that movement cues produced in order to support joint action may also act as teaching cues (as demonstrated in Chapter 2), the role that informative haptic cues produced in order to coordinate play in teaching and learning is an open question (McEllin et al. 2018). Perhaps by being haptically coupled to a novice, an expert can help constrain the novice's action exploration space, constraining the many degrees of freedom a novice is faced with when learning a new action (Bernstein, 1967; McEllin et al. 2018). Moreover, considering the evidence that haptic guidance yields advantages over visual guidance when learning the temporal structure of an action sequence, being haptically coupled to a novice may allow the expert to transmit information about the timing of an action in ways that are not afforded by visual demonstration (Feygin, Keehner & Tendick, 2002; McEllin et al. 2018). Additional research investigating learning from haptically coupled joint action could be useful in order to understand how cues through the haptic channel can be used to scaffold a novice's learning, and even how these cues may yield advantages over visual demonstration.

5.2.3 Groups teaching groups

In addition to one to one joint actions, when learning how to conduct a joint action, often people are taught by groups, or have to participate with groups who are experts in their particular domain. For example, when taking a tango class, teachers will demonstrate the to be reproduced moves as a dyad. Likewise, students need to observe and reproduce the demonstrated movements as a dyad. This requires teachers to produce informative movement cues for their co-actors in order to successfully coordinate, but also to produce informative movement cues for the audience that they are demonstrating to. Moreover, not

only do these teaching cues need to convey information about how each student should execute their part of the dance, but the teaching cues also need to convey information about how these actions should relate to each other, such as when they should and should not be synchronized.

Considering the above, we believe that two questions can be raised out of this issue, which relate to the work done in this thesis. The first question concerns how actors modulate the kinematics of their actions differently when they have multiple informative intentions, compared to when they only have one informative intention. For example, when either teaching or coordinating, actors may rely on exaggerating their kinematics in order to produce informative cues for their co-actors, as demonstrated in chapter 2. However, when having to coordinate with a co-actor but also having to demonstrate to an audience of students, actors may use kinematic exaggeration in order to produce teaching cues for the audience and rely on other strategies such as reducing temporal variability (Vesper, van der Wel, Knoblich & Sebanz, 2011) in order to produce coordination cues for their co-actors. Alternatively, the fact that cues produced in teaching and coordination can be overlapping (also demonstrated in chapter 2), the same cues that are used to achieve coordination with a co-actor may also double up as learning cues for the audience.

The second question concerns how dyads teach other dyads about how their movements should relate to each other. As well as teaching how to execute an action, novices also need to be taught how to coordinate their actions with a partner. This could entail something like joint sensorimotor communication, with dyads exaggerating the kinematics of their individual actions, or even exaggerating how their actions relate to each other in order to signal an important coordination point in a performance. For example, in a dance sequence which requires actors to wave their right arms from side to side in

synchrony, actors may exaggerate the distance between their arms right before changing the direction of their waving movement in order to signal that they will shift from in-phase waving to anti-phase waving. This could function to allow for another dyad to predict and coordinate with this movement, or could function to provide learning relevant information to a dyad watching a demonstration. Future research should aim to investigate how dyads jointly exaggerate aspects of their performances in order to either coordinate with other dyads, or to teach other dyads.

5.2.4 Cultural transmission

We have considered the roles that movement cues play in teaching on the individual level and on the level of small groups, but what role do movement cues produced in joint actions play in teaching on a societal scale? Traditions and practices of a given culture are a result of generations of transmission and innovation, with demonstration and imitation thought of as the key mechanism by which information is transmitted within cultures (Heyes, Huber, Gergely & Brass, 2009).

Considering the evidence that movement cues produced to support joint action may also support learning (as demonstrated in chapter 2), the role that joint action coordination plays as a mechanism involved in cultural transmission needs to be considered. In addition to demonstration, knowledge is shared through joint action, with people teaching and learning through participation. Compared to demonstrative contexts, joint actions are diverse in terms of how actors are coupled (e.g. visually or haptically) and how many co-actors are involved, and laden with cues produced by actors in order to support coordination. In addition, movement variability associated with interacting with a partner may also increase an actor's action exploration space, thus fostering creativity and innovation (Wilf, 2013; Wu, Miyamotor, Castro, Olveczky & Smith, 2014). Thus, it is likely that with regards to cultural

evolution, demonstration and joint action differ with regards to how information is transmitted from generation to generation. Moreover, the type of information transmitted through generations is likely to differ depending on the type of joint action through which it was transmitted. Thus, understanding the manner in which information is transmitted has the potential to advance our understanding of cultural evolution, and why cultural practices vary so much.

Let's consider a dance sequence being taught by teachers to students in three dance schools (school A, B and C), which have three different teaching methods. In school A, the teacher demonstrates a sequence, and the student reproduces that sequence. In school B, the teacher and the student perform the sequence together in synchrony. In school C, the teacher holds the student and guides them through the sequence haptically. Now, let's imagine that the students of schools A, B and C become teachers and teach their students the dance sequence in the same way that they were taught it. And now their students become teachers and teach the dance sequence in the same way they learnt it, and so on for several generations. It is possible that after these several generations, what started off as one dance sequence is now three distinct dance sequences performed differently by the three schools. Although many other factors could explain this divergence, the manner in which the sequence was transmitted through the generations is likely to have had an influence on how these dance sequences diverged. For example, it could be that school A's version of the dance sequence resembles the original sequence spatially, with the spatial aspects of the sequence being preserved due to the emphasis that learning from demonstration places on visual cues. Because of the need to exaggerate one’s movements and slow down in order to achieve interpersonal coordination, school B's version of the dance has got slower and more exaggerated. School C's version of the dance however, could bear little resemblance to the original dance spatially because visual cues are not used, but could be very similar to the

original dance temporally, because of superior temporal resolution of the haptic channel. Translating this idea into a transmission chain style experiment could yield interesting insights into the role of sensorimotor communication in cultural transmission.

5.3 The potential role of informative and relational cues in human-robot interaction

Our research may also be applied to human-robot interaction, in order to help design robots that can interact and coordinate safely and seamlessly with humans, both for coordination and teaching purposes. As robots are interacting with humans more often, the big challenge of human-robot interaction is to develop systems that can both communicate and recognize intentions, in order to allow for effective coordination with an end user.

There is already some evidence that robots can produce informative movement cues in order to allow observers to discriminate between the robot's intentions. One study by Dragan, Lee and Srinivisa (2013) successfully designed a policy which allowed robots to disambiguate the goal of their actions, by deviating from the optimal trajectory in a sensorimotor communication like manner, thus allowing an end user to anticipate the goal of the robot's motion earlier and more effectively. Developing robots that are able to produce informative movement cues which make their actions easier to predict is important in order to make it possible to effectively coordinate with them. Likewise, developing robots that can produce cues that draw an observer's attention to learning relevant parts of an action sequence is important in order to have robots that can effectively teach. Research like ours, which quantifies the kinematics of teaching and coordination is important in order to achieve these goals.

With regards to detecting informative intentions from movement cues, the development of methods that can accurately identify the kinematic signatures of different informative intentions and classify actions on the basis of these kinematic signatures is possible. The methods used in order to investigate the 'observability' of instrumental intentions first identify the movement features that discriminate between different instrumental intentions, and then classify the movements on the basis of these discriminant features (Cavallo et al. 2016). It has been suggested that these methods can be leveraged in order to develop robots that can accurately identify an actor's instrumental intentions on the basis of their reach to grasp movements (Becchio et al. 2017; Sciutti et al. 2015). In addition to these methods, it is possible that our methods can also be leveraged in order to design robots that can effectively recognize and discriminate between different informative intentions. This would allow robots to know, firstly whether or not to respond to a human partner's actions (e.g. to engage the end user in presence of informative movement cues, and not to engage the end user in absence of informative movement cues), and secondly how to respond appropriately to an end user's actions (e.g. to imitate and attend to the learning relevant parts of the end user's action sequence in the presence of teaching cues, and to predict and synchronize with the end user in presence of coordination cues).

As well as recognizing a robot's intentions, successful human-robot interaction will also require that end users harbor a positive attitude towards the robots that they are interacting with. Not only does this require robots to appear emotionally expressive (Hortensius, Heleke & Cross, 2018), it also requires that robots appear as competent interaction partners which are enjoyable to interact with. By understanding relational cues that reflect the ability to coordinate and the level of affiliation between two actors, we can design robots that appear as competent interaction partners, who are even able to build some sort of rapport with their interaction partners. Moreover, understanding the movement cues

which drive aesthetic experience will allow for designing robots that can move in such a manner that is enjoyable to watch and interact with. Thus, we believe that our findings can contribute to the development of robots which display the ability to effectively engage in social interactions as partners' that end users enjoy interacting with.

References

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience, 11*(9), 1109-1116.

Ansuini, C., Cavallo, A., Bertone, C., & Becchio, C. (2015). Intentions in the brain: the unveiling of Mister Hyde. *The Neuroscientist*, *21*(2), 126-135.

Becchio, C., Cavallo, A., Begliomini, C., Sartori, L., Feltrin, G., & Castiello, U. (2012). Social grasping: From mirroring to mentalizing. *Neuroimage*, *61*(1), 240-248.

Becchio, C., Koul, A., Ansuini, C., Bertone, C., & Cavallo, A. (2017). Seeing mental states: an experimental strategy for measuring the observability of other minds. *Physics of Life Reviews*.

Becchio, C., Manera, V., Sartori, L., Cavallo, A., & Castiello, U. (2012). Grasping intentions: from thought experiments to empirical evidence. *Frontiers in Human Neuroscience*, *6*.

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

Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. *Cognition, 106*(2), 894-912.

Becchio, C., Zanatto, D., Straulino, E., Cavallo, A., Sartori, G., & Castiello, U. (2014). The kinematic signature of voluntary actions. *Neuropsychologia*, *64*, 169-175,

Bernstein, N . (1967) *The Coordination and Regulation of Movement*. Pergamon Press, Oxford.

Brand, R. J., & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science*, *11*(6), 853-861.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for ‘motionese’: modifications in mothers’ infant-directed action. *Developmental Science, 5*(1), 72-83.

Buchsbaum, D., Griffiths, T.L., Plunkett, D., Gopnik, A., & Baldwin, D. (2015).Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology*, *76*, 30-77.

Calvo-Merino, B., Jola, C., Glaser, D. E., & Haggard, P. (2008). Towards a sensorimotor aesthetics of performing art. *Consciousness and cognition*, *17*(3), 911-922.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14-27.

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports, 6*.

Clark, H. H. (1996). Communities, commonalities, and communication. *Rethinking Linguistic Relativity*, *17*, 324-355.

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

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, *13*(1991), 127-149.

Cross, E. S., Kirsch, L., Ticini, L. F., & Schütz-Bosbach, S. (2011). The impact of aesthetic evaluation and physical ability on dance perception. *Frontiers in Human Neuroscience*, *5*, 102.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148-153.

Curioni, A., Minio-Paluello, I., Sacheli, L. M., Candidi, M., & Aglioti, S. M. (2017). Autistic traits affect interpersonal motor coordination by modulating strategic use of role-based behavior. *Molecular Autism*, *8*(1), 23.

Dahan, A., Noy, L., Hart, Y., Mayo, A., & Alon, U. (2016). Exit from synchrony in joint improvised motion. *PloS One*, *11*(10), e0160747.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727-738.

Dragan, A. D., Lee, K. C., & Srinivasa, S. S. (2013, March). Legibility and predictability of robot motion. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction* (pp. 301-308). IEEE Press.

Dunst, C. J., Gorman, E., & Hamby, D. W. (2012). Child-directed motionese with infants and toddlers with and without hearing impairments. *Center for Early Literacy Learning*, *5*(8).

Eskenazi, T., Rueschemeyer, S. A., de Lange, F. P., Knoblich, G., & Sebanz, N. (2015). Neural correlates of observing joint actions with shared intentions. *Cortex*, *70*, 90-100.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104.

Fessler, D. M., Holbrook, C., & Dashoff, D. (2016). Dressed to kill? Visible markers of coalitional affiliation enhance conceptualized formidability. *Aggressive Behavior*, *42*(3), 299-309

Feygin, D., Keehner, M., & Tendick, R. (2002). Haptic guidance: Experimental evaluation of a haptic training method for a perceptual motor skill. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on* (pp. 40-47). IEEE.

Fukuyama, H., Qin, S., Kanakogi, Y., Nagai, Y., Asada, M., & Myowa-Yamakoshi, M. (2015). Infant's action skill dynamically modulates parental action demonstration in the dyadic interaction. *Developmental Science*, *18*(6), 1006-1013.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17*(2), 535-543.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, *4*(1), 23-48.

Georgiou I., Becchio C., Glover S., Castiello U. (2007). Different action patterns for cooperative and competitive behaviour. *Cognition*, 102, 415- 433

Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy. *Interaction Studies*, *6*(3), 463-481.

Gergely, G., & Csibra, G. (2013). Natural pedagogy. *Navigating the social world: What infants, children, and other species can teach us*, 127-132.

Gibson, J. J. (1979). The theory of affordances The Ecological Approach to Visual Perception (pp. 127-143).

Grèzes, J., Frith, C., & Passingham, R. E. (2004). Brain mechanisms for inferring deceit in the actions of others. *Journal of Neuroscience*, *24*(24), 5500-5505.

Haken, H., Kelso, J. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*(5), 347-356.

Hart, Y., Noy, L., Feniger-Schaal, R., Mayo, A. E., & Alon, U. (2014). Individuality and togetherness in joint improvised motion. *PloS One*, *9*(2), e87213.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243-259.

Heyes, C. M., Huber, L., Gergely, G., & Brass, M. (2009). Evolution, development and intentional control of imitation. *Philosophical Transactions of the Royal Society B, 364*, 2291–2443.

Hortensius, R., Hekele, F., & Crossy, E. S. (2018). The perception of emotion in artificial agents. *IEEE Transactions on Cognitive and Developmental Systems*.

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

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in Cognitive Sciences*, *9*(1), 21-25.

Jacobsen, T., & Höfel, L. (2003). Descriptive and evaluative judgment processes: Behavioral and electrophysiological indices of processing symmetry and aesthetics. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 289-299.

Jacobsen, T., Schubotz, R. I., Höfel, L., & Cramon, D. Y. V. (2006). Brain correlates of aesthetic judgment of beauty. *Neuroimage*, *29*(1), 276-285.

Kirsch, L. P., Urgesi, C., & Cross, E. S. (2016). Shaping and reshaping the aesthetic brain: Emerging perspectives on the neurobiology of embodied aesthetics. *Neuroscience & Biobehavioral Reviews*, *62*, 56-68.

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

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.

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.

Koterba , E. A., & Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants' attention and object exploration. *Infant Behavior and Development*, *32*(4), 437-444.

Kourtis, D., Knoblich, G., Wozniak, M., & Sebanz, N. (2014). Attention allocation and task representation during joint action planning. *Journal of Cognitive Neuroscience, 26(10), 2275-86*.

Lakens, D., & Stel, M. (2011). If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. *Social Cognition*, *29*(1), 1-14.

Manera, V., Becchio, C., Cavallo, A., Sartori, L., & Castiello, U. (2011). Cooperation or competition? Discriminating between social intentions by observing prehensile movements. *Experimental Brain Research*, *211*(3), 547-556.

Manera, V., Schouten, B., Becchio, C., Bara, B. G., & Verfaillie, K. (2010). Inferring intentions from biological motion: A stimulus set of point-light communicative interactions. *Behavior Research Methods*, *42*(1), 168-178.

Marsh, K. L., Richardson, M. J., Baron, R. M., & Schmidt, R. C. (2006). Contrasting approaches to perceiving and acting with others. *Ecological Psychology*, *18*(1), 1-38.

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

McEllin, L., Knoblich, G., & Sebanz, N. Imitation from a joint action perspective. *Mind & Language*.

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

McManus, I. C., Edmondson, D., & Rodger, J. (1985). Balance in pictures. *British Journal of Psychology*, *76*(3), 311-324.

McManus, I. C., Jones, A. L., & Cottrell, J. (1981). The aesthetics of colour. *Perception, 10*(6), 651-666.

Miles, L. K., Nind, L. K., & Macrae, C. N. (2009). The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of Experimental Social Psychology*, *45*(3), 585-589.

Nagai, Y., & Rohlfing, K. J. (2009). Computational analysis of motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development*, *1*(1), 44-54.

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.

Noy, L., Levit-Binun, N., & Golland, Y. (2015). Being in the zone: physiological markers of togetherness in joint improvisation. *Frontiers in Human Neuroscience*, *9*, 187.

Patel, D., Fleming, S. M., & Kilner, J. M. (2012). Inferring subjective states through the observation of actions. *Proc. R. Soc. B*, *279*(1748), 4853-4860.

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. (2013). Human sensorimotor communication: A theory of signaling in online social interactions. *PLoS One*, *8*(11), e79876.

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*.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*(2), 129-154.

Repp, B. H., & Keller, P. E. (2008). Sensorimotor synchronization with adaptively timed sequences. *Human Movement Science*, *27*(3), 423-456.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, *27*, 169-192.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), 264.

Rogoff, B. (2003). *The Cultural Nature of Human Development*. Oxford university press.

Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology*, *54*(1), 175-203.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, *112*(4), 585.

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

Sartori, L., Becchio, C., Bara, B. G., & Castiello, U. (2009). Does the intention to communicate affect action kinematics?. *Consciousness and Cognition*, *18*(3), 766-772.

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

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (in press). Co-representation of others' task constraints in joint action. *Journal of Experimental Psychology: Human Perception and Performance.*

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299-309.

Sciutti, A., Ansuini, C., Becchio, C., & Sandini, G. (2015). Investigating the ability to read others' intentions using humanoid robots. *Frontiers in Psychology*, *6*, 1362.

Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic Bulletin & Review*, *16*(1), 170-175.

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., & Prinz, W. (2005). How two share a task: Corepresenting Stimulus–Response mappings. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 1234–1246.

Slepian, M. L., Young, S. G., Rutchick, A. M., & Ambady, N. (2013). Quality of professional players' poker hands is perceived accurately from arm motions. *Psychological Science*, *24*(11), 2335-2338

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*(1-2), 3-23.

Sperber, D., & Wilson, D. (2004). Relevance theory. *Handbook of Pragmatics. Oxford: Blackwell*, 607-632.

Torrents, C., Castañer, M., Jofre, T., Morey, G., & Reverter, F. (2013). Kinematic parameters that influence the aesthetic perception of beauty in contemporary dance. *Perception*, *42*(4), 447-458.

Trujillo, J. P., Simanova, I., Bekkering, H., & Ozyurek, A. (2018). Communicative intent modulates production and perception of actions and gestures: A Kinect study. *Cognition*, *180*, 38-51.

Tsai, J. C. C., Sebanz, N., & Knoblich, G. (2011). The GROOP effect: Groups mimic group actions. *Cognition*, *118*(1), 135-140.

van der Wel, R. P., Knoblich, G., & Sebanz, N. (2011). Let the force be with us: dyads exploit haptic coupling for coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1420.

Vesper, C., & Richardson, M. J. (2014). Strategic communication and behavioral coupling in asymmetric joint action. *Experimental Brain Research*, *232*(9), 2945-2956.

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

Vesper, C., Knoblich, G., & Sebanz, N. (2014). Our actions in my mind: Motor imagery of joint action. *Neuropsychologia, 55*,115-121.

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., van der Wel, R. P. R. D., 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., van der Wel, R. P., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, *211*(3-4), 517-530.

Vicary, S., Sperling, M., Von Zimmermann, J., Richardson, D. C., & Orgs, G. (2017). Joint action aesthetics. *PloS One*, 1*2*(7), e0180101.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*(3), 34-41.

Wilf, E. (2013). Sociable robots, jazz music, and divination: Contingency as a cultural resource for negotiating problems of intentionality. *American Ethnologist*, *40*(4), 605-618.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, *38*(10), 1559-1579.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science, 20*(1), 1-5.

Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: an instance of the ideomotor principle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*(1431), 501-515.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729-R732.

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.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89-100.

Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., & Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature neuroscience*, *17*(2), 312.

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, *112*(2), 201-216.

Zeki, S. (2001). Artistic creativity and the brain. *Science*, *293*(5527), 51-52

ACKNOWLEDGEMENT

TO EXTERNAL FUNDING AGENCIES

CONTRIBUTING TO PHD DISSERTATIONS

Name of doctoral candidate: Luke McEllin

Title of dissertation: INVESTIGATING THE PRODUCTION AND PERCEPTION OF MOVEMENT CUES IN JOINT ACTION

Name of supervisor(s): Natalie Sebanz, Gunther Klaus Knoblich

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 616072 – JAXPERTISE.**

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.**