

**THE SOCIAL IS PREDICTIVE:
HUMAN SENSITIVITY TO ATTENTION CONTROL IN ACTION PREDICTION**

by

Ana Pesquita

MSc., University of Lisbon, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2016

© Ana Pesquita, 2016

Abstract

Observing others is predicting others. Humans have a natural tendency to make predictions about other people's future behavior. This predisposition sits at the basis of social cognition: others become accessible to us because we are able to simulate their internal states, and in this way make predictions about their future behavior (Blakemore & Decety, 2001). In this thesis, I examine prediction in the social realm through three main contributions. The first contribution is of a theoretical nature, the second is methodological, and the third contribution is empirical. On the theoretical plane, I present a new framework for cooperative social interactions – the predictive joint-action model, which extends previous models of social interaction (Wolpert, Doya, & Kawato, 2003) to include the higher level goals of joint action and planning (Vesper, Butterfill, Knoblich, & Sebanz, 2010). Action prediction is central to joint-action. A recent theory proposes that social awareness to someone else's attentional states underlies our ability to predict their future actions (Graziano, 2013). In the methodological realm, I developed a procedure for investigating the role of sensitivity to other's attention control states in action prediction. This method offers a way to test the hypothesis that humans are sensitive to whether someone's spatial attention was endogenously controlled (as in the case of choosing to attend towards a particular event) or exogenously controlled (as in the case of attention being prompted by an external event), independent of their sensitivity to the spatial location of that person's attentional focus. On the empirical front, I present new evidence supporting the hypothesis that social cognition involves the predictive modeling of other's attentional states. In particular, a series of experiments showed that observers are sensitive to someone else's attention control and that

this sensitivity occurs through an implicit kinematic process linked to social aptitude. In conclusion, I bring these contributions together. I do this by offering an interpretation of the empirical findings through the lens of the theoretical framework, by discussing several limitations of the present work, and by pointing to several questions that emerge from the new findings, thereby outlining avenues for future research on social cognition.

Preface

This thesis describes a novel theoretical framework for cooperative social interactions and presents a new methodology utilized in seven experiments testing sensitivity to attention control in action prediction. The theoretical framework was developed by the author in collaboration with James T. Enns and Robert Whitwell. The author of this thesis was the primary contributor to the identification and design of the methodology supporting the experimental research program in roughly equal collaboration with James T. Enns and Craig C. Chapman. The experiments took place at the University of British Columbia during 2013-16. Data analysis was performed in equal collaboration between the author and James T. Enns. The author collected the data presented here in collaboration with Emily Ryan, Jacob Shieh, Jessica Leung, Mallika Khanijon, Nathan Wispinski, Nessa Bryson, Puneet Sandhu and Tracy Lam. Ulysses Bernardet developed custom software for video recording. All of the writing in this thesis is the author's own, and incorporates suggestions given by James T. Enns. A modified version of Chapter 3 authored by A. Pesquita, C.S. Chapman & J.T. Enns is currently in press in the *Proceedings of the National Academy of Sciences* journal. The same chapter was presented at the "Interactive Social Cognition: An Emerging Science" Symposium at the 25th Annual Meeting of the Canadian Society for Brain, Behavior and Cognitive Science, Ottawa, Canada. This research was approved by the University of British Columbia Behavioral Research Ethics Board (Human Attention while reaching H11-00946).

Table of Contents

Abstract	ii
Preface.....	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
Acknowledgements.....	xi
Dedication.....	xii
1 General introduction	1
1.1 Social predictive processing	3
1.2 Social attention in action prediction	6
1.3 Thesis overview	8
2 Predictive joint-action model (pJAM).....	12
2.1 Introduction	12
2.2 A hierarchical predictive approach to joint-action.....	18
2.2.1 Fundaments of hierarchical predictive processing.....	18
2.2.2 Applying hierarchical processing to the social domain	22
2.2.3 Predictive Joint-Action Model (pJAM).....	25
2.2.4 Implementation challenges addressed by pJAM.....	30
2.3 Model predictions	35
2.3.1 Goal representation layer	35
2.3.2 Action planning layer.....	37
2.3.3 Sensory routing layer	53
2.4 Discussion.....	55
3 Sensitivity to attention control in action prediction	59
3.1 Methodology.....	61
3.1.1 Stimuli recording	62
3.1.2 Stimuli selection	64
3.1.3 Manipulation check.....	66
3.1.4 Summary	73
3.2 Are humans' sensitive to attention control in others?.....	74
3.2.1 Method.....	75
3.2.2 Results	76

3.2.3 Discussion.....	79
3.3 Does sensitivity to attention control contributes to a reactive advantage in social interactions?	81
3.3.1 Method.....	83
3.3.2 Results	85
3.3.3 Discussion.....	86
3.4 Is sensitivity to attention control consciously accessible to observers? ..	88
3.4.1 Method.....	88
3.4.2 Results	89
3.4.3 Discussion.....	91
3.5 Where on the actors' body can the attention control signal be seen? ..	93
3.5.1 Method.....	94
3.5.2 Results	95
3.5.3 Discussion.....	97
3.6 How early in the time-course of an observed action is the attention control signal available?	98
3.6.1 Method.....	98
3.6.2 Results	99
3.6.3 Discussion.....	100
3.7 Is sensitivity to attention control linked to social aptitude?	101
3.7.1 Social aptitude and sensitivity to attention control	102
3.7.2 The kinematics of human sensitivity to attention control	104
3.8 Summary and discussion of the empirical studies	108
4 General discussion.....	115
4.1 Theoretical framework.....	115
4.2 Empirical findings	117
4.3 Bringing theory and findings together	120
4.3.1 Initial state of the predictive architecture.....	122
4.3.2 Probabilistic predictions during action observation.....	124
4.3.3 Prompting observers' prediction responses	129
4.4 Conclusion	131
Bibliography.....	132

List of Tables

Table 1. Means of eight kinematic measures taken on the distribution of reaches used as stimulus materials in the Experiments.	69
Table 2. Correlations between temporal and kinematic measurements.	71
Table 3. Principal component analysis, first component weights.	73

List of Figures

Figure 1 Diagram from Vesper and colleagues (2010, p.999) representing the minimal components for a joint-action architecture. The outer circle represents shared goals. Co-tasks divide the inner circle. Monitoring and prediction processes act on representations of the shared goal and partner's co-tasks.	14
Figure 2. Comparison of sensorimotor and social interaction loops from Wolpert, Doya and Kawato (2003, p.594).	15
Figure 3 Joint-action of two young boys carrying a table down some stairs. Image by James Aldridge retrieved from http://jamesaldrige-artist.co.uk/blog/?p=346	26
Figure 4 The diagram illustrates the predictive joint-action model (pJAM), which is minimally composed of three layers: goal-representation, action-planning, and sensory routing. The framework assumes that each partner in a joint-action maintains internal models of both themselves and their co-partners. The goal-representation layer is responsible for maintaining and updating the shared goals guiding the interaction. The action-planning layer outputs motor commands that take into account both the desired states of oneself and one's partners in the interaction. The sensory routing layer receives the inflow of sensory input and compares it to internal model predictions pertaining to each partner's action outcomes within the interaction. Each layer generates predictions of the information that it expects to observe in the layer below. Continuous comparison between adjacent layers results in error signals that are sent up to optimize subsequent predictions in the layer above.	28
Figure 5 Goal representation layer in pJAM.	36
Figure 6 Action-planning layer in pJAM.	38
Figure 7 Sensory routing layer in pJAM.	54
Figure 8 Illustration of the method from the actors' perspective. Actors were filmed through plexiglass reaching to two possible targets. On "chosen" trials both locations were lit and actors had to choose (not shown), on "directed" trials only one location was lit and actors were directed to reach to that location (as shown).	63

Figure 9 An example of the video-clips framing. Critically, the LEDs are not visible in the stimuli.	64
Figure 10 (A) Overall means of the actors' movement initiation times for "choice" and "direct" actions. (B) Means of the movement initiation times for each of the 4 actors (A1 to A4) for "choice" and "direct" actions. (C) Overall means of the movement times for "choice" and "direct" actions. (D) Means of the movement times for each of the 4 actors (A1 to A4) for "choice" and "direct" actions. Error bars correspond to one standard error of the mean.	66
Figure 11 Illustration of the method from the observers' perspective. Observers respond to each video by pressing a spatially mapped key press as rapidly as possible to indicate where the actor was reaching.	75
Figure 12 (A) Mean correct response time (RT) in the experiment reported in Chapter 3.2. Error bars are +/- 1 standard error. (B) Mean correct RT for each of the four actors. (C-D) The data in A-B after each observer's correct RT has been converted to z-scores in order to standardize the distributions for individual differences in mean speed and variance.	78
Figure 13 Illustration of the method from the observers' perspective. Observers attempt to beat the actor to the target.	84
Figure 14 Proportion of times the observer reaches the correct target faster than the actor in chosen and directed conditions, collapsed across the four actors. Error bars are one standard error of the mean. Values above .50 indicate that the observer was faster than the actor more often than the opposite.	86
Figure 15 (A) Mean z-scores of correct RT in the experiment reported in Chapter 3.4. Error bars are +/- 1 standard error. (B) The proportion of hits and false alarms of observers trying to discriminate chosen from directed trials, after rank ordering observer's response biases from conservative (reluctant to respond "choice") to liberal (reluctant to respond "direct").	90
Figure 16 Representative drawings of the masked video-clips. (A) The actors' head was masked leaving the torso and limbs visible. (B) The actors' body was masked leaving only the head and neck visible.	94

Figure 17 Mean z-scores of correct RT in the experiment reported in Chapter 3.5, separately for trials in which the body and limbs were visible versus when only the head was visible. Error bars are +/- 1 standard error.....	96
Figure 18 Mean proportion correct response in the temporal occlusion experiment reported in Chapter 3.6. Error bars are +/- 1 standard error.....	100
Figure 19 Scatterplot of the relation between observer's speeded sensitivity scores in the experiments reported in chapter 3.2 to 3.5 and their Autism Quotient scores.	104
Figure 20 Scatterplot of movement initiation vs. movement duration trade-off in reach responses of participants reporting higher and lower social aptitude levels.	105
Figure 21 Scatterplot of observers' sensitivity to social attention at the reach initiation stage against sensitivity at the reach duration stage.....	107
Figure 22 Action prediction cycle in pJAM.	121
Figure 23 Predictive architecture state at the start of the trial. Before the activation of bottom-up swipes of information, prediction is at chance level - 50% left and 50% right.	123
Figure 24 Predictive architecture state while minimizing the error between top-down predictions and bottom-up information.	125
Figure 25 Illustration of response triggering. Probabilistic predictions about the actor weight on the observer motor plans. Once the bias towards one side reaches the response threshold, the motor plan is executed.....	130

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Jim Enns, for the support he has offered to my development as a scientist. There is much to learn with Jim: His unmatched capacity to distil the important ideas from the redundant ones; the quick swing with which he approaches experimental design and data analysis; and last but not the least, his ability to “blue-pen” any rambling text into a poignant paragraph. For the last five years, I have been observing and hopefully also absorbing some of these skills. Thank you to Rebecca Todd and Mark Schaller, who were the committee members for this thesis. Their questions, comments and suggestions have been very helpful in shaping this manuscript. I also thank Craig Chapman for his support and contributions to the empirical work presented in this thesis. Thank you to Robert Whitwell, Alan Kingstone, and Rebecca Todd for providing valuable comments and suggestions to my comprehensive exam that has metamorphosed into a considerable part of this thesis. A heartfelt thanks to all my colleagues at the Vision Lab and the B.A.R. Lab, and all the research assistants that I had the pleasure to work with. Without this social context, everything would have been less colorful. A special thank you is owed to Ulysses Bernardet for his continuous all-round and unconditional support. His ability to problem-solve is in a class of its own, and I have had the privilege to count on his help throughout the completion of this work. Finally, thank you to my family back in Portugal; your “saudade” is always present. Funding for this research came from a Ph.D. scholarship to the author A. Pesquita from the Portuguese Fundação para a Ciência e Tecnologia (SFRH/BD/76087/2011), and by a Discovery Grant to J. T. Enns from the Natural Science and Engineering Council of Canada.

Dedication

To my dear husband Ulysses,

Five years ago we decided to trade our honeymoon package for two one-way tickets to Vancouver. This was the start of our great adventure. We have worked hard, no doubt, but we have also explored this new land and marveled together at the Canadian wild. Your passion for life and science inspires me. Here's to our next five years, and the ones after that.

Love,

Ana

1 General introduction

We have only so much as to glance at another human being and we at once begin to read beneath the surface.

-Nicholas Humphrey (2002)

Prediction is not just one of the things your brain does. It is (...) the foundation of intelligence.

-Jeff Hawkins (2005)

Anthropologists have long considered that the evolution of larger, more powerful and complex brains was triggered by the early hominids' need to venture into new lands (Martin, 1983). The brain adapted through natural selection to match new survival needs, and practical skills such as tool-making, fire-lighting, and spear-throwing, emerged as the key accomplishments of the new impressive cognitive powers of early hominids' brains (Brown et al., 2012).

More recent views suggest that the increasing complexity of the social environment was also a pivotal evolutionary pressure contributing to the development of modern brains. The emergence of social structures, with individuals who were both able to play the role of peerless collaborators and ruthless competitors, could have only occurred hand-in-hand with the evolution of a powerful brain, able to process the intricate nuances of social relationships (Marean, 2015).

This trend in anthropological science offers a critical insight to the modern study of human cognition. If we are to understand human cognition, we must not only

consider how we process symbols and physical information but also understand how we process other individuals (Blakemore & Decety, 2001; de Gelder, 2006). This, undoubtedly, comes with a new set of challenges. A considerable part of other individuals' existence is inaccessible. Their thoughts, memories, intentions and emotions all take place in an inner theater that is closed to us. And inside which we can only peek in through the distorting window of language communication and the ambiguous window of observable behavior. So, how are we able to by-pass this inherent separation from one another and experience the rich and diverse types of social interactions that animate our lives?

Current answers to this question propose that we rely on predictions about the hidden dimensions of our social counterparts to sustain successful social interactions (Manera, Schouten, Verfaillie, & Becchio, 2013; Ramnani & Miall, 2004; Sparenberg, Springer, & Prinz, 2012; Springer, Hamilton, & Cross, 2012). The idea is that we use our own cognitive resources to build internal models of other people's cognitions. Simulations about someone else's cognitive states (e.g. about what they are feeling, thinking, and attending to) guide our expectations about their future behavior, and in this way contribute to the viability of social interactions.

From the many hidden cognitive processes that can support our predictions of someone else's behavior, one is considered to have a special revealing quality – attention. Attention is the “data-handling method in the brain” (Graziano & Kastner, 2011; Graziano, 2013; Webb & Graziano, 2015). Thus decoding someone else's focus of attention provides us with palpable clues about which information is engaging their inner cognitive mechanisms. This largely

contributes to our ability to make predictions about someone else's future actions (Baron-cohen, 1995; Baron-Cohen, 2000; Calder et al., 2002).

In this thesis, I will address predictive mechanisms underlying social cognition both at the theoretical and empirical level. On the theoretical side, I will present a framework for cooperative social interactions termed the predictive joint-action model (pJAM). On the empirical side, I will present new evidence suggesting that human's perceptual sensitivity to someone else's internal attentional states facilitates action prediction.

1.1 Social predictive processing

Historically, the idea of prediction as a central mechanism of human cognition emerged conjointly with early elaborations about perception and motor control. In the 19th century, William James suggested that "every mental representation of a movement awakens to some degree the actual movement which is its object" (James, 1890, pp.293). Key to this insight was the notion that by merely imagining a future action one can anticipate its motor and sensory outcomes. This insight was later formalized in what has come to be known as *ideomotor theory*, which posits the existence of a common code linking action and perception (Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz, 1990).

The idea that action and perception are coupled in this way releases "information processing" from having to wait for an action or a sensory event to actually occur, and brings them both into the realm of prediction, i.e., mental simulations about future actions and sensations that are yet to occur. Anticipation thus becomes the key to understanding how relatively slow neural processes are able to coordinate perception and action with external events that

are happening in real time. It is important to note that although prediction mechanisms such as these were already being contemplated in very early stages of both psychology and neuroscience, it was not up until recently that these ideas coalesced into wide acceptance. In the past, the majority of theoretical models of cognition portrayed a serial succession of processes. This meant that information processing started with the reception of sensory input, which was then passed on to the “black-box” of higher cognitive functions, and ended with the output of overt behavior. Such reasoning stemmed from original behaviorist approaches and was passed on to early cognitivist theories (Bubic, von Cramon, & Schubotz, 2010; Cisek, 1999).

Fast-forward to current times. The human brain is broadly accepted to be a prediction machine (Clark, 2013; Hawkins & Blakeslee, 2007). Current research in neuroscience and psychology indicates that prediction is a fundamental principle of neural processing and cognition (Brown & Brüne, 2012; Bubic et al., 2010). The general consensus is that the way we perceive and act upon the world is not only a result of incoming sensory information (i.e. bottom-up information), but also integrates our internal biases, knowledge, and previous experiences (i.e. top-down predictions). This feat is accomplished in the brain by a hierarchy of computational events that sequentially try to reduce discrepancies between bottom-up and top-down swipes of information (Clark, 2013). This kind of predictive processing is central to both perception (Enns & Lleras, 2008) and motor control (Wolpert & Flanagan, 2001).

Can the predictive principles underlying perception and action in individual cognition be extended to explain social interactions? If so, to what extent, and with which limitations? The answer to these questions is still in its infancy. One

positive consideration comes from theories suggesting that the human ability to infer the goals and intentions of someone else's actions can be explained by predictive coding (Jacob & Jeannerod, 2005; Kilner, Friston, & Frith, 2007; Wolpert et al., 2003). Understanding someone else's actions via predictive coding goes beyond merely asserting that we use our own motor substrate to encode models of others' actions (a concept famously introduced by the discovery of the mirror neuron system; Gallese & Goldman, 1998). In addition to that, it proposes that we generate active predictions about the consequences of observed actions, even before they occur so that these expectations can be compared to how people's intentions (generated in higher-levels of the processing hierarchy) are translated into motor events when they do occur. It is proposed that the most probable cause of the observed action will be inferred by minimizing the prediction error at all levels of the hierarchy (Jacob & Jeannerod, 2005; Kilner, Friston, & Frith, 2007; Wolpert et al., 2003). Additional support for the prediction hypothesis is given by a wide range of empirical observations of prediction during social perception (Kilner, Vargas, Duval, Blakemore, & Sirigu, 2004), social interaction (Sebanz & Knoblich, 2009), and social learning (Abernethy, Zawi, & Jackson, 2008).

In this thesis, I will propose that the well-established principles underlying predictive individual cognition can help us better understand the inner-workings of cooperative social interactions, i.e. joint-actions. Sebanz, Bekkering, & Knoblich (2006) define joint-action as "a social interaction whereby two or more individuals coordinate their actions in space and time to bring about change in the environment." I will propose a predictive framework that attempts to account for the interlocking of partners' intentions, actions, and perceptions occurring during joint-actions.

1.2 Social attention in action prediction

Imagine you are sitting in a cafe. A girl enters. You both recognize each other as the unflattering versions of your profile photographs. Your date has started. As the first minutes pass you wonder: Will she leave or stay? Knowing the focus of her attention will considerably narrow down your predictions about her future behavior. If her eyes are fixed on you, there is some hope. If her eyes wander, not so much. Alas, she glances languidly at her phone. You decide to beat her to the punch, and politely announce how good it was to have met her.

As illustrated by this example, we track the focus of someone else's attention to predict their future behavior and adapt accordingly. Several studies indicate that humans are remarkably sensitive to where someone is attending (Bayliss, Schuch, & Tipper, 2010; Bayliss & Tipper, 2005, 2006; Friesen & Kingstone, 1998; Langton & Bruce, 2000; Rogers et al., 2014). Furthermore, it has been suggested that this ability contributes to rough representations of the others' mental state (Simon Baron-Cohen, 1995, 2000; Calder et al., 2002). But do these representations only fill out the content of the other's mind, or do they also hold information on the control of that content?

Let's get back to the example of the cafe encounter. Whether your date intentionally decided to direct her attention away from you or merely turned to the phone because it unexpectedly blinked tells you different things about her mental state, and possibly about the success of your date. Thus, there is important social information in knowing whether someone's spatial attention was endogenously controlled (as in the case of choosing to glance at the phone in search for a distraction) or exogenously controlled (as in the case of reacting to an unexpected phone blink). This leads to the question of whether humans

are able to distinguish between these two kinds of attention control in the observed actions of others?

A positive answer to this question is expressed in a recent theory that social awareness involves the predictive (forward) kinematic modeling of other people's attention (Graziano & Kastner, 2011; Graziano, 2013). According to this proposal, humans constantly construct and update sophisticated models of other people's attentional states. As well as representing the perceived location of someone else's attention, these models are posited to comprise rich representations of how attentional resources are deployed, including the spatial and temporal consequences of attention on action. Thus, these models are posited to include the nature of control so that the spatial and temporal consequences of an attentional state can be predicted in the actions of others before they occur. In this view, social attention modeling allows observers to make conscious elaborations about someone else's attentional states, contributing to the ability to make sense of other's actions, and predict what they might do next (Graziano & Kastner, 2011; Graziano, 2013, 2015).

In close pursuit of Graziano's (2013) theoretical proposition, in this thesis, I present an empirical study investigating human sensitivity to attention control. The study is structured around the central question - Are observers sensitive to someone else's attention control? A positive answer to this question is then followed up with branching questions aimed at characterizing human sensitivity to attention control in terms of conscious processing, temporal and spatial features, as well as its link to social skill.

1.3 Thesis overview

This thesis aims at making three contributions to the current understanding of predictive mechanisms in social cognition. The first contribution is of a theoretical nature, the second is methodological, and the third contribution is empirical. The thesis comprises two main sections corresponding to Chapter 2 (presenting the theoretical framework) and Chapter 3 (reporting a new methodological approach and the associated empirical research).

In Chapter 2 I present a new theoretical framework for human cooperative action – the predictive joint-action model (pJAM). Chapter 2.1. introduces the motivations behind the development of the framework. In recent years, there has been a proliferation of research about human cooperative behavior. Yet, the development of theoretical frameworks in this field has not kept pace with the increasing number of research findings (Knoblich, Butterfill, & Sebanz, 2011). In response to this identified need, Chapter 2.2 outlines a hierarchical predictive framework for joint-action. In Chapter 2.3 I discuss pJAMs' predictions in light of evidence from the current literature on joint-action. Chapter 2.3 concludes the theoretical section of the thesis. There I will discuss the overall success of utilizing a hierarchical predictive approach to account for the implementation challenges of joint-action.

Chapter 3 is dedicated to a series of empirical studies investigating human sensitivity to social attention control. In Chapter 3.1, I start by describing the general methodological approach I developed to generate new data in this area. This methodology is composed of two stages. In the first stage, I developed and tested stimuli sets composed of video clips of actors reaching for one of two possible targets while either choosing (endogenous control) or being directed

(exogenous control) to one target. In the second stage, this stimulus set was used in a series of experiments addressing the questions:

- Are observers sensitive to someone else's attention control?
- Does sensitivity to attention control contribute to a reactive advantage in social interactions?
- Is sensitivity to attention control a conscious process?
- Where on the actors' body is the attention control signal available?
- How early in the time-course of an observed action is the attention control signal available?
- Is sensitivity to attention control linked to social aptitude?

In Chapter 3.2, I describe one experiment designed to test observers' sensitivity to someone else's attention control. In this experiment, we asked observers to predict the development of chosen vs directed actions. The findings from this experiment provided initial evidence indicating that observers are sensitive to someone else's attention control. Moreover, the results show that there is a "choice advantage", i.e. observers are faster at predicting the end-target of chosen actions compared to directed ones. The following sub-chapters are dedicated to characterizing the observed human sensitivity to attention control.

In Chapter 3.3, I report one experiment probing whether sensitivity to someone else's attention control can offer observers a motor advantage in social interaction settings. In this experiment, observers are asked to compete with the video recorded actors, by attempting to reach the end-target before the actors do. The findings showed that observers could quickly harness their sensitivity to attention control in order to generate an adaptive motor response.

In Chapter 3.4, I present two experiments to test whether sensitivity to the attention control of a social other is or is not a conscious process. In these experiments, participants were asked to guess whether each observed action was chosen or directed. The two experiments differed in whether or not participants received feedback about the accuracy of their responses. The findings from both experiments indicated that sensitivity to attention control was not accessible to the observer's conscious awareness.

In Chapter 3.5, I report findings from an experiment investigating whether the control signal is coming from the head or the body of the actors. The results showed that observers' sensitivity to attention control cues was robustly resistant to the occlusion of actors' body parts, suggesting that the cues to attention control are distributed throughout the body.

In Chapter 3.6, I report the findings from one experiment examining the time course of sensitivity to attention control. The findings revealed that sensitivity to attention control was only observable in the early stages of movement observation. This supports its value in action prediction mechanisms.

In Chapter 3.7, I present analyses indicating that observers with higher social aptitude also exhibit stronger sensitivity to attention control states in their responses. These analyses also address differences in the kinematic profiles of sensitivity to attention control between individuals with higher and lower social skills. These observations bolster the hypothesis that sensitivity to attention control arises from the involuntary tendency for humans to model the attentional states of others.

In Chapter 3.8, I review the findings from this research while discussing their implications for the field of social cognition. The main conclusion discussed is the observation that humans are sensitive to attention control through an implicit kinematic process linked to empathy. An interpretation for the 'choice advantage' is proposed based on the fluency of kinematic cues. At last, the limitations of the research project are discussed, leading to proposals for future work.

Finally, in the General Discussion (Chapter 4) I bring together the two streams of this thesis. I will use theoretical concepts of predictive processing modeling, described in Chapter 2, to frame the new evidence of sensitivity to attention control, reported in Chapter 3. I note that the empirical part of this thesis was not directly designed to test the joint-action theoretical model. However, the task shares some core similarities with joint-action tasks (i.e. participants are required to monitoring and predicting of someone else's actions and the subsequent execution of an appropriate motor response). Therefore, pJAM has proven itself useful as a framework to interpret the findings, and identify limitations of the empirical studies presented in this thesis.

These elaborations will offer some support to the hypothesis that social cognition involves the predictive modeling of other's attentional states.

2 Predictive joint-action model (pJAM)

Research in a number of related fields has recently begun to focus on the social, perceptual, cognitive, and motor workings of cooperative behavior. Indeed, there now appears to be enough coherence in these efforts to refer to the study of the mechanisms underlying human cooperative behavior as the field of joint-action (Knoblich, Butterfill, & Sebanz, 2011; Sebanz, Bekkering, & Knoblich, 2006). Yet, the development of theoretical frameworks in this field has not kept pace with the proliferation of research findings. In this chapter, I propose a hierarchical predictive framework for the study of joint-action termed the predictive joint-action model (pJAM). Afterward, I will derive predictions from the model, and juxtapose these predictions with empirical evidence from the current literature on joint-action. In the process, I will identify where new empirical evidence is necessary to test the models' predictions. Finally, I discuss the overall success of the hierarchical predictive approach to account for the implementation challenges of joint-actions. This is done with the larger goal of uncovering the theoretical pieces that are still missing in a comprehensive understanding of joint action.

2.1 Introduction

The ability of humans to cooperate with one another vastly increases the range of their potential actions (Clark, 1996). It is through cooperation that we achieve goals unattainable to the single individual, “whether it be carrying a log, or building a skyscraper” (Stix, 2014). Hence, cooperation is seen to be of central importance to our species' evolutionary success (Tomasello, 2009). In recent years, the field of cognitive science has turned its spotlight on cognition in the

social milieu (Semin & Cacioppo, 2006). As a result of this increased interest, the study of the cognitive processes underlying human cooperative behavior is now a field of research in its own right (Knoblich, Butterfill, & Sebanz, 2011).

These new studies investigating the perceptual, cognitive, and motor components of cooperation have also recently converged on a consensual operational definition of “joint-action”. Sebanz, Bekkering, & Knoblich (2006) define joint-action as “a social interaction whereby two or more individuals coordinate their actions in space and time to bring about change in the environment.” Every joint-action, therefore, requires an interlocking of two or more individuals’ intentions, actions, and perceptions (Sebanz & Knoblich, 2009). This attunement between partners is what enables ensemble musicians to create a unified sound texture and tango dancers to move together so swiftly that it seems difficult to imagine them apart. Effortless as it might seem on the surface, however, even the simplest instances of joint action, such as playing catch or carrying an object together, require a diverse ensemble of cognitive processes to be coordinated.

The minimal requirements for a joint-action architecture have been defined by Vesper and colleagues (2010) in the following way. An architecture for joint-action must minimally support the capacity to (1) represent a shared goal and corresponding individual tasks, (2) monitor and predict each partner’s actions, and (3) allow for continuous coordination. Figure 1 illustrates the components of the proposed minimal architecture for joint-action. Although instrumental for mapping the requirements of a joint-action model, this proposal does not specify in any detail how these requirements might be implemented.

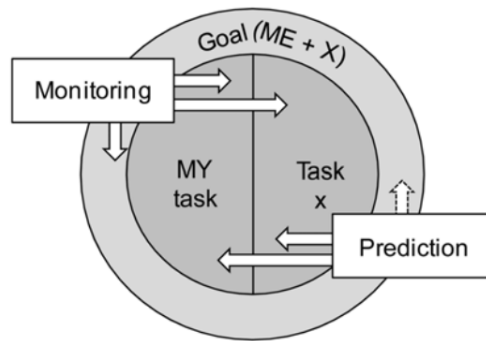


Figure 1 Diagram from Vesper and colleagues (2010, p.999) representing the minimal components for a joint-action architecture. The outer circle represents shared goals. Co-tasks divide the inner circle. Monitoring and prediction processes act on representations of the shared goal and partner's co-tasks.

Closer to a computational solution for a theory of joint-action is Wolpert, Doya and Kawato's (2003) proposal, which is premised on the possibility of close parallels between individual and social motor control. In 1996 Wolpert & Miall presented a model of sensorimotor computation with the aim of formalizing the mechanisms behind skilled motor control. The goal was to explain how an organism is able to act optimally towards a goal despite uncertain and changing environmental circumstances (e.g., the unknown properties of objects that are the targets of action in the face of continuously changing environmental conditions). In 2003 Wolpert, Doya and Kawato (2003) were the first to suggest that there might be a computational parallel between motor control and social interaction. Specifically, they proposed that the sensorimotor computations involved in acting on one's own body during individual motor control are comparable to the communicative signals involved in acting on other people's behavior during social interactions. Figure 2 illustrates the proposed parallelism between individual motor control and social motor control. Wolpert et al.s' (2003) proposal has received some notice in the joint-action research

community, with the framework being often cited as a useful approximation of the mechanisms sustaining joint-actions (e.g., Becchio, Sartori, & Castiello, 2010; Doerrfeld, Sebanz, & Shiffrar, 2012; Häberle, Schütz-Bosbach, Laboissière, & Prinz, 2008; Knoblich et al., 2011; Loehr, Kourtis, Vesper, Sebanz, & Knoblich, 2013; Pecenka & Keller, 2011; Ramenzoni, Sebanz, & Knoblich, 2014; Sartori, Becchio, & Castiello, 2011; Sebanz & Shiffrar, 2009; Vesper, Butterfill, Knoblich, & Sebanz, 2010; Vesper, van der Wel, Knoblich, & Sebanz, 2013).

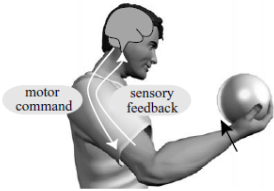
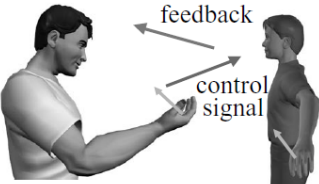
	motor control	social interaction
loop	(a) 	(b) 
control signal	motor command	communicative actions e.g. speech, gesture
consequences	change in my body's state	change in your mental state
state	configuration of my body	mental state of your mind

Figure 2. Comparison of sensorimotor and social interaction loops from Wolpert, Doya and Kawato (2003, p.594).

However, in our view Wolpert et al.s' (2003) proposal has not garnered the full attention it deserves, perhaps because it appeared prior to the most recent surge of interest in the problem of joint-action. It is also perhaps for the same reason — the theory appearing slightly ahead of its time — that several key aspects of joint-action, such as goal sharing, task co-representation, and interpersonal coordination, were not addressed by Wolpert et al. (2003). In summary then, Wolpert et al.s' (2003) proposal does not meet the minimal requirements for a joint-action framework as delineated by Vesper and

colleagues (2010), just as in complementary fashion, Vesper et al. (2010) does not match the computational rigor of Wolpert et al.'s (2003) model for joint motor control. In the present review, I will seek to bridge these gaps by presenting a hierarchical predictive framework for joint-action, the predictive joint-action model (pJAM). This framework is fully compatible with Wolpert et al.'s (2003) computational notions, but in addition it incorporates the necessary higher-order organization to deal with the specifics of joint-action implementation.

Joint-actions are a pervasive part of our daily lives (e.g. shaking hands, playing soccer, washing dishes together) and seem to come about without much effort. However, when examined more closely, the implementation of even the simplest of joint-actions reveals itself to be a complex and dynamic process. Joint-actions are marked by high degrees of freedom (i.e. at any given moment each partner can act in a multitude of ways) and hidden states (i.e. partners don't have direct access to each other's internal states). I propose that a hierarchical predictive processing approach might be appropriate to solve the implementation challenges inherent to joint-actions.

One of the two core ideas underlying hierarchical prediction is that the brain is a prediction machine, meaning it continuously tries to match sensory and motoric information with predictions based on goals and intentions. As a result of this bidirectional exchange of predictions from the top (goals and intentions) and signals from the bottom (sensory and motor signals), the system is able to find computational solutions to complex and underspecified problems, such as the ones posed by joint-action (Clark, 2013; Hawkins & Blakeslee, 2007). The second core idea is that this exchange of signals (predictions and sensorimotor signals) is

hierarchical, meaning that it occurs at each of several levels in a multi-layered system. This allows the architecture to respond appropriately to a much wider range of conditions than would be possible if there were only two layers. For example, while one's own and others' contributions to a joint action are incorporated at the higher-level of goal representations, lower layers at the motor level can flexibly simulate separately the expected actions pertaining to self and other. In addition, the first levels of sensory processing can attribute the convoluted sensory outcomes of a joint-action to each agent in the interaction.

In the following section, I will present the predictive joint-action model (pJAM). I will start by introducing the general principles of hierarchical predictive processing. I then use these notions to describe pJAM. Subsequently, I will discuss the challenges that are specific to the implementation of joint-action and describe how pJAM tackles each of these challenges.

Afterward, I will juxtapose predictions derived from pJAM with empirical evidence from the current joint-action literature. In doing so, I aim to integrate the various cognitive processes that sustain joint-action (e.g., goal setting and sharing, action prediction, coordination strategies and interpersonal sensory processing) into one overarching framework. In the conclusion, I will evaluate the overall success of using the hierarchical predictive approach in capturing the complexity of joint-action. My hope is that this evaluation will reveal the theoretical pieces that are still missing for a comprehensive understanding of joint action.

2.2 A hierarchical predictive approach to joint-action

I suggest that joint-action can be best understood within a hierarchical predictive processing framework. Before presenting the proposed application of this approach to joint-action, we will briefly summarize the general principles of the hierarchical predictive processing approach, and highlight Wolpert's et al. (2003) hierarchical approach to motor control and social processing.

2.2.1 Fundamentals of hierarchical predictive processing

The core idea is that the brain is a predictive machine, which continuously tries to match bottom-up information with top-down predictions. In perception the main task of this predictive machine is to infer external causes from their bodily effects (i.e. motor and sensory signals including proprioceptive information). This is a complex and costly computational task, because many different causes can result in similar effects, and moreover, a solution must be found in a very short time. Hierarchical predictive processing in perception, also known as predictive coding, offers insights about how the brain solves these computational problems (Bar, 2009).

According to hierarchical prediction principles, processing is distributed in a multi-level hierarchical cascade of events. The lowest layer in the hierarchy corresponds to sensory input, and the higher levels correspond to internal simulations of that input. Processing is marked by a bidirectional swipe of information between hierarchical levels. Each level in the system both receives ascending signals from the lower levels (or the external world, at the first layer), while concurrently generating downward predictions about these same signals arriving from lower-levels.

Generative models, at each level of the hierarchy, output predictions about the information on the level below. However, since many different potential causes can be consistent with the incoming information from the subordinate level, each layer maintains several parallel generative models. Each of the generative models represents a state probability. The predictions outputted by these probabilistic models are continuously compared against the flow of incoming information from the subordinate level in the processing hierarchy, resulting in prediction errors. In turn, prediction errors are sent back to the higher level, via forward connections, which sharpen the fit of the probabilistic models, approximating their next predictions to the information represented in the lower level. Cycles of concurrent predictions and error-correction occur throughout the hierarchy. As a result of this bidirectional exchange of predictions from above and signals from below, errors are minimized at both lower and higher levels, giving rise to a structure of activations that represents the most likely cause of the sensorial input, “a kind of virtual version of the sensory data” (Clark, 2013).

In addition to offering a computational solution to the fast pace and robust nature of human perception, the hierarchical predictive approach also offers an account of perception and action interactions. The main task of motor control is to process the events needed to take an organism from its current motor state to the desired motor state by Wolpert & Miall (1996). Through the lens of hierarchical predictive processing, the desired action goal is treated as an actual state of affairs, causing a cascading downward prediction of what should be experienced next in the layers below. Error signals from each layer are sent back up and thus are used to adapt the movement output as it unfolds. These adjustments, in turn, change the sensory input, thus continuously minimizing the error throughout the hierarchy. In this fashion, hierarchical predictive

mechanisms are proposed to iteratively lead to a solution that will take the organism from its current motor state to the desired one (Hawkins & Blakeslee, 2007). This iterative account of the perception-action loop offers an elegant solution to how organisms successfully act within complex and ever-changing environments. The most important underlying idea is that cognitive systems can infer solutions to the problem of how to get from motor state A to motor state B, by minimizing internal predictive errors in resemblance to Bayesian inferential processes (Friston, Mattout, & Kilner, 2011; Friston, 2003; Todorov, 2004).

Haruno, Wolpert, and Kawato's (2003) proposal of hierarchical modular selection and identification for control (HMOSAIC) overlap, at least in some fundamental aspects, with the overarching principles of hierarchic predictive processing. The HMOSAIC was originally put forward as a model of motor learning and production. The HMOSAIC posits a multi-level hierarchical architecture for motor control. Each vertical level of the HMOSAIC comprises parallel modules. These modules correspond to processors for generating predictions (forward models) and those for generating control signals (inverse models). Let's consider the example of reaching to grasp a coffee mug. Modules embedded in the upper levels of the hierarchy represent information of a more abstract and symbolic nature. In the example, these modules would correspond to a symbolic representation of the task of reaching for a coffee mug and its associated object semantics. Modules in the lower levels of the hierarchy represent low-level dynamics, such as movement elements, and object sensory features. In the example, these modules would represent e.g. information about limb position and velocity. Modules in the middle-levels represent different ways to structure and organize movement elements for a range of different purposes, such as e.g. different movement trajectories for reaching towards the coffee mug.

Modules within a given level of the hierarchy, operating in parallel, are dedicated to different possible states. Within any given level, the modules are evaluated on the basis of how well their predictions (termed priors) fit the signals arriving from the level underneath (termed responsibilities). Possible abstract goals represented by parallel modules at high-levels of the hierarchy (e.g. grasp the sugar pot vs. grasp the coffee mug vs. grasp the spoon) output predictions to the adjacent lower level, comprising its own parallel models including information about the possible different trajectories of the limb while reaching for its end-target. The different higher-level task-goal modules are activated according to how well their predictions (priors) fit the information represented in modules at the immediately lower level (responsibilities). In a similar fashion, these mid-level module predictions of arm trajectory are evaluated against sensory information arriving from the lower-levels, and the modules with the best fit are activated. This bi-directional flow of information between levels permits the reentrant and recursive processes that underlie module updating and selection at different levels of the hierarchy.

The HMOSAIC proposal has several similarities with the general principles of hierarchical predictive coding. Both approaches agree that: (1) Processing occurs through a multi-level hierarchy ranging from abstract symbolic representations (higher levels) to sensory input representations (lower levels); (2) Bidirectional comparison of information occurs between vertical levels, allowing the system to use Bayesian-like computations find a solution for the motor control problem of transitioning from a current motor state to the desired one; (3) Parallel modules at each level of the hierarchy represent different possible states, allowing the system to cover a large level of possible realities and decrease processing times.

Here, I will focus on the commonalities between the hierarchical predictive approach as presented by Clark (2013) and the hierarchical approach to motor control as presented by Wolpert et al. (2003). However, I note that these approaches differ in their fine-grained implementation details. For example, the approaches differ in their description of the parallel processing occurring within each level. Whereas Wolpert et al. (2003) propose that pairs of inverse-forward models are responsible for generating priors (predictions about the layer below), Clark (2013) does not describe the computation details of the parallel representations comprised in each level of the processing hierarchy. In this paper, our focus is not to propose solutions to the computational implementation of motor control. Our focus is to use the general principles of hierarchical processing to motor control, which are mostly common to both approaches, to address the specific case of joint-action.

2.2.2 Applying hierarchical processing to the social domain

Wolpert et al. (2003) were the first to suggest that there might be a computational parallel between motor control and social interaction. The authors posit the HMOSAIC as an overarching framework for both individual and social motor control. They posit that HMOSAIC, initially devised to account for individual motor control, can also sustain social operations. In particular, they describe how this architecture can support social action recognition and social mimicry.

Suppose now that you are watching someone else reach out to pick up a cup of coffee. The HMOSAIC structure can be dedicated to the process of recognizing the goal of someone else's action (i.e. action recognition). The modules at different levels of the hierarchy represent different levels of description of the

observed action. The lower level modules in the sensory modalities represent different observable action elements. The middle level modules represent different sequences of those elements. The highest-level modules would represent different goals and intentions. As the observation of the movement unfolds, the predictions from lower and middle-level modules representing “reaching out to pick up a cup of coffee” are born out in the observations over those that signal “reaching out to move the cup of coffee away” and in so doing strengthen their responsibility signals. These responsibility signals are propagated to higher-level modules where they activate the modules that reflect the goals and intentions (“take a sip of coffee” or “pass the coffee mug”) that are structurally associated with generating the behavior “reaching out to pick up a cup of coffee”. These authors further suggest that a similar process can sustain social action mimicry. In this case, two HMOSAIC structures would be necessary. One dedicated to planning and executing ones’ own actions, and another dedicated to processing someone else actions would be involved. Mimicry could occur through an attunement between both HMOSAICS at the lower and middle-levels of the hierarchy in the absence of attunement at the goal or intention levels. Wolpert et al. (2003) speculate that how well one comes to understand or reproduce another’s actions will depend on the similarity of the HMOSAIC that generated an actor’s behavior and the HMOSAIC of the observer that interprets it. Consistent with Wolpert et al. (2003) proposal, Kilner, Friston, and Frith, (2007) offer an account of “mind reading” as a predictive coding process. The authors propose that brain areas involved in processing others’ intentions are reciprocally connect in a hierarchical fashion, with the pre-supplementary motor area receiving low-level inputs from visual areas, and parietal and pre-frontal areas responsible for processing motor and symbolic processing. This proposal

offers further support to the idea of using hierarchical predictive processing to study social processes.

However, Wolpert et al. (2003) aimed at more than proposing a model for action recognition and mimicry. Their claim was that a wide spectrum of social interactions obeys to the basic principles of hierarchical sensorimotor computations. Yet, this is where the explanatory power of their framework finds some challenges. Wolpert et al (2003) state that what makes social interactions difficult to capture in a computational model is their open-ended nature. The authors highlight two general difficulties: (1) Time delays between communicative actions and their social consequences can range from seconds to days; (2) The space of possible responses to a communicative action is very large, and therefore responses are not easily predicted. This makes it difficult to concretely relate the proposals of the model to a big section of real-world social interactions, which are often open-ended and multifaceted. In sharp contrast to this open-ended dilemma of many real-world interactions, the specific case of joint-action is restricted by the existence of a shared goal. In particular, the multitude of time delays and possible responses is capped by the assumption that both partners behave towards the achievement of a mutually agreed upon interaction goal. Therefore, if I reduce social interaction to the specific case of joint-action, I reduce the complexity of the interaction to a level at which I can usefully apply the principles of hierarchical processing.

I propose an architecture for joint-action that harnesses the principles of hierarchical predictive processing, compatible with the computational notions proposed by Wolpert et al. (2003), to match the requirements for joint-action implementation delineated by Vesper et al. (2010) (i.e. represent a shared goal

and corresponding individual tasks, monitor and predict each partner's actions, and allow for continuous coordination). Admittedly, using a framework intended to organize research on motor control in an individual to better understand joint-action between two or more individuals will likely be incomplete. Nonetheless, this is still worth doing in our opinion, because it will help to reveal in a very concrete way, where new theoretical ideas are needed to extend what is currently known about individual actions to the newer domain of joint-action.

2.2.3 Predictive Joint-Action Model (pJAM)

How are two or more independent individuals able to infer and implement a joint motor solution that will lead them to achieve their shared goal? To help ground this problem, we will use the scenario of two young brothers trying to carry a table down a set of stairs (Figure 3). The boys have agreed to move the table from the terrace to the garden's corner. But although the boys share the same goal, each one must contribute differently to the task. The younger brother lifts the back of the table, while the older one supports the weight at the front. The boys must continuously adapt to each other's movements while carrying the table and navigating their way towards their desired destination. As we will see, this seemingly simple task implies a complex interlocking of intentions, actions, and perceptions between the two boys. In the following sections, we use this example to delineate, layer-by-layer, pJAM's hierarchical predictive framework for joint-action.



Figure 3 Joint-action of two young boys carrying a table down some stairs. Image by James Aldridge retrieved from <http://jamesaldrige-artist.co.uk/blog/?p=346>

The predictive joint-action model (pJAM) is conceptualized minimally as three hierarchical processing layers: *goal representation*, *action-planning*, and *sensory routing*, as illustrated in Figure 4. The hierarchical organization allows the architecture to represent and find solutions for the joint-action process at different levels of abstraction, from high-level symbolic representations of the goal to lower-level chunking of movement elements (e.g. musculoskeletal dynamics).

The goal representation level is at the top of the hierarchy. It is responsible for symbolic representations of shared goals. Parallel probable shared goals co-exist at this level. In our example, each boy has the goal of cooperating with one another to carry the table from the terrace to the yard. The processing hierarchy treats the shared goal as an actual state of affairs, causing a cascading downward prediction of what should be experienced next in the layers' bellow. The goal representation layer will output an abstract representation of the 'desired joint-

state' to the layer bellow (e.g. a symbolic representation of moving the table together), and receive as input information about the currently estimated joint-state (e.g. abstract representation of the ongoing cooperation) provided by the layer bellow. The continuous comparison between these two sources of information will generate a signal indexing the deviation between the actual state of the system and its desired state (i.e. joint-state error). This error signal is fed-back into the goal representation layer, where it is used to update and prune the shared goal models. Thus at each iteration, the output of the 'shared goal models' will be more precise and specific, providing the layers bellow more precise guidelines of what should be experienced next. It will be more effective in allowing the brother to execute actions that will lead them closer to their goal.

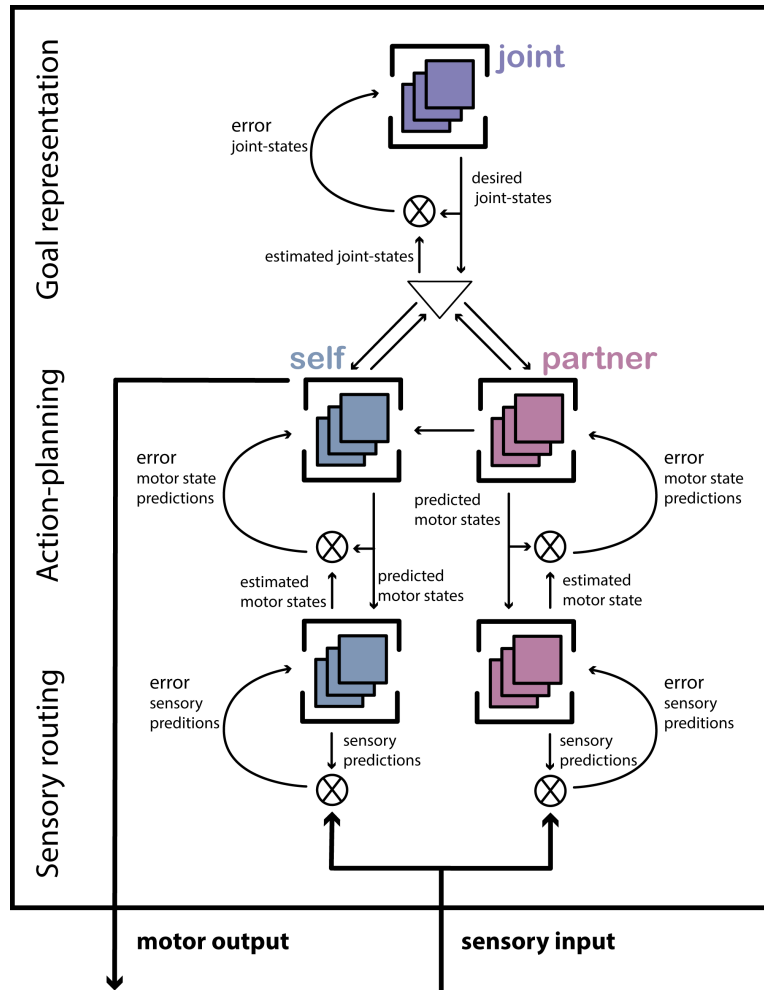


Figure 4 The diagram illustrates the predictive joint-action model (pJAM), which is minimally composed of three layers: goal-representation, action-planning, and sensory routing. The framework assumes that each partner in a joint-action maintains internal models of both themselves and their co-partners. The goal-representation layer is responsible for maintaining and updating the shared goals guiding the interaction. The action-planning layer outputs motor commands that take into account both the desired states of oneself and one's partners in the interaction. The sensory routing layer receives the inflow of sensory input and compares it to internal model predictions pertaining to each partner's action outcomes within the interaction. Each layer generates predictions of the information that it expects to observe in the layer below. Continuous comparison between adjacent layers results in error signals that are sent up to optimize subsequent predictions in the layer above.

pJAM proposes a bifurcation between action partners at the action-planning layer. This bifurcation accounts for the distribution of task load between the individual partners in the interaction. The load of carrying the table can be divided between the brothers in many different ways. For example, the older brother can hold the front of the table and the younger brother the back or vice-versa, the brothers can both face forward or face each other, etc. pJAM proposes that sets of parallel self-other models represent possible individual contributors to the joint action. In other words, at this level, the joint-state models are broken down into models encoding each individual's expected contribution to the desired joint-state, i.e. co-task models. This bifurcation is expressed both at the action-planning layer and sensory routing layer. This means that parallel cascades of downward predictions about motor states and upwards state estimations based on sensory information are maintained for each partner in the interaction.

Parallel co-task models of one's own contributions and the partner's contributions to the joint-action produce the desired motor state signals, which are compared to the actual motor state estimates arriving from the subordinate level in the hierarchy. These comparisons generate error signals indexing the deviation between desired and estimated individual motor states. These error signals are fed back into the upper layer to help calibrate, prune and sharpen the co-task generative models pertaining to each partner in the joint-action. The continuous optimization of co-task models will allow for one to iteratively compensate for deviations between the current state and the desired state of the interaction. Thus, this continuous process of minimizing error at the action-planning layer gives rise to compensatory coordination.

The action-planning layer is also responsible for outputting a motor command for action execution. The generation of this motor command is informed by our internal predictions about our partner's next motor states. The process of integrating predictions about one's partners in addition to one's own motor planning leads to anticipatory coordination. In this way, partners are able to bypass the temporal delays that would otherwise arise from waiting to receive information about each other's actual motor states before being able to plan and execute their own actions.

At the level of the sensory routing layer, the bulk of the incoming sensory information reflecting the outcomes of the joint-action is compared to independent sensory predictions referring to one's own and partner's expected sensory action outcomes. This comparison will serve as a split gateway that parses 'own' and 'partner' sensory information into their corresponding predictive streams. Deviations between sensory input and sensory predictions (i.e. sensory predictive errors) are fed-back to sensory predictive models and continuously improve sensory parsing (i.e. the system's ability to direct received sensory information to their corresponding predictive cascades). This process will allow the predictive system to attribute external consequences to each individual's actions.

2.2.4 Implementation challenges addressed by pJAM

Next, I will describe each layer of the pJAM in further detail by defining the implementation challenge that the layer addresses, and positing how the challenge is met by the pJAM.

2.2.4.1 Goal representation layer

Implementation challenge It is commonly accepted that partners in a joint-action have similar internal representations of the interaction goal and that this shared goal representation guides each partners' actions (Sebanz et al., 2006). Here, already, we find the first challenge to joint-action implementation: How can internal goal representations be shared with enough detail to guide the unfolding of an interaction in space and time (with all the possible variations that this entails)? Language is crucial for how people agree to pursue a goal together (Clark, 1996). However, verbal exchanges are often too slow and too processing heavy to guide the fast-paced adaptations necessary to accomplish most joint-actions. Imagine, for example, how difficult it would be for tango dancers to coordinate their movements if they would have to continuously verbally inform each other about their upcoming movements. In contrast, action-perception mechanisms offer a faster route to coordination (Sebanz & Knoblich, 2009). Verbal communication, cultural conventions, and common sense knowledge are all crucial elements that restrict variation in goal representation between partners (Clark, 1996). However, it would be difficult to imagine that these factors alone could lead to enough specification to account for the full unfolding of a joint-action. But even if it would be possible for partners to a priori construct very similar and detailed shared goal representations (i.e. each partner would have a copy of the same step-by-step blueprint for the interaction), the shared representation wouldn't be of much help once confronted with the variability that the actual joint-action execution entails. This can be simply illustrated through our situational example. Let's say that there is a rock in the place where the brothers were initially aiming at positioning the table. Are the brothers doomed to behave like mindless robots and lay down the table where they

initially intended, even if one table leg will be unstable on top of a rock? No. They will adapt, and they will adapt together. To sharpen the question at hand: How are goals shared with enough detail to guide joint-action, but also with enough flexibility to allow for adaptation?

pJAM solution pJAM suggests that each partner keeps several parallel ‘shared goal models’, a sort of halo of probable variations of the shared goal. These ‘shared goal models’ represent the desired joint-states as if existed and discharge expectations (priors down the hierarchy). These discharged top-down predictions are sent to the lower level of the processing hierarchy where they are compared to estimations of the actual joint-action state. Continuous comparison between these adjacent layers produces error signals pertaining to the predictions of the ‘shared goal models’. These error signals are fed back to the *goal representation layer* where they are used to prune the ‘shared goal models’, and in this way sharpen the individual’s representation of the shared goal.

The mechanism described above can account for the necessary goal flexibility in joint-action. However, for this iterative flexibility to be useful in joint-action, each partner’s individual ‘shared goal models’ have to converge into similar states. How can this occur? Apart from the initial loose representations of the shared goal informed by communication, social norms, etc., partners also share the outcomes of their joint-actions. Thus, each individual partner is exposed to similar streams of bottom-up sensory information as a consequence of their combined actions. Thus through hierarchical predictive mechanisms, partner’s individual systems have a good chance of continuously converging into a close-

enough internal representation of the shared goal, leading to the ability to take to term successful joint-actions.

2.2.4.2 Action planning layer

Implementation challenge Successful joint-actions imply a continuous counterbalance of each individual's contributions to the goal. In the example, imagine that one brother lost strength for a moment and let the table swing to the left, in optimal coordination, the other brother will respond by compensating for this deviation to bring the table back to the desired course. However, in many cases, a posteriori compensation is not a viable option, due to the fast temporal constraints of most joint-actions. Thus, partners must be able to anticipate each other and accommodate for each other's movement changes even before they occur. In our example, imagine that one brother is about to lose grip of the table. The other brother might be able to predict what is about to happen, and quickly lift the table higher to regain control. The critical question here is: How are the two young brothers able to plan their individual actions to optimally compensate for and anticipate each other's actions under changing conditions?

pJAM solution PJAM accounts for both compensatory and anticipatory coordination. *Compensatory coordination* comes about through continuous error minimization at the action-planning layer. This is achieved by using error signals (generated by the comparison of desired motor states and estimated motor states) to improve one's models of both one's own and partner's co-tasks. In pJAM horizontal connections between 'own' and 'partners' models support anticipatory coordination. One's motor commands will be informed by one's

internal predictions about the partner's desired next states leading to *anticipatory coordination* (i.e., coordination that is based on a prediction of partner's next actions; Keller, 2013).

2.2.4.3 Sensory routing layer

Implementation challenge In the case of the two brothers carrying the table, the sensory feedback combines information about the consequences of both brothers' actions. How is this combined sensory information parsed into the individual outcomes of one's own and partner's actions?

pJAM solution In pJAM sensory input is routed into self or partner's hierarchical processing streams by comparing the prediction of the sensory outcomes of both self and partner's to the received sensory input. This comparison will allow the system to attribute external consequences to each individual's actions. In addition, it will lead to the percolation through the system of prediction errors that are specific to each partner in the joint-action, ultimately serving to train internal models of both oneself and the other. Attributing sensory consequences to oneself vs. others results in the sense of agency (Obhi, 2012; Schüür & Haggard, 2011), that might help joint-action adaptation by contributing to the division of joint-tasks into individual co-tasks. In addition, it is also expectable that the quality of the sensory feedback pertaining to one's actions will be more detailed, richer and accurate than the sensory feedback pertaining partner's actions. Thus, our internal predictive streams about ourselves will be more accurate than our predictive streams about our partners. This is in line with the observation that it is easier to coordinate with oneself than with others (Keller, Knoblich, & Repp, 2007). Nonetheless, by continuously minimizing error across

partner's predictive streams, our internal models of others will improve, which is expressed by the observation that coordination with others improves with practice (van der Wel, Sebanz, & Knoblich, 2012).

2.3 Model predictions

pJAM's architecture offers several predictions about the processes underlying joint-action. Some of these predictions have been addressed in the current empirical literature, while others remain to be tested. Next, I will juxtapose the model predictions with evidence from joint-action literature, and identify the areas where further empirical studies are necessary. This will serve to support the usefulness of the proposed model.

2.3.1 Goal representation layer

The *goal-representation layer* in pJAM is posited to maintain probabilistic shared-goal models, which output predictions about the desired joint-state. In turn, these predictions are compared with estimations of the current joint-state that come about by merging the estimates of each individual's motor contributions to the joint-action (supplied by the *action-planning layer* beneath). Figure 5 shows a diagram of the goal representation layer in pJAM. This organization implies the following: pJAM predicts that individuals in a joint-action have the capacity to monitor both joint and individual goals.

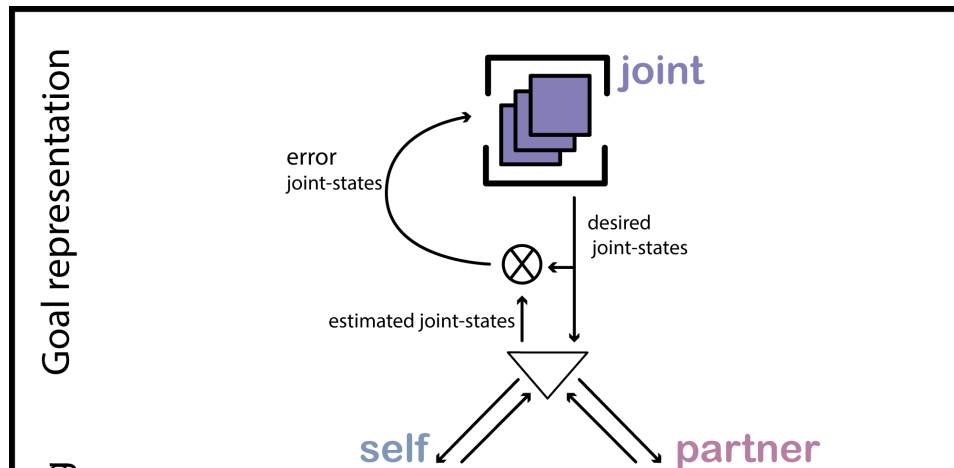


Figure 5 Goal representation layer in pJAM.

This prediction is supported by evidence arriving from musical ensemble studies. When playing together musicians must simultaneously maintain a representation of the desired unified sound of the ensemble as well as a representation of their own musical contributions to the overall sound (Keller, 2014). This observation was put to the test by a recent experimental study. In this study, EEG was recorded from pairs of pianists playing a previously memorized duet. During the performances, some of the keystrokes were programmed to originate altered pitches that did or did not change the joint auditory outcome (i.e., the harmony of a chord resulting from the two pianists' combined pitches). ERPs revealed that feedback-related negativity was elicited during altered auditory outcomes when these affected one's own, one's partner's, and joint-action outcomes. Thus indicating that partners in musical joint actions monitor not only the joint outcomes of their actions, but also their own and their partner's contributions to the joint-goal (Loehr, Kourtis, Vesper, Sebanz, & Knoblich, 2013).

The findings from Loehr and colleagues (2013) support pJAMs *goal representation layer* features: (a) They support the prediction that partner's in a joint-action have the capacity to monitor both the shared and individual goals; (b) They highlight the functional role of prediction error in joint-action, because joint-action outcomes are indexed by feedback-related negativity ERPs, known to encode unexpected events; (c) They support pJAMs proposed organization in which the goal representation layer is distinct from the action planning layer. This is because, in Loehr and colleagues (2013) task, the motor actions of the participants are consistent between changed and unchanged pitches (i.e. the pianists play the same key with the same finger) what changes is the sound outcome (i.e. whether the pitch is key consistent or not). Thus, the study supports that goal encoding and monitoring can be independent of motor processes.

One aspect of the *goal representation layer* that misses empirical support is the splitting of the shared goal representation (and corresponding desired joint-states) into individual co-goals (and correspondent desired individual states in the interaction). Although current findings show that partners in the interaction maintain both shared and individual representations (Keller, 2013; Loehr, Kourtis, Vesper, Sebanz, Günther, et al., 2013), the processes that moderate between shared and individual goals have been difficult to capture empirically.

2.3.2 Action planning layer

The action-planning layer is responsible for generating predictions about one's own and partner's motor contributions to the joint-task. Information about deviations between one's own predicted and estimated states (i.e. prediction errors) gives rise to compensatory coordination. In simultaneous, continuous

information about one's partners' predicted next states allows for anticipatory coordination. All these processes are supported by the action-planning layer in pJAM, represented by the diagram in Figure 6. The following predictions can be derived from the organization of pJAM's action planning layer:

- i. *We keep representations of our partner's expected contributions to the task (i.e. co-task models).*
- ii. *We generate predictions about our partners' future motor states.*
- iii. *We encode deviations between partners' motor predictions and action states, i.e. prediction errors.*
- iv. *The better our models of our partners are, the better we are able to coordinate with them.*
- v. *We can both anticipate and compensate for partners' actions.*

Next, I will present evidence that corroborates some of these predictions, and identify which predictions have not been tested empirically.

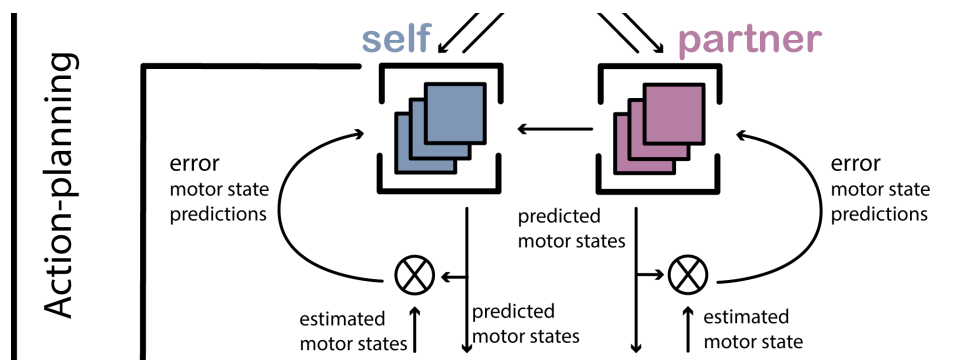


Figure 6 Action-planning layer in pJAM.

- i. *We keep representations of our partners expected contributions to the joint-task*

pJAM predicts that we keep motor representations of our co-partners action states. This prediction is derived from the proposed organization of the *action-planning layer*, where parallel probabilistic models are posited to represent one's own and partner's expected contributions to the joint-action (i.e. probabilistic co-task models). There is extensive evidence in the joint-action literature supporting this prediction. Next, I will highlight some of this literature.

It is now generally accepted that partners in a joint-action keep models of each other's expected roles in the interaction (Atmaca, Sebanz, Prinz, & Knoblich, 2008; Sebanz et al., 2006; Sebanz, Knoblich, & Prinz, 2003, 2005). The most prominent methods used to address this process are adaptations of well-known stimulus-response competition tasks to the social context. For example, the "joint Simon task" (Sebanz, Knoblich, & Prinz, 2003) compares individuals' performance in the Simon task when executed alone (Simon, 1969) with the performance in collaboration with a partner (joint Simon task). Results from the individual Simon task show that responses are faster when stimulus and response are spatially compatible, whereas non-corresponding stimulus-response pairs result in slower responses (Kornblum, Hasbroucq, & Osman, 1990). Notably, if we eliminate the stimulus-response feature overlap, by presenting a task with only one response location (i.e. a go/no-go task), the effect disappears (Liepelt & Prinz, 2011). This pattern of results is known as the Simon effect.

In the joint Simon task, responses are distributed across two participants, so that each individual is in control of pressing one of the two keys (right or left) in response to their assigned stimulus (e.g., red or blue cue). It is important to highlight that each participant is only responsible for half of the responses. This transforms the Simon task in a go/no-go task at the individual level. Similarly to the standard Simon task, red and blue cues are presented to the left or right side of the participants, and stimuli location is irrelevant to response selection (Sebanz et al., 2003). The critical question is, can the stimulus-response competition effect be observed when two participants perform the task together? In the social setting, the ideal strategy is for participants to ignore each other's part of the task. If individuals adhere to this ideal strategy, the results from the Joint Simon task should resemble the results of an individual go/no-go task. However, empirical evidence shows stimuli-response competition in the joint-Simon task, suggesting that participants internally model both their own and their partner's expected contributions to the joint-action (Knoblich & Sebanz, 2006; Sebanz et al., 2003).

Converging evidence suggesting that individuals keep internal models of both their own and their partners expected contributions to a joint-action comes from similar adaptations of other classical stimuli-response compatibility tasks to the social realm. For example, the "joint flanker" effect demonstrates that co-representation is not restricted to tasks initiating spatial interference, but generalizes to tasks involving arbitrary stimulus-response associations (Atmaca, Sebanz, & Knoblich, 2011). Additionally, the compatibility effect between numerical and spatial stimuli termed SNARC effect, has also been observed in the social transformation of this task - joint SNARC effect (Atmaca et al., 2008).

It is, however, important to note that stimuli-response compatibility effects are found both, when sharing a task with a social partner, and when in the presence of salient non-social factors. For instance, Dolk, Hommel, Prinz and Liepelt (2013) substituted the social partner in the Joint Simon task for a Chinese cat statue. The authors showed that if the statue were made to be a sufficiently salient event to provide a strong spatial reference, participants would start allocating task co-representations to the inanimate object. This shows that although at work during joint-action, the mechanism underlying task co-representations is not specifically social. This conception is in line with the view that social and non-social events are processed in similar ways, though social events often are more salient, recruiting more cognitive resources (Friesen & Kingstone, 1998; Langton & Bruce, 2000) .

Are co-task representations encoded at the motor level? A positive indication is offered by Holländer, Jung, and Prinz (2011). The findings from this study showed that lateralized readiness potential ERPs, not only when participants prepared to act themselves but also when it was the partner's turn to respond. This observation suggests that each partner maintains covert motor activations relating to the expected contributions their co-partners in the interaction. Given that we represent others motor plans, what prevents us from executing these plans? Following studies have shown that neural inhibition mechanisms are at work to ensure that one does not execute other people's expected contributions to the task (Sebanz et al., 2006; Tsai, Kuo, Jing, Hung, & Tzeng, 2006).

Recent studies show that our internal models of partner's expected roles in the interaction are influenced by contextual and personal factors. Regarding contextual factors, Kuhbandner, Pekrun, & Maier (2010) have shown that under

negative mood induction the extent to which partners encoded each other expected tasks in the interaction was diminished compared to when partners underwent positive mood induction. This finding can offer insight on previous observations that individuals in positive moods are more likely to like a stranger (Baron, 1987), and show higher cooperative tendencies (Forgas, 1998). Similar observations were made when asking participants to complete the joint-Simon task under the cooperative and the competitive scenarios.

Regarding person-specific factors, de Bruijn, Miedl and Bekkering (2008) report individual differences in the extent to which individuals tend to represent their partner's side of the task. These individual differences were nonetheless sensitive to manipulation and training. Colzato, de Bruijn and Hommel (2012) primed participants' self-concept as individualistic or social before testing the extent to which participants model their partner's co-tasks using the joint-Simon method. The results from this study showed that the joint Simon effect was more pronounced in the group primed with social affiliation words. This finding suggests that modeling other's roles in an interaction can be manipulated by increasing the relevancy of social factors. A finding that was substantiated by a related study (Colzato et al., 2012) testing pairs of Buddhists and atheists in the joint Simon task. Buddhist religion integrates a world-view in which compassion is a central teaching. The results showed that Buddhist's responses tended towards stronger integration of the partner's side of the task. Taken together, these studies seem to suggest that the extent to which we model other's expected contributions to an interaction can be modulated by personal factors such as self-construal and social beliefs.

In sum, pJAM's prediction of parallel representations of both one's own and one's partners expected contributions to the task (i.e. co-task models) is well grounded in current literature, which indicates that partners maintain co-task representations of one another expressed at the motor preparation level, which is modulated by social salience.

ii. We generate predictions about our partners' future motor states

pJAM proposes that co-task models (at the action-planning layer) output predictions of the desired motor states for both oneself and one's partners in the joint-action. Thus, the model predicts that partners continuously update predictions of each other's next motor states during the interaction. Next, I will highlight empirical findings that support this prediction.

Partners in a joint-action task must attend to one another while simultaneously predicting each other's next actions. That observing another person is not a passive process was elegantly shown in a study measuring participants' eye-gaze patterns while they were performing a block-stacking task alone versus while they were observing another person stacking blocks. Similar predictive eye-gaze patterns were found in advance of critical hand grips, both in those grips that were executed and in those that were only observed actions (Flanagan & Johansson, 2003). This finding indicated that during action observations we actively engage in action prediction.

Detailed predictions of kinematic features are believed to be implemented through internal action simulations (Graf et al., 2007; Parkinson, Springer, & Prinz, 2012; Sparenberg et al., 2012). In an especially ingenious study of this phenomenon, Graf and colleagues (2007) asked participants to observe an action

sequence of walkers rendered as point-light displays. At the end of each action sequence, the point-light walker disappeared behind a screen. After a determined interval of time, the walker reappeared on the other side of the screen as a static image. Participants were then asked to decide whether the static posture depicted a continuation of the walking cycle. Results showed that people were better at correctly identifying a posture as part of the walking sequence when the occlusion time matched the time that would take the walker to reach the specific static position presented at the end. Thus suggesting that predicting the unfolding of others actions is dependent on internal simulations that integrate the temporal and spatial constraints of action execution.

It is currently accepted that knowledge about the natural statistics of human action is used to predict the spatiotemporal unfolding of observed actions (Sebanz & Knoblich, 2009). In agreement with this view, Neri, Luu, and Levi (2006) hypothesized that expectations about the unfolding of social interactions should facilitate action perception in social settings. To test this hypothesis, the authors invited participants to observe point-light videos of pairs dancing or fighting. Noise dots were scattered around the point-light displays affecting the reliability of visual cues. Crucially, one of the point-display agents (target agent) was either synchronized or desynchronized with its partner. Participants were tasked with detecting the presence of the target agent in the interaction (dancing or fighting). The results showed that visual detection performance was better in interaction sequences where agents were acting synchronously compared to acting asynchronously. The authors interpreted the better detection rate for synchronous agents as resulting from the close match between the observer's internal simulation of the interaction and the actual unfolding of the interaction.

Manera, Schouten, Verfaillie and Becchio (2013) closely replicated this previous study, but instead of dancing or fighting, the pairs of point-light agents executed a communicative interaction. For example, one agent pointed to something on the ground, prompting the other to pick it up. Similarly, to what was observed in the initial study, participants were more successful at detecting agents when the interaction maintained its natural temporal dynamics. Taken together the studies above suggest that the human brain integrates previous expectations about the spatiotemporal dynamics of action execution to generate precise action predictions about the unfolding of social interactions.

There is strong evidence showing that internal predictions of others' actions borrow ones' own motor system. For example, a study using Functional Magnetic Resonance imaging (fMRI) reported increased BOLD responses in premotor areas both when participants prepared to perform the actions themselves and when participants anticipated that the confederate would perform the action (Ramnani & Miall, 2004). Along the same lines, electrophysiological markers of motor preparation, i.e. readiness potential, have been found to precede the observation of movement onset Kilner, Vargas, Duval, Blakemore and Sirigu (2004). These findings suggest that expectations about what others will do next are coded, or at least available, at the motor system level.

In the context of joint-action, motor involvement during action anticipation of interaction partners (as measured by anticipatory ERPs) has been shown to be higher than the activation relating to the motor involvement occurring when anticipating bystanders' actions (Kourtis, Sebanz, & Knoblich, 2010). Therefore, suggesting that motor involvement in prediction is modulated by the social

relevance of the other. Furthermore, recent evidence suggests that the success of interpersonal coordination between partners in a joint-action is supported by partner's motor representations of each other's actions. In this regard, Novembre, Ticini, Schütz-Bosbach and Keller (2012) showed that successful temporal coordination during a musical duet was positively related to the extent to which musicians internally encode each other's actions (measured by corticospinal activation after Transcranial Magnetic Stimulation). Additionally, in this study, self-reported empathy was positively related to the extent to which musicians encoded their partner's actions. Therefore suggesting that the ability to maintain rich motor representations of our partner's actions improves interpersonal coordination.

The studies described above suggest that (a) when passively observing or interacting with another person we build internal models of their actions, which are instrumental for generating predictions about their next motor states, and (b) we use our own motor substrate to support these predictions. In particular, some recent findings suggest that the richness of our motor encodings of others is increased when we engage in social interactions (Kourtis et al., 2010) and is related to our social aptitude traits (Novembre et al., 2012). This evidence is coherent with the pJAMs prediction that joint-action is achieved by generating continuous predictions about the actions of our cooperators.

iii. We encode deviations between partners' motor predictions and action states, i.e. prediction errors.

In pJAM prediction errors are posited to be instrumental in approximating the joint-state of the interaction to its desired goal. Thus, the model predicts that we

monitor deviations between the expected and the estimated motor states of both ourselves and our counterparts in the interaction. Next, I will highlight recent insights into the functional role of prediction errors in social interactions that offer support to pJAM's prediction.

Research comparing the neural processing of mistakes made by oneself with observed errors of other people, suggests that *self* and *other* error monitoring is supported by overlapping neural resources (as measured by error-related potentials on the medial frontal cortex and the motor cortices; van Schie, Mars, Coles, & Bekkering, 2004). However, not all mistakes receive the same amount of monitoring. Recent research indicates that error monitoring is influenced by the social affiliation between the observer and the person that makes the mistake.

In a study carried out by Kang, Hirsh and Chasteen (2010), participants were paired with strangers or friends and observed their partners performing a Stroop task. The results from this experiment showed stronger amplitude of error-related potentials for participants who were paired with a friend compared with participants who were paired with a stranger. This suggests that social closeness enhances the salience of other people's errors. A related study suggests that the impact that social closeness has on observed error monitoring is not specific to the long-term bond that exists between friends. In particular, Carp, Halenar, Quandt, Sklar and Compton (2009) artificially manipulated the closeness between pairs of participants by deceiving pairs of participants about their degree of world-view similarity. Error-related brain potentials, measured while observing the partner perform the Flanker task (Atmaca et al. 2011), were influenced by the perceived closeness to the partner. This study shows that increased monitoring of another person's mistakes is not specific to long-term

social bonds, and can be successfully biased by temporary manipulations of social closeness between individuals. The social context in which we observe another person make a mistake has also been observed to influence error monitoring. For instance, Koban, Pourtois, Vocat and Vuilleumier (2010) investigated the processing of observed errors in cooperative and competitive social interactions. The results revealed higher error-related negativity (ERN) responses occurring when observing a cooperator's mistakes compared to observing a competitor's mistakes. This supports the view that cooperators mistakes are more salient. Thus, this observation is in line with the notion that social error monitoring plays a functional role in cooperative behavior.

This empirical research indicates that humans are able to encode both ones' own and their partners' errors. Most importantly, the literature suggests that this process is extremely permeable to social factors. Taken together these observations are in line with pJAM's prediction of error minimization strategies during cooperative social interactions.

However, one important prediction that follows from pJAMs *action planning layer* has not been empirically observed. This prediction is that the computation of prediction errors about co-partners is fundamental to interpersonal coordination. Although current literature points to the computation of prediction errors during joint-action, these have not been functionally linked to optimal coordination. Therefore, future studies are necessary to test this prediction. Such studies would have to manipulate prediction errors and measure the effect that such manipulation would have on interpersonal coordination during a joint-action. Prediction errors can be modulated either by manipulating expectations (top-down manipulation) or manipulating the sensory

input (bottom-up manipulation). Amplitude measurements of error-related ERPs would allow checking these manipulations. PJAM predicts a link between prediction errors and the ability to coordinate with a partner. Thus, it predicts that the amplitude of prediction errors would have an effect on coordination over time.

iv. The better our models of our partners are, the better we are able to coordinate with them.

pJAM architecture is based on an error minimization strategy, in which prediction errors are used as learning signals to improve predictions about action outcomes. It is commonly observed that we get better at cooperating with others the more we experience it. For example, in team sports, the ability to predict the behavior of teammates greatly contributes to cooperative success (Savelsbergh, Williams, Van der Kamp, & Ward, 2002), whereas in competitive sports predicting the opponent can give competitors the extra edge (Jones & Miles, 1978). Another striking example of specialized interpersonal prediction is the case of ensemble music performances, where musicians need to predict each other's actions to generate a unified sound (Keller, 2014). Three complementary pieces of evidence support the notion that expertise in action coordination is supported by improving the internal models of one's partners (i.e. internal models of their future motor commands, and the sensory consequences of these commands). Firstly, training improves prediction; experts need less information to make accurate predictions and are proficient in anticipating other's errors and deception attempts (Mori & Shimada, 2013). The second finding is that experts show higher levels of motor activation during prediction compared to novices. This finding supports the idea that predictions

are implemented by internal models encoded at the motor level (Aglioti, Cesari, Romani, & Urgesi, 2008). The third finding is that it is easier to coordinate with oneself than to coordinate with the another person. One possible interpretation of this finding is that our models of ourselves are more accurate than our models of others. In other words, our models of ourselves generate better predictions about the sensory consequences of our actions, than our models of others are able to predict the sensory outcomes of others actions (Keller et al., 2007). Taken together these findings are in line with the idea that internal models of our co-partners, which are continuously improved through experience, support interaction.

pJAM offers a framework to encompass evidence of learning and acquired expertise in action prediction. In specific, pJAM comprises a hierarchical predictive stream dedicated to modeling one's interaction partners. Through successive error minimization (achieved by comparing downwards predicted states with upwards estimated states) the theoretical framework is in line with the empirically observed improvement of action prediction through practice.

v. Partners can both anticipate and compensate for each other's actions.

pJAM encompasses the implementation of both compensatory and anticipatory coordination strategies. Next, I will highlight studies from the joint-action literature that reveal the implementation of such coordination strategies.

Compensatory coordination In pJAM this coordination strategy is proposed to be the result of error minimization at the *action-planning level*. It is suggested that continuous optimization of co-task models will iteratively contribute to compensate for deviations between the current state and the desired state of

the interaction. Behavioral evidence for the tendency to compensate for someone else's movements is found in a few experimental studies. Sebanz and Shiffrar (2007) asked participants to watch someone trying to balance on a slippery surface. The authors measured participants' spontaneous body tilt during action observation. The results showed that participants made small movements compensating for the actor's imbalance. For example, participants tilted to the left when the actor was about to fall to the right side. These findings suggest that individuals involuntarily execute compensatory movements when observing an action that does not match the desired goal, thus supporting the possibility that partners in a joint-action compensate for each other's deviations from the shared goal. It should be noted, that this study doesn't show that compensatory strategies are used in joint-action. Rather it shows that individuals have the tendency to complete each other's actions, thus giving preliminary support to the idea that such compensatory tendencies could be harnessed to cope with the interpersonal coordination demands of joint-actions.

Relevantly, a follow-up study shows that spontaneous compensatory movements during action observation were modulated by whether observers share the same goal as the observee (Häberle et al., 2008). Findings from this study indicated that while observing a cooperator (i.e. a participant who shares the same goal) tended to perform small movements congruent with goal achievement. However, when observing a competitor (i.e. a participant who has an opposite goal) the spontaneous compensatory movements were incongruent with goal achievement. This evidence further supports the notion that partners who share the same goal compensate for each other's deviations from the desired goal. Thus further supporting the potential value of compensatory strategies in joint-action.

Anticipatory coordination In many joint-action situations, one partner has to prepare or even initiate a complementary movement before fully receiving information about the co-partner's behavior. This requires partners to integrate into their motor planning predictions about what the other will do next (Keller, 2007). Thus leading to anticipatory coordination.

In pJAM, predictions about the partner's desired next state (discharged by co-task generative models) flow down the hierarchy for comparison with bottom-up information, but importantly, these predictions are also relayed horizontally to the models responsible for generating a motor command (i.e. self co-task models). Thus, motor commands integrate predictions about the partner's next actions. This horizontal sharing of information, between 'partner' and 'self' predictive cascades, allows for anticipatory coordination.

Anticipatory coordination has been widely reported in joint-action studies (Sebanz & Knoblich, 2009). For example, Pecenka & Keller (2011) observed that when asked to tap in synchrony with auditory sequences, some participants revealed a tendency to adapt their tempo to predicted auditory events (anticipatory strategy), while other participants followed the strategy of tracking past events (compensatory strategy). The authors subsequently tested how these individual tendencies influence interpersonal coordination. They hypothesized that if temporal prediction improves interpersonal coordination, pairs of people that show predictive tendencies would coordinate better than pairs of people that showed the tendency to track past events. The results showed that pairs who tended to anticipate each other's actions, instead of tracking what each other do at each moment in time, have a better ability to synchronize. This supports the notion that predicting is a better strategy than

compensating when it comes to interpersonal coordination. In fact, studies suggest that the best coordination partners capitalize on the relationship between action predictability and interpersonal coordination by exaggerating their behavior (Goebel & Palmer, 2009) and diminishing the variability in their actions (Vesper, van der Wel, Knoblich, & Sebanz, 2011), thus making themselves easier to predict.

pJAM proposes a theoretical framework for the implementation of anticipatory and compensatory coordinative structures. These coordinative strategies are endogenous to the overall functioning of the hierarchical predictive system. Hence, pJAM offers a processing structure for the empirical evidence described above.

2.3.3 Sensory routing layer

pJAM predicts that received sensory input is continuously compared with sensory predictions about each partner's action outcomes. The diagram in Figure 7 represents the sensory routing layer in pJAM. Support for the notion that sensory predictions about oneself and one's partners are compared in parallel to the incoming sensory input comes from studies observing interpersonal sensory cancellation. It is well known that the process of matching between received and predicted sensory action outcomes is sometimes used to filter out the expected sensory results of an action, a phenomenon known as sensory cancellation (Blakemore, Frith, & Wolpert, 1999). One famous observation of the sensory cancellation effect lies on the fact that it is hard, if not impossible, for one to tickle oneself (Blakemore, Wolpert, & Frith, 2000).

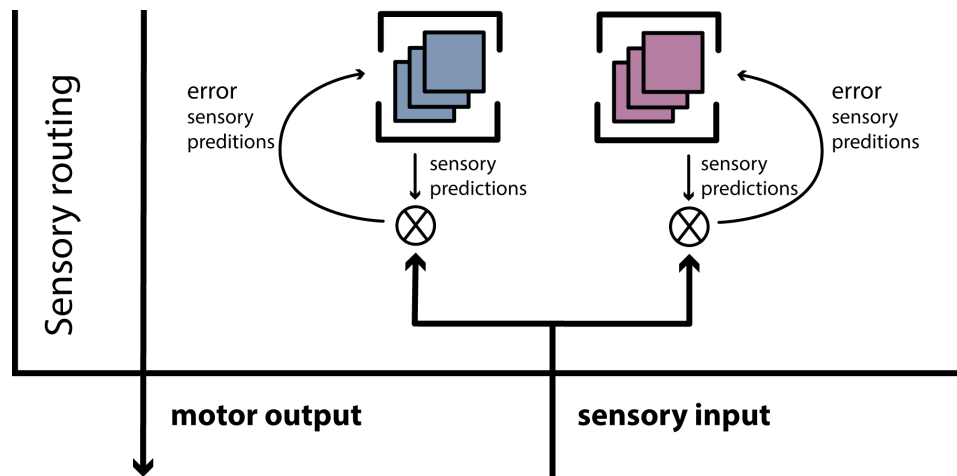


Figure 7 Sensory routing layer in pJAM.

Sato (2008) investigated sensory cancellation in the social realm. The authors presented participants with an auditory tone after themselves pressing a key, after observing another person pressing the key, or unexpectedly. Participant's perceived the tone as being less intense both when they or the other person pressed the key, compared to the unexpected condition. The authors interpreted that this attenuation of sensation occurred because participants maintained predictions about the outcomes of other people's actions, which were used to attenuate the sensation of sensory events.

Furthermore, binding sensory information to action outcomes is considered to be the basis of the sense of agency (Obhi, 2012; Schüür & Haggard, 2011). Studies of the sense of agency in joint-action settings suggest that partners in a joint action can differentiate between their own and another person's contributions to the sensory outcomes of a joint-action (Loehr, 2013), and that this effect is influenced by partner's experience with the task (van der Wel et al., 2012).

These observations indicate that the sensory inflow of information during a social situation is allocated to being an outcome of one's or other's actions. This process is crucial for the pJAM proposal because the model relies on the upward swipe of reliable sensory information to train one's models and predictions about oneself and one's partners. In pJAM, the sense of agency might be instrumental at the higher -level of goal representation where desired joint-states are broken down into models representing different probabilistic options through which the joint task can be shared across the partners.

However, these previous studies have not tested whether sensory routing (i.e. linking sensory outcomes to individual actions) is necessary for optimal joint-action implementation. One possible way of testing this would be to devise a task that manipulates how easy or difficult it would be to allocate sensory consequences to individual actions, and measure if this modulation would have an effect on partner's ability to achieve a shared goal.

2.4 Discussion

The main goal of this chapter was to capture the inner workings of joint-action employing hierarchical predictive notions. The overall success of this endeavor can be assessed by asking whether pJAM meets the minimum requirements for an architecture of joint-action as proposed by Vesper and colleagues (2010). Next, I will summarize how pJAM addresses each of the proposed minimal requirements.

(1) The architecture must support shared goal and corresponding individual task representations. pJAM proposes that shared goals are represented in a probabilistic fashion at the higher-level layer of the hierarchy - *goal*

representation layer. In the layer immediately below, the *action-planning layer*, the framework comprises parallel probabilistic models of both one's own and partner's motor contributions to the desired joint-state. Comparisons between these adjacent layers lead to the continuous actualization of shared goal representations.

(2) The architecture must support the processes of monitoring and predicting each partner's actions. pJAM comprises parallel predictive cascades for each intervenient in the joint-action. The core cognitive process of this model is *prediction*, which is posited to occur at different levels of abstraction (i.e. from goal representation, to motor planning and sensory expectations) for each intervenient in the interaction.

(3) The architecture must allow for continuous coordination. pJAM supports the implementation of anticipatory and compensatory coordination strategies in an endogenous way to the overall functioning of the hierarchical predictive system.

Overall I consider that pJAM successfully matches the minimal requirements for an architecture of joint-action as defined by Vesper and colleagues (2010). I consider that pJAM extends previous sensorimotor accounts of motor control in social situations (Wolpert et al., 2003), by proposing a framework that attempts to address implementation challenges that are specific to joint-action situations (e.g. shared goals, action prediction and coordination).

Furthermore pJAM offers a preliminary insight into one long-standing open question identified in previous joint-action literature reviews: "One main challenge for future work seems to be to understand how lower-level processes

like action simulation and higher-level processes like verbal communication and mental state attribution work in concert, and under which circumstances they can overrule each other” (Natalie Sebanz & Knoblich, 2009; p. 365). This question remains far from being fully answered. However, hierarchical predictive mechanisms offer a promising solution that links processes occurring at different levels of representation.

Another joint-action implementation challenge to which pJAM offers insight is the question of how joint-actions are led to successful completion given that shared goals and task divisions are initially underspecified. Hierarchical predictive processing allows a system to reach a solution through Bayesian inference (Friston et al., 2011; Friston, 2003; Todorov, 2004) and thus represents a powerful way to deal with under-specified and mutable problems such as the unfolding of a joint-action in space and time. It is important to note that the proposed framework (pJAM) does not attempt to make illations about brain organization, but rather to use knowledge about how the brain solves action-perception computational problems to think about join-action.

Finally, the exercise of structuring current evidence according to the predictions proposed by pJAMs organization has revealed that both contextual factors (e.g. interaction goal, relationship between partners) and personal factors (e.g. personality traits, beliefs) modulate the predictive hierarchy cascades (Colzato, de Bruijn, et al., 2012; de Bruijn et al., 2008; Iani, Anelli, Nicoletti, Arcuri, & Rubichi, 2011; Kuhbandner et al., 2010). The current sensorimotor framework does not offer an account of these factors interact with joint-action mechanisms. The challenge of understanding the two-way influences between what are commonly considered to be *social phenomena* (e.g. social relationships and

personality) and what are considered to be *cognitive phenomena* (e.g. motor control) goes well beyond the narrow domain of joint-action addressed here. In fact, core fields of cognitive psychology research, such as attention (Ristic & Enns, 2015), are in search for new theoretical ideas that can better capture cognition in its personal and social environment.

3 Sensitivity to attention control in action prediction

A recent theory suggests that social cognition involves a predictive model of other people's attentional states (Webb & Graziano, 2015). In this conceptualization attention is defined as a data-handling mechanism. Allocating attention means prioritizing one information processing operation rather than others. Graziano (2013) notes that the allocation of attention is inherently linked to behavioral control. The author proposes that attention often has a quality of control on behavior. He further extrapolates that this quality of attention is at the root of social cognition. In this view, modeling other people's attentional states is one of the most important cognitive mechanisms we use to predict their future behavior (which is crucial to maintaining social interactions). The idea is that we continuously build and actualize sophisticated models of other people's attentional states. Graziano (2013) proposes that different sources of information can contribute to social models of attention, such as contextual information, facial expressions, gaze allocation, movement kinematics, etc. According to this proposal, we continuously gather cues that allow us to internally simulate someone else's attentional states and in this way make predictions about their future behavior.

Previous studies of social perception report acute human sensitivity to where another's attention is aimed. Here I present evidence that human social understanding involves not only knowing where someone else is attending but also sensitivity to how the other's attention has been controlled. The control of attention is among the most widely studied topics in all of cognitive science

(Corbetta & Shulman, 2002; Posner & Rothbart, 2007; Posner, 1980). Attention is *endogenous* when controlled voluntarily, such as the goal-directed intention to attend to a particular event in the environment. Attention is *exogenous* when controlled by environmental factors, such as a spatially local change in appearance or sound. In this thesis, I will present a series of experiments designed to probe third-person perception of attention control states. These studies followed a two-stage methodology. In a first stage, presented in section 3.1, I created video-clips of actors performing actions under exogenous and endogenous control. The exogenous control condition was created by externally directing actor's actions to a specific target (directed actions). The endogenous control condition was created by letting actors choose the target of their actions (chosen actions). This manipulation follows Graziano's (2013) conceptualization of attention as data prioritization. In directed actions, the external stimulus is prioritized (exogenous attention control). In chosen actions, the internal decision-making is prioritized (endogenous attention control). In the second stage of this project, I used these two categories of video-clips to test observers' sensitivity to actors' attention control states as expressed through their reaching actions.

A first experiment revealed that observers were faster at predicting the end-target of someone else's actions when the actor choose the action's target (endogenous attention control) compared to when the actor was directed to the target (exogenous attention control). Thus suggesting that humans are able to capitalize on subtle differences in bodily cues that occur when someone else's attention is controlled by an internal choice versus an external signal in order to improve their predictions about someone else's actions (presented in section 3.2). Follow-up experiments showed that (1) sensitivity to attention control gives

observers a reactive advantage in social interactions (presented in section 3.3), (2) sensitivity to attention control is not consciously accessible (presented in section 3.4), (3) attention control signals were widely distributed over the actor's body, though stronger in the torso and limbs than in the head (presented in section 3.5), (4) the signal was available early on in the movement (presented in section 3.6), (5) and finally that sensitivity in the kinematic responses of observers was correlated with observer's social aptitude, as measured by the Autism Quotient Scale (Baron-Cohen et al., 2001; Ruzich et al., 2015; presented in section 3.7). Together these experiments suggest that social cognition involves the predictive modeling of other's attentional states. Next, I will detail the methods and results of these experiments, and discuss the corresponding findings in light of the current literature on social cognition.

3.1 Methodology

This research project probed whether observers were sensitive to someone else's attentional control states. My colleagues and I operationalized this research question using a two-stage methodology. In the first stage – stimuli construction stage - I recorded videos of actors reaching to one of two possible targets while either choosing (endogenous attention control) or being directed (exogenous attention control) to one target. For simplicity, from now on I will refer to the endogenous attention control condition as the “chosen” condition and the exogenous attention control condition as the “directed” condition. In the second stage – experimental stage – I presented observers with videos of both conditions (chosen and directed) in randomized order and measured their responses. This was done with the goal of assessing observers' sensitivity to actors' attentional states. The experimental stage will be addressed further on in

Chapters 3.2 to 3.7. Next, I will focus on the stimuli construction stage. I will start by describing the procedures followed during the stimuli recording. Afterward, I will explain the process used to construct a stimuli set that was equated in temporal cues between conditions. And finally, I will present a manipulation check showing that the stimuli set portrays subtle kinematic differences between conditions.

3.1.1 Stimuli recording

As shown in Figure 8, actors were filmed reaching to two possible targets after choosing (endogenous control) or being directed (exogenous control) to a target. To assist in the creation of the set of videos, actors were recruited from the same population as observers. A total of 11 potential actors were filmed. Five actors were excluded due to technical difficulties with the recordings. From the remaining 6 actors, we selected 4 (2 females, ages 19-21) that followed instructions in all respects and consented to have their reaches recorded for presentation to other participants as stimulus materials.

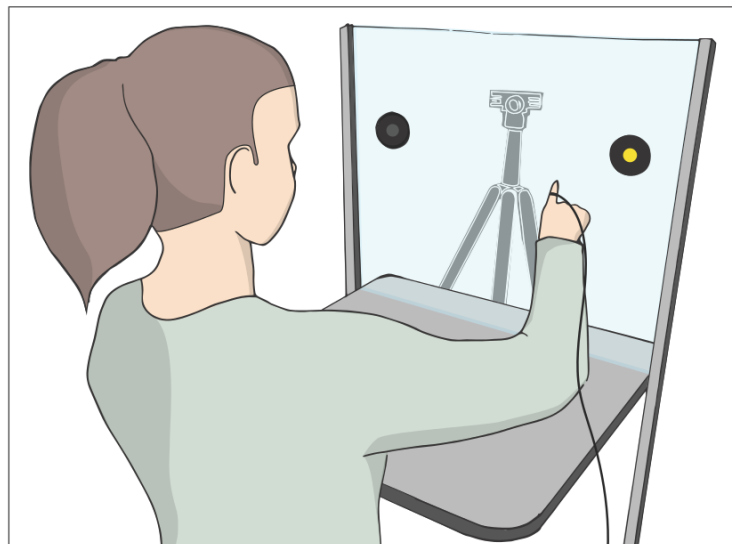


Figure 8 Illustration of the method from the actors' perspective. Actors were filmed through plexiglass reaching to two possible targets. On "chosen" trials both locations were lit and actors had to choose (not shown), on "directed" trials only one location was lit and actors were directed to reach to that location (as shown).

Actors were seated at a table facing a Plexiglas panel positioned 56 cm from the table edge. Actors were filmed at 50 fps, 800x800 pixels, using a Flea3 camera placed on the opposite of the Plexiglass frame. Two LED lights facing the actor served as cues. The LED lights were positioned 20 cm to the right and the left side of a central fixation point located at the average actor's eye-level. The videos start at cue presentation and end after the reach is completed. Actors were instructed to begin each trial by fixating the central point. This was followed by the simultaneous onset of an auditory beep and the visual cue(s). On directed trials, one of the two LEDs was illuminated randomly, and actors were instructed to reach and touch it as rapidly as possible; on chosen trials both LEDs were lit and the instructions were to rapidly choose one LED to touch. Previous studies have characterized how choices are expressed in reaching movements (Gallivan & Chapman, 2014). Using reaching movements, as stimuli will allow us to study observers' sensitivity to action control. Actors were instructed to make each choice in the moment and to try to select the left and right LEDs about equally often, which they did (50.87% right overall). The inter-trial interval was kept deliberately short (1000 ms following each response) in order to prevent strategic choosing in advance of the cue. Each actor completed a total of 100 trials in both the chosen and directed conditions. Importantly, the LEDs were not visible in the videos. Figure 9 shows an example of the video-clips framing.

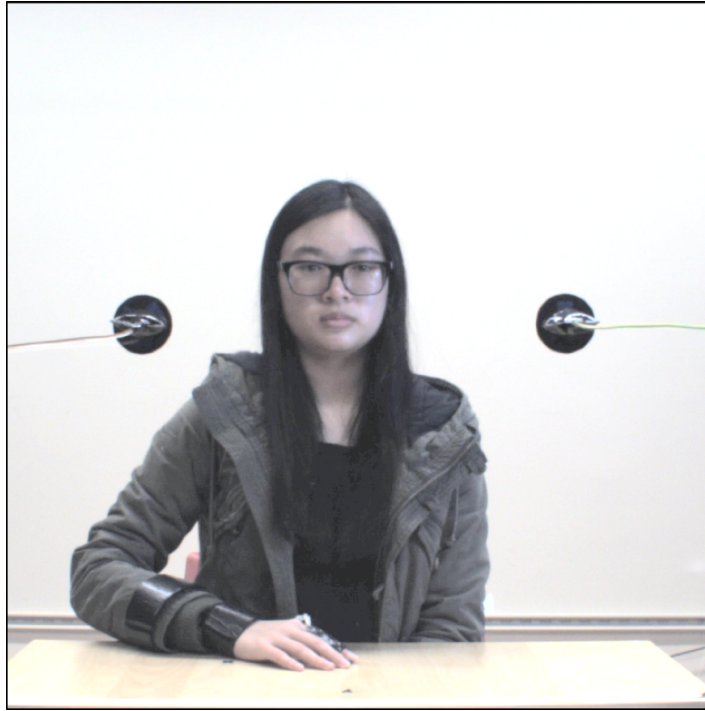


Figure 9 An example of the video-clips framing. Critically, the LEDs are not visible in the stimuli.

3.1.2 Stimuli selection

Because our goal was to test for sensitivity to how the reaches were controlled, not sensitivity to overt differences in the onsets or movement times of the reaches, we first eliminated temporal cues that might distinguish chosen from directed actions. From a pool of 800 video-clips (4 actors x 200 trials), we first selected 100 videos at random from each actor and ranked them according to their initiation and movement times. t-tests evaluated whether there were significant differences in either initiation or movement times. If a test was positive, the videos in the tails of the distribution were replaced by randomly selected from the remaining videos until no differences remained.

This resulted in 100 test clips for each actor (400 total), with an equal number of chosen and directed reaches. Figure 10 shows the initiation time and movement time for the four actors. Initiation time was not significantly different between conditions, $t(49) = -0.81; 1.49; 0.29; -0.95$, nor was movement time, $t(49) = 0.06; -0.87; -0.78; 0.10$, for actors 1 to 4 respectively. In specific, mean differences between conditions in initiation time ranged from - 6.61 to 16.66 ms and were not significant for any of the actors, $t(49) = -0.81; 1.49; 0.29; -0.95$, for actors 1 to 4 respectively. Mean differences in movement times between conditions ranged from -11 to 1.3 ms and were also not significant, $t(49) = 0.06; -0.87; -0.78; 0.10$, for actors 1 to 4 respectively. However, there were still naturally occurring differences between actors, both in their overall initiation time, $F(3,392)=75.09$, $p < .001$, $\eta^2 = .363$ (means in rank order $A3=302$ ms, $A2=299$ ms, $A4=282$ ms, and $A1=205$ ms), and in their movement time, $F(3,392)=771.23$, $p < .001$, $\eta^2 = .855$ (means in rank order $A2=757$ ms, $A4=619$ ms, $A3 =589$ ms, and $A1=387$ ms).

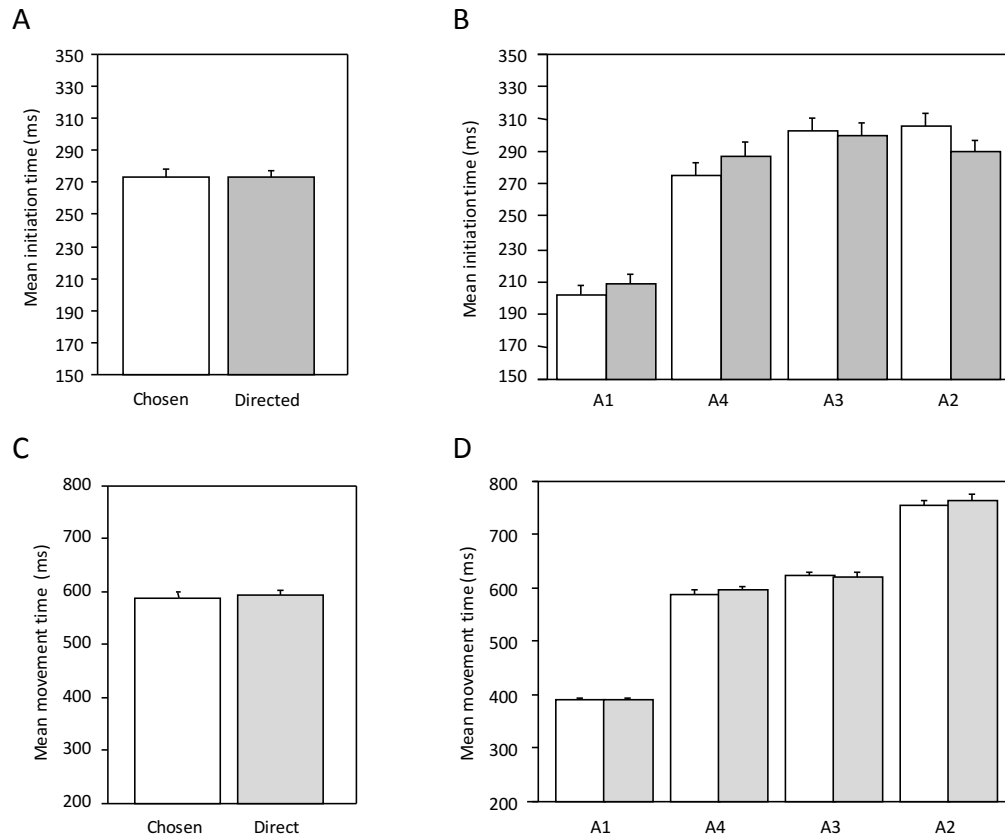


Figure 10 (A) Overall means of the actors' movement initiation times for "choice" and "direct" actions. **(B)** Means of the movement initiation times for each of the 4 actors (A1 to A4) for "choice" and "direct" actions. **(C)** Overall means of the movement times for "choice" and "direct" actions. **(D)** Means of the movement times for each of the 4 actors (A1 to A4) for "choice" and "direct" actions. Error bars correspond to one standard error of the mean.

3.1.3 Manipulation check

Attention is *endogenous* when controlled voluntarily, such as the goal-directed intention to attend at a particular event in the environment. Such control is relatively slow, effortful, and can be sustained. Attention is *exogenous* when controlled by environmental factors, such as a spatially local change in appearance or sound. By way of contrast, this mode of control is fast, effortless,

and short-lived (Corbetta & Shulman, 2002; Posner & Rothbart, 2007; Posner, 1980).

In order to determine whether actors' reaches were influenced by attention control, we tested for subtle kinematic differences between conditions. Our hypothesis was that chosen reaches would express the decision required on those trials, with longer times to peak acceleration and curved trajectories reflecting the process of choosing versus reacting (Gallivan & Chapman, 2014). To test this hypothesis, we compared chosen and directed reaches on eight kinematic measures as shown in Table 1.

The eight kinematic measures were the following: *Peak velocity*, indexing the maximum velocity achieved during the reach; *Time to peak velocity*, indicating the time elapsed from movement initiation (i.e. finger lift-off) until the peak velocity was achieved. *Horizontal trajectory curvature*, this measurement indexed the amount of inward curvature in the horizontal trajectory. Reaches that followed a central path before committing to the end-side, had higher horizontal trajectory curvature values, compared to movements that directly followed a path to the end-side; *Side-commitment angle* indexes the angle depicting the transition from neutrality (i.e. the most central point in the trajectory) to side selection (i.e. the most outwards point to the end-side of the trajectory). Thus, higher side-commitment angles correspond to more marked transitions from neutrality to side-selection compared to movements with lower side-commitment angles; *Side-commitment distance* corresponds to the length of the side-commitment angle. Longer angles depict movements in which the decision-making is distributed through the reach, whereas shorter angles depict faster decision-making transitions; *Vertical trajectory curvature* indexes the

amount of vertical curvature in the movement. High vertical trajectory curvatures correspond to movements that deviate from the straight trajectory from home to target by making an upwards curve in the vertical dimension; *Ascending angle* corresponds to the angle between lift-off point and the utmost point in the vertical trajectory, higher ascending angles correspond to more abrupt vertical traveling reaches; *Ascending distance* corresponds to the length of the ascending angle. Reaches that take more time to achieve their maximum vertical location peak depict longer ascending distances.

Four of the eight kinematic measures were consistent with the hypothesis, and none trended in the opposite direction. In comparison to directed reaches, chosen reaches had a marginally larger *time to peak velocity*, a higher mean *vertical trajectory curvature*, a larger mean *ascending angle*, and a longer *mean ascending distance*. These findings support the hypothesis that longer times to peak acceleration and curved trajectories reflect the process of choosing a target location compared to simply being directed to the same location. This is because choosing a target entails the added process of deciding which action plan to implement (reach left vs. right). Previous studies show that in fast arm reaches, directional decisions are expressed in the curvature of trajectories and their velocity profiles (Gallivan & Chapman, 2014).

Table 1. Means of eight kinematic measures taken on the distribution of reaches used as stimulus materials in the Experiments.

	ANOVA test				
	Chosen (Mean)	Direct (Mean)	F-test	p	η^2
Peak velocity	2256	2200	3.471	.063	.001
Time to peak velocity	191.3	187.1	3.065	.081	.003
Horizontal trajectory curvature	18390	18703	0.808	.369	.002
Side-commitment angle	60.26	59.55	0.927	.336	.002
Side-commitment distance	121.5	120.8	0.047	.829	.000
Vertical trajectory curvature	42138	41035	9.564	.002	.015
Ascending angle	62.57	62.34	3.396	.066	.007
Ascending distance	412.8	411.0	5.451	.020	.010

Notes. *Peak velocity* = maximum velocity achieved during the reach. *Time to peak velocity* = time elapsed from movement initiation until peak velocity. *Horizontal trajectory curvature* = index of the amount of inward curvature in the horizontal trajectory. *Side-commitment angle* = angle depicting the transition from neutrality (most central point in the trajectory) to side selection (most outwards point to the end-side of the trajectory). *Side-commitment distance* = length of the side-commitment angle. *Vertical trajectory curvature* = index of the amount of curvature in the vertical trajectory. *Ascending angle* = angle between lift-off point and the utmost point in the vertical trajectory. *Ascending distance* = length of the ascending angle.

In addition to these kinematic differences between conditions, each of the eight measures differed significantly between actors, as one might expect, given each actor's individual style of responding. However, with only one exception, these differences in individual actor style did not interact significantly with the reported main effects for chosen versus direct reaches. The exception was that peak velocity was significantly higher for chosen than directed reaches for actor 1, $t(49) = 3.15$, $p = .01$, whereas the other actors did not differ on this measure.

Two types of trade-offs are typically observed in reaching movements: trade-offs between initiation and movement times, and trade-offs between movement time and trajectory curvatures (Schmidt & Lee, 2011). Consistent with this expectation, Table 2 shows several medium to strong significant correlations between temporal and kinematic measures. To help us understand whether these relationships pointed to a common underlying factor, we submitted the eight kinematic measures along with the temporal features for each reach in the stimuli set to a principal component analysis (PCA). To further focus this analysis on only those kinematic effects that distinguished chosen from direct reaches, we performed the PCA after first computing z-scores for each measure. These z-scores were computed by dividing the difference between the measurement value and the mean of that measurement for the correspondent actor per the standard deviation of that measurement for that actor. This meant that there were no longer any differences between actors in these measures, nor interactions between actor and condition.

Table 2. Correlations between temporal and kinematic measurements.

	Initiation time	Movement time	Total time	Peak velocity	Time to peak velocity	Horizontal trajectory curvature	Side-commitment angle	Vertical trajectory curvature	Ascending angle	Ascending distance
Initiation time	–	–	–	–	–	–	–	–	–	–
Movement time	.50**	–	–	–	–	–	–	–	–	–
Total time	.74**	.95**	–	–	–	–	–	–	–	–
Peak velocity	-.54**	-.78**	-.80**	–	–	–	–	–	–	–
Time to peak velocity	.44**	.45**	.51**	-.65**	–	–	–	–	–	–
Horizontal trajectory curvature	-.07	-.04	-.06	.12*	-.13**	–	–	–	–	–
Side-commitment angle	-.20**	-.12*	-.15**	.133*	-.07	-.19**	–	–	–	–
Side-commitment distance	-.00	.2**	.12*	-.03	-.07	.11*	-.44**	–	–	–
Vertical trajectory curvature	.22**	.54**	.50**	-.34**	.05	.13*	-.01	.10*	–	–
Ascending angle	-.20**	-.23**	-.25**	.25**	-.22**	-.12*	.34**	.04	.16**	–
Ascending distance	-.25**	-.36**	-.37**	.45**	-.45**	-.04	-.01	.01	.15**	.22**

Note. Degrees of freedom were 498 for all correlations. ** corresponds to p-values <.001; * corresponds to p-values <.05.

Visual inspection of a scree plot, showing the total variance accounted for by the PCA as a function of an increasing number of potential components revealed a plateau after the first component. The first component alone accounted for 21.98% of the kinematic variability. The measurement loadings on this component were generally positive for chosen reaches and negative for directed

reaches, leading to a significant difference overall, $F(1,392)=5.20$, $p=.02$, $\eta^2 = .01$. Thus showing that the first PCA component successfully distinguished between chosen and directed reaches.

Table 3 shows the first component weights associated with each measure. Inspection of these component loadings shows positive weights ($\geq .3$) for movement time, total time, side-selection angle, vertical area under the curve (v-AUC), and ascending angle. No negative loadings were relevant ($\leq -.3$). This pattern supports the hypothesis that chosen reaches reflect endogenous orienting by portraying a reaching pattern in which slower movements take longer to achieve peak velocity, have marked transitions from center to end-side, and display arched vertical trajectories. Whereas exogenous orienting has a reactive nature, which is reflected by a relationship between faster reaches which tend to quickly achieve peak velocity and have straighter trajectories from home to target.

Table 3. Principal component analysis, first component weights.

	Weights
Initiation time	-.108
Movement time	.674
Total time	.397
Peak velocity	.032
Time to peak velocity	-.010
Horizontal trajectory curvature	.093
Side-commitment angle	.389
Side-commitment distance	.107
Vertical trajectory curvature	.794
Ascending angle	.745
Ascending distance	.244

3.1.4 Summary

In this section, I presented a new methodological approach that allows for the dissociation between observers' sensitivity to an action end-location and observers sensitivity to action control. I have described the procedures followed in the stimuli construction stage. These resulted in a video-library consisting of 100 video-clips of 4 different actors. For each actor, the library has an equal number of chosen and directed reaches. Importantly, whereas the temporal differences were equated between conditions, the stimuli portrayed kinematic differences between conditions, indicating greater decisional cues in chosen reaches compared to direct reaches. In the next sections, I will describe a series of experiments that used this stimulus set to test whether observers, blind to the condition under which the actors were reaching, were nonetheless sensitive to actors' attention control states.

3.2 Are humans' sensitive to attention control in others?

In a series of experiments my colleagues and I set out to study third-person perception of attention control. The first experiment tested observers' sensitivity to actors' attentional states. Figure 11 illustrates the person perception task from the observer's perspective. I presented chosen and directed videos to observers. Observers were asked to rapidly indicate the target of the actor's reach. Two alternative hypotheses were considered. If observers based their predictions solely on the kinematic cues of the reaching actions, they should fare better on directed trials since those reaches take less time to reach peak acceleration and moved more directly through space to the target location. I call this the physical signal hypothesis and contrast it with what I call the social prediction hypothesis. In this later hypothesis, if observers can capitalize on bodily cues reflecting the actors' internal process of choosing a target, they should be faster to predict chosen actions compared to directed ones. Thus, the results would show a "choice advantage".

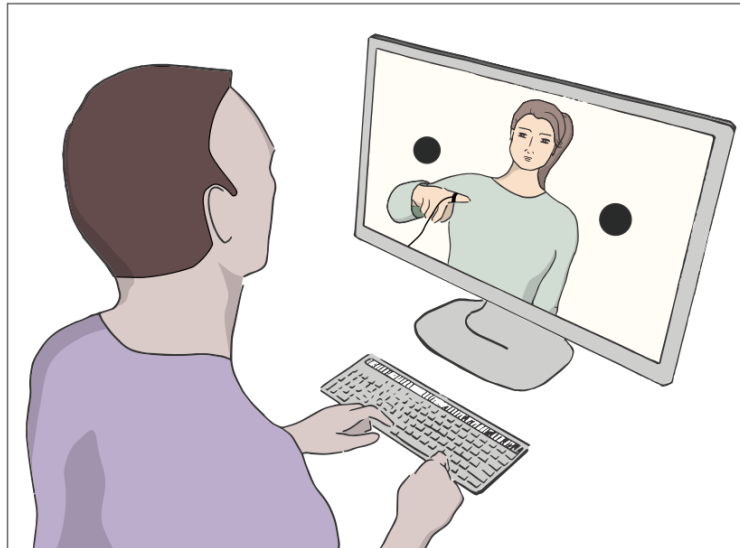


Figure 11 Illustration of the method from the observers' perspective. Observers respond to each video by pressing a spatially mapped key press as rapidly as possible to indicate where the actor was reaching.

3.2.1 Method

Observers. Thirty participants (18 female, 4 left-handed) with a mean age of 21.9 ($SD = 4.6$) were recruited from the University of British Columbia Human Subject Pool to serve as observers. The only exclusion criterion was failing to report normal or corrected to normal vision. Observers received partial course credit in exchange for one hour of time, as approved by the UBC Behavioral Research Ethics Board. All participants read and signed a written informed consent document prior to testing. The document described the procedures, informed participants they would receive partial credit in a qualifying Psychology course, and that they could withdraw from participation at any point without

penalty. A power analysis indicated that with 30 observers there is a 76.35% chance of detecting an effect size of 0.5 with a two-tailed t-test and alpha at .05.

Procedure. Figure 11 illustrates the experiment from the observer's perspective. Observers were simply asked to respond to each actor's reach with a spatially mapped speeded key press to indicate whether the actor was reaching to the left or right as rapidly as possible. However, they were also told to minimize their errors by making no more than 10-20%. Accuracy feedback was not provided to the observers. Each trial began with the observer's index fingers resting on these keys and their eyes on a fixation cross for 1-1.5 seconds. This was followed by a video-clip showing an actor reaching for a target, and the observer's response. Each video played to completion independently of the actors' responses. Critically, observers could not see the cues for action that were visible to the actors.

The session began with 8 practice trials, involving an actor that was not used in the main test. Observers were told that actors would reach left and right an equal number of times and at random. The 100 trials for each actor were shown in a single block, in counterbalanced order across observers, and observers were given a short break between each of the four blocks of trials. At the conclusion of the session, observers completed the 50-item Autism-Spectrum Quotient (AQ) (Baron-Cohen, et al., 2001).

3.2.2 Results

Figure 12 shows the mean correct response time (RT) in the chosen and directed conditions overall (panel A) and for each of the four actors ranked by the speed with which observers could discriminate whether they were pointing left or right

(panel B). Panels C and D show the data after each observer's correct RT had been converted to z-scores in order to control for the larger differences in the mean speed and variance of the four actors' reaches (panel B). Both of these analyses make it clear that RT was faster in the chosen than in the directed condition for each of the four actors (A1 to A4). This conclusion was supported by the following analyses.

Incorrect trials and responses more than 3 standard deviations from the mean were excluded. Response accuracy, correct RT, and z-scores of correct RT were each subjected to repeated-measures ANOVA examining the effect of condition (chosen, directed) and actor (A1 to A4). Z-scores were computed on the correct RT values by subtracting each observer RTs from the mean RTs of that observer to the corresponding actor, and dividing this by the standard deviation of the observer's RTs for this actor.

Observers responded correctly on 81% of trials (standard error of the mean = 0.7%), with significant differences in accuracy between actor videos, $F(3,87) = 15.31$, $p < .001$, $\eta^2 = .346$ (in rank order A3 = 85%, A2 = 83%, A4 = 81%, and A1 = 75%), but no differences between condition ($p > .25$), nor an interaction ($p > .09$). The observation that observers have a rate of incorrect responses close to 20%, which is relatively high for a movement direction task, suggests that participants were following the instructions by responding before the full unfolding of the actors' reach. Analysis of correct RT indicated significant main effects of condition, with responses to chosen reaches made significantly faster than responses to directed reaches, $F(1,29) = 70.39$, $p < .001$, $\eta^2 = .708$, and actor $F(3,87) = 31.48$, $p < .001$, $\eta^2 = .521$, and an interaction, $F(3, 87) = 3.21$, $p < .03$, $\eta^2 = .100$.

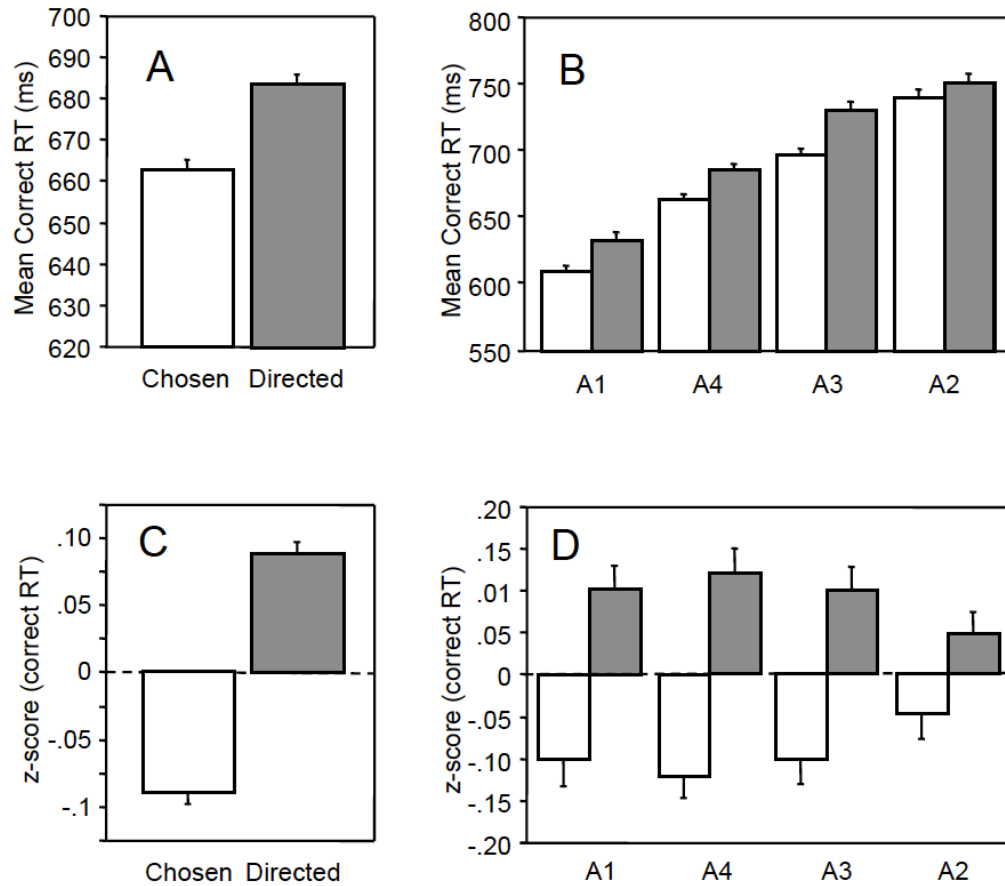


Figure 12 (A) Mean correct response time (RT) in the experiment reported in Chapter 3.2. Error bars are +/- 1 standard error. (B) Mean correct RT for each of the four actors. (C-D) The data in A-B after each observer's correct RT has been converted to z-scores in order to standardize the distributions for individual differences in mean speed and variance.

To test whether the choice advantage was influenced by observer accuracy, we included overall accuracy as a between-subjects factor, after dividing the participants into more accurate (mean accuracy = 93% correct) and less accurate (mean accuracy = 69% correct) halves. This indicated no interaction of condition x accuracy, $F(1,28) = 1.38$, $p < .25$, $\eta^2 = .012$, with both groups showing a 19 ms advantage in the chosen condition. This indicates that the difference between

responses to chosen and directed actions is not due to a speed-accuracy trade-off.

Z-scores were computed by subtracting the mean of observer correct responses to each actor from observers' raw scores and then dividing the difference by the correspondent standard deviation. This allows me to consider the effect of condition after controlling for the large variability in reaching behavior between actors. In these analyses, the main effect of actor was no longer significant, but there was a main effect of condition, $F(1,29) = 80.51$, $p < .001$, $\eta^2 = .735$. In the experiments that follow we undertake a similar analysis of accuracy, correct response time, and z-scores, but for simplicity we will only present graphs showing the mean z-scores and their standard errors. None of the conclusions differed depending on whether an analysis was based on raw RT or on z-scores.

3.2.3 Discussion

The results showed that observers were faster to discriminate the location of an actor's reach when it was chosen than when it was directed. Thus indicating a "choice advantage": predicting a chosen action is easier than predicting a directed action. This suggests that observers are sensitive to actors' bodily cues reflecting the internal process of intentionally choosing the end-target of the action, thus, the results are in accordance with the *social prediction* hypothesis. Overall, these findings are consistent with the claim that social awareness involves a predictive model of the attentional state of others, and that modeling others' attention includes not only information about *where* the other is attending, but whether the control of attention is endogenous or exogenous (Graziano & Kastner, 2011; Graziano, 2013).

The observation that humans are sensitive to social attentional states fits well with a larger number of recent findings indicating that the perception of another persons' inner states goes far beyond the most obvious, widely investigated, cues of facial expression, posture, paralinguistic movements and gestural conventions (Johnson & Shiffrar, 2013). These new studies reveal that, on one hand, seemingly neutral movements encode relevant social information, and on the other hand, observers are sensitive to this information. For example, the way one reaches for and grabs a Lego piece during a game has been shown to encode an individual's intention to cooperate or compete with a partner, changing the way the partner moves when it's their turn to play (Cristina Becchio, Sartori, Bulgheroni, & Castiello, 2008; Manera, Becchio, Cavallo, Sartori, & Castiello, 2011); the kinematics of running gives away one's intention to deceive a sports opponent (Mori & Shimada, 2013); and that despite the conventional wisdom of maintaining a neutral face while playing poker, the value of the poker hand is unconsciously expressed in arm movement kinematics during the game and can be picked up by observing players (Slepian, Young, Rutchick, & Ambady, 2013). Each of these findings implies that when we execute actions, we are far less opaque than we thought ourselves to be. Inner cognitive processes are constantly being expressed and are thus available in the public realm as relevant stimuli during action observation. And when we observe others' actions, we are remarkably sensitive to subtle body cues revealing their inner states. This raises the question of how these cues are integrated with other information in making a response to observed behavior during social interactions. Next, I will report an experiment investigating whether social sensitivity to attention control offers observer's an advantage in a social interaction setting.

3.3 Does sensitivity to attention control contributes to a reactive advantage in social interactions?

The findings reported in Chapter 3.2 indicate that observers are sensitive to someone else's attentional states. But can this perceptual sensitivity be utilized in social interaction settings? To investigate whether sensitivity to attention control could be translated into a motor response advantage during social interactions, my colleagues and I gathered inspiration from a series of experiments investigating the 'reactive advantage' phenomenon. This phenomenon implies that reacting to another person's actions is faster than initiating an action (La Delfa et al., 2013; Pinto, Otten, Cohen, Wolfe, & Horowitz, 2011; Welchman, Stanley, Schomers, Miall, & Bühlhoff, 2010).

The history of research on the reactive advantage phenomenon offers an interesting interlude. The Physics Nobel laureate Niels Bohr was a Western movie aficionado. In his spare time, it is written, he mused that in Hollywood gun duels, good cowboys always win in spite of the fact that the villains drew first. Bohr's acute intuition led him to suggest that this was something more than a Hollywood plot twist. Indeed, it reflected a psychophysical principle – human reactions to events are faster than human actions that are self-initiated (Cline, 1987).

The physicists' insight was recently put to the test. Welchman and colleagues (2010) devised a laboratory version of a gun-fight, where participants sat face to face and competed against each other in being the first to finish a pre-defined sequence of button presses. The authors observed that opponents who started the movement last were faster in completing the full sequence, compared with

the ones that first initiate the movement. Following Bohr, they referred to this as a reactive advantage. This result was later replicated in several studies that studied details of the kinematic characteristics of reactive actions (La Delfa et al., 2013), and that found that the advantage was restricted to pre-programmed ballistic movements (Pinto et al., 2011).

Current interpretations of the reactive advantage phenomenon give emphasis to motor differences between reacting and acting. Whereas initiated actions involve a considerable allocation of resources in motor planning and preparation, reactive actions require less sophisticated planning. This is consistent with Wolpert et al (2003) framework for social interactions. According to this proposal, observing other's actions activates one's own action representations, thus facilitating the execution of reactions to another person's action. The lighter processing cost of reactive actions is considered to be at the root of the reactive advantage effect (La Delfa et al., 2013; Pinto et al., 2011; Welchman et al., 2010). This behavioral finding converges with neuroscientific evidence pointing to a differentiation in the neural processes underlying these two types of movement. A striking illustration of the dissociation between reactive and initiated movements comes from the observation that some Parkinson patients experience severe difficulty in initiating an action themselves, but can swiftly perform that action when it is in reaction to an external trigger (Siegert, Harper, Cameron, & Abernethy, 2002).

My colleagues and I hypothesized that social aspects, such as the sensitivity to the attention control of an opponent, may offer a contribution to the reactive advantage over and above any benefits derived from differences in motor preparation. Grounded on the findings from the previous experiment, we

predicted that, opponents' reactions to chosen actions would be faster than reactions to directed actions. This would indicate that perceptual sensitivity to attention control can be quickly transferred into a motor response, further offering an advantage to the reactive opponent over the one that initiates the movement.

3.3.1 Method

Observers. Thirty participants (20 female, 3 left-handed) with a mean age of 23.8 (SD = 4.1) were recruited from the University of British Columbia Human Subject Pool to serve as observers. The only exclusion criterion was failing to report normal or corrected to normal vision. Observers received partial course credit in exchange for one hour of time, as approved by the UBC Behavioral Research Ethics Board. All participants read and signed a written informed consent document prior to testing.

Procedure. In this experiment, my colleagues and I aimed at creating a competition scenario between actors and observers. Therefore, we asked participants to perform similar actions as the actors, so that they could directly try to be faster than the actors in reaching the end-target. We positioned our participants in the same reaching apparatus used previously to record the actor videos. We presented the actor's videos on a large display monitor (83 cm x 67 cm), such that the actor videos were approximately life size. Figure 13 illustrates the experiment from the observer's perspective. The session began with 8 practice trials, involving an actor that was not used in the main test. During the experiment, videos were presented in four blocks in randomized order. Each block presented 100 trials of one actor in random order. Chosen and directed trials were presented in equal proportion. Each trial began with the observer's

index fingers resting on these keys and their eyes on a fixation cross for 1-1.5 seconds. Observers were allowed to make small self-paced breaks in-between blocks. Observers began each trial with the index finger of their right hand at a center home position marked on the table. Observers responded to each video by reaching as rapidly as possible to the target location they thought the actor was reaching toward. We framed the task as a competitive scenario. Observers were instructed to treat this as a game in which they could “beat the actor” by reaching to the actor’s target location before the actor himself, without making more than 10-20% errors. We recorded the observer’s reach initiation time and movement time on each trial using Optotrack to sample the 3D position of the right index finger at 200Hz. At the conclusion of the session, observers completed the 50-item Autism-Spectrum Quotient (AQ) (Baron-Cohen, et al., 2001).

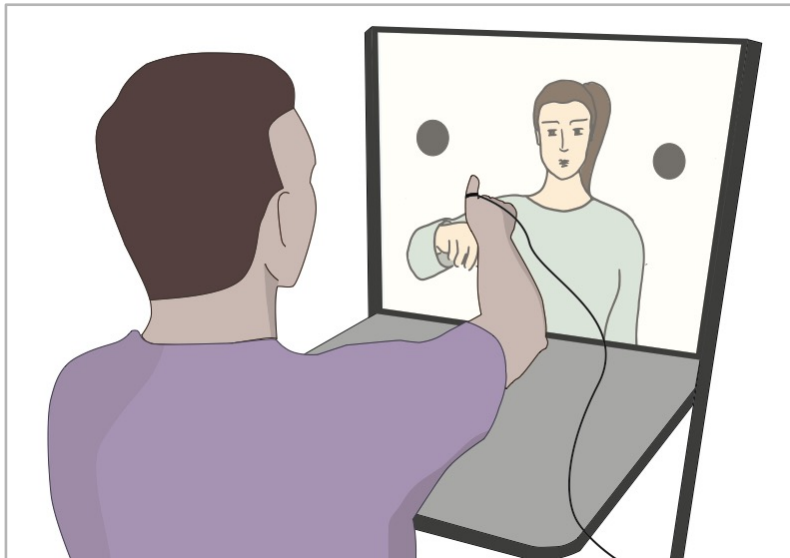


Figure 13 Illustration of the method from the observers' perspective. Observers attempt to beat the actor to the target.

3.3.2 Results

Observers responded correctly on 96% of trials (standard error of the mean = 0.03 %). Incorrect trials and movement time responses more than 3 standard deviations from the mean were excluded before computing the analysis. The reactive advantage was indexed as the proportion of trials in which the observer beat the actor to the target. To do so we compared the total movement time between observer and actor for each correct trial (i.e. the observer reached for the same side as the actor). The results showed that there was a reactive advantage, with observers beating the actors 59% of the times (standard error of the mean = 0.014%), which is significantly above the 50% benchmark, $t(29) = 4.25$, $p < 0.001$. This result was true also when considering each actor individually ($t(29) = t = -43.53, 21.15, 2.9732, 9.28$, $p < .001$ with Bonferroni correction, for actors 1 to 4 respectively). The observation of a reactive advantage is not surprising in our set-up because the actors' movements were previously recorded giving observers an unnatural advantage. Despite this limitation, we considered that any variations in reactive advantage between the chosen and direct conditions would provide information regarding the main question: Does sensitivity to attention control contribute to a reactive advantage in social interactions?

Figure 14 shows the proportion of observer wins in both the chosen and the direct condition. Competition proportions were subjected to repeated-measures ANOVA examining the effect of condition (chosen, directed) and actor (A1 to A4). This analysis indicated significant main effects of condition, with observers beating actors to the target more often when reacting to chosen reaches compared directed reaches, $F(1,29) = 4.732$, $p = .03$, and actor $F(3,203) = 1092$, p

< .001, with actor 2, 4, 3 and 1 in descending order of overall reactive advantage, and no significant interaction between condition and actor.

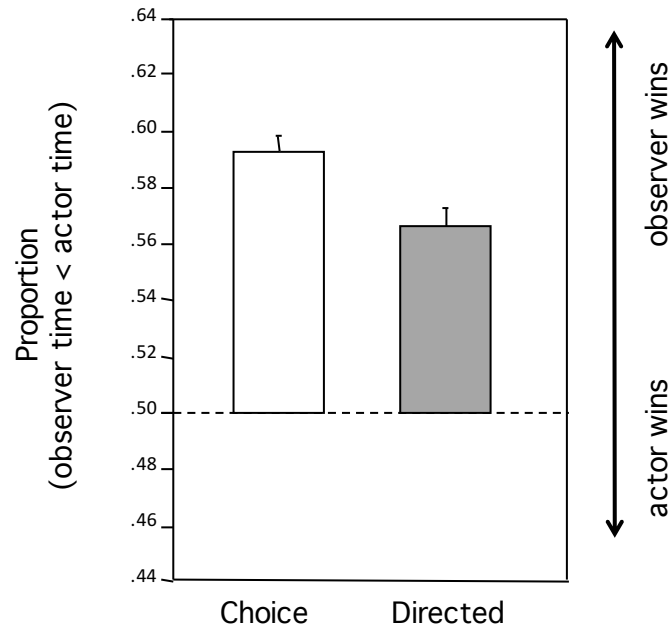


Figure 14 Proportion of times the observer reaches the correct target faster than the actor in chosen and directed conditions, collapsed across the four actors. Error bars are one standard error of the mean. Values above .50 indicate that the observer was faster than the actor more often than the opposite.

3.3.3 Discussion

The results showed that observers were generally faster than actors, documenting the reactive advantage in the boundaries of our specific experimental setting. More importantly, the results also showed that observers had a greater advantage when reacting to an opponent who was making a choice than when their opponent's action was directed by an unseen cue. This finding supports the hypothesis that sensitivity to the attention control of others

(a social signal) contributes to the reactive advantage over and above any benefits derived from the slower initiation times of a decision-making actor relative to a reacting observer (a physical head start).

Previous studies investigating whether social factors contributed to the reactive advantage had concluded that the phenomenon is not inherently social. This was demonstrated by observations of the reaction advantage in non-social settings (i.e. participants opposed graphical computer stimuli; Pinto et al., 2011; Welchman et al., 2010), and in social settings where the richness of social cues had been considerably deteriorated (i.e. opponents didn't have visual access to one another; Welchman et al., 2010). The approach of these previous tests was to remove the social dimension from the task and measure if the reactive advantage would still subsist. Thus, their results successfully show that a social dimension is not a necessary condition for the phenomenon. My approach was quite the opposite, we modulated the social dimension in the task, and measured whether the richness of the social signal contributed to the reactive advantage. Taken together these observations suggest that, albeit not necessary, social signals contribute the reactive advantage. However, this interpretation is limited by the fact that our methodology does not allow us to compare self-initiated actions from reactive actions. This is because observers always react to a previously videotaped actor. Nonetheless, the study suggests that social perception might be relevant to how individuals perform reactive actions.

In addition, the findings suggest that perceptual sensitivity to someone else's attention control can be swiftly transformed into an appropriate motor response. Thus supporting the idea that the ease of social interactions is sustained by our ability to use predictive models of our social counterparts to

quickly guide our responses during social interactions (Graziano & Kastner, 2011; Knoblich & Flach, 2001).

3.4 Is sensitivity to attention control consciously accessible to observers?

Previous experiments (reported in Chapters 3.2 and 3.3) showed that observers' speeded responses to actors' reaches are faster when the target of the reach is chosen rather than directed. But it is one thing for a social prediction model to influence kinematic behavior (i.e., the observer's spatially mapped response); it is another to have this information accessible at a conscious level. In the following experiment, my colleagues and I asked whether information about other's attention control is accessible at the conscious level or is used implicitly by observers. According to Graziano (2013), observers consciously perceive the attentional state of other individuals. This is illustrated by the author through a scenario involving observer Abel and actor Bill. Abel sees that Bill's gaze is directed toward a coffee mug. Abel then constructs a model of Bill that includes not only the spatial target of Bill's attention (the mug) but a model of Bill's intention that "Bill wants to have a sip of coffee." (Graziano & Kastner, 2011a; Graziano, 2013). Next, we will present a new experiment that replicated the conditions of the previous experiment, but in addition probed whether observers could discriminate the attentional state of actors after they had responded to the target location of the actor's reach.

3.4.1 Method

Observers. Thirty participants (10 female, 2 left-handed) with a mean age of 23.1 (SD = 4.3) were recruited from the University of British Columbia Human

Subject Pool. Participants received partial course credit in exchange for one hour of their time. All participants reported normal or corrected to normal vision. The UBC Behavioral Research Ethics Board approved student participation for credit in this study.

Stimuli and Procedure. This experiment used the same pool of 400 videos as the previous experiments. This experiment repeats the procedure of the experiment reported in Chapter 3.2, with the added feature that after making each location prediction response, participants judged whether the actor had made the choice of which target to point to. Before commencing the experiment, the experimenter informed the participants that they would be watching videos in which actors pointed to one of two potential targets (left or right). Participants were further informed that 50% of the trials corresponded to movements in which the actor reached to a target of their own choosing (endogenous orienting), and the remaining 50% trials corresponded to reaches to an externally cued target (exogenous orienting). The experimenter told participants that trials in each block would be presented in random order. At each trial, after the participants indicated their prediction of the side to which the actor was reaching (left or right), the following question appeared on screen “Did the actor choose where to point?” Participants were instructed to respond by pressing one of two specially marked keys indicating “yes” and “no.” Upon completing the experiment, participants filled in the 50-item Autism-Spectrum Quotient (Baron-Cohen, et al., 2001).

3.4.2 Results

Figure 15 shows the mean z-scores of correct RT in the chosen versus directed conditions (panel A) and shows the proportion of hits and false alarms observers

made in response to the question of whether the video they had just responded to represented a chosen or directed trial (panel B), after rank ordering observers in terms of their response biases from conservative (reluctant to respond “chosen”) to liberal (reluctant to respond “direct”). These data show that the main finding of the experiment reported in Chapter 3.2 replicated under these conditions (i.e., correct responses were faster on chosen than directed trials) but that observers were unable to report whether the actors they were responding to were chosen or not. These conclusions were supported by the following analyses.

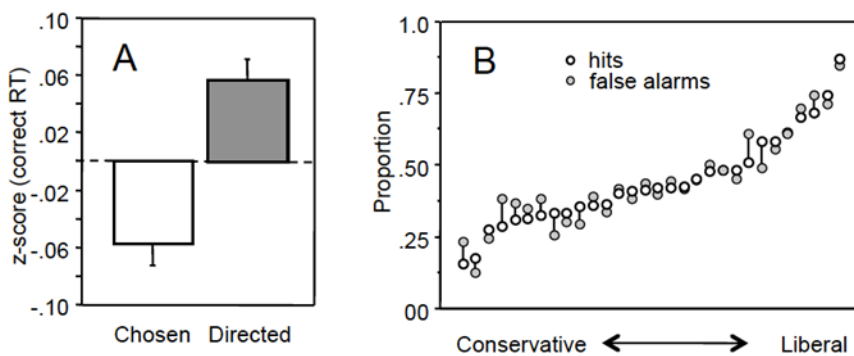


Figure 15 (A) Mean z-scores of correct RT in the experiment reported in Chapter 3.4. Error bars are +/- 1 standard error. (B) The proportion of hits and false alarms of observers trying to discriminate chosen from directed trials, after rank ordering observer’s response biases from conservative (reluctant to respond “choice”) to liberal (reluctant to respond “direct”).

Observers responded correctly on 78% of trials (standard error of the mean = 0.8%), with significant differences in accuracy between actor videos, $F(3,87) = 5.84$, $p < .001$, $\eta^2 = .169$ (in rank order A3 = 79%, A2 = 79%, A4 = 78%, and A1 = 73%), but no differences between condition or any interaction ($p > .50$). Analysis of correct RT indicated significant main effects of condition, $F(1,29) = 23.42$, $p <$

.001, $\eta^2 = .447$, and actor $F(3,87) = 34.67$, $p < .001$, $\eta^2 = .545$. Examination of the relation between the choice advantage and accuracy indicated the mean choice advantage was 21 ms for the 15 participants who were most accurate (mean accuracy = 92% correct) and only 7 ms for the 15 participants who were least accurate (mean accuracy = 64% correct), $F(1, 28) = 7.24$, $p < .01$, $\eta^2 = .136$. Analysis of z-scores also indicated a main effect of condition, $F(1, 29) = 14.74$, $p < .001$, $\eta^2 = .337$.

Analyses of the proportion of hits and false alarms in response to the question of whether a video represented a chosen or directed trial revealed no significant differences, either when the data were aggregated as a group or for any observer individually (all $p > .25$). We also replicated this insensitivity in explicit reports in a new sample of 30 observers, who were (1) not asked to predict the target locations and (2) were given trial-by-trial accuracy feedback on their guesses about whether the observer was choosing or reacting on each trial, so that they could devote their full attention to the task. The results were the same. Not a single one of the observers had a hit rate that differed significantly from their false alarms rate.

3.4.3 Discussion

Contrary to the expectation based on Graziano (2013), we found no evidence, either in the observers as a group, or among individual observers, that their explicit attempts to discriminate chosen from directed actions exceeded the chance level of guessing. This observation departs from the conceptualization that social awareness arises from an attention modeling mechanism (Graziano & Kastner, 2011; Graziano, 2013). According to which one of the consequences of having a predictive model of someone else's attention is that it allows us to

become consciously aware of the other's attentional state. The results of our test of that claim, however, were not positive. Yet, the observers in the experiment were able to distinguish these two types of reaches in their speeded kinematic responses. This pattern of findings implies that sensitivity to attention control influences an observer's action, but that it is not accessible to the observer's conscious awareness.

Taken together this pattern of findings implies that sensitivity to attention control measured in this study is signaled through implicit mechanisms (i.e., they are not accessible to consciousness). As such, it is another example of a dissociation consistent with dual processing streams (Goodale & Milner, 1992; Goodale, 2011), this time between visually-guided action that is informed by someone else's control state and conscious awareness of that state. The dual streams hypothesis proposes a general division of labor between visual processing involved in action control (dorsal stream) and visual processing leading to conscious perception (ventral stream). In the present context, we speculate that visual cues reflecting action control are processed rapidly through the dorsal stream in order to guide observer' reactions. Such fast vision-for-action processing is likely essential for the predictive aspect of social modeling, which is time sensitive. That is, the predictions must by necessity be complete in advance of both the modeled actions of an actor and any appropriate responses, if required, by the observer. Nonetheless, it is important to consider that recent studies suggest that the idea of two streams of visual perception that only converge signals until they reach very late stages of cortical analysis (e.g. superior temporal sulcus (STS), extrastriate and fusiform body areas (EBA and FBA)) may be an oversimplification (Mather, Pavan, Bellacosa Marotti, Campana, & Casco, 2013). Thus, it is probable that also in our task the ventral stream

processing is involved, to some extent, however not reaching conscious formulations.

3.5 Where on the actors' body can the attention control signal be seen?

The experiments described in previous chapters indicated that observers were sensitive to actors' attentional states expressed in the actors' body postures and movements. Next, I ask where can the attention control be seen in the body. Extant theories of social cognition have focused on the eyes as the primary source of information about social attention (Simon Baron-Cohen, 1995; Perrett & Emery, 1994). More recent evidence suggests that head and body position also play a role (Graziano, 2013; Langten, Watt, & Bruce, 2000). In this experiment, we investigated where the control signal is coming from in the video-clips of the actors. Specifically, we asked whether the signal differentiating chosen from directed actions is signaled through the actor's head and eye movements, the kinematics of the body and limbs, or a combination of both. To do so we selectively masked either the head (leaving the torso and limbs visible) or the body of the actors (leaving only the head visible), as portrayed in Figure 16, while again asking observers to make a speeded response to the target of the actor's reach.

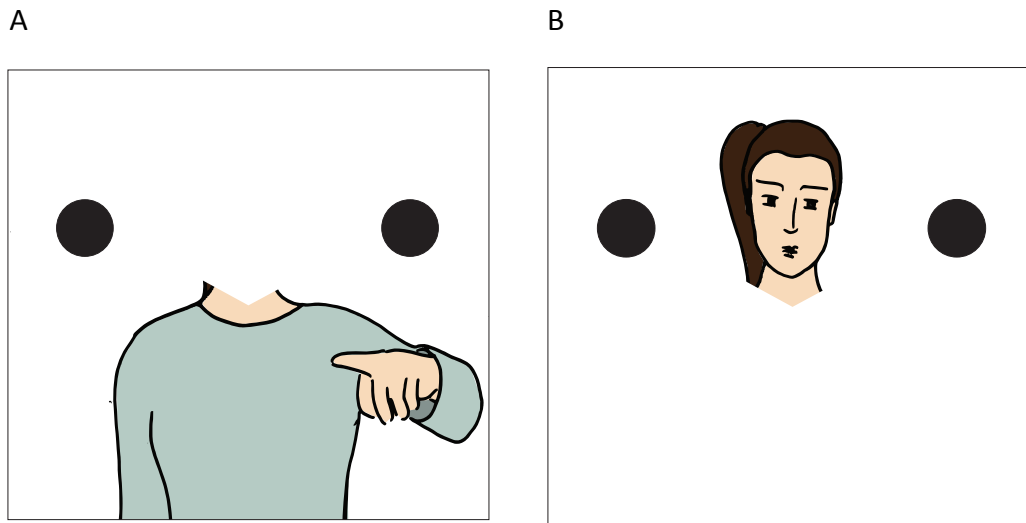


Figure 16 Representative drawings of the masked video-clips. (A) The actors' head was masked leaving the torso and limbs visible. (B) The actors' body was masked leaving only the head and neck visible.

3.5.1 Method

The method in this experiment was identical to the one described in Chapter 3.3 with the following exceptions:

- (1) Thirty different observers (24 female, all right-handed) with a mean age of 21.1 years old ($sd = 2.17$).
- (2) The 400 videos were each shown twice, once showing only the actors' head (including face, neck, and eyes) and once showing only the actors' body (torso and arms). Head and body videos were randomly interspersed in each block of trials.

3.5.2 Results

Figure 17 shows the mean z-scores of correct RT in the chosen versus directed conditions, separately for trials in which only the body and limbs were visible versus when only the head was visible. These data show that observers were more sensitive to the difference between chosen and directed trials when the body and limbs were visible than when the head was visible. These conclusions were supported by the following analyses.

While the results showed that the head alone conveyed a weak signal concerning the attentional state of the actor, consistent with the eyes as a channel to another's attentional state (Simon Baron-Cohen, 1995; Perrett & Emery, 1994), the results revealed a stronger signal when only the torso and limbs were visible, consistent with more widely distributed signals over the body indicating the attentional state of actors (Graziano & Kastner, 2011a; Graziano, 2013; Langten et al., 2000).

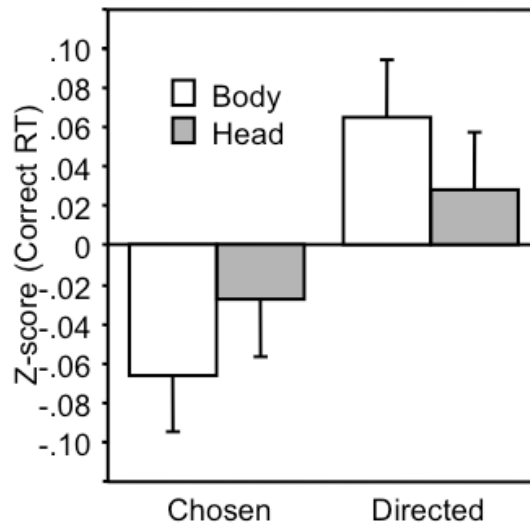


Figure 17 Mean z-scores of correct RT in the experiment reported in Chapter 3.5, separately for trials in which the body and limbs were visible versus when only the head was visible. Error bars are +/- 1 standard error.

Observers responded correctly on 78% of trials (standard error of the mean = 0.6%), with significant differences in accuracy between actor videos $F(3,87) = 21.23$, $p < .001$, $\eta^2 = .423$ (in rank order A4 = 81%, A3 = 81%, A2 = 74%, and A1 = 74%), but not between conditions ($p > .09$). Response accuracy was also significantly greater when the body was visible (mean = 82%) than when the head was visible (mean = 73%), $F(1,29) = 37.06$, $p < .001$, $\eta^2 = .561$.

Analysis of correct RT indicated that chosen trials were faster by 11 ms than direct trials, $F(1, 29) = 12.76$, $p < .02$, $\eta^2 = .306$, there were actor differences, $F(3, 87) = 43.35$, $p < .001$, $\eta^2 = .599$, and responses when the body was visible were faster by 134 ms than when only the head was visible, $F(1, 29) = 76.48$, $p < .001$, $\eta^2 = .725$. Responses to choice movements were faster than responses to direct movements by 14 ms when the body was visible and 8 ms when the head was

visible, $F(1,29) = 1.44$, $p < .25$, $\eta^2 = .047$, but the responses on head trials were also slower (134 ms) and more variable (standard error of 12 ms versus only 6 ms for body trials). Analysis of z-scores, which controlled for these differences, indicated a significant advantage on chosen over direct trials, $F(1,29) = 18.89$, $p < .001$, $\eta^2 = .394$, with this effect being significantly larger when the body was visible than when only the head was visible, $F(1,29) = 5.84$, $p < .02$, $\eta^2 = .168$. Examination of the relation between the choice advantage and accuracy indicated the choice advantage was larger for the 15 participants who were most accurate (mean accuracy = 86%, mean z-score difference = .134) than for the 15 participants who were least accurate (mean accuracy = 69%, mean z-score difference = .050), $F(1,28) = 4.50$, $p < .04$, $\eta^2 = .084$.

3.5.3 Discussion

The results showed that actor's heads alone conveyed only a weak signal concerning actors' attentional state. This is somewhat at odds with the widespread view that the eyes are the most important channel to another's attentional state (Simon Baron-Cohen, 1995; Perrett & Emery, 1994). In contrast, the results indicated a stronger signal when only the body was visible, consistent with more widely distributed signals over the body indicating the attentional state of actors (Graziano & Kastner, 2011a; Graziano, 2013; Langten et al., 2000).

This result is consistent with other recent research probing bodily kinematics for clues about people's intentions. For example, how one reaches for a Lego piece predicts the intention to cooperate or compete with a partner during a game (Manera et al., 2011). The kinematics of running reveals the intention to deceive a sports opponent (Mori & Shimada, 2013). The value of the poker hand is unconsciously expressed in arm kinematics that can be perceived by opponents

(Slepian et al., 2013). The present results add to this literature by showing that observers are sensitive to behavioral cues reflecting processes of attention control. It will be important in future studies to record the body kinematics of actors in greater detail, perhaps by using point-light-displays to isolate features of bodily movements that carry the signal of attention control.

3.6 How early in the time-course of an observed action is the attention control signal available?

In the previous chapters, I have presented evidence supporting human sensitivity to attentional states in action prediction. Next, I will present an experiment investigating the timeline of social sensitivity to attention control. Graziano (2013) highlights the predictive kinematic function of modeling another's attentional state. Such a forward model allows an observer's response to an actor to begin even before the actor's actions have been completed. Early prediction is even essential in some situations of joint action, for example in moving heavy furniture, where agents must coordinate their actions under strict temporal constraints (Sebanz & Knoblich, 2009). In this experiment, we examined the time course of sensitivity to attention control by using a temporal occlusion task. Videos of the actors' reaches were cut at 6 different lengths from the onset of the cue. Observers were asked to indicate the likely end target of the actor's actions after watching each of these brief video segments in random order.

3.6.1 Method

The method in this experiment was identical to the one described in Chapter 3.2 with the following exceptions:

(1) Thirty different observers (17 female, 3 left-handed) with a mean age of 22.71 years old ($sd = 3.43$) served as observers.

(2) Using the same pool of videos as in previous experiments, we cut each video at 6 different lengths from the onset of the cue (0-100 ms to 0-600 ms, in 100 ms steps). Videos were randomly sampled from this pool on each trial.

(3) Observers reported the likely end target of the actor's reach, and so percentage correct became the dependent measure. Because this involved guessing on many trials when the segments were short, the speed of responding was not emphasized.

(4) Observers completed 2 blocks of 600 trials, separated by a short break. Each block consisted of the presentation of 100 videos from a single actor, and the two actors selected for each observer were counterbalanced across observers.

3.6.2 Results

Figure 18 shows the mean proportion correct responses in the chosen and directed conditions as a function of the time from the onset of the actor's cue. These data show that observers can predict the target location more accurately for the chosen than the directed condition at the shortest two video lengths. This conclusion was supported by an ANOVA indicating significant main effects of condition, $F(1,29) = 23.90$, $p < .001$, $\eta^2 = .452$, and time, $F(5,145) = 1149.99$, $p < .001$, $\eta^2 = .975$, and an interaction, $F(5,145) = 27.54$, $p < .001$, $\eta^2 = .487$. Simple effects testing indicated that the chosen advantage in accuracy was significant at 100ms and 200ms (both $p < .01$) but not at the longer time bins (all $p > .15$).

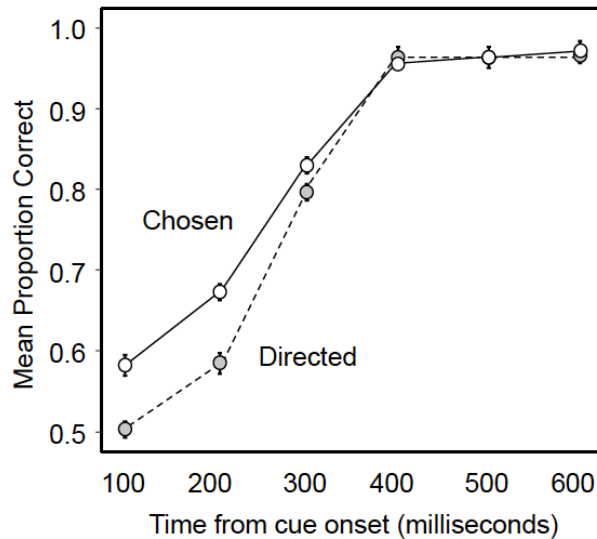


Figure 18 Mean proportion correct response in the temporal occlusion experiment reported in Chapter 3.6. Error bars are +/- 1 standard error.

3.6.3 Discussion

Graziano (2012) emphasizes the predictive nature of modeling social attention. As such, the sooner one can predict another's action, the more time one will have to consider and execute appropriate reactions (Konvalinka, Vuust, Roepstorff, & Frith, 2010; Manera, Schouten, Verfaillie, & Becchio, 2013; Sebanz et al., 2006; Sebanz & Knoblich, 2009). The results of this experiment showed that the advantage in responding to a chosen versus directed reach of an actor is already evident in the first 100 to 200 ms of processing following cue onset. This implies that observers are able to use the preparatory movements that preceded the actor's reach to make a target location prediction, such as small shifts in body balance supporting the arm motion.

The musculoskeletal constraints of the body require that moving one limb often engages the activation of other body parts. For example, initiating an arm-

reaching movement requires the engagement of the shoulders, torso, and even the lower limbs in order to make the necessary postural adjustments to stabilize the body (Hollerbach & Flash, 1982). Humans appear to have implicit knowledge of these biomechanical principles, and use this knowledge to predict others actions. For example, basketball experts are able to predict the end result of a shot before the ball leaves the athletes hand (Aglioti et al., 2008). Observers of a soccer player are able to predict the kick direction prior to the foot-to-ball contact (Diaz, Fajen, & Phillips, 2012). Deception in sports is detected above chance before the runner changes direction (Mori & Shimada, 2013). More closely related to the present task, a competitive reaching study showed that preparatory cues (i.e. movements and postural configurations preceding the lift-off of the finger) give opponents an advantage (Cormiea, Vaziri-Pashkam, & K., 2015). This is consistent with theories emphasizing the predictive nature of modeling social attention (Graziano & Kastner, 2011b; Graziano, 2013; Webb & Graziano, 2015).

3.7 Is sensitivity to attention control linked to social aptitude?

In the previous sub-chapters, I have presented evidence indicating that social perception involves sensitivity to someone else's attentional states. If the sensitivity of observers' responses to the attentional state of actors reflects the mental modeling of social attention, then individual differences in the strength of this sensitivity may be related to social aptitude on a broad scale. To test this hypothesis, my colleagues and I correlated individual differences in social sensitivity to attention control with self-reported social aptitude, as measured by the Autism Quotient (Baron-Cohen et al., 2001; Ruzich et al., 2015). Next, I will

report two sets of analysis. The first one probes overall trends in the relationship between social aptitude and social attention modeling. The second one hones in on the kinematics of sensitivity to social attention control.

3.7.1 Social aptitude and sensitivity to attention control

In each of the previously reported experiments participants fill out the 50-item Autism Quotient (AQ) (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), which captures variation in the tendency toward autistic traits in the general population. Individuals with a higher level of autistic-like traits show a non-clinical propensity to empathize less strongly with others and to engage in systemized thinking (e.g. great attention to detail, rigid interests), whereas individuals with lower levels of autistic traits display the opposite cognitive profile. To provide context, an AQ score of 32 or more points is suggested by (Baron-Cohen et al., 2001) to be a useful cut-off for distinguishing individuals with clinical levels of autistic traits. Almost all observers in this study were in the range of 5-35 and it is important to caution that this scale is not intended for exclusive use in clinical diagnoses.

To examine possible relations between observer's social aptitude and their sensitivity to the attentional state of actors, we assigned each observer a sensitivity score based on their mean difference in z-scores between the directed and chosen conditions. In experiments where observers made quick key presses responses predicting the end-target of actors' movements (reported in chapters 3.2 and 3.4), this was a mean difference score across all four actors. In the reactive advantage experiment (reported in chapter 3.3) we used mean difference score in movement initiation time across all four actors. In the experiment the body-part occlusion experiment (reported in chapter 3.5) we

used the mean difference score only for the Body condition, which provided a stronger and more reliable signal than the Head condition. One observer in this experiment did not complete the AQ questionnaire. In the temporal occlusion experiment (reported in chapter 3.6) we used the mean difference score in the 100ms and 200ms time bins, where the signal was strongest.

Figure 19 shows a scatterplot of observer's speeded sensitivity score in the experiments of chapters 3.2 to 3.5 and their AQ scores. These experiments each had a negative correlation between the measure of speeded response sensitivity and the AQ, $r(28) = -.284$, $p > .1$, $r(28) = -.4$, $p < .05$, $r(28) = -.478$, $p < .01$, and $r(27) = -.387$, $p < .05$, respectively. The correlation over all observers in these experiments was $r(117) = -.322$, $p < .001$. However, there was almost no correlation in Experiment 4, where response sensitivity was measured in accuracy rather than speed, $r(28) = -.004$. This is consistent with observers with greater social aptitude being able to respond more rapidly to an actor who is selecting their reach with intention rather than being directed.

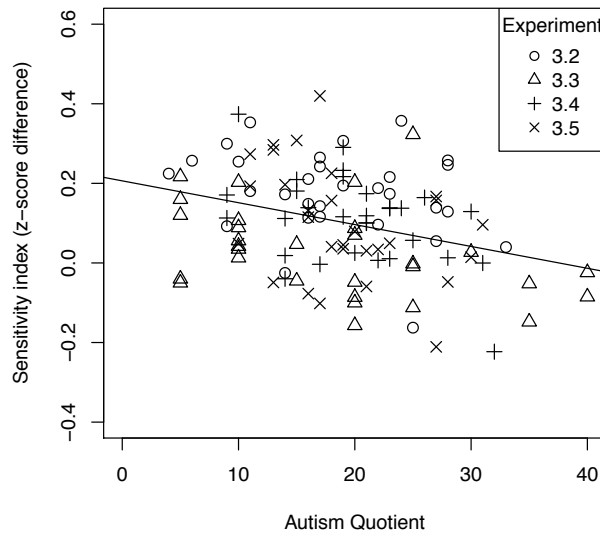


Figure 19 Scatterplot of the relation between observer's speeded sensitivity scores in the experiments reported in chapter 3.2 to 3.5 and their Autism Quotient scores.

These results showed that for four independent groups of observers (Experiments reported in chapters 3.2 to 3.5), sensitivity to actors' attentional states, as measured in speeded responses to the targets of the actors' reaches, were negatively correlated with scores on the Autism Quotient. This implies that individuals with higher levels of empathy tended to also show the greatest sensitivity to actor choice in their speeded responses.

3.7.2 The kinematics of human sensitivity to attention control

A commonly observed kinematic signature of rapid arm-reaches is the trade-off between movement initiation time and movement duration. The distribution between the duration of time passed before finger lift-off and the duration of the reach itself reflect underlying cognitive strategies. Longer initiation times followed by faster movement times reveal a tendency towards performing the

bulk of processing before initiating the reach, whereas shorter initiation times paired with longer movement times indicate a bias towards in-flight cognitive processing (Schmidt & Lee, 2011). Is social sensitivity to attention control regulated by these kinematic strategies? Next, I will report an analysis of movement initiation vs. movement duration trade-offs in light of individual differences in social aptitude. This analysis was performed on the data collected in the reactive advantage experiment reported in chapter 3.3.

Figure 20 shows an overall tendency for reaching initiation time vs. reach duration trade-offs in our sample. This is expressed by a strong negative correlation between observers mean initiation and mean duration times (i.e. participants who were fast to lift-off took longer to get to the end-target, and vice-versa), $r(28) = -.719$, $p < .001$.

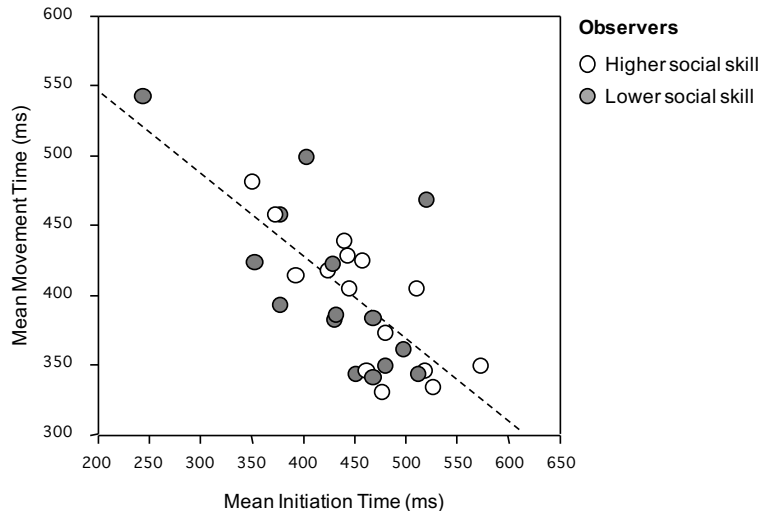


Figure 20 Scatterplot of movement initiation vs. movement duration trade-off in reach responses of participants reporting higher and lower social aptitude levels.

To examine whether the cognitive strategies underlying reach initiation vs. duration time trade-offs were related to social aptitude, we mean-split our

sample into higher and lower social aptitude and compared response times between these two two-sub-groups. Visual inspection of Figure 19 suggests that observers who reported higher social aptitude tended to take longer to initiate the reach (mean=453 ms, s.e.=15.13) compared to observers with lower social aptitude (mean= 424.13 ms, s.e.=18.47). In compensation, the same observers with high social aptitude, who were slow at initiating their reaches, tended to be faster at getting to the target (mean=396.13 ms, s.e.=12.26) than observers who reported lower social aptitude counterparts (mean=405.6 ms, s.e.=15.96). This observation offers preliminary support to the notion that kinematic trade-offs are linked to individual differences in social aptitude. These analyses suggest that more socially apt observers, who were slower to begin moving, had more time to observe the unfolding action and form predictions before moving. This might explain why, as a consequence, high social observers show higher sensitivity to attention control in their response times (as indicated by the correlations between AQ and speeded response times reported in the previous sub-section; Figure 18).

To examine the relationship between sensitivity to social attentional states and kinematic strategies, we independently computed initiation time sensitivity scores and movement time sensitivity scores for each observer. These sensitivity scores corresponded to mean differences between the directed and chosen conditions. Figure 21 shows the relationship between observers' sensitivity to social attention at the movement initiation stage and at the movement duration stage. A marked negative relationship indicates that social sensitivity tends to be portrayed either at the reach initiation stage or at the movement duration stage, $r(28) = -.742$, $p < .001$. The scatterplot also shows that all observers have some degree of sensitivity to social attention control in their responses, what varies

between individuals is when in the unfolding of the reaching response this sensitivity is displayed.

Individuals with higher social aptitude tended to show more social sensitivity at the reach initiation stage (mean=13.2 ms, s.e.=2.11) compared to individuals with lower aptitude (mean=8 ms, s.e.=2.22). In compensation, the same individuals who reported lower social aptitude showed more sensitivity later in their movement times (mean=8.47 ms, s.e.=2.58) compared to the socially apt (mean=2.53 ms, s.e.=1.54). This suggests that observers with higher social skills are able to utilize their sensitivity to attention control earlier in their motor responses than observers with lower social skills.

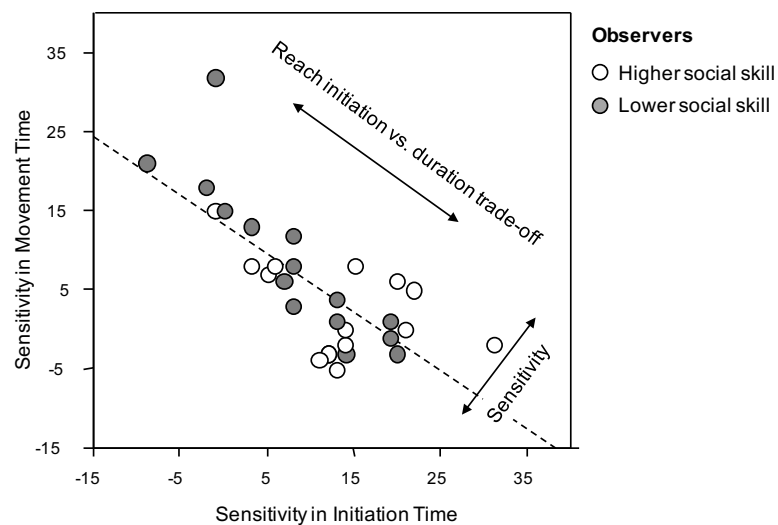


Figure 21 Scatterplot of observers' sensitivity to social attention at the reach initiation stage against sensitivity at the reach duration stage.

Taken together these analyses indicated that observers with higher autistic traits (i.e. lower social aptitude) initiate their movement responses before profiting from the available social cues, and as a result, only integrate this information later on in their responses. Observers with higher social aptitude, delay the

initiation of their responses to early on gather relevant social cues, and consequently are able to utilize this information more promptly in their responses. This evidence suggests a kinematic pattern for social aptitude that is consistent with recent research showing that persons with autistic behavioral tendencies show lower response inhibition and error processing (Kana, Keller, Minshew, & Just, 2007; Larson, Fair, Good, & Baldwin, 2010; Larson, South, Krauskopf, Clawson, & Crowley, 2011). In addition, these findings add validity to the proposal that observers were using predictive modeling of social attention (Graziano & Kastner, 2011a; Graziano, 2013; Webb & Graziano, 2015). Although the predictive modeling of social attention may be a core mechanism of human observers, social experts appear to be more fluent in using it.

3.8 Summary and discussion of the empirical studies

The main goal of this study was to investigate human sensitivity to social attention control. To do so my colleagues and I realized a series of experiments that aimed at addressing the following questions:

- Are observers sensitive to someone else's attention control?
- Does sensitivity to attention control contributes to a reactive advantage in social interactions?
- Is sensitivity to attention control a conscious process?
- Where on the actors' body is the attention control signal available?
- How early in the time-course of an observed action is the attention control signal available?
- Is sensitivity to attention control linked to social aptitude?

Next, I will discuss the findings obtained while investigating each of these questions, and address how they relate to current literature.

Are observers sensitive to someone else's attention control? Graziano

(2013) posits that social awareness is a predictive model of someone else's attention. This conceptualization implies that the attentional states underlying observed actions should have some measurable effect on observers' responses. Offering empirical support to Graziano's theoretical proposition (Graziano & Kastner, 2011; Graziano, 2013), the experiment reported in chapter 3.2 showed that observers were faster at correctly predicting the target of an action driven by intentional attention allocation (endogenous orienting) compared to predicting an externally guided action (exogenous orienting). These results indicate that observers are particularly sensitive to whether attention orienting is intentional or not. This observation is aligned with the well-accepted notion that the social perception of attention is a relevant process contributing to the human ability to infer someone else's inner intentions, i.e. Theory of Mind (Simon Baron-Cohen, 1995, 2000; Calder et al., 2002). Intention inference is often considered as a high-level cognitive process (Jacob & Jeannerod, 2005), our findings are evidence of what are likely early low-level inputs to Theory of Mind processes.

Does sensitivity to attention control contributes to a reactive advantage in social interactions? Findings reported in chapter 3.3 indicate that observers in a fast-reaching competitive setting are better at beating their opponent to the end-target when the opponent is choosing where to reach compared to when the opponent is being directed. This suggests that observers are able to harness perceptual cues reflecting attention control to generate fast behavioral

responses. These findings revive the discussion on whether the reactive advantage phenomenon (i.e, reacting to another person's actions is faster than initiating an action) is purely motor or whether it also has a social component. Unlike previous studies that conclude that the reactive advantage is independent of its social setting, the results from our study indicate that social cues about the opponent's attentional state contribute to the reactive advantage over and above any motor benefits (La Delfa et al., 2013; Pinto et al., 2011; Welchman et al., 2010).

Is sensitivity to attention control a conscious process? According to Graziano's (2013) theory, observers consciously perceive someone else's attentional state. Departing from this expectation, findings reported in chapter 3.4 showed that, although sensitivity to attention control guides behavioral responses, this process is not reflected in verbal reports. This pattern of results implies that human sensitivity to attention control is signaled through mechanisms of implicit perception that are not accessible to consciousness. Our findings make the case for the dissociation between the visual processing of attentional states and its awareness. The dual-stream model of visual processing offers a framework for the interpretation of such dissociation. This model describes a division of labor between visual processing involved in action control (dorsal stream), and visual processing leading to conscious perception (ventral stream) (Goodale, 2011). We speculate that visual cues reflecting attention control are quickly processed through the dorsal stream and are used to guide observer's reactions. This type of fast vision-for-action processing is essential for the predictive aspect of social attention modeling, which must by necessity run rapidly in advance of both actor's actions and observer's reactions, and cannot afford the slower time constraints of conscious elaboration.

Where on the actors' body is the attention control signal available?

Whereas most research on the social perception of spatial attention has focused on eye-gaze cues (Langten et al., 2000), recent evidence reveals that other bodily cues might be just as important (Nummenmaa & Calder, 2009). Our study shows that this is also the case for sensitivity to attention control. Findings reported in chapter 3.5 revealed sensitivity to attention control both when viewing head-cues only (including the neck, face, and eyes) and when viewing body-cues only (including torso and arms). Thus indicating that the attention control signal is widely distributed. Furthermore, observers' response time advantage at predicting the target of endogenous orienting actions compared to exogenous orienting was greater when only the body was visible. This suggests observers are particularly sensitive to body cues reflecting intentional attention orienting.

Previous evidence indicates that human observers are well apt at reading other people's hidden intentions from observed bodily kinematics. For example, the way one reaches for and grabs a Lego piece during a game has been shown to reveal an individual's hidden intention to cooperate or compete with a partner (Manera et al., 2011); the kinematics of running portrays one's intention to deceive a sports opponent (Mori & Shimada, 2013); and that despite the conventional wisdom of maintaining a neutral face while playing poker, the value of the poker hand is unconsciously expressed in arm movement kinematics during the game that can be perceived by opponents (Slepian et al., 2013). Each of these previous findings supports the notion that inner intentions are constantly being expressed through bodily behavior, and are thus available in the public realm as relevant stimuli during action observation. Our study adds to the previous literature by showing that observers are sensitive to behavioral cues

reflecting inner cognitive processes of attention control. In future experiments, the complete body kinematics of the actors could be recorded. This would allow us to manipulate point-light-displays of the actors to probe in further detail which features of the bodily movements carry the attention control signal.

How early in the time-course of an observed action is the attention control signal available?

The ability to quickly predict someone else's actions is essential for most social interactions. The sooner individuals can predict the actions of their social counterparts, the more range they will have to generate appropriate reactions (Sebanz et al., 2006; Sebanz & Knoblich, 2009). Findings reported in chapter 3.6 showed that the advantage in responding to endogenous orienting vs. exogenous orienting actions occurs promptly within the first 200 ms of observing actor's responses. Thus suggesting that observers are sensitive to attention control during preparatory movements that precede the unfolding of the reaching action. Action prediction based on preparatory movements has been previously reported in competitive scenarios. For example, basketball experts are able to predict the end result of shoot before the ball leaves the athletes hand (Aglioti et al., 2008); observers in a goal-keeper scenario are able to predict the kick direction prior to the foot-to-ball contact (Diaz et al., 2012); deception in rugby runners is perceived above chance before the runner changes direction (Mori & Shimada, 2013); and very closely related to our task, in a competitive arm reaching scenario attackers' preparatory movements give opponents a reactive advantage (Cormiea et al., 2015). Our study converges with this previous evidence by showing that observers can leverage information from preparatory movements, and in addition, advances that preparatory motion is more informative when intentional orienting underlies action execution. In conclusion, the early availability of the attention control signal in preparatory

movements substantiates the predictive aspect of social attention modeling (Graziano & Kastner, 2011a; Graziano, 2013).

Is sensitivity to attention control linked to social aptitude? Individuals with Autistic Disorders experience difficulties in attributing perceived spatial attention orienting to an inner mental state, leading to a wide range of social impairments (Baron-Cohen, 1994, 2000). Extrapolating from clinical knowledge, it would be expected that autistic-like traits in normal population would relate to the extent individuals are sensitive to other's attentional states. Indeed, the findings from the reported experiments showed that participants with higher social aptitude, as measured by the Autism Quotient Scale (Baron-Cohen et al., 2001), were more fluent at utilizing their social sensitivity to quickly predict the end-target of endogenous orienting actions over exogenous orienting ones. Therefore indicating that sensitivity to the attention control of a social other is linked to one's general level of social aptitude. This pattern of results is consistent with the view that modeling someone else's attention is a core mechanism supporting general social abilities in humans (Simon Baron-Cohen, 1995, 2000; Calder et al., 2002; Graziano & Kastner, 2011a; Graziano, 2013). As a core faculty, sensitivity to attention control is available to all, but experts are more fluent at it.

In conclusion, our study contributes to current knowledge about the perceptual mechanisms underlying social cognition by showing that the social perception of attention is more sophisticated than previously thought: More than perceiving *where* others are attending to, we showed that humans are also implicitly sensitivity to *how* attention is deployed to a spatial location. This observation gives empirical support to current theoretical views that propose that human

social cognition involves the predictive modeling of our social counterparts' attentional states (Graziano & Kastner, 2011; Graziano, 2013).

4 General discussion

In this concluding section, I bring together the theoretical and empirical streams of this thesis, described in Chapters 2 and 3 respectively. I will start by summarizing and discussing the outcomes of each chapter independently. Then I will utilize the theoretical concepts of predictive processing to interpret new evidence of human sensitivity to attention control. Finally, I will identify several important new questions that have been raised by these findings and outline avenues for future research on social cognition that will provide answers to these questions.

4.1 Theoretical framework

Recently there has been a surge of interest in studying cognition in its social milieu. As part of this trend, an increasing number of research findings on the perceptual and motor workings of cooperative behavior have been reported, constituting joint-action as a field of research in its own right (Knoblich, Butterfill, & Sebanz, 2011; Sebanz, Bekkering, & Knoblich, 2006). Yet, the development of theoretical frameworks for joint-action has not kept pace with the proliferation of research findings. In this thesis, I proposed a hierarchical predictive approach to joint-action implementation, the predictive joint-action model (pJAM). Previous frameworks had either addressed the phenomenon by describing the high-level cognitive processes involved in joint-action (Vesper et al., 2010) or by focusing on the sensor-motor level of joint-action implementation (Wolpert et al., 2003). pJAM is an improvement over these previous accounts of joint-action because it addresses joint-action simultaneously at the symbolic and sensorimotor level.

pJAM assumes three layers of decreasing processing abstraction, from higher-level processing to lower-level processing. In specific, the model assumes a predictive cascade comprising a goal representation layer, an action-planning layer, and a sensory routing layer. The general idea of the framework is that each layer encodes parallel state probabilities about the information in the layer below, at several spatial and time scales. Continuous comparison between adjacent layers, and subsequent error minimization, ultimately contributes to the successful implementation of joint-actions.

This architecture offers concrete insights about three open questions previously identified in joint-action literature reviews:

- (1) How are high-level (e.g. goal sharing and verbal agreements) and low-level processing (e.g. interpersonal motor adaptation) integrated into joint-actions (Knoblich et al., 2011)? The hierarchical organization of pJAM offers a computational structure that accounts for processing at different levels of abstraction. In specific, through its distributed processing cascade, the framework binds symbolic representations with motor plans and perceptual processing (Clark, 2013).
- (2) How are joint-actions successfully taken to term given the inherent under specification of goals and tasks between partners (Vesper et al., 2010)? The Bayesian-like functioning of the hierarchical predictive cascade offers a solution to this problem. Partners share similar top-down and bottom-up information streams. They share a rough representation of the joint-goal and respective co-tasks (top-down). And they also receive similar sensorial inputs (bottom-up). By relying on an iterative error-reduction process between top-down expectations and bottom-up information, it is

probable that partners' internal models of the necessary joint-action states to achieve the shared goal will increasingly converge into similar representations (Vesper & Richardson, 2014).

- (3) How are 'self' and 'other' representations managed in joint-actions? How does agency emerge in joint-actions (Pacherie, 2007, 2012)? In pJAM sensory routing, occurring at the first level of bottom-up processing assigns sensory outcomes to parallel streams of information processing pertaining to one's own and other's actions. This allows the framework to account for the emergence of a subjective experience of agency in joint-actions (Stenzel et al., 2014).

Apart from advancing the current state-of-the-art in joint-action theoretical frameworks and offering insight into long considered questions in the field, I posit that pJAM also offers a structured way to think about empirical evidence. Next, I will discuss the empirical findings reported in this thesis and consider how they relate to the theoretical framework proposed.

4.2 Empirical findings

The human ability to make predictions about someone else's actions is central to our social lives. It has recently been proposed that attention is central to social prediction. Knowing where and how someone else is directing their attention, can provide us with valuable clues about what they intend to do next (Graziano & Kastner, 2011; Graziano, 2013; Webb & Graziano, 2015). Previous studies of social perception report acute human sensitivity to where another's attention is aimed. In Chapter 3 I start by presenting a new method to study social sensitivity to attention control in action prediction. This method is divided into two stages – stimuli recording stage and experimental stage. The two-stage design allowed

me to isolate observers' sensitivity to actors' spatial orienting from observers' sensitivity to actors' attentional control. This represents an improvement of previous methodologies in which actors and observers states were not decoupled (Welchman et al., 2010). Experiments using the new methodology showed that human social understanding involves not only knowing where someone else is attending but also sensitivity to how the other's attention has been oriented to that location. When observers were given the opportunity to predict the location of a videotaped actor's reach, they were faster to do so when the actor was deciding where to reach (endogenous attention control) than when the actor was being directed by an external cue (exogenous attention control). This was true despite our care in removing all temporal cues from the sampling of the actor's reaches and in randomizing the two types of reaches shown to observers. This implies that the decision undertaken by the actor is visible to the observer before it being executed by the actor. Yet tests of whether the observer's sensitivity to the actor's choice was consciously accessible were negative. Tests of where the signals about the actor's choices were coming from indicated that the signals were widely distributed over the body, though stronger in the torso and limbs than in the head. Tests of when the signal was available indicated it was influential even before the actor's limb started moving. Finally, sensitivity in the speeded decisions of observers was correlated with a paper-and-pencil measure of social aptitude.

In sum, the main finding of this study is that action prediction is easier for most observers when actors are choosing to act rather than being directed externally, a finding I have termed as the "choice advantage". The secondary findings were (a) that sensitivity to choice in the kinematics of others is not consciously accessible to observers, but (b) that it is correlated with an independent

measure of social aptitude in everyday life. This bolsters the view that social action observation is a fast and implicit kinematic process linked to empathy. Taken together, these observations are consistent with recent theoretical proposals claiming that social awareness involves the predictive (forward) kinematic modeling of the action consequences of others' attentional states (Graziano & Kastner, 2011b; Graziano, 2013; Webb & Graziano, 2015).

However, I would like to highlight one specific limitation of these findings. The results might be specific to the competitive nature of the task. Observers were asked to guess the actor's action goal (reach to the left vs. right target) as fast as possible before the actor. Framing the task as a competition might motivate observers to more closely process any intentional cues portrayed in the actor's behavior because observers need to predict the actor's hidden action goal in order to be successful competitors. It is possible, that in cooperation scenarios, sensitivity to attentional control is not as relevant. Cooperation entails that both partners share the same action goal (Knoblich, Butterfill, & Sebanz, 2011; Sebanz, Bekkering, & Knoblich, 2006). Therefore, partners assume that they share the same action goal. This may potentially decrease the relevance of processing the control cues in observed actions.

Next, I will utilize the hierarchical predictive framework described in Chapter 2 to discuss the observed empirical findings, further identify limitations in the studies, and propose future research about the social perception of attentional states.

4.3 Bringing theory and findings together

In this section, I utilize the theoretical concepts introduced in Chapter 2 to discuss the empirical findings reported in Chapter 3. But before that, I will address an initial shortcoming of this endeavor. Whereas pJAM is directed at joint-action phenomena, the empirical studies in this thesis do not fully qualify as joint-actions. This is because actors and observers did not share the same goal, and did not act together to exert a change in the environment (Knoblich & Sebanz, 2006). Instead, the studies employed an action prediction task, where observers attempted to predict the unfolding of actors' actions. Nevertheless, I propose that the empirical findings in this thesis fall within the hierarchical predictive approach followed by pJAM. Several aspects of the studies support the viability of this idea. Concretely, the experimental task required the monitoring and predicting of someone else's actions and the subsequent execution of an appropriate motor response. All of these aspects are central to joint-actions and are featured in the pJAM architecture. Figure 22 highlights the parts of pJAM that will be used to discuss the empirical findings reported in this thesis.

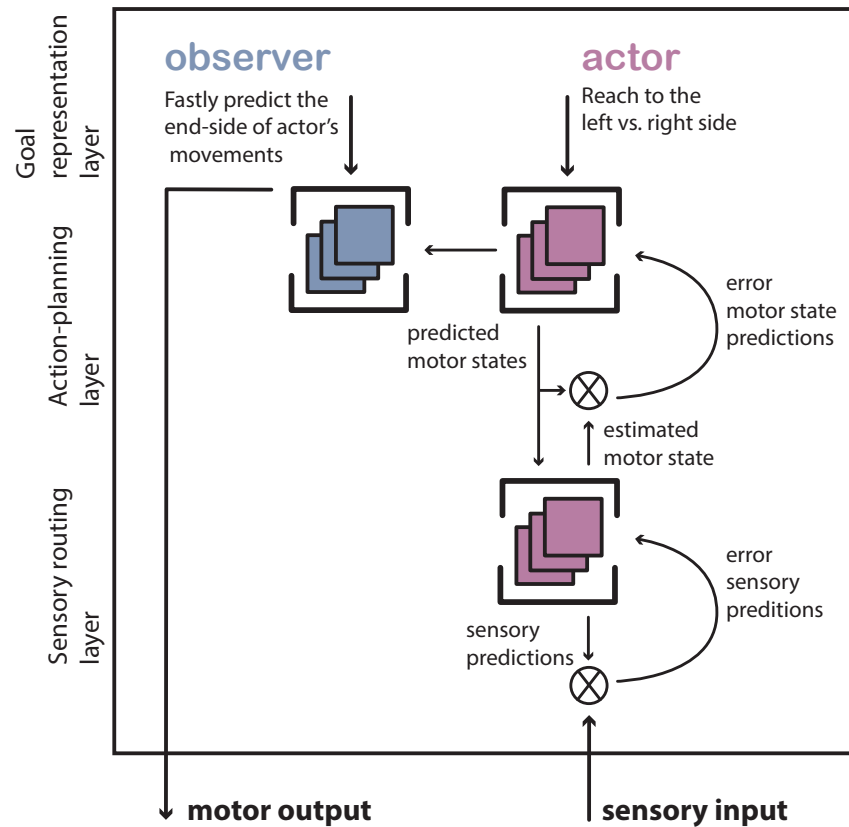


Figure 22 Action prediction cycle in pJAM.

Now that we have identified the useful parts of the model for the task at hand, let's simulate how action prediction in the experimental task is supported by the hierarchical architecture of pJAM. I will guide you through this simulation in three stages. First, I will describe the expected state of the predictive hierarchy before action observation (i.e. at the beginning of the trial, before observing the actor). Afterward, I will give an account of how the system might function once action observation commences. This will include the minimization of deviations between the observer's predictions about the actor's actions and incoming information from action observation. Finally, I will give an account of how

observer's responses are triggered. At each stage, I will juxtapose the observed empirical evidence to the functioning of the predictive hierarchy.

4.3.1 Initial state of the predictive architecture

Figure 23 illustrates the starting state of the predictive architecture. At the action-planning layer, probabilistic models encode parallel predictions about the future unfolding of the actor's action. At the start of each trial, before commencing action observation, the state probabilities about the actor's future movement end-side are at chance-level, i.e. there is a 50%-50% split between right and left predictions.

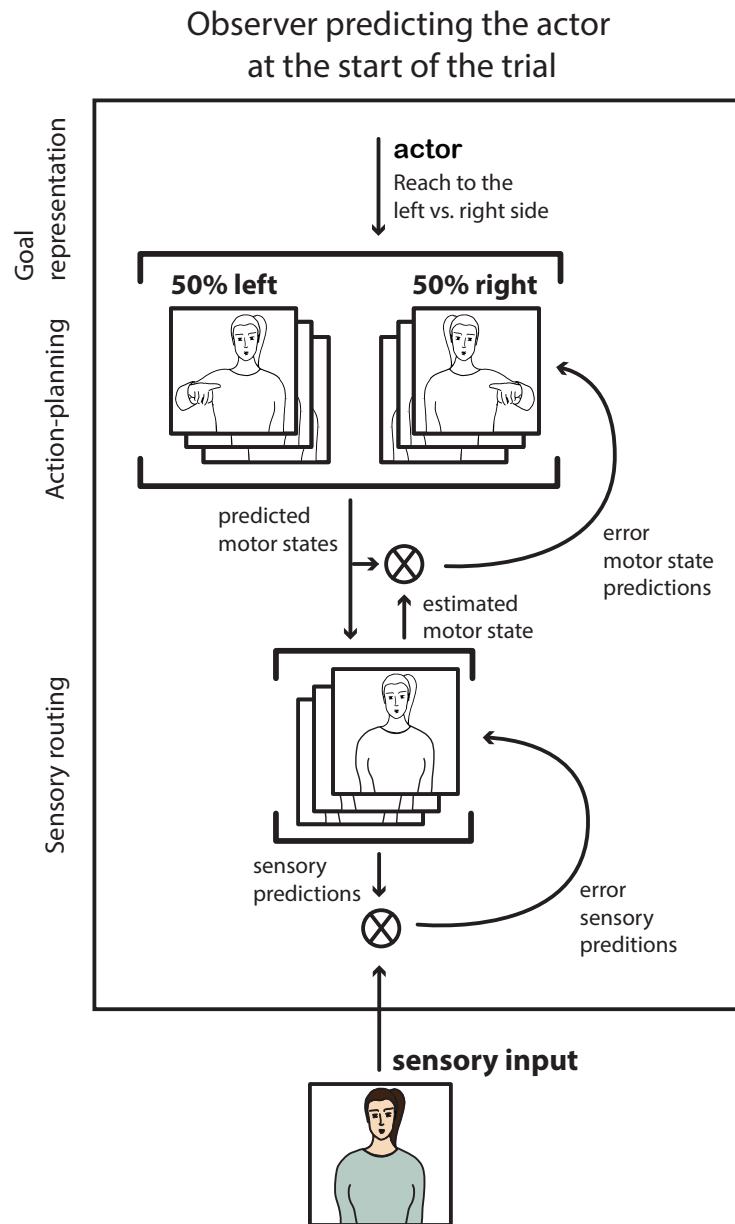


Figure 23 Predictive architecture state at the start of the trial. Before the activation of bottom-up swipes of information, prediction is at chance level - 50% left and 50% right.

4.3.2 Probabilistic predictions during action observation

As the video-clip of the actor starts playing, observers start gathering information to continuously update their probabilistic models about the actor's biases towards each possible end-side (Clark, 2013; Graziano, 2013). Figure 24 represents the changes occurring in the predictive architecture as top-down (i.e. predictions about the end-side of actors actions) and bottom-up information (observed movement cues) start traveling through the processing hierarchy. Once the video starts - revealing actors' early movements - sensory information starts traveling up the predictive cascade. Comparisons between incoming sensory information and the corresponding predicted states are continuously made. Errors between predicted and received information are used to improve the probabilistic predictive models, at the action-planning layer. In this way, early movement cues start shifting the probabilistic models to bias one side over the other. In an effort to minimize deviations between predicted and observed states, the initial 50%-50% distribution of end-side probabilities is shifted to favor one side, e.g. 70% left and 30% right.

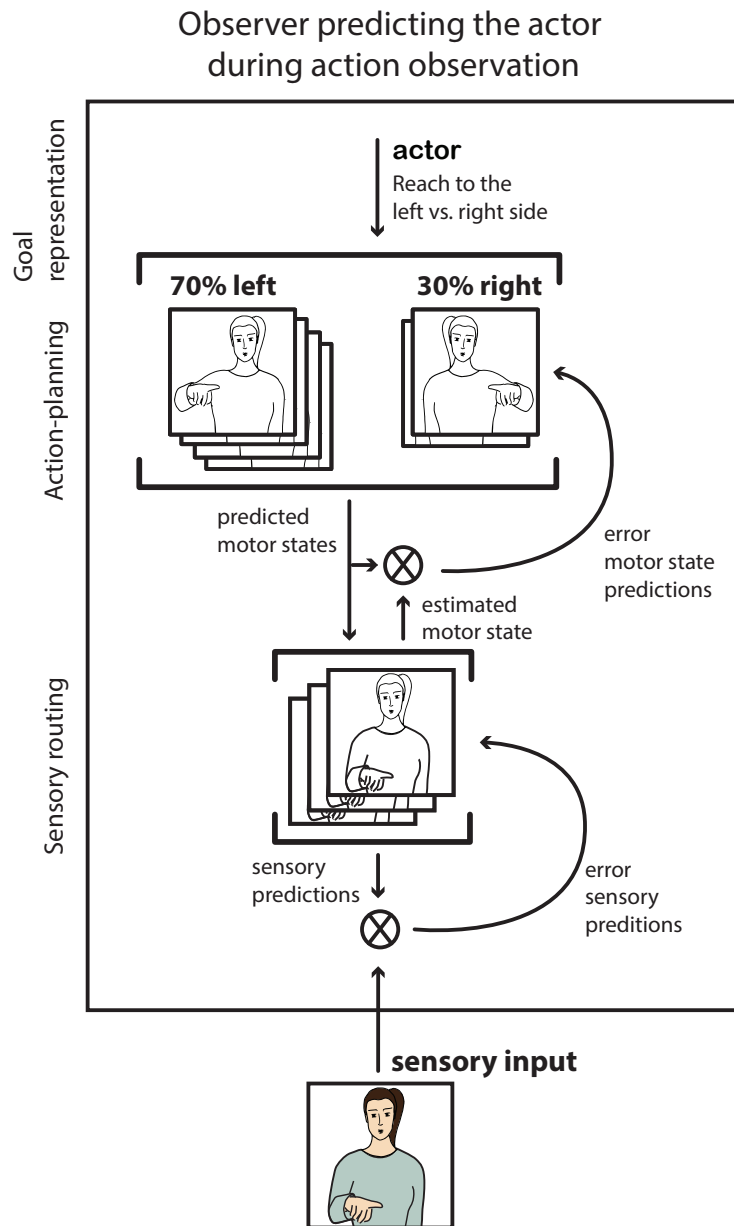


Figure 24 Predictive architecture state while minimizing the error between top-down predictions and bottom-up information.

Now let's consider how the empirical findings relate to the described framework states. The findings showed a "choice advantage": Observers were faster at

predicting chosen versus directed actions (as reported in Chapters 3.2-3.5). In convergence, observers were also more accurate when predicting chosen actions compared to directed actions, when only the initial parts of actors' actions were available (as reported in Chapter 3.6). Seen through the lens of the framework, the "choice advantage" means that when the sensory input corresponded to chosen actions the comparisons between action prediction states and sensory information gave rise to a faster shift of the state probabilities in favor of one side, leading to faster predictions. The framework offers two alternative explanations for the "choice advantage". One hypothetical explanation puts the emphasis on the quality of the incoming information. I will call this the *preparatory cues hypothesis*. Conversely, the other explanation puts the emphasizes on the nature of the predictive models. This hypothesis is termed the *models of intentional control hypothesis*. Next, I will consider each of these hypothesis separately, and provide a description of future studies designed to test them.

The preparatory cues hypothesis According to this hypothesis, the observed choice advantage occurs because early kinematic cues in the execution of chosen actions carry predictive information about the actor's ultimate choice. This conceptualization is consistent with evidence indicating that action components are not independent of one another; at any moment in time, internal mental biases and existing bodily states unconsciously influence the unfolding of the subsequent movements in a sequence (Rosenbaum, Herbort, van der Wel, & Weiss, 2014). It follows from this that choice actions should follow more naturally and predictably from the pre-choice mental and postural states of the actor than directed actions. Actions that are directed by an external signal — and so are not chosen — are much less likely to follow smoothly from an actor's

recent mental and postural history. Thus, the kinematic cues available from an actor have greater predictive value for subsequent action when the actor is choosing the target of a reach than when the actor is responding to an unpredictable external signal. Observing the stream of consistent kinematic cues in an actor's chosen behavior can explain observers' ability to predict the outcome of the reach earlier in time.

The models of intentional control hypothesis An alternative explanatory hypothesis to the choice advantage is that our internal models of other's behavior assume the nature of intentional control (Jacob & Jeannerod, 2005). The goal of social predictive models is to anticipate what others will do next (Brown & Brüne, 2012; Bubic, von Cramon, & Schubotz, 2010; Sebanz & Knoblich, 2009). Therefore, it is not unreasonable to consider that these models integrate the effects of intentional control on action execution. According to this hypothesis, observers' internal predictive models of actor's actions are inherently closer to chosen actions than to directed actions. Thus, when matching actors incoming movements to observers internal predictions of these movements, chosen actions will be a closer match, and will faster tip the probabilistic predictions towards one end-side, ultimately leading to faster predictions.

How might we disambiguate between these two possible interpretations? Both hypothesis can be tested in future empirical studies. Testing the *preparatory cues hypothesis* can be achieved by manipulating bottom-up information, i.e. actor's actions. To test the influence of preparatory cues on sensitivity to attention control, actors could be filmed either when preparing their choice ahead of time or not. This new stimuli set would support a 2x2 experimental

design - preparation before cue (yes, no) x attentional control state (chosen, directed). Analysis of observers' responses would disambiguate sensitivity to action preparation from sensitivity to attention control. One potential result would be quite conclusive. If the choice advantage remains in trials where actors prepared ahead of time, but disappears when actors avoided preparation ahead of time, then the choice advantage is driven by observers' sensitivity to actors' strategic preparation during the recording task. However, if the choice advantage is maintained in both conditions, then this indicates that sensitivity to attention control is not fully driven by preparation cues.

Testing *the models of intentional control hypothesis* can be achieved by manipulating top-down information, i.e. observers' expectations. An independent group design could be applied. Some observers would be informed before the start of the experiment that the actions they will try to predict were executed according to the actors own choice, while others would be told that the actors were executed in response to an external signal. This manipulation aims at biasing internal models to encode endogenous control or exogenous control. If the manipulation is successful, then directed actions would be easier to predict when observers expect the actor to be directed by an external stimulus (exogenous control), and choice actions would be easier to predict when the observers expected the actors be in control of their end-target side (endogenous control). This pattern of results would indicate that a match between attention control expectations and observed attention control is at the basis of human sensitivity to someone else's attentional states.

4.3.3 Prompting observers' prediction responses

Let us return to the description of the experiment using the hierarchical predictive framework. At some point during action observation, the probabilistic bias will be strong enough to prompt the observer to execute a motor response. This can be conceptualized as a decisional threshold. The bias towards one side, encoded by internal models of actors' states, has to reach a certain threshold in order to lead observers to respond. This is illustrated in Figure 25. The findings reported in Chapter 3.7 indicate that individuals with lower social aptitude as measured by the Autism Quotient Scale (Baron-Cohen et al., 2001), are more impulsive in initiating their motor responses, compared to individuals with higher social aptitude. Thus the threshold for response is lower for individuals with lower social aptitude. Putting this observation in the context of the hierarchical framework brings to clarity that, in these studies, observers with lower social aptitude were at a disadvantage due to lower response inhibition (Kana et al., 2007; Larson et al., 2010, 2011), rather than being impaired at the sensory layer (Blake, Turner, Smoski, Pozdol, & Stone, 2003) or action modeling layer (Natalie Sebanz, Knoblich, Stumpf, & Prinz, 2005).

In sum, looking at the empirical findings through the lens of the theoretical framework showed where future studies are necessary to further our understanding of human sensitivity to attention control, and allowed for the integration of observed behavioral findings within a cognitive processing structure.

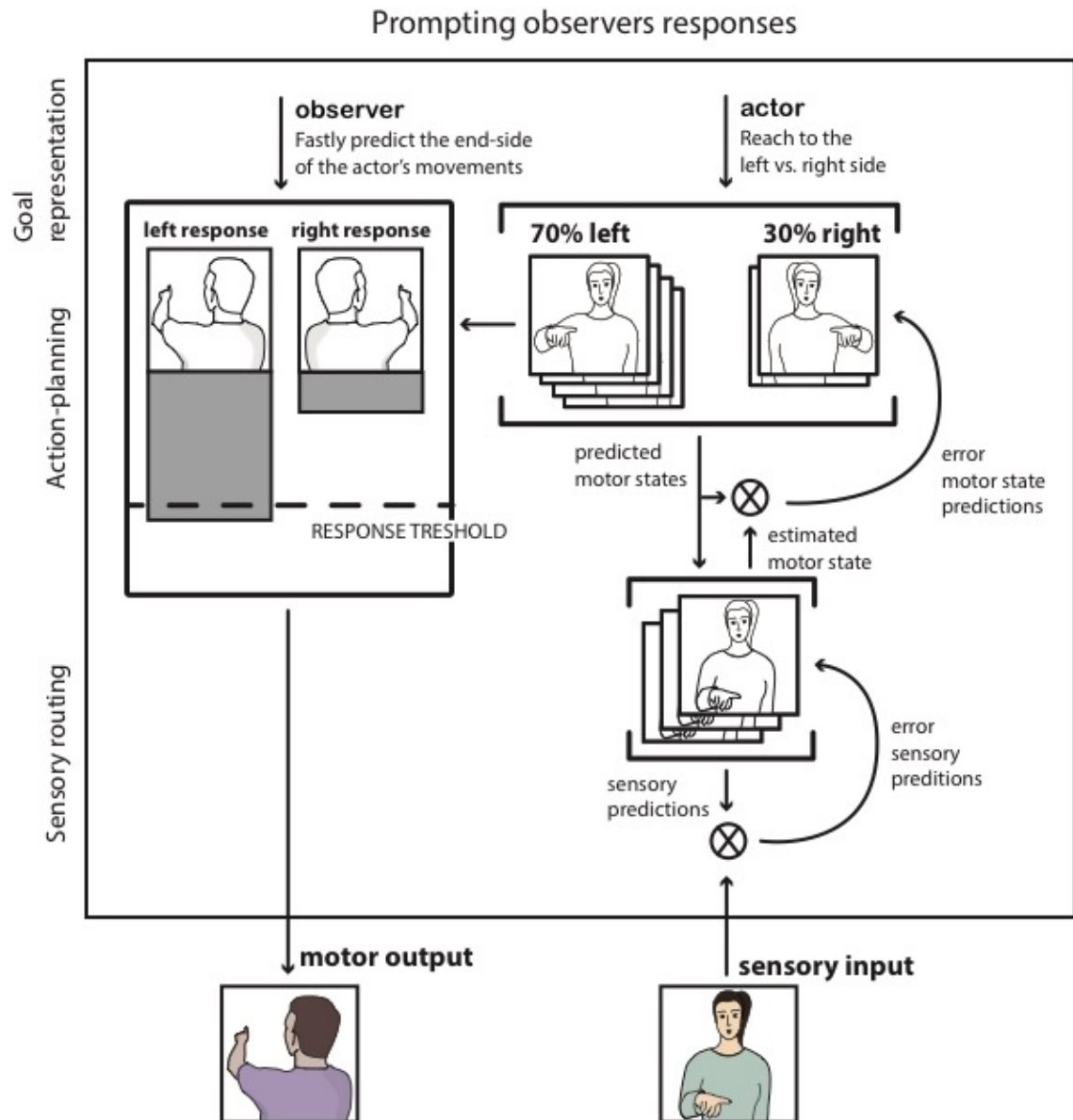


Figure 25 Illustration of response triggering. Probabilistic predictions about the actor weight on the observer motor plans. Once the bias towards one side reaches the response threshold, the motor plan is executed.

4.4 Conclusion

I started this thesis by suggesting that prediction is at the core of social cognition. Human prediction abilities are a bridge between self and other: others become accessible to us because we are able to internally model them and predict their future behavior (Blakemore & Decety, 2001). This speaks to the importance of understanding social predictive mechanisms in human cognition. This thesis offers three contributions to this effect. First, it posits a new theoretical approach to the study of social cooperative interactions. Second, it develops a methodological framework in which an observer's sensitivity to an actor's attentional control can be isolated from that observer's sensitivity to the target of the actor's attention. Third, it presents new evidence in support of the hypothesis that social cognition involves the predictive modeling of other's attentional states (Graziano & Kastner, 2011; Graziano, 2013; Webb & Graziano, 2015). I hope that these contributions represent stepping-stones to further our understanding of the impressive human social abilities.

Bibliography

- Abernethy, B., Zawi, K., & Jackson, R. C. (2008). Expertise and attunement to kinematic constraints. *Perception*, 37(6), 931–948.
- 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.
- Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: sharing tasks with real and imagined co-actors. *Experimental Brain Research*, 211(3-4), 371–85.
- Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action co-representation: the joint SNARC effect. *Social Neuroscience*, 3(3-4), 410–20.
- Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1521), 1235–43.
- Baron-Cohen, S. (1994). How to build a baby that can read minds: Cognitive mechanisms in mind reading. *Curr. Psychol. Cogn.*, 13(5), 513–552.
- Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore, J. D. Philip, & P. Dunham (Eds.), *Joint attention Its origins and role in development* (pp. 41–59). New York: Lawrence Erlbaum Associates, Inc. Publishers.
- Baron-Cohen, S. (2000). Theory of Mind and Autism : A Review. In *International Review of Research in Mental Retardation* (Vol. 23, pp. 169–184).
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, 6(6), 248–254.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17.

- Baron, R. A. (1987). Interviewer's Moods and Reactions to Job Applicants: The Influence of Affective States on Applied Social Judgments. *Journal of Applied Social Psychology*, 17, 911–926.
- Bayliss, A. P., Schuch, S., & Tipper, S. P. (2010). Gaze cueing elicited by emotional faces is influenced by affective context, 18(8), 1214–1232.
- Bayliss, A. P., & Tipper, S. P. (2005). Gaze and arrow cueing of attention reveals individual differences along the autism spectrum as a function of target context. *British Journal of Psychology (London, England : 1953)*, 96(Pt 1).
- Bayliss, A. P., & Tipper, S. P. (2006). Predictive gaze cues and personality judgments: Should eye trust you? *Psychological Science*, 17(6), 514–520.
- 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., Sartori, L., & Castiello, U. (2010). Toward You: The Social Side of Actions. *Current Directions in Psychological Science*, 19(3), 183–188.
- Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychological Science*, 14(2), 151–157.
- Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews. Neuroscience*, 2(8), 561–7.
- Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport*, 11(11), 11–16.
- Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. *Journal of Cognitive Neuroscience*, 11(5), 551–559.
- Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience*, 6(May), 147.
- Brown, K. S., Marean, C. W., Jacobs, Z., Schoville, B. J., Oestmo, S., Fisher, E. C., ... Matthews, T. (2012). An early and enduring advanced technology originating 71,000 years ago in South Africa. *Nature*, 491(7425), 590–3.

- Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4(March), 25.
- Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia*, 40(8), 1129–1138.
- Carp, J., Halenar, M. J., Quandt, L. C., Sklar, A., & Compton, R. J. (2009). Perceived similarity and neural mirroring: evidence from vicarious error processing. *Social Neuroscience*, 4(1), 85–96.
- Cisek, P. (1999). Beyond The Computer Methaphor: Behavior as interaction. *Journal of Consciousness Studies*, 6(12), 125–142.
- Clark, A. (2013). Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–253.
- Clark, H. H. (1996). Using language. *Computational Linguistics*, 23, 425.
- Cline, B. L. (1987). *Men who made a new physics: physicists and the quantum theory*. Chicago, IL: University of Chicago Press.
- Colzato, L. S., de Bruijn, E. R. a, & Hommel, B. (2012). Up to “me” or up to “us”? The impact of self-construal priming on cognitive self-other integration. *Frontiers in Psychology*, 3(September), 341.
- Colzato, L. S., Zech, H., Hommel, B., Verdonshot, R., van den Wildenberg, W. P. M., & Hsieh, S. (2012). Loving-kindness brings loving-kindness: the impact of Buddhism on cognitive self-other integration. *Psychonomic Bulletin & Review*, 19(3), 541–5.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, 3(3), 201–215.
- Cormiea, S., Vaziri-Pashkam, M., & K., N. (2015). Unconscious reading of an opponent’s goal. *Journal of Vision*, 15(12), 43–43.
- de Bruijn, E. R. a, Miedl, S. F., & Bekkering, H. (2008). Fast responders have blinders on: ERP correlates of response inhibition in competition. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 44(5), 580–6.

- de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews. Neuroscience*, 7(3), 242–9.
- Diaz, G. J., Fajen, B. R., & Phillips, F. (2012). Anticipation from biological motion: the goalkeeper problem. *Journal of Experimental Psychology. Human Perception and Performance*, 38(4), 848–64.
- Doerrfeld, A., Sebanz, N., & Shiffrar, M. (2012). Expecting to lift a box together makes the load look lighter. *Psychological Research*, 76(4), 467–75.
- Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2013). The (not so) social Simon effect: a referential coding account. *Journal of Experimental Psychology. Human Perception and Performance*, 39(5), 1248–60.
- Enns, J. T., & Lleras, A. (2008). What's next? New evidence for prediction in human vision. *Trends in Cognitive Sciences*, 12(9), 327–33.
- Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, 424(6950), 769–71.
- Forgas, J. P. (1998). On feeling good and getting your way: mood effects on negotiator cognition and bargaining strategies. *Journal of Personality and Social Psychology*, 74, 565–577.
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3), 490–495.
- Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16(9), 1325-1352.
- Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, 104(1-2), 137–60.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493-501.
- Gallivan, J. P., & Chapman, C. S. (2014). Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets. *Frontiers in Neuroscience*, 8(215), 10–3389.

- Goebel, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, 26(5), 427–438.
- Goodale, M. A. (2011). Transforming vision into action. *Vision Research*, 51(13), 1567-1587.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.
- Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., & Prinz, W. (2007). Predicting point-light actions in real-time. *NeuroImage*, 36 Suppl 2, T22–32.
- Graziano, M. S. A. (2013). *Consciousness and the Social Brain*. New York: Oxford University Press.
- Graziano, M. S. A., & Kastner, S. (2011). Awareness as a perceptual model of attention. *Cognitive Neuroscience*, 2(2), 125-127.
- Häberle, A., Schütz-Bosbach, S., Laboissière, R., & Prinz, W. (2008). Ideomotor action in cooperative and competitive settings. *Social Neuroscience*, 3(1), 26–36.
- Haruno, M., Wolpert, D. M., & Kawato, M. (2003). Hierarchical MOSAIC for movement generation. *International Symposium on Limbic and Association Cortical Systems*, 1250, 575–590.
- Hawkins, J., & Blakeslee, S. (2007). *On intelligence*. Macmillan.
- Holländer, A., Jung, C., & Prinz, W. (2011). Covert motor activity on NoGo trials in a task sharing paradigm: evidence from the lateralized readiness potential. *Experimental Brain Research*, 211(3-4), 345–56.
- Hollerbach, J. M., & Flash, T. (1982). Dynamic interactions between limb segments during planar arm movement. *Biological Cybernetics*, 44(1), 67–77.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *The Behavioral and Brain Sciences*, 24(5), 849–78; discussion 878–937.
- Iani, C., Anelli, F., Nicoletti, R., Arcuri, L., & Rubichi, S. (2011). The role of group

- membership on the modulation of joint action. *Experimental Brain Research*, 211(3-4), 439–45.
- Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in Cognitive Sciences*, 9(1), 21–5.
- James, W. (1890). The principles of psychology. *New York Holt*, 118, 688.
- Johnson, K. L., & Shiffrar, M. (Eds.). (2013). *People Watching: Social, Perceptual, and Neurophysiological Studies of Body Perception*. New York.
- Jones, C. M., & Miles, T. R. (1978). Use of advance cues in predicting the flight of a lawn tennis ball. *Journal of Human Movement Studies*, 4, 231–235.
- Kana, R. K., Keller, T. A., Minshew, N. J., & Just, M. A. (2007). Inhibitory Control in High-Functioning Autism: Decreased Activation and Underconnectivity in Inhibition Networks. *Biological Psychiatry*, 62(3), 198–206.
- Kang, S. K., Hirsh, J. B., & Chasteen, A. L. (2010). Your mistakes are mine: Self-other overlap predicts neural response to observed errors. *Journal of Experimental Social Psychology*, 46(1), 229–232.
- Keller, P. E. (2007). Musical ensemble synchronisation, (December), 80–83.
- Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Consciousness and Cognition*, 16(1), 102–11.
- Keller, P. E. (2013). Ensemble performance : Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 1–69). Oxford: Oxford University Press.
- Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, 8(3), 159–66.
- Kilner, J. M., Vargas, C., Duval, S., Blakemore, S.-J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, 7(12), 1299–301.
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological Research on Joint

- Action: Theory and Data. In WDK2003 (Ed.), *The Psychology of Learning and Motivation* (Vol. 54, pp. 59–101). Burlington: Academic Press.
- Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: interactions of perception and action. *Psychological Science*, 12(6), 467–72.
- Knoblich, G., & Sebanz, N. (2006). The Social Nature of Perception and Action. *Current Directions in Psychological Science*, 15(3), 99–104.
- Koban, L., Pourtois, G., Vocat, R., & Vuilleumier, P. (2010). When your errors make me lose or win: event-related potentials to observed errors of cooperators and competitors. *Social Neuroscience*, 5(4), 360–74.
- 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* (2006), 63(11), 2220–30.
- Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility--a model and taxonomy. *Psychological Review*, 97, 253–270.
- Kourtis, D., Sebanz, N., & Knoblich, G. (2010). Favouritism in the motor system: social interaction modulates action simulation. *Biology Letters*, 6(6), 758–61.
- Kuhbandner, C., Pekrun, R., & Maier, M. a. (2010). The role of positive and negative affect in the “mirroring” of other persons’ actions. *Cognition & Emotion*, 24(7), 1182–1190.
- La Delfa, N. J., Garcia, D. B. L., Cappelletto, J. a M., McDonald, A. C., Lyons, J. L., & Lee, T. D. (2013). The gunslinger effect: why are movements made faster when responding to versus initiating an action? *Journal of Motor Behavior*, 45(January 2015), 85–90.
- Langten, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4(2), 50–59.
- Langton, S. R., & Bruce, V. (2000). You must see the point: automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology. Human Perception and Performance*, 26, 747–757.

- Larson, M. J., Fair, J. E., Good, D. A., & Baldwin, S. A. (2010). Empathy and error processing. *Psychophysiology*, 47(3), 415–424.
- Larson, M. J., South, M., Krauskopf, E., Clawson, A., & Crowley, M. J. (2011). Feedback and reward processing in high-functioning autism. *Psychiatry Research*, 187(1-2), 198–203.
- Liepelt, R., & Prinz, W. (2011). How two share two tasks: evidence of a social psychological refractory period effect. *Experimental Brain Research*, 211(3-4), 387–96.
- Loehr, J. D. (2013). Sensory attenuation for jointly produced action effects. *Frontiers in Psychology*, 4(April), 172.
- Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., Günther, K., & Abs. (2013). Monitoring Individual and Joint Action Outcomes in Duet Music Performance. *Journal of Cognitive Neuroscience*, 25(7), 1049–1061.
- Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience*, 25(7), 1049–61.
- Lotze, R. H. (1852). *Medicinische Psychologie oder Physiologie der Seele*. Leipzig, Germany: Weidmann'sche Buchhandlung.
- Manera, V., Becchio, C., Cavallo, A., Sartori, L., & Castiello, U. (2011). Cooperation or competition? Discriminating between social intentions by observing prehensile movements. *Experimental Brain Research*, 211(3-4), 547–56.
- Manera, V., Schouten, B., Verfaillie, K., & Becchio, C. (2013). Time will show: real time predictions during interpersonal action perception. *PloS One*, 8(1), e54949.
- Marean, C. W. (2015). The Most Invasive Species of All. *Scientific American*, 313(2), 32–39.
- Martin, R. D. (1983). *Human Brain Evolution In An Ecological Context* (Vol. 93). New York: Columbia University Press.
- Mather, G., Pavan, A., Bellacosa Marotti, R., Campana, G., & Casco, C. (2013).

- Interactions between motion and form processing in the human visual system. *Frontiers in Computational Neuroscience*, 7(May), 65.
- Mori, S., & Shimada, T. (2013). Expert anticipation from deceptive action. *Attention, Perception & Psychophysics*, 75(4), 751–70.
- Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, 9(9), 1186–92.
- Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2012). Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cerebral Cortex (New York, N.Y. : 1991)*, 22(12), 2894–903.
- Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, 13(3), 135–143.
- Obhi, S. S. (2012). The troublesome distinction between self-generated and externally triggered action: a commentary on Schüür and Haggard. *Consciousness and Cognition*, 21(1), 587–8.
- Pacherie, E. (2007). The Sense of Control and the Sense of Agency. *Psyche*, 13(1), 1–30.
- Pacherie, E. (2012). The Phenomenology of Joint Action : Self-Agency vs . Joint-Agency. In A. Seemann (Ed.), *Joint Attention: New Developments* (Vol. 93, pp. 343–389). MIT Press.
- Parkinson, J., Springer, A., & Prinz, W. (2012). Before, during and after you disappear: Aspects of timing and dynamic updating of the real-time action simulation of human motions. *Psychological Research*, 76, 421–433.
- Pecenka, N., & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research*, 211(3-4), 505–15.
- Perrett, D. I., & Emery, N. J. (1994). Understanding the intentions of others from visual signals: Neurophysiological evidence. *Current Psychology of Cognition*, 13, 683–694.
- Pinto, Y., Otten, M., Cohen, M. a, Wolfe, J. M., & Horowitz, T. S. (2011). The boundary conditions for Bohr’s law: when is reacting faster than acting?

Attention, Perception & Psychophysics, 73(2), 613–20.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3-25.

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, 58, 1–23.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships Between Perception and Action* (pp. 167–201). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ramenzoni, V. C., Sebanz, N., & Knoblich, G. (2014). Scaling up perception-action links: Evidence from synchronization with individual and joint action. *Journal of Experimental Psychology. Human Perception and Performance*, 40(4), 1551–65.

Ramnani, N., & Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nature Neuroscience*, 7(1), 85–90.

Ristic, J., & Enns, J. T. (2015). The changing face of attentional development. *Current Directions in Psychological Science*, 24(1), 24–31.

Rogers, R. D., Bayliss, A. P., Szepietowska, A., Dale, L., Reeder, L., Pizzamiglio, G., ... Tipper, S. P. (2014). I want to help you, but I am not sure why: gaze-cuing induces altruistic giving. *Journal of Experimental Psychology. General*, 143(2), 763–77.

Rosenbaum, D. A., Herbort, O., van der Wel, R., & Weiss, D. J. (2014). What's in a Grasp. *American Scientist*, 102(5), 366.

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism*, 6(1), 2.

Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: the role of movement information. *Cognition*, 119(2), 242–52.

- Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and Cognition*, 17(4), 1219–27.
- Savelsbergh, G. J. P., Williams, A. M., Van der Kamp, J., & Ward, P. (2002). Visual search, anticipation and expertise in soccer goalkeepers. *Journal of Sports Sciences*, 20, 279–287.
- Schmidt, R., & Lee, T. (2011). *Motor Control and Learning: A Behavioral Emphasis. Human Kinetics*. pp.592
- Schüür, F., & Haggard, P. (2011). What are self-generated actions? *Consciousness and Cognition*, 20(4), 1697–704.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–6.
- Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science*, 1(2), 353–367.
- Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, 88(3), B11–21.
- 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(6), 1234–46.
- Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: an ERP study of action planning and control in co-acting individuals. *Journal of Cognitive Neuroscience*, 18(5), 859–70.
- Sebanz, N., Knoblich, G., Stumpf, L., & Prinz, W. (2005). Far from action-blind: Representation of others' actions in individuals with Autism. *Cognitive Neuropsychology*, 22(3), 433–54.
- Sebanz, N., & Shiffrar, M. (2007). Bodily bonds: Effects of social context on ideomotor movements. In Y. Rossetti, M. Kawato, & P. Haggard (Eds.), *Sensorimotor foundations of higher cognition (attention and performance, XXII)*. Oxford, UK: Oxford University Press.
- Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: the role

- of expertise. *Psychonomic Bulletin & Review*, 16(1), 170–5.
- Semin, R., & Cacioppo, J. T. (2006). Synchronization , Coordination , and Co-Regulation. In *Grounding Social Cognition* (pp. 119–128).
- Siebert, R. J., Harper, D. N., Cameron, F. B., & Abernethy, D. (2002). Self-initiated versus externally cued reaction times in Parkinson’s disease. *Journal of Clinical and Experimental Neuropsychology*, 24, 146–153.
- Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 81(1), 174–176.
- 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), 2335–8.
- Sparenberg, P., Springer, A., & Prinz, W. (2012). Predicting others’ actions: evidence for a constant time delay in action simulation. *Psychological Research*, 76(1), 41–9.
- Springer, A., Hamilton, A. F. D. C., & Cross, E. S. (2012). Simulating and predicting others’ actions. *Psychological Research*, 76(4), 383–7.
- Stenzel, A., Dolk, T., Colzato, L. S., Sellaro, R., Hommel, B., & Liepelt, R. (2014). The joint Simon effect depends on perceived agency, but not intentionality, of the alternative action. *Frontiers in Human Neuroscience*, 8(August), 595.
- Stix, G. (2014). The “It” Factor. *Scientific American*, 311(3), 72–79.
- Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, 7(9), 907–15.
- Tomasello, M. (2009). *Why we cooperate*. *Human Resource Management*, 49(6), 206.
- Tsai, C.C., Kuo, W.J., Jing, J.T., Hung, D. L., & Tzeng, O. J.L. (2006). A common coding framework in self-other interaction: evidence from joint action task. *Experimental Brain Research*, 175(2), 353–62.
- van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2012). The sense of agency during skill learning in individuals and dyads. *Consciousness and Cognition*,

21(3), 1267–79.

van Schie, H. T., Mars, R. B., Coles, M. G. H., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, 7(5), 549–54.

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks : The Official Journal of the International Neural Network Society*, 23(8-9), 998–1003.

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

Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, 211(3-4), 517–30.

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.

Webb, T. W., & Graziano, M. S. A. (2015). The attention schema theory: a mechanistic account of subjective awareness. *Frontiers in Psychology*, 06(April), 1–11.

Welchman, A. E., Stanley, J., Schomers, M. R., Miall, R. C., & Bühlhoff, H. H. (2010). The quick and the dead: when reaction beats intention. *Proceedings. Biological Sciences / The Royal Society*, 277(1688), 1667–74. 3

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

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

Wolpert, D. M., & Miall, R. C. (1996). Forward Models for Physiological Motor Control. *Neural Networks : The Official Journal of the International Neural*

Network Society, 9(8), 1265–1279.