The Geppetto's Dilemma:
Social Sense of Agency and Artificial Agents

Luca Pascolini
100217694
University of East Anglia
School of Psychology
August 2021

A thesis submitted in partial fulfilment of the requirements of the University of East Anglia for the degree of Doctor of Philosophy.

Abstract

The experience of controlling the outcomes produced by our actions is known as sense of agency. During social interaction, fundamental importance is given to monitoring actions performed towards other people, and determining whether an environmental outcome is resulted from our actions or someone else's. If we cause an outcome in the environment, our perception of the time elapsed between our action and the outcome produced will be compressed. This phenomenon has been defined "temporal binding" and has been widely accepted as implicit evidence of the sense of agency. Using a temporal binding paradigm, Experiments 1-3 show evidence for an implicit sense of agency emerging in participants that executed an indirect social action (i.e., a vocal command). Furthermore, data provide evidence that a temporal binding effect can be generated when observing physical and social actions performed by other people, but only if visual access to their actions is allowed.

Given that temporal binding emerges when we observe other humans, the question arises whether such effects are limited to human agents alone. Experiments 4-5 deployed the same temporal binding paradigm to assess subjective time compression experienced in relation to robot-generated actions. Findings suggest that temporal compression was experienced for robotic actions only when the robot was perceived to be independent from human control. Furthermore, experiencing the robot as independent led participants to confer it a higher degree of mental representation, and they were more likely to refer to its components adopting human terms.

Experiments 6-7 adopted a spatial alignment paradigm to investigate the contribution of bodily and mental representation towards action representation of other agents. The results of these experiments indicated that humanoid appearance was not a crucial feature to enable action representation. However, data suggests that action representation for other agents may involve higher level mechanisms, such as explicit belief and joint action representation. This thesis combines novel findings with previous scientific literature to expand current cognitive models of agency and action representation, where the perception of mental activity of other agents gains higher relevance compared to their physical appearance.

Contents

List of Figures

Acknowledgements

When I firstly expressed interest in pursuing a Ph.D., all the people I knew working in Academia heavily warned me to think this through. It would have been most likely the hardest challenge of my studies. Now with hindsight, I can acknowledge what kind of challenge they were talking about: take a ship, in the dark, and go exploring somewhere nobody has ever been to. Luckily, I was not alone sailing this ocean.

Firstly, I would like to thank my Supervisory Team. Dr. Natalie Wyer provided invaluable support during all the time I have been working with her. She strongly encouraged my independence, without ever letting me feel lost in case of uncertainty. She was the nautical chart in my hands, guiding me where I could not orientate by my own. Dr. Andrew Bayliss contributed with unexpected passion to my research. He always promoted a scientific approach based on high-standard and open science. His dedication was inspiring, and helped me to develop a solid methodology and research design skills. He was the compass in my hand, helping me to plot the route to sail on.

However, a ship does not go much afar by its own. It is only thanks to the help of my fellow peers that I could led it forward. I am grateful to Dr. Michela Caldano, Nicholas Menghi, Laura Manno, Francesco Silvestrin, Greta Melega, and Alice Bush, which were always available to discuss research protocols and results interpretations. The friendship we built inside and outside the workspace was a solid scaffolding that prevented mental breakdowns. I would also like to thank Dr. Gareth Edwards and Dr. Lisa Stephenson, which as senior students introduced me to the professional research practice, showing me the attitudes of a good researcher. We have all been comrades in this expedition.

Not to be forgotten is the support provided by the University staff, with special thanks to Jackie Orford and Scott Steward-Smith as part of the Local Support team, and Richard Carey-Knight, Malcolm Rae, and Elisabeth Watson as members of the Technical team. Without them the Ph.D. route would have been much less smooth.

Last, but not least, I want to thank my family and especially my partner Alessia Battistini, for their unconditional love and support, which gives me a reason to go back home every night, and happily wake up the next morning.

Author's declaration and publications

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas, and contributions from the work of others.

The research presented in Chapter 2 has been communicated to the scientific community during Conferences Poster Presentations, and it is currently been reviewed for publication in the *Journal of Experimental Psychology: Human Perception and Performance*.

The research presented in Chapter 3 has been communicated to the scientific community during Conferences Poster Presentations, and it is in preparation for publication, to be submitted to *Cognition*.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the School of Psychology Ethics Committee at the University of East Anglia.

SECTION 1: Introduction and Aims

Chapter 1: Introduction

"Agency used righteously allows light to dispel the darkness
and enables us to live with joy and happiness."
Robert D. Hales, 1968 -

Aims and Objectives

The aim of this thesis is to examine the cognitive mechanisms involved in action representation during human interaction with other agents, when they are natural (i.e., other humans) or artificial (i.e., vocal assistants, robots). Namely, the effects of agency detection for actions exerted by different agents and action modalities will be explored. This chapter will start with a detailed outline of the aims and purpose of the thesis in a comprehensive summary. Successively, the reasons beyond the imminent importance of human interaction with artificial agents will be explained, followed by a review of the relevant literature.

Thesis Overview

The primary aim of this work is to explore the social dimension of the sense of agency, which has generally been investigated in the sole context of direct physical actions. To do this, the means by which indirect social actions generate temporal compression (an implicit index of the sense of agency) will be explored. As second point, the thesis will expand related knowledge about action observation, testing whether other-generated physical and social actions are computed analogously by the observer. These data will set a milestone towards a third aim of this work, which is to compare how instances of agency are detected when observing human and robotic actions, further exploring the fundamental features involved in this dissociation. Lastly, the thesis will aim to investigate the effects elicited on our motor system by the representation of actions performed by other agents, and if this element is based on bodily representations (e.g., anthropomorphic shape) or mental instances (e.g., action opportunity).

The importance of understanding artificial agents

In the famous novel *The adventures of Pinocchio,* Carlo Collodi narrates the story of a woodcarver named Geppetto who, in the impossibility of being father, takes the initiative and carves himself a puppet son out of a solid pine log. The novel then drifts on fantastic elements, involving a fairy that animates the puppet and eventually makes him a real boy after a long redemption journey. Overlooking the narrative's fantastic features, we can wear Geppetto's shoes, and see what he likely asked himself: is Pinocchio an object, or is he human? Does he have goals and act to accomplish them? Is he agentic?

In recent times, wooden puppets are not frequently found in our daily lives, but we do have experience with other artificial agents. As a matter of fact, robots and vocal assistants have started to accompany our daily routines, assisting us in a wide range of activities not necessarily limited by tasks that are repetitive, hazardous, or prone to errors (Takayama et al., 2008). On the contrary, there are a multitude of other activities where artificial agents can help us, such as teaching (Church et al., 2010), and health care. For example, some artificial gents have been successfully deployed with elderly patients to increase social communication and improve cognitive performance (Tapus et al., 2007; Birks et al., 2016; Yamazaki et al., 2016), leading also to positive outcomes on social bonding and emotional expression among dementia patients (Martin et al., 2013). Younger patients can also benefit from artificially assisted clinical intervention. People falling within the Autistic Spectrum Disorders (ASD) can be helped in the practice of their basic social skills, such as emotion understanding and joint attention (Tapus et al., 2012; Kajopoulos et al., 2015; Warren et al., 2015).

Given the swift diffusion of such artificial agents, it is not hard to envisage how their presence will exponentially increase in the next decades. Thus, it becomes of crucial importance to explore the cognitive nature of the social interaction between humans and robots, and to provide insightful research findings (Wiese et al., 2017). In the first instance, it is of fundamental importance to investigate the mechanisms of agency detection, to understand what makes a robot to be perceived as an artificial agent, rather than a tool at our disposal. As such, robot manufacturers and software developers will be guided towards the design of artificial agents that can best interact with us, and be widely accepted by the public – somewhat what we regret it did not happen every time we watch a dystopic robot uprising movie, such as *Blade Runner* or *2001: A Space Odyssey*.

Literature Review

In our daily activities, and particularly during our interactions with other people, it is essential for us to efficiently detect which events are the result of our own (or others') actions, and dissociate them for the events in the environment for which we are not responsible. For example, if we press the doorbell when visiting a friend's house and hear the bell ring, we know that our action was the cause of the sound. Self-produced events are generally easy to detect, and are followed by a conscious feeling of efficacy over causing a change in the environment, known as sense of agency (Jeannerod, 2003). The sense of agency is usually generated automatically, granting us quick and informative feedback about our influence on the environment. Since the first investigations into voluntary actions and sense of agency (Haggard et al., 2002), the interest of the scientific community has been growing noticeably. In this Chapter, the existing research relevant to understanding the sense of agency will be discussed, with a clear distinction between feeling of agency and judgements of agency, and how they are theorised to be generated. Successively, different methods by which agency can be measured will be reviewed, with a specific focus on subjective temporal compression as an implicit measure of the sense of agency. Following a discussion about the cognitive effects exerted by action execution and perception, this Chapter will examine motor effects induced by our motor system during action representation.

The experience of agency

The sense of agency is generated through the integration of two distinct components: feeling of agency and judgements of agency (Synofzik et al., 2008). The feeling of agency is based on low-level mechanisms generated through sensorimotor signals. By contrast, the judgement of agency refers to high-level processes regarding the authorship of an action, and is generated when agency becomes explicitly conscious. Prior research has shown that the judgement of agency can be affected by top-down cognitive mechanisms, like prior beliefs and action related awareness, making it more sensitive to contingency between events (Gallagher, 2000). Moreover, more recent evidence suggests that the feeling of agency can also be influenced by such higher-level cognitive processes, such as prior beliefs (Desantis et al., 2011). This element will be further addressed in Chapter 3, where we show how different knowledge and expectations about a robot's behaviour can affect participants' agency detection. According to this model, feeling of agency and judgements of agency are deeply interconnected, with feeling of agency generating a necessary but not sufficient condition for a judgement of agency to be computed (Haggard & Tsakiris, 2009). Therefore, even though feelings of agency and judgements of agency are usually arising together, they can also be dissociated. Moore et al. (2012) delved into this dissociation, testing whether feelings of agency and judgements of agency were modulated differently through sequential patterns of action and outcome. Their findings support a model where feelings of agency and judgements of agency can be thought of as two dissociable systems, and yet not completely independent from each other. Additional evidence for a dissociation between feelings of agency and judgements of agency was reported by Saito et al. (2015), who found no correlation between the two. Taken together, previous studies point towards a dissociation between feelings of agency and judgements of agency, but how these two systems may influence each other is not yet clearly understood. Data from Chapter 2 will help to shed some light on the topic, as we showed that feelings of agency can be experienced for observed actions performed by another agent, whilst judgements of agency were not.

Models for agency detection

The cognitive mechanisms underlying the computation of the sense of agency have been widely debated, and more than one theoretical model has been proposed to the scientific community. On one hand, it has been suggested that the sense of agency could be generated by external cues that produce reflective post-hoc inferences about the environment (Wegner, 2003; Wegner & Wheatley, 1999). On the other hand, following the forward models of motor control (Blakemore et al., 2001) the sense of agency could arise as a consequence of internal cues of pre-reflective motor and sensory predictions (Blakemore et al., 2002). In more recent times, the optimal cue integration account has been proposed, posing a challenge to the pre-existing theoretical framework (Moore et al., 2009). According to this model, both external and internal cues are supposed to promote the formation of a sense of agency (Synofzik et al., 2010). These three models will be further discussed below, with critical consideration of their function in the generation of the sense of agency.

The apparent mental causation model provides a reconstructionist and postdictive interpretation of agency detection (Wegner, 2002). According to this model, agency detection for self-generated actions depends on explicit consideration of the intention-action and action-outcome connections, based on three main factors: consistency, priority, and exclusivity. Consistency refers to the fact that executed actions need be explained by an appropriate intention. Priority refers to the need of the intention to precede the action, in terms of prior planning and execution.

Exclusivity refers to the exclusion of other plausible causes for the outcome to be produced. In other words, if we ring a doorbell, and we had the specific intention to ring the bell (consistency), the button press happens after our intention to perform the action (priority), and the bell does not ring in response of other events (exclusivity), agency can be retrospectively consciously detected. More recently, this model has been expanded to account not only for a deep intention-action connection, but for an action-outcome connection as well. As such, consistency also imply a congruency between the executed action and the perceived outcome, while priority also refers to the need of the action to precede the outcome. According to this model, motor predictions do not concur for agency detection. To support this model, research provided evidence that explicit agency judgements can be generated with no contribution of internal motor signals (Wegner et al., 2004). During this experiment, participants were paired with a confederate partner, positioned behind them, with their arms rested in a position that conformed to the participants' natural posture (See Fig. 1). Participants were asked to observe their mirror reflections, while the confederate executed arm and hand gestures. Results showed that when researchers asked the confederate to make specific movements, participants were likely to report a sense of control over those movements. This finding was interpreted as evidence that judgements of agency can be experienced if the three components (i.e., consistency, priority, and exclusivity) are met. As a matter of facts, participants heard an instruction that was congruent with the observed action (consistency), when the action happened after the instruction (priority) and no other means could explain the action (exclusivity). However, it is worth to emphasise that participants observed their mirror reflections after hearing the instructions provided by the experimenter. As such, participants developed a conscious representation of the action before they saw the confederate's arms performing that action, which could have influenced their motor system. Yet, this finding was corroborated with records of increased galvanic skin responses when the confederate's hand was approached with a source of pain, providing insights into participants' sense of ownership to the confederate's hand. As interpreted by the authors, these findings suggest that agency can be detected with no involvement of motor systems. Authors therefore argued that perceptions of having conciously willed an action are illusory, as judgements of agency are postictively generated once that the outcomes of that action have been experienced.

Figure 1. Experimental setup in the Wegner et al. (2004) experiment. Participants saw themselves in the mirror as a confederate hidden from view performed arm movements and hand gestures.

By contrast, according to the comparator model (see Fig. 2), agency detection occurs as the result of making predictions about the state of the motor system (internal forward model) and sensory processing of action outcomes (forward dynamic model: Blakemore et al., 2002, Wolpert & Ghahramani, 2000). Once the goal of an action has been set, the inverse model computes the adequate motor program to perform the appropriate action, which will then be sent to the relative effector. At the same time, an efference copy of such motor program is sent to the forward model, which will estimate the predicted sensory outcomes for the executed action. After the action has resolved, actual sensory feedback and predicted outcomes will be compared. If the two are congruent, the action outcomes are perceived to result from the action performed, and sense of agency is experienced. On the other hand, if there is a mismatch, it means that the executed action needs correction, as the selected motor program was not adequate, or because external factors interfered during the action execution (Haggard, 2017). For example, when practicing basketball free throws in one's garden, a player's goal is to throw the ball in the basket. As such, the appropriate motor program is selected and sent to the relative effectors (core, legs, arms, and hands). At the same time, an efference copy of that motor program is used to predict the outcomes of the imminent action. If the predicted outcomes (e.g., ball in) then matches the actual sensory feedback, agency is attributed. If it does not, it means that the motor program needs correction (perhaps to account for the wind). As a result, people who developed highly refined motor programs (i.e., elite players) are capable to predict whether a throw will be successful even before that the ball is thrown (Aglioti et al., 2008).

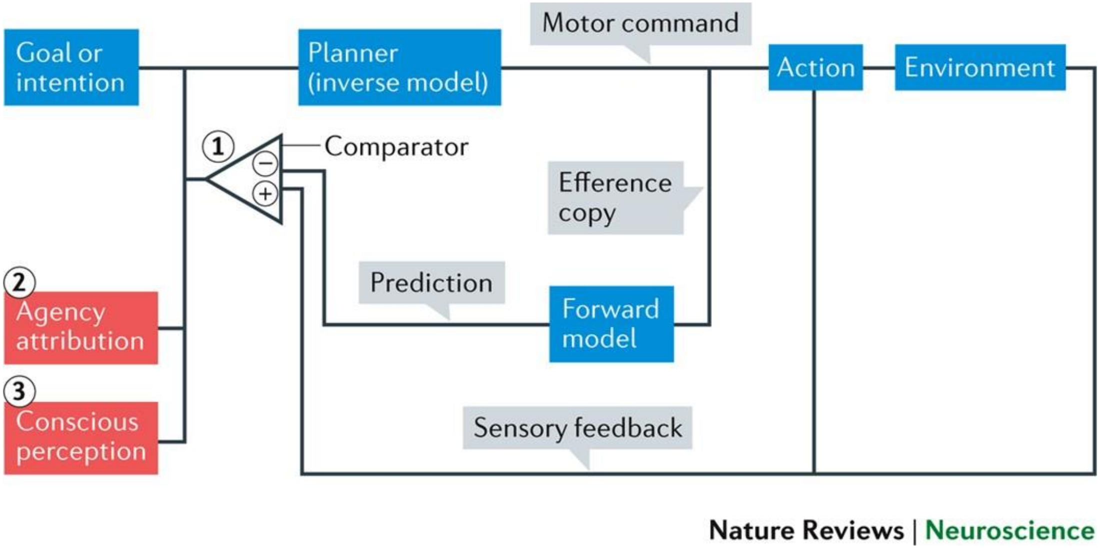


Figure 2. The comparator model of action control from Haggard (2017). The comparation between predicted and actual sensory outcomes can lead to: (1) adjustments of the motor program deployed (incongruent match), (2) agency detection for the action performed (congruent match), and (3) sensory attenuation of predictable outcomes.

As a secondary element, since the sensory system is already deployed during the feedback prediction, the perfect congruency between actual and predicted outcomes leads to a sensory attenuation phenomenon, where the sensory response is perceived as weaker than it actually is. This model received vast empirical support. For example, Blakemore et al. (1999) developed a device capable of replicating the exact movement executed with the index finger with a robotic hand (see Fig. 3). This device was further capable of delaying or rotating the movement trajectories. Researchers asked participants to execute light movements with their finger, as the robotic hand replicated them on the participants' other palm. In different trials, the mirrored moments were delayed (100, 200, 300 ms) or rotated (30°, 60°, or 90°), to effectively manipulate the actual sensory feedback participants experienced. After each trial, participants were requested to report the tickling experienced in a 1-10 scale. Results showed an increased ticking feeling when the action was delayed or rotated, following a linear trend. In other words, when there was no delay or rotation the forward model correctly predicted the sensory consequences of the movement.

On the other hand, as the sensory feedback deviated from the model predictions (increasing delay or rotation), the sensory discrepancy between predicted and actual feedback increased, leading to a decreased sensory attenuation, and thus resulting in the participants experiencing higher degrees of ticking sensations.

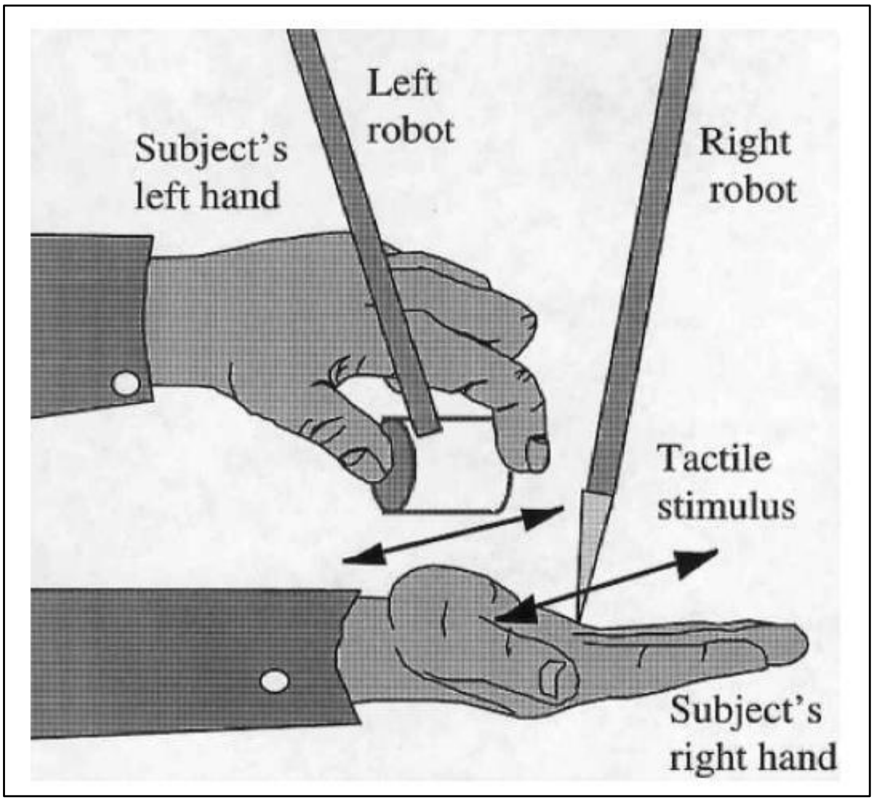


Figure 3. The motion replication device deployed in Blakemore et al. (1999). Movements executed with the index finger were mirrored by the robot hand.

Since the sense of agency is detected with minimal cognitive effort, researchers wondered about the reliability of the perceptual system. In order to reliably detect agency, it is expected that a vast range of features are examined. The optimal cue integration model (Moore et al., 2009) theorises that the sense of agency is related to multiple sources, and identifies the effects of both predictive and postdictive accounts of agency. Analogous to the comparator model, sensorimotor predictions serve as predictive cues, while postdictive cues can be identified in sensory feedback. At the same time, prior knowledge about the environment and explicit beliefs can serve as either predictive or postdictive cues, hence including reconstructive components of the apparent mental causation model (Synofzik et al., 2013). The optimal cue integration model theorises that different cues are integrated to generate an appropriate estimate of agency through the computation of each discrete sensory cue, weighting them on the basis of their reliability and importance (Vosgerau & Synofzik, 2012, see Fig. 4 for a graphical representation). This results in a flexible approach to agency detection, and can also explain differences in perceived agency across different contexts (Moore & Fletcher, 2012). This model has received empirical support from Moore and Haggard (2008), who manipulated the probabilistic occurrence of the action outcome. They showed that with a high chance of outcome occurrence, a sense of agency was generated even if the action did not cause an outcome at all. By contrast, when the occurrence chance was low, agency was only reported when the action was followed by its outcome. According to this evidence, it is supposed that agency detection was affected by prior knowledge about the chances for the outcome to occur.

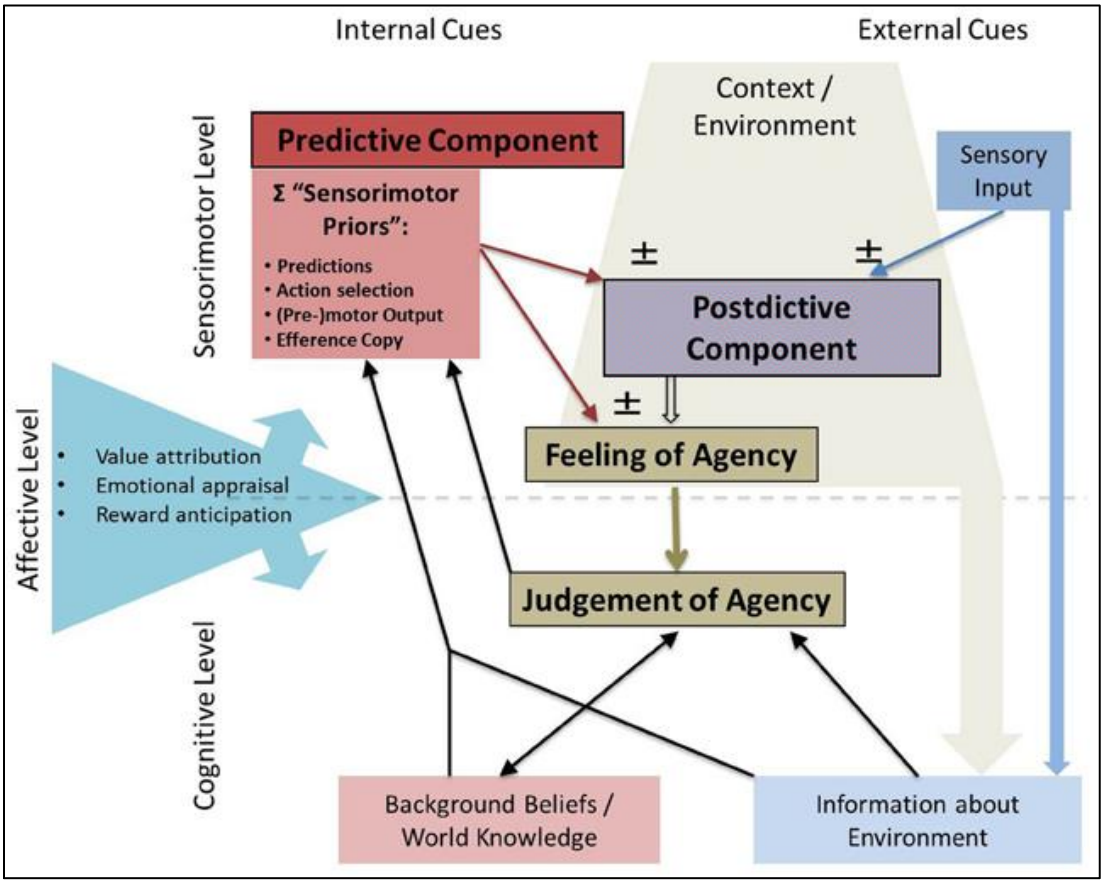


Figure 4. The optimal cue integration model triggering the experience of agency, as theorised by Synofzik et al. (2013).

Sense of agency and time perception

Existing literature has consistently reported a fascinating phenomenon that occurs when an executed action produces an outcome, which has been described as a subjective temporal compression of the time elapsed between the action and the outcome. Action and outcome become bound together in time, hence this effect has been named as temporal binding (Barlas & Obhi, 2013; Engbert & Wohlschläger, 2007; Engbert et al., 2007; see Moore & Obhi, 2012 for a review). Different authors have proposed different nomenclatures to the binding effect, some of which imply that mere causality between action and outcome is sufficient to trigger temporal compression (causal binding: Buehner & Humphreys, 2009), whilst others have theorised that pre-existing intention and action planning are necessary to experience time compression (intentional binding: Haggard et al., 2002). Since it is not the primary goal of this thesis to address the discrepancy between these models, the more inclusive term of temporal binding will be adopted. However, indirect evidence that intention is a crucial element to generate a binding effect will be provided in Chapters 2 and 3.

Since temporal binding tends to occur for actions that are followed by predictable outcomes, it is thought to reflect an implicit dimension of the sense of agency. For example, it has been shown that volitional and intentional actions generate a binding effect, as opposed to involuntary and unintentional actions (Barlas & Obhi, 2013; Engbert et al., 2007; Engbert & Wohlschläger, 2007; Haggard et al., 2002; Tsakiris & Haggard, 2003; Tsakiris et al., 2005). A common interpretation of these findings is that the generation of a binding effect is due to the human inability to accurately quantify elapsed time, and as such, time perception relies on internal processes to estimate temporal duration.

Different models have tried to account for internal processes involved in time perception, which can be classified as intrinsic and dedicated models (Wittman, 2013). On one hand, intrinsic models theorise that cognitive and sensory mechanisms serve as internal reference points for time perception. For example, the degree of cognitive resources allocated during a task can indicate its time duration. On the other hand, dedicated models refer to pacemaker-accumulator processes, where a build-up of 'beats' generated by an internal clock scans the time spent (Gibbon et al., 1984; Wearden et al., 1999). The pace of the beats is affected by motor activity and arousal degree, as reflected by athletes entering a competition trance, who report delayed temporal perceptions (Hagura et al., 2012). When the pace of the internal beats hastens, more pacing beats are gathered, and time is perceived to be delayed. On the contrary, perceptions of time are shortened when the pace of the beats slows. The internal clock's pace is thought to decelerate when a motor prediction generates agency. This process leads to a subjective temporal compression (temporal binding), and thereby reflects implicit agency detection (Wenke & Haggard, 2009).

Measures of the sense of agency

The sense of agency can be measured through different methods, including the explicit experience of agency, sensory prediction cues, and the temporal binding effect. Specific methods developed and adopted to investigate different components of the sense of agency will be reviewed throughout the next paragraphs.

Since the sense of agency manifests itself as the explicit experience of *I did that*, some authors have adopted explicit ratings to measure the sense of agency. As such, these measurements provide an assessment for judgements of agency, rather that feelings of agency. Traditionally, paradigms measuring explicit agency ratings lead participants into an introspective evaluation, asking them to provide a dichotomous answer of agent identification, or alternatively a rating of the degree of agency experienced for a specific sensory outcome. For example, participants can be asked whether they, another person, or a computer caused the outcome (Aarts et al., 2005; Dewey & Carr, 2013), or to report the degree of control they had over the performed action (Hon et al., 2013; Sato & Yasuda, 2005; Wegner et al., 2004). However, given the explicit nature of these responses, participants become particularly prone to social desirability effects, leading to an increased chance of acting in a way that would be preferred by the experimenter. Additional confounds could be found in the tendency to polarize scale ratings to match an integer value or the expected behaviour (Hornik, 1981; Tversky & Kahneman, 1974). It is worth noting that previous studies using these methods have shown a consistent cognitive bias: participants tend to underestimate others' agency in favour of their own, misattributing to themselves outcomes that are not actually related to their actions (Tsakiris et al., 2005; Wegner & Wheatley, 1999). Surprisingly, this bias is found to be particularly strong when the action outcome is positive, rather than negative or neutral, suggesting that explicit agency ratings in social contexts could be influenced by a self-enhancement bias as well (Bandura, 1982; Obhi, 2012). Furthermore, self-reported feelings of control over an action are shown to depend on the agent's own skill of introspection (Barlas & Obhi, 2013, Sebanz & Lackner, 2007). Taken together, these elements suggest that explicit agency ratings may not reflect the actual experience of agency.

As introduced above, the match between predicted and actual sensory information can serve as cue for the sense of agency. When predicted feedback matches the sensory afferent information, a sensory attenuation effect occurs, leading to a reduction of the outcome's salience. For example, a self-produced tickling movement results in a decreased tickling feeling, self-generated sounds are perceived as less loud, and self-produced speech results in attenuated neural activity (Bays et al., 2006; Blakemore et al., 1999; Blakemore et al., 1998; Blakemore et al., 2000; Ford et al., 2007; Timm et al., 2014; Weiss et al., 2011). As such, these sensory effects can be exploited as measures for the sense of agency, where decreased sensory salience and neural activity provide an evidence of agency.

The documented phenomenon of subjective time compression between action and outcome led to the development of a vast range of paradigms that could be adopted to measure the sense of agency. This subjective temporal compression between executed action and produced outcome is believed to be due to the deceleration of the internal clock beats, triggered by the motor prediction cues found in agentic actions (Wenke & Haggard, 2009). Classically, these paradigms compare the subjective amount of elapsed time to the actual delay between an action performed and the outcome it produced. Previous results are consistent in reporting a shorter perceived duration of time between the events, which become temporally bound (see Moore & Obhi, 2012 for a review). Such temporal binding measurements are theorised to reflect a reliable implicit degree of feelings of agency. Different paradigms have been proposed to assess temporal compression throughout the years, including the Libet clock method, the interval duration estimate, and the interval reproduction. In the following paragraphs, each paradigm will be discussed in detail.

The exploitation of subjective time compression as a measure for implicit sense of agency was introduced by Haggard et al. (2002), who adopted an existing method developed by Libet et al. (1983) to investigate pre-action readiness potentials (which will be discussed in detail below). The pioneering work from Haggard and colleagues (2002) involved participants looking at a clock on a monitor, with a rotating hand. In the agentic condition, participants were asked to press a button at a time of their choosing, which in turned generated a tone after a fixated delay (250 ms). Participants' task was to report the clock hand position either when the button was pressed or when the tone was heard. Their performance was then compared to two separate baseline conditions, in which the action led to no outcome or the tone was automatically produced, and participants reported the hand position at the event of interest. Findings showed that while the clock hand position was reported accurately when the events were independent (baseline condition), time compression was achieved when participants executed an action that produced the tone. More specifically, participants reported that the clock hand's position was delayed for the action (e.g., position 8 when it was 5), while it was anticipated for the tone (e.g., position 13 when it was 15). Fig. 5 shows a typical trial procedure of the Libet clock method.

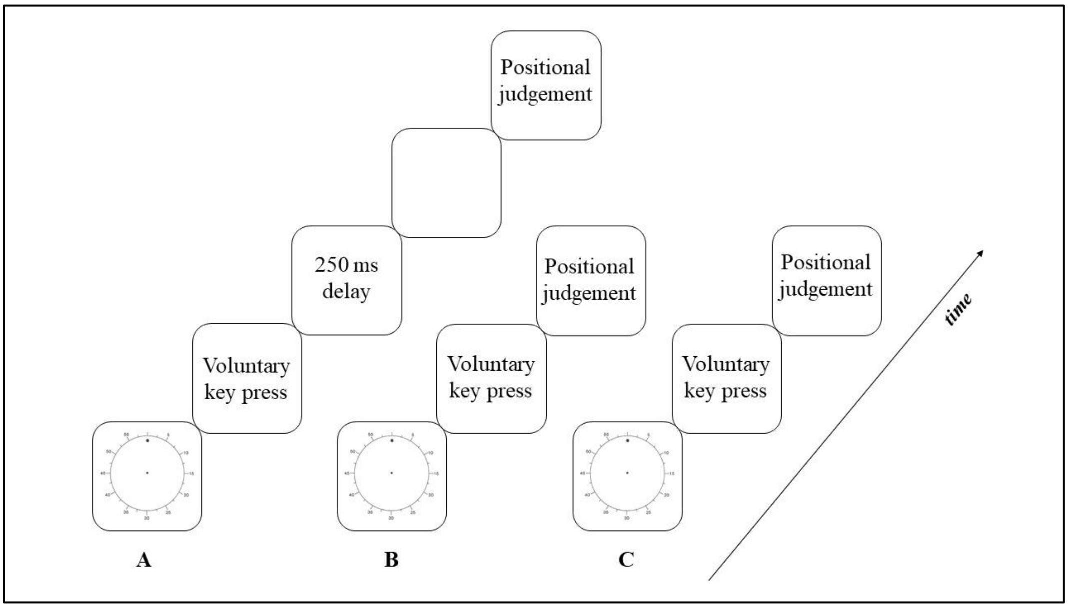


Figure 5. Trial procedure for agentic conditions (A) and baseline conditions (B & C). In the agentic condition participants executed a voluntary key press which release a tone after a 250 ms delay. Participants then reported positional judgements of either the key press or the tone. In baseline conditions, participants either performed a key press (B) or heard an automatically generated tone (C), then reported positional judgements of the event of interest in distinct blocks.

The Libet clock method is by far the most widely adopted measure to assess temporal binding (e.g., Engbert & Wohlschläger, 2007; Haggard et al., 2002; Moore & Haggard, 2008; Obhi & Hall, 2011; Strother et al., 2010). This paradigm provides great value in the assessment in the perceived onset of the specific events, allowing a direct investigation of the specific contributions of action and outcome in the generation of temporal compression. Moreover, reporting the clock hand position also prevents social desirability effects, as the aim of the task in unknown by participants. Nevertheless, it is worth noting that this paradigm measures the time distortion of one specific event, rather than the actual elapsed time between action and outcome, providing an indirect measure of the temporal binding effect. Moreover, the mechanism underlying action control can be additionally disrupted by the demand to divide attention between the action and the clock (Engbert et al., 2007). Also, participants could tend to report rounded values of positional judgements (e.g., 30 instead of 29), leading to an inaccurate reflection of the actual judgment.

An explicit estimate of the duration of a time interval is a kind of magnitude assessment where participants are asked to translate perceived time intervals into conventional units of time (e.g., milliseconds), which can then be compared across agentic and control conditions. Typically, the two events defining the beginning and the end of the interval are an action (e.g., a key press) and a consequential auditory stimulus (e.g., a tone) (Engbert et al., 2007; Engbert, et al., 2008; Moore et al., 2009). The task's explicit nature reveals similar disadvantages as explicit agency ratings. For example, participants are more likely to report rounded numbers or specific multiples (Hornik, 1981). Furthermore, participants could provide biased responses due to an anchoring effect, wherein short durations can decrease temporal estimates while long durations can increase them (Tversky & Kahneman, 1974). Hence, reported temporal estimates may not reliably reflect the time intervals actually perceived.

More recently, a new paradigm to measure temporal binding effects has been developed by Buehner and Humphreys (2009), and has been widely adopted and adapted by the scientific community. This paradigm consists of presenting participants with variable temporal durations. In control conditions, the temporal interval is defined by two independent events, not related by agency or causal relationship (e.g., two tones). In agentic conditions, participants perform an action (e.g., a key press), which after a variable time delay releases the outcome (e.g., a tone). Participants are then asked to reproduce the perceived temporal duration between the events, by pressing a key for the same duration of time. As opposed to explicit verbal estimates, time interval reproductions allow a direct measurement of the perceived elapsed time, and an experiential reproduction of time intervals. Hence, the chance of biased responses is reduced, because temporal intervals do not need to be translated to conventional temporal units. The inclusion of an appropriate control condition accounting for the sequence of two events allows this paradigm to consider the deep connection between agency, causality, and contiguity between the events (Buehner & Humphreys, 2009) providing an advantage compared to the Libet method where temporal compression is assessed on singular events. It is important to emphasise that in control conditions, the target time interval begins after the first tone. As such, participants could include the duration of the stimulus inside the target time interval, increasing its duration. This confound could account for most of the binding effects eventually found in agentic conditions. However, to the author's knowledge no evidence has been reported to support this claim, and it could be possible that if participants include the duration of the starting stimulus inside the target time interval for the control condition, they may do the same for agentic conditions. Experiments 1-3 reported in Chapter 2 will account for this possible confound, as across different conditions participants executed and heard vocal actions. Given the congruent modality (i.e., auditory) between control and agentic conditions, it possible to minimise this confound.

Sensory and neural attenuation in agency detection

As discussed above, agency detection is theorised to function through forward models and predictive mechanisms. Forward models take advantage of efference copies of motor programs to predict the sensory feedback resulting from those motor acts (Blakemore et al., 2002; David et al., 2008; Davidson & Wolpert, 2005; Synofzik et al., 2008; Wolpert, 1997; Wolpert & Ghahramani, 2000; Wolpert et al., 1995). As a consequence of sensory prediction, the salience of the perceived stimulus is reduced, when compared to sensory stimuli that are externally generated (thus, not predictable). For example, ratings of intensity for somatic effects are reduced when generated by voluntary movements, compared to TMS-induced involuntary movements (Tsakiris & Haggard, 2003), while self-generated tones are perceived to be quieter than other- or computer-generated tones (Weiss et al., 2011). Recent evidence showed that tones are perceived as quieter also when produced by an observed agent, suggesting that forward models can be triggered by the observation of other-generated actions, through the activation of the mirror neuron system (Sato, 2008). Such sensory attenuation following motor predictions is believed to serve an adaptive function to reduce sensory processing for predictable stimuli, in order to optimise cognitive resources that can best be allocated to the processing of external stimuli, like unexpected events that could require immediate attention and reaction (Bays et al., 2006; Hesse et al., 2010). This idea is widely supported by evidence of reduced awareness for visual stimuli preceded by self-produced changes (Berberian & Cleeremans, 2010).

Sensory attenuation phenomena have also been shown on a neural level, whereby self-generated tactile, visual, auditory, and speech stimuli resulted in reduced neural responses (Eliades & Wang, 2003; Haggard & Whitford, 2004; Hesse et al., 2010; Hughes & Waszak, 2011; Jo, et al., 2014; Martikainen et al., 2005; Poonian et al., 2015). Event-related potentials (ERPs) are electrophysiological measures to stimuli responses. They can be measured using electroencephalography (EEG), a non-invasive research method capable of detecting electrical brain activity using electrodes positioned on the scalp surface (Luck & Kappenman, 2012). Among the several ERP waveforms indicating neurocognitive processes, studies on the sense of agency focussed their interest on the pre-action readiness potential (as a reflection of neural activity involved in motor preparation), while research on sensory attenuation has investigated the N1 and P2 components, which are responsive to agency contexts in action-outcome processing.

The pre-action readiness potential is a transitory increase of slow negative potentials that begins around 2000 milliseconds prior to a conscious decision to perform an action (Dirnberger et al., 1998; Libet et al., 1983; Soon et al., 2008; see Shibasaki & Hallett, 2006 for a review). Readiness potentials are theorised to indicate the phases of action initiation and preparation, and to contribute to the formation of a sense of agency. As such, readiness potentials have been used as a measure to investigate voluntary actions (Haggard, 2008; Shibasaki & Hallett, 2006). The amplitude of negative potentials has been shown to reflect the degree of implicit agency measured over an action related outcome. Namely, Jo et al. (2014) reported increased negative amplitudes associated with greater degrees of temporal binding effects. Interestingly, this association has been found only in early readiness potentials (circa 2000 ms prior to the action), but not in late readiness potentials (circa 500 ms prior to the action). Also, TMS-induced disruptions of the pre-supplementary motor area (SMA; a brain region associated with the conscious intent of acting) resulted in reduced temporal binding effects in relation to the outcome produced, whilst temporal binding effects over the action performed were not affected (Moore et al., 2010). Taken together, this evidence suggests that readiness potentials are involved in the generation of the temporal binding effect, and could serve a critical role in the magnitude of agency experienced.

Crucial to the purpose of this thesis, when considering actions executed by other agents, motor predictions are hypothesised to rely on the simulation of the observed action in the same areas that are involved when that action is self-generated (Rizzolatti et al., 1996). Previous research investigating neural correlates of action observation showed evidence that slow-wave activity (i.e., readiness potentials) may be generated prior to both self-produced and other-produced observed actions (Kilner et al., 2007). However, whether such neural activity reflects agency detection has yet to be determined.

The N1 component is a negative waveform which has been linked to auditory ERPs. It generally occurs around 80-120 ms after the presentation of an auditory stimulus. Analogously, the P2 component is a positive waveform linked to auditory ERPs, with maximum amplitude reached around 150-275 ms after the presentation of an auditory stimulus. Together, these two components are known as the N1/P2 complex. Research on sensory attenuation has consistently reported reduced N1 and P2 amplitudes for auditory stimuli resulting from self-generated voluntary actions, when compared to auditory stimuli generated independently from participants' will (Ford et al., 2013; Horváth et al., 2012; Knolle et al., 2012; Knolle et al., 2013; Kuhn et al., 2014; Martikainen et al., 2005; Sowman et al., 2012). Analogously, N1 amplitudes are reduced for self-generated speech (Ford et al, 2007; Timm et al., 2014). Therefore, N1 and P2 amplitudes appear to respond sensitively to agency contexts, as a neural reflection of predictive forward models. Additionally, N1 amplitudes are reported to be reduced for repeated auditory stimuli (Grill-Spector et al., 2006). The suppression of ERPs for both self-generated stimuli and repeated stimuli is theorised to be associated with sensory attenuation, which results in attenuated sensory processing of the stimuli driven by predictive forward models (Bays et al., 2008; Hesse et al., 2010).

Neural mechanisms of the sense of agency

Several attempts have been made from the scientific community to investigate the neural mechanisms of the sense of agency. Failures in agency detection have been associated with greater activation of the angular gyrus (AG; Farrer et al., 2008). Successive research reported consistent evidence (Chambon et al., 2012). In this fMRI study, action selection mechanisms were dissociated from action-outcome matching, through a subliminal priming which linked participants' responses to a target (either compatible or incompatible). Results showed increased AG activations associated with decreased control ratings in incompatible trials (when there was no match between prime and target), while no increase was found in compatible trials (see Fig. 6).

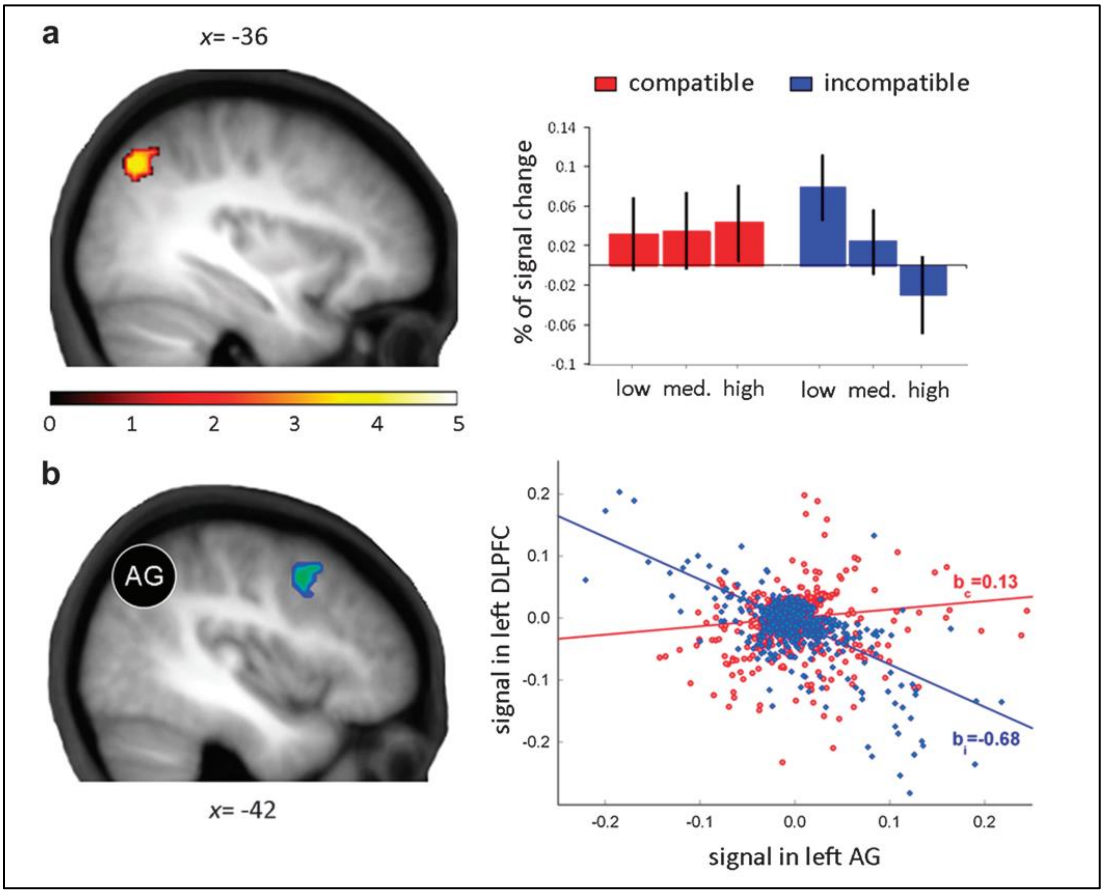


Figure 6. AG increased activation associated with decreased control ratings in incompatible, but not compatible trials (as shown in Chambon et al., 2012). Incompatible trials are also shown (bottom right) to decrease functional connectivity between the angular gyrus and the left dorsolateral prefrontal cortex.

Whereas the AG has been found to detect agency disruptions, the dorsolateral prefrontal cortex (DLPFC) has been shown to influence the fluency of action selection mechanisms. This evidence indicates that on a neural level, predicted outcomes are retrospectively compared with actual outcomes (Chambon et al., 2012). The influence of the DLPFC in the generation of the sense of agency has been supported by Khalighinejad et al. (2016), who reported increased DLPFC activity during the selection of alternative actions. These findings were replicated by later research, showing agency contexts modulating AG activity (Beyer et al., 2018). Interestingly, no association was reported between AG activity and temporal binding effects (Kühn et al., 2013), although different magnitudes of temporal compression were shown to positively correlate with increased activity of the left SMA. As discussed above, motor acts have been shown to affect time perception, and recent work from Merchant & Yarrow (2016) indicate the SMA as critical to the execution of timing-related actions. Taken together, this evidence suggests that the brain regions involved with the generation of the sense of agency are the EG, the DLPFC, and the left SMA (see Haggard, 2017 for a review).

Agency and affective coding

To this point, the ability to determine which outcomes in the environment are produced by one's own premeditated actions has been examined. This process generates a sense of agency over the actions performed, comparable to a feeling of responsibility towards the outcomes produced. However, research has reported that agency is not a dichotomous experience (i.e., either it is experienced or not). Rather, agency can be experienced in different magnitudes depending on whether those outcomes are favourable or not. A possible application of this finding relates to action outcomes that can emotionally affect other individuals. For example, producing an outcome that leads someone to smile could be experienced differently from an outcome that leads them to cry. Contexts eliciting positive reactions are likely to increase the sense of agency experienced, whereas actions leading to negative outcomes are more likely to be rejected, as if the feeling of responsibility towards those outcomes is reduced (Yoshie & Haggard, 2013).

Initially, agency detection was theorised to be generated solely through predictive internal forward models. Successively, this account has been expanded to take into consideration contributions of reflective postdictive cues, weighted on their salience and reliability to determine agency detection (Synofzik, et al., 2008; Synofzik et al., 2013; Vosgerau & Synofzik, 2012). More recent research has evidenced that affective coding could be among those cues (Gentsch, & Synofzik, 2014). The affective coding account proposes three emotional factors of agency: prospective, immediate, and retrospective affective coding (Gentsch, & Synofzik, 2014). Prospective affective coding theorises that affective traits or context variables can influence prospective representations of the sense of agency. Immediate affective coding is intended as quick and implicit processing of salient emotional stimuli. Retrospective affective coding indicates post-hoc evaluations of agency based on affectivity towards the outcome produced. Affective coding can supposedly be mediated by different individual affective styles (for example, tolerance, affect suppression, and emotion regulation), which can also influence the reliability-weighting process in agency detection. This mechanism can also account for individual differences in agency detection and agency disruptions in affective disorders as schizophrenia (Voss et al., 2010) and depression (Ratcliffe, 2013).

Given the importance of time perception as the basis on which binding effects can be used to measure the sense of agency, it is fundamental to consider if and how affect plays a role in time perception. As discussed above, given the absence of objective cognitive measures to sample time, subjective perceptions of time are not an accurate reflection of time. Temporal judgements can therefore be influenced by external and internal contexts, where affect could pose a great impact on perceived time. Previous research has investigated affective modulation of time perception, consistently reporting that actions producing negative outcomes led to subjective temporal dilation (or inverse binding; Yoshie & Haggard, 2013; see Droit-Volet et al., 2013 for a review). For example, multisensory stimuli inducing negative emotions (i.e., sadness, fear, anger, disgust) resulted in reduced temporal compression when compared to neutral and positive stimuli (Bar-Haim et al., 2010; Doi & Shinohara, 2009; Droit-Volet et al., 2004; Gil & Droit-Volet, 2012; Grommet et al., 2011; Mella et al., 2011; Yamada & Kawabe, 2011). Negative stimuli are believed to affect temporal elongations as a result of an arousal-based process, which hastens the pace of the internal clock mechanism. Since the number of beats built up directly influences subjective time perception, increasing the beats leads to temporal dilation. This mechanism could be due to an arousal-based preparation of the body for a fight-or-flight response to external threat, or can be interpreted as a dedicated adaptive process intended to delay subjective time as a mean for an efficient response preparation (Droit-Volet & Meck, 2007; Hagura et al., 2012).

Actions and emotional states are bonded in a bidirectional relationship (Gentsch & Synofzik, 2014). The actions performed on the environment can influence our emotional states, while emotions can shape the way we act with the environment. For example, feelings of responsibility over self-generated negative or positive outcomes can induce feelings of shame or pride, whilst shame and pride can influence whether the same action in the future will be avoided or repeated. Evidence about how action and affect are connected can be provided by affective disorders. For example, depressive states are often associated with fewer interactions with the environment, caused by a sense of reduced feeling of volition known as anhedonia (Ratcliff, 2013). This evidence has been reported also on an implicit level by Obhi et al. (2013), who found reduced temporal compression in participants induced with a temporary depressive state. One of the processes playing a major role in the connection between actions and emotional states could be the self-serving bias (Greenberg et al., 1982; Mezulis et al., 2004), defined as the tendency to confer positive outcomes to the self, while rejecting negative outcomes as due to the environment. The self-serving bias is believed to serve an adaptive function as to preserve healthy self-esteem (Greenberg et al., 1982; Mezulis et al., 2004). Previous research has reported consistent evidence of the presence of a self-serving bias towards positive outcomes (Yoshie & Haggard, 2013; Wilke et al., 2012). A reflection of this self-serving bias can also be observed in the tendency to reject negative outcomes, as shown by Takahata et al. (2012) who reported reduced temporal binding for actions that led to monetary loss, compared to actions leading to monetary gain. Interestingly, it has been shown that increasing the magnitude of outcomes' negativity entails greater temporal compression compared to less severe negative outcomes (Moretto et al., 2011). Although this element is seemingly inconsistent with the notion of a self-serving bias reducing agency for negative outcomes, it is worth noting that the authors did not investigate contributions of positive outcomes, hence a direct comparison is inappropriate. On the other hand, the magnitude of agency experienced for positive or negative outcomes could be an indication of the degree of association between the actions performed and the self-identity.

Sense of agency in social contexts

Throughout the millennia, humans developed complex cerebral structures in order to adapt to the vast range of social interactions we constantly experience given our intrinsically social nature (Adolphs, 2009; Blakemore, 2008). Human behaviour is therefore greatly influenced by the relationship between the actions we perform and the social outcomes they produce. Consequently, it is of critical importance to account for the contributions of social outcomes in the generation of the sense of agency. The scientific community has recently become interested in the social domain of the sense of agency, but the evidence reported has not always been consistent. For instance, Yoshie & Haggard (2013) explored how the social and emotional valence of the outcomes produced by our actions affect the implicit degree of agency experienced. Using a temporal binding paradigm, they found that social negative auditory stimuli (e.g., a voice screaming in fear) resulted from action execution led to decreased temporal compression, when compared to neutral and positive auditory stimuli. According to the authors, this evidence fits nicely with idea of rejecting negative events from the self to the environment, and suggests that action and affect could be connected for the adaptive reason to facilitate social interaction. This mechanism could help to promote further social interaction leading to positive outcomes, whilst discouraging the reiteration of actions producing negative outcomes (Yoshie & Haggard, 2013). However, successive attempts to replicate these effects have reported that positive and negative auditory stimuli generate a comparable degree of temporal compression (Christensen et al., 2016). Overall, the evidence reported regarding the specific agency mechanisms in the social domain are still blurred. Moreover, although some studies investigated the social valence of the outcomes produced, little attention has been given to the social valence of the action. In Chapter 2, the specific difference between physical actions (i.e., actions producing a direct change in the environment) and social actions (i.e., action producing a change in someone else's behaviour) will be addressed, to further explore the social domain of the sense of agency.

Sense of agency and joint action

Most of the literature discussed above involved measures of implicit components of the sense of agency (i.e., temporal binding) in non-social settings where participants performed actions alone. However, in many environments it is uncommon for humans to act by themselves, because social interaction is a dominant feature of human behaviour. This results in actions performed in social contexts, such as performing joint actions with other individuals and observing actions performed by others.

To successfully perform a joint action, it is not sufficient for the agents to control their own actions (i.e., correctly predicting their outcomes). In addition, they have to coordinate their actions with those performed by other agents, in order to achieve a shared goal. To this end, each agent needs to represent others' actions and predict their outcomes as well, to be able to adjust what they are doing to what others are doing (dyadic adjustment; Pacherie, 2014). Such dyadic adjustment between agents is necessary for joint action, but not yet sufficient. They also need to share a goal, and to understand the mutual influence of each other's intentions and actions on that goal. For this to be possible, all agents need to represent the combined outcomes of their actions and those of others, whilst using their predictions of these joint outcomes to assess and monitor progress towards the shared goal and adjust on the next moves (triadic adjustments; Pacherie, 2014)

The possibility to engage in joint action allows us to realise outcomes that as a single agent we could not (or at least not as easily) realise by ourselves. For example, two people may lift together a heavy object, whereas neither of them would have been able to do it alone. Thus, joint action is a way to increase agency scope. However, this poses the question of whether this increased agency is experienced as self-agency (i.e., an expansion of one's agency boundaries), or as we-agency (i.e., a merging of one's agency in a group's collective agency). In the previous sections, it was suggested that for individual action, the sense of agency depends to a large extent on the compatibility between predicted and actual outcomes. However, in joint action it is also critical to predict outcomes derived from others' actions, in order to facilitate interaction and coordination, whilst simultaneously distinguishing between self and other to avoid conflicts and interference effects (Sebanz et al., 2006; Wenke et al., 2011). Consequently, in joint action contexts, agents may be increasingly focused on self-other discrimination, while some of the principal cues used in individual action contexts (e.g., compatibility between predicted and actual outcomes) become much less reliable as markers of self-agency. Nevertheless, the strictness of this challenge may depend on the structural properties of the action performed (Pacherie, 2014).

Previous research has explored how social contexts and joint action can influence agency detection (Capozzi et al., 2016; Engbert et al., 2007; Engbert et al., 2008; Obhi & Hall, 2011a; Pfister et al., 214; Stephenson et al., 2018; Strother et al., 2010; Wohlschläger et al., 2003; see Moore & Obhi, 2012 for a review). This research aimed to investigate how the sense of agency is affected while interacting with other people, and if a sense of agency can be experienced over observed actions. Capozzi et al. (2016) used a Libet clock method to assess temporal compression when two individuals concurred in a joint action task. The experiment involved a naïve participant and a confederate, each of whom completed the joint action task under instructions to cooperate or compete. More specifically, participants were asked to press a key, after which a tone was presented. At this point, the confederate would have pressed another key, with the intention to ‘harmonize’ (cooperate) or ‘wipe out’ (compete) the previous tone, and a different tone followed. The confederate’s action was fake, as the temporal interval between the two tones was always constant. Temporal binding measures were recorded either on the tone participants caused, or on the one they believed the confederate caused. Results showed a subjective time compression for the tone that resulted from self-generated actions, but an inverse binding (i.e., temporal dilation) was found on the second tone (that participants believed was initiated by their partner), regardless the intention to cooperate or to compete. The authors suggested that the repulsion effect could contribute to the ability to discriminate self-generated events from other events in the environment and to preserve a self-related sense of agency.

Seemingly contrasting evidence has been reported by Strother et al. (2010), who asked their participants to place their index fingers on one end of a space bar each, and instructed to press the key at a time of their own choosing. Crucial for this experiment, if the other participant executed the action first, the other should have let their finger move with the space bar without exerting any force. Interestingly, temporal compression was found to be comparable between the initiator and the passive participant, and the same effect was found in an experimental variation where one participant was instructed not to move at all. The fact that participants experienced time compression even when they were aware of not having performed an action provides further indication that explicit and implicit dimensions of agency are dissociable, as later supported by further research (Obhi & Hall, 2011b).

However, it is worth noting that although both studies involved two agents, Capozzi et al. (2016) had two actions executed together, while Strother at al. (2010) had only one action, in which a participant always responded as a passive observer. A possible explanation for the inconsistency between these findings could lie in the need to preserve self-efficacy in ambiguous contexts. In fact, participants could experience increased magnitude of binding for their own actions when they needed an additional cue to dissociate that action from other actions in more complex situations (Kunde et al., 2018).

Sense of agency and action observation

Previous research has also investigated the mechanisms involved in agency detection during the observation of actions executed by someone else. Within this context, Engbert et al. (2008) used explicit temporal estimates to compare binding effects for self- and other-generated voluntary or involuntary actions. In the active condition, participants depressed a lever, after which a tone was presented. In the passive condition, a motor was used to depress the lever where participants' finger was just resting. In the active-observation condition, the same motor was used to depress the lever where the experimenter's finger was laying, pretending he intentionally moved. In the passive-observation condition, the same procedure was used on a rubber hand positioned on the lever. Results showed that participants judged the time interval between their voluntary actions and the tone as shorter than the interval between their involuntary movements and the tone (as consistent with what was originally reported by Haggard et al., 2002). On the other hand, no difference was found comparing the active and passive observed conditions, showing that temporal compression wasn't experienced when observing (illusory) other-generated actions.

By contrast, Wohlschläger et al. (2003) used the Libet clock method to compare the binding effects between self-generated actions, other-generated actions, and a control condition. In one of three blocks, participants were asked to press a lever, after which a tone was presented. In the other two blocks participants observed the experimenter completing the very same task, or the lever being automatically depressed using a solenoid. Results showed a clear binding effect for self-generated actions and (to a lesser extent) for other-generated actions. In a similar fashion, Poonian & Cunnington (2013) tested the same research question using the interval reproduction method. Their experiment involved three conditions. In the control condition, a tone was presented and after a variable time interval, the same tone was played again. In the action condition, participants were asked to perform a key press, after which a tone was presented. In the observation condition, participants watched video clips showing a hand pressing a key, after which a tone was presented. In all the three conditions, the time interval between the two events was variable. After each trial, participants were instructed to replicate the time duration between the two events, pressing and holding a key for the same amount of time. Results indicated a similar and consistent binding effect in the action and observation conditions, but no temporal compression in the control condition.

One possibility that might account for such discrepancies is offered by considering how temporal binding for observed actions might relate to action observation more generally. Prior investigations into the mirror neuron system (Di Pellegrino et al., 1992; Gallese et al., 1996; see Rizzolatti & Craighero, 2004 for a review) suggest that observing someone else executing an action may trigger motor simulation processes, where self-generated and observed actions are computed analogously. In this way, observing an action could engage the comparator model, resulting in the generation of a temporal compression which is not associated with a concurrent explicit sense of agency (Sato, 2008). Importantly, studies that did not detect temporal binding for observed actions employed a method where, in order to promote experimental control, the other's action was not visible to participants (or indeed was not actually performed by the other actor). For example, Obhi & Hall (2011) had participants and confederates separated by a curtain, and the action was always (in reality) produced by the participant. Another example can be found in Engbert et al. (2007), which had participants observe an illusory action, where the hand of the experimenter was just lying on a self-depressing lever. The evidence reported by previous research, showing that fake and illusory actions failed to produce a binding effect, led us to wonder whether time compression might occur only for observed actions through the activation of the mirror neuron system and would thus necessitate direct observation of the action. Given the crucial importance of action observation in the human interaction domain, we aimed to shed some light on this debate, which is further explored in Chapter 2.

Sense of agency and artificial agents

Nowadays, artificial agents such as vocal assistants are becoming increasingly popular. These artificial agents are capable of a large number of functions, some of which can be executed autonomously: playing our favourite music, suggesting the best path to avoid traffic, or reminding us a specific event in the calendar. However, since these artificial agents don't have an embodied presence, it may be difficult to interact with them as actual agents, capable of conscious intentions and executing actions. The increasingly rapid development of artificial agents suggests that soon different kinds of robots will be part of our lives, assisting us both in domestic and working environments (Glasauer et al., 2010). Compared to the aforementioned vocal assistants, embodied robots could not only satisfy our requests, but also act directly on and manipulate the environment (Wykowska et al., 2016). In this sense, such robots are already (or soon will be) capable of functions normally performed by other humans – an eventuality that raises important questions about the nature of human-robot interactions.

Previous research has shown that artificial agents (and social robots in particular) can influence our cognitive processes, including decision making (Shinozawa et al., 2005), spatial representation (Bainbridge et al., 2008) and joint attention (Kompatisari et al., 2018) in a similar fashion as do other humans (Frischen et al., 2007). However, only in more recent times the scientific community showed interest into agency detection during human-robot interaction. Pioneering findings showed that the sense of agency experienced towards our own actions can be reduced when such actions are executed together with a social robot (Ciardo et al., 2020). Specifically, participants reported lower agency ratings when they executed an action jointly with a social robot (compared to by themselves alone) if that action was associated with a negative cost (losing a variable amount of points). These findings are consistent with previous research that showed a strong self-serving bias in attributing actions with a positive outcome to ourselves, while ceding responsibility for negative actions to others (Wegner & Wheatley, 1999; Yoshie & Haggard, 2013). Consistent results were also found using implicit agency measures. Roselli et al. (2019) compared temporal binding effects when participants acted alone or interacted with a robot. In this study, participants used a Libet clock method to report the timing of actions they executed alone (individual condition) or jointly with a robot (social condition). During the social trials, either participants or the robot could initiate the action. Their data indicated that participants showed weaker binding effects towards their actions when they were executed in the presence of the robot, compared to when participants acted alone, in a similar fashion to what has been reported to happen in presence of a human agent (Beyer et al., 2018). In other words, their actions were less bound to the outcome when that outcome could also be attributed to the robot. These results are consistent with what was reported on an explicit level by Ciardo et al. (2020), providing further evidence that the temporal binding effect could reflect an implicit dimension of the sense of agency. Taken together, these findings suggest that responsibility for the outcomes of our actions can be displaced towards another agent, that is also capable of executing the action, regardless of it being human or robotic.

As discussed above, interacting with social robots can trigger specific mechanisms typically involved in the interaction between humans. However, to our knowledge, the instances of agency detection for observed robotic-generated actions have not been addressed specifically. In Chapter 3 this topic will be expanded, investigating how observed human- and robotic-generated actions differs on agency detection, and what factors can influence this process.

Agency and action representation

In the previous sections, action production and perception have been discussed with specific focus on their influence on our cognitive systems. However, agency detection is not the only cognitive mechanism that can be used to assess intentionality and volition of our actions. In fact, the relationship between actions and the environment is believed to be driven by cognitive constructs known as affordances (Gibson, 1979). Affordances were initially theorised to reflect environmental offerings as perceived by a sentient being, but in more recent times the scientific community adopted the affordance construct to indicate action-oriented representations and dispositions (Sakreida et al., 2016). For instance, if a chair is seen, it will afford the action of sitting, rather than lifting. By these means, affordances provide strong cues to the operations of objects (Norman, 1996), but are inevitably dependent from the actor's experience of those objects. As a consequence, newly experienced objects would not be able to afford an action, or they could even afford an incorrect action (as in the Little Mermaid, the first time Ariel sees a fork on a table, uses it to brush her hair). In the following paragraphs, the extent to which agency and affordances can be measured through motor resonance effects will be discussed.

Recent evidence has shown that perceived affordances can affect our motor system. For example, specific actions can be potentiated (i.e., more easily executed) if participants are primed with the image of an object that affords that action (Ellis & Tucker, 2000; Symes et al., 2007; Tucker & Ellis, 2004). In a pioneer investigation, Tucker and Ellis (1998) showed their participants upright or inverted images of ordinary graspable tools, asking them to categorise the images by pressing a key with their right or left hand. They reported a spatial alignment effect, as reaction times (RTs) were shorter for tools presented in a congruent position to the response hand (e.g., tools oriented to the right were categorised more rapidly with the right hand). They interpreted their findings as evidence that the depiction of a visual object provides not only a description of its visual features, but also encodings of actions related to that object. As such, task-irrelevant information (e.g., right or left orientation) can potentiate (i.e., facilitate) the execution of congruent actions when the responding hand is spatially aligned to the orientation of the affording feature of the object. In a later study, Ellis and Tucker (2001) showed that specific object attributes (e.g., dimension) can suggest (or even demand) specific motor acts. The authors designed this experiment with object dimension (large/small) and object nature (natural/manufactured) as independent variables. Participants were primed with one the image of an object on each trial, and their task was to categorise them as natural or manufactured executing a specific motor act (power grasp or precision grasp). Results showed that error rate and RTs were reduced when the grasping response was compatible to the dimension of the object presented, regardless of its nature. Taken together, this evidence suggests that during the execution of a voluntary action, our motor system is sensible to object information, and can use it to facilitate action selection mechanisms.

Contextual and social affordances

Further research has used this spatial alignment paradigm (Bub & Masson, 2010) to investigate different contextual features that enable action facilitation effects. It has been shown that the capability of an object (e.g., a handled mug) to afford a congruent action (e.g., precision grasp) is not intrinsic to the object, but depends on its spatial location in respect to the agent (Costantini et al, 2010). In this study, participants were primed with a task-irrelevant stimulus (i.e., a right- or left-handled mug), which could have been presented either within or outside participants' reachable space. Participants were asked to perform a precision grasp action with their right or left hand as soon as the stimulus was presented, regardless of its location. Results showed that a spatial alignment effect (i.e., reduced RTs for congruent actions) was found only when the object was presented within participants' reachable space. In contrast, when the object was presented outside their reachable space, there was no motor facilitation effect, with data showing no difference in RTs between congruent (e.g., right-handed grasp executed in response to a right-handled mug) and incongruent (e.g., right-handed grasp executed in response to a left-handled mug) actions. This finding has also been consistently replicated with neurophysiological methodologies. Follow-up research (Cardellicchio et al., 2001) has used TMS techniques to induce motor evoked potentials (MEPs) on the hand muscles deployed during precision grasp action. MEPs are electric fluctuations within a specific muscular district, recorded through electrodes placed in contact with the skin. In this study, MEPs were recorded on the opponens pollicis and the first dorsal interosseus (the main muscles involved in precision grasp actions), while participants were presented with the same stimuli and completed the same procedure of the previous study. Their data suggest that being primed with a stimulus congruent to the action performed induces higher MEPs, but only when the object was presented within participants' reachable space. Taken together, this evidence indicates that spatial location of the object can affect its power to afford an action.

Nevertheless, in ecological contexts humans rarely perceive and act upon objects by themselves. Scientific literature reports that in social contexts, joint attention (Bayliss et al., 2006) and emotional facial expressions (Bayliss et al., 2007) can influence affective evaluations of objects. By consequence, prior research investigated whether motor facilitation mechanisms could be influenced by the presence of other agents in the social context. Using a spatial alignment paradigm, Costantini et al. (2011) showed that object presentation can potentiate congruent actions not only when the object is presented within participants' reachable space, but also when it is presented within someone else's reachable space. Again, the behavioural effects found reflection in increased MEPs for the same manipulations (Cardellicchio et al., 2013). The authors did not interpret these data as evidence of an increase of participants' reachable space, but rather as an altercentric re-mapping of the environment, consistent with spatial perspective-taking mechanisms (Tversky & Hard, 2009). Consequently, perceived objects can afford a congruent motor act either directly (when they are presented within our reachable space) or indirectly (when they are presented within the reachable space of other agents). However, this key finding and the mechanisms involved in motor facilitations for others' actions received no further interest. In Chapter 4 this topic will be expanded, with the aim to provide insights about the physical features possibly involved in the generation of motor facilitation effects for other's actions. Namely, contributions of anthropomorphic body representations will be investigated, in order to assess whether humanoid embodiment can provide artificial agents with enhanced motor resonance.

Thesis outline

In recent times, the diffusion of artificial agents increased exponentially, with great progress achieved in relation of technical development. However, it is still unclear which mechanisms can facilitate the relationship between natural (i.e., humans) and artificial agents, that is to interact in a social and intuitive way. In order to promote a successful interaction with others, humans need to understand and predict their behaviours, a mechanism that finds its roots in the ascription of agency. Nevertheless, while agency detection has been widely investigated in it most common physical domain, the extent of which we can detect and ascribe a social sense of agency is still unclear.

Chapter 2 will examine this topic, exploring whether a sense of agency and its implicit components (i.e., subjective time compression) can be experienced in the social domain of action, and how agency is ascribed to other agents when they are human or artificial. Namely, the capability of indirect social actions (e.g., vocalisations) of producing temporal binding effects will be investigated (Experiment 1), with further expansion also on observed manual and vocal actions as executed by humans or artificial agents (Experiment 2), and the relevance of visual access to kinematics information (Experiment 3).

Successively, Chapter 3 will explore implicit agency detection on human- and robot-generated actions (Experiment 4), investigating how explicit attitudes and prior knowledge about the robot can moderate the experienced agency perception and mentalisation towards the artificial agent (Experiment 5).

Chapter 4 will focus on motor facilitation effects, investigating the features that enable action representation and potentiation in social contexts, and how these may contribute to a more intuitive robot design. More specifically, contributions of bodily anthropomorphic representation (Experiment 6) and actual action opportunity (Experiment 7) will be explored.

Chapter 5 will address a general discussion of the results reported in the experimental chapters and will provide a novel neurocognitive model of agency detection also accounting for empirical evidence of temporal compression experienced over observed actions performed by other actors, with a possible involvement of the mirror neuron system.

Statement on the impact of the Covid-19 pandemic

Part of the original doctoral program was impacted by the restrictions resulted from the COVID-19 pandemic. Experiments 1-5 were unaffected, as data collection took place before the virus outbreak. According to the he original doctoral program, Experiment 5 should have been followed by an EEG study, investigating for the first time the N1 component of auditory ERPs resulted from observed robotic actions. However, UK and UEA efforts to prevent further spread of the infection resulted in national lockdowns and laboratory closures, starting from March 2020, and ending in May 2021. As a direct consequence, any EEG data was impossible to collect. Yet, Chapter 3 will include a comprehensive addendum, where neural correlates of the sense of agency will be addressed and the planned EEG experiment will be described. Data in experiments 6 and 7 were collected during the pandemic, optimizing on-line testing methods. In order to minimise confounds due to the impossibility of completing the procedures in controlled environments, calculated sample sizes were doubled. Further implications and limitations of on-line testing will be discussed in Chapter 4.

SECTION 2

-

Experimental Chapters

Chapter 2

Sense of agency and social actions

"Words have incredible power. They can make people's hearts soar,
or they can make people's hearts sore."
Mardy Grothe, 1995

We daily observe changes in our environment (e.g., a lamp lighting up). Such changes occur for a vast multitude of reasons: because we acted to make the change (i.e., I flicked the switch), because someone else did something to make the change (i.e., you flicked the switch), or because of factors that we can't directly observe (i.e., the switch has a timer). As human agents, our own action system is a primary cause of changes to the environment. This system has its neural basis in the premotor cortex and the primary motor cortex (Rizzolatti & Luppino, 2001), and operates to change the environment from its current state to a new state according to our goals and intentions (Ridderinkhof et al., 2004). Following the earlier example, we may move our hand to flick a light switch in order to achieve the goal of lighting up a room. In this way, our own actions produce some physical outcome in the environment. Our environment is dynamic, however, with many things changing at a given time, only some of which are caused by our own actions. Our motor system therefore monitors action outcomes to optimise action control and also to correctly assume authorship over events in the environment where our action is the cause of the effect (Gallagher, 2012). This results in a sense of agency: a recognition of ourselves as agents of goal-directed actions that produce changes in the external environment. A correct sense of agency helps us to distinguish *what we are doing* from *what is happening*.

Implicit measures of agency can be explored by the means of an effect known as temporal binding (Haggard et al., 2002), whereby perception of the temporal distance between actions and outcomes is compressed for causally associated actions and outcomes, whilst remaining relatively accurate when judging the temporal distance between two independent events. This is why the temporal binding effect is theorised to measure an implicit dimension of the sense of agency (see Haggard, 2017 for a review). Implicit (i.e., time compression) and explicit (i.e., feeling of responsibility towards an outcome) agency can be thought of as dissociable, but the two are not thought to be completely independent (Moore et al., 2012). This is consistent with previous work from Synofzik et al. (2013) who proposed an optimal cue integration account where implicit agency is supposed to operate at a sensorimotor level, whereas explicit agency occurs as result of higher-level processing. Following this evidence, this study aimed to measure the temporal binding effect as an implicit measure of the sense of agency, and to collect explicit self-reported ratings of agency as a manipulation check.

Social domain of action

The actions we use to directly modify the environment are generally thought to be motor actions such as reaching, grasping, or throwing. However, our ability to influence the environment is not limited to these physical manipulations. On the contrary, we also execute a broad range of social actions (i.e., facial expressions, eye gaze, speech) that, while unable to directly affect the physical environment, have the power to change the environment indirectly by influencing the behaviour of others. Such social actions convey a codified signal to a designated receiver, and can change other people’s behaviour (e.g., Caspar et al., 2016; Stephenson et al., 2018). For example, while using our voice to achieve the goal of plucking guitar strings may be ineffective, using our voice to ask a friend (or an artificial assistant) to play a song is likely to succeed. In this way, social actions allow us to produce an intended change from the current state of the environment to our desired one, in the same way as physical actions do. An open question, however, is the extent to which the indirect outcomes of non-physical actions may also elicit similar signatures of the elicitation of a sense of agency. Furthermore, being able to detect the outcomes produced on other people, could be essential skill to understand their actions and to ascribe them appropriate mental states (Happé et al., 2016). Hence, the importance of comprehending the nature of agency involved in social actions is crucial for the comprehension of social cognition.

However, since manual and vocal actions differ from each other, they may not necessarily recruit the same mechanisms underlying agency. Still, having a strong agency detection system accounting for both physical and social self-generated actions would be more efficient than having two separate and dedicated systems. When considering earlier research on social actions, interesting findings have been reported. Stephenson et al. (2018) adopted a temporal binding paradigm to test whether social actions such as initiating joint attention could lead to an implicit sense of agency towards the social outcome produced (i.e., the follower gazing at the same object). They found that leading the gaze of an on-screen face induced an underestimation of the temporal gap between action and outcome. This is consistent with Pfeiffer et al. (2012), who reported an increased feeling of control over congruent gaze responses induced in other people.

Alternatively, there could be reasons to believe that the sense of agency could operate in different ways between the physical and social domains. As a matter of fact, we continuously experience physical actions producing immediate outcomes. For example, when pressing the start button on a computer, we will almost instantly see LEDs flashing and hear rotors whirling. The temporal gap in which we can assess whether the action was successful is very short, and leaves little or no ambiguity: if the action did not produce the intended change, it needs correction. In contrast, when considering social actions which are capable of producing changes in another person, the temporal delay to produce an outcome could be much longer, making it more difficult to predict the outcome (Kunde et al., 2018). For example, when we call someone's name, they may not immediately turn to us, if distracted or busy with other tasks. Hence, there could be much more ambiguity in assessing the effectiveness of our social actions, due to the larger and more variable amount of time to experience their outcomes. This difference may be a crucial one for how the sense of agency functions across physical or social contexts. The greater variability in the delay between social actions and outcomes may lead to no binding effect at all, as social agency detection may be ascribed to higher-level processes as the generation of a theory of mind (Premack & Woodruff, 1978). In fact, even if no implicit marker of agency is generated (namely, subjective time compression), we could still rely on explicit higher-level mechanisms to efficiently detect whether our social actions affected someone's behaviour. On the other hand, given the importance of social agency detection, the greater unpredictability of social actions could actually lead to stronger implicit effects. Such implicit effects could rely on a flexible system, able to account for the intrinsic variability of time. Hence, whether vocal actions could generate a temporal binding effect (associated with an implicit dimension of the sense of agency) is a fascinating question for social cognition.

Speech as action

Speech and other vocalisations serve as social actions to indirectly affect several aspects of the environment by communicating ideas to others, and are fundamental to human social and cognitive development (Luria & Yudovich, 1971).

Producing a verbal utterance is one of the most common actions humans execute: in a fluent conversation we pronounce approximately two to three words each second (corresponding averagely to four syllables, or ten to twelve phonemes) out of or vocabulary, which typically contains 10 to 100 thousand words (Levelt, 1999). In this sense, vocal actions can be performed to produce a vast range of outcomes, such as influencing others' motor actions (that is, a purely physical influence, for example asking someone to flick a light switch to turn the light on), or influencing others' cognitive and emotional states (that is, a purely social influence, for example a lecturer presenting new concepts to a student, changing their knowledge states), or influencing other's social behaviour (that is, a hybrid between social and physical influence, for example telling someone to look at a specific object to redirect their attention). In fact, when talking to other people, we are able to influence them in ways that can be directly or indirectly observed. Thus, vocalisations are actions, but their outcomes can be direct or indirect.

Motor control for verbal speech production in right-handed people is ascribed mostly to areas in the left cerebral hemisphere (Indefrey & Levelt, 2004). These areas include the left posterior inferior frontal gyrus, the left temporal cortex, the left insula, the left primary motor cortex, and the bilateral SMA. There are also subcortical areas involved, such as the basal ganglia (Booth et al., 2007) and cerebellum (Ackermann, 2008), which supports the sequencing of speech syllables into smooth, fast, and rhythmically organized words and longer sounds. Many of the aforementioned areas are also involved in the motor control for physical actions. Notably, patients diagnosed with motor neuron disease (MND), a neurodegenerative disease specifically impairing the motor system (Bak & Hodges, 1999), show deficits also in the production and comprehension of action verbs. Post-mortem examinations of MND patients revealed damage in the inferior frontal gyrus, as well as in the motor and premotor cortex (Aziz-Zadeh & Damasio, 2008). This evidence suggests that these areas are not only fundamentals for motor processing, but also for processing of action verbs. Furthermore, previous research reported evidence suggesting a possible cognitive representational overlap between speech and action, as observing vocal actions recruited brain regions known to be involved in action perception, such as the premotor and the adjacent primary motor cortex (Andric et al, 2013; Skipper et al., 2005).

The question to whether vocalisations elicit an implicit sense of agency over their outcomes, as physical actions do, has been previously addressed by Limerick et al. (2015), who deployed a Libet clock method and reported reduced temporal compression for vocal actions, compared to manual actions. With this study we aim to further expand knowledge on the topic, implementing a control condition where no action (neither vocal nor manual) is performed. On a further element, vocal actions, being able to be perceived through action observation (i.e., looking at someone as they speak) or in the absence of it (i.e., just hearing their voice) can help us to understand the mechanisms beyond the generation of temporal compression for actions performed by others. In fact, if temporal binding effects results as response to a causal relationship between two events, we should expect participants to underestimate the temporal gap in all conditions where they know that an action produced an outcome. On the other hand, if temporal binding effects are informative about intentionality through motor simulation processes, we should expect a discrepancy between vocal actions that can be genuinely observed and vocal actions that can only be heard. To the author's knowledge this topic has never been investigated before, and as such this study aimed to directly compare executed and observed physical and vocal actions producing the same kind of outcome commonly used in temporal binding paradigms (i.e., an auditory tone).

Overview of Experiments 1-3

In three experiments, the hypothesis that vocal actions (like motor actions) give rise to a sense of agency was tested, as measured implicitly by the temporal binding effect (see Haggard, 2017 for a review). Participants' reproductions of the time interval between their actions and the outcomes produced by those actions were compared across different conditions: when they performed a physical or vocal action, when they observed the execution of a physical or vocal action, and when they were witnessing two independent events. Hence, it was predicted that reliable temporal compression would be achieved over self-generated physical actions and self-generated vocal actions. According to previous literature (Poonian & Cunnington, 2013; Strother et al., 2010; Wohlschläger et al., 2003) it was also predicted that genuine and observable physical and vocal actions performed by other people would result in temporal underestimation towards their outcomes. To pre-empt the findings, the data showed that performing a vocal action induced an underestimation of the time interval between action and outcome, an index of an implicit sense of agency (Experiment 1). Successively, it is shown how observing others producing vocal actions produced by others elicited a binding effect (Experiment 2), but only with direct visual access to the vocal action execution (Experiment 3). In all experiments detailed below, it is reported how sample size were determined, all data exclusions (if any), all manipulations, and all measures.

Experiment 1

In Experiment 1, participants completed an interval reproduction task under four different conditions, manipulated within-subjects. Each condition featured a different start stimulus, each of which produced an identical outcome. First, in order to replicate previous findings using this task (e.g., Howard et al., 2016; Humphreys & Buehner, 2010; Poonian & Cunnington, 2013; Stephenson et al., 2018) the experiment included an Operant Manual condition where participants reproduced the time interval between their own manual action (pressing a key on the computer keyboard) and a tone produced in response to that action. To test whether vocal actions were also capable of producing a similar level of temporal compression, it was included an Operant Vocal condition where participants reproduced the time interval between a vocal action (i.e., saying the word ‘Go’, as in Limerick et al., 2015) and a tone produced in response to that action. Lastly, it was aimed to provide further evidence relating to the long-term debated possibility that observing actions performed by others can also produce a temporal binding effect. To do this, an Observed Manual condition was included, where participants reproduced the time interval between a manual action performed by the experimenter (i.e., the participant observed the experimenter pressing a key on the computer keyboard) and a tone produced in response to his action. This condition provides a conceptual replication of previous studies using an interval reproduction task in lieu of the Libet clock method (e.g., Wohlschläger et al., 2003). As is typical for temporal binding paradigms, performance in the three experimental conditions (Operant Manual, Operant Vocal, and Observed Manual) were compared with a Control condition in which no action was performed by any agent. Here, participants reproduced the time interval between two tones that were automatically produced by a computer. At the end of each experimental block, participants were also asked to report how much they felt in control of the action outcome in the different conditions, providing an explicit index of their sense of agency.

Method

Participants

A power analysis (carried out with G*Power 3; Faul et al., 2007) indicated that to detect a medium-large effect size (as reported by Stephenson et al., 2018) d_z = 0.69, with 1 – β = 0.95 at α = 0.05, a minimum sample size of 30 would be required. Therefore, 32 participants (a sample size capable of detecting an effect size d_z = 0.66, with 1 – β = 0.95 at α = 0.05) aged 18-33 years (*M* = 20.24, *SD* = 2.65, 28 were females), recruited from the University of East Anglia, completed the experiment. Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported having normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Design

The temporal binding effect was measured using an interval reproduction task (Humphreys & Buehner, 2010), where participants were asked to reproduce the duration of the time interval between two events by pressing and holding down the central key on a response box for the same amount of time. In order to effectively manipulate participants' agency perceptions, the nature of the first event varied in each block, while the second event remained unaltered.

Stimuli

A first low-pitch tone (150ms, 440Hz sine wave, sample rate: 44100Hz, bitrate 16: Poonian & Cunnington, 2013) was created as the start stimulus in the Control condition. A second high-pitch tone (100ms, 1 KHz sine wave, sample rate: 44100Hz, bitrate 16: Humphreys & Buehner, 2010) was created as the end stimulus in the Control condition and as the action's outcome in the Operant Manual, Operant Vocal and Observed Manual conditions. The experiment was run using E-prime version 3.0 (Psychology Software Tools, Inc., Sharpsburg, PA, USA). All auditory stimuli were created using MATLAB (MathWorks, Natick, MA, USA).

Apparatus and materials

The experimental setting consisted of two adjacent chairs and a table in a dimly lit room. A computer monitor (BENQ XL2411: size: 24”; resolution: 1920x1080; refresh rate: 60Hz) and a set of external speakers (Bose Companion 2 Series III) were placed on the table and used to display experimental stimuli. A Chronos multifunctional response and stimulus device (Psychology Software Tools, Inc., Sharpsburg, PA, USA) with microphone, and an external keyboard (Kensington KP400) were placed on the table and used to collect participants’ responses. The height of the table was 80cm. The position of the monitor was centred to participant’s body midline, 60cm from the edge of the table. The position of the speakers was respectively 30cm on the right (right speaker) and 30cm on the left (left speaker) of participants’ body midline, 50cm from the edge of the table. The position of the keyboard was centred to participants’ body midline, 25cm from the edge of the table. The position of the response box was 40cm on the right of participants’ body midline, 25cm from the front edge of the table. The position of the microphone (only used in the Operant Vocal condition) was centred to participants’ body midline, 5cm from the edge of the table.

Procedure

Participants completed four blocks of trials each featuring a different start stimulus. Each block consisted of 5 practice trials and 30 experimental trials, for a total of 120 experimental trials. Block order was pseudorandomized across participants. After each block of trials, participants self-reported the degree to which they felt control over the high-pitch tone. The instruction on screen was “Please rate 1 to 8 how much control you felt over the high tone, 1 meaning no control and 8 meaning a lot of control” (Beyer et al., 2018).

Participants were invited to the laboratory and welcomed. They read an information sheet and provided informed consent to join the study. Prior to beginning the experiment, participants were introduced to the experimental equipment. They completed a sample of ten experimental trials (five Operant Manual trials and five Operant Vocal trials, without the reproduction task) in order to demonstrate that when they executed an action (whether manual or vocal), a high tone would be released. This served to ensure that they understood the causal relationship between the two events.

Control trials began with a white fixation cross displayed in the centre of the screen (1500ms) after which, at a random interval (1500-2000ms) a low-pitch tone (150ms, 440Hz) was presented. Then, after the target interval of time (randomised between 500 and 1500ms) a high-pitch tone (100ms, 1kHz) was presented. Participants were then asked to reproduce the duration of the interval between the two tones by pressing and holding down the central key on the response box for the same amount of time. After participants completed the reproduction task for the considered trial, they were asked to press the space bar to begin the following trial.

In the Operant Manual trials, participants were instructed that after the fixation cross disappeared (1500ms) performing a specific action (i.e., press the 0 key on the number pad at any moment of their choosing) would produce, after a variable time interval (randomized between 500 and 1500ms) a high-pitch tone (100ms, 1kHz). Participants were encouraged to avoid rushing in giving the action, or to start the action at any predetermined fixed time (e.g., counting to 3 and perform the action). Following the tone participants were instructed to press and hold down the central key on the response box for a duration equal to that of the time interval between their action and the tone. The target time interval was set to begin after the end of participants' action (i.e., when the 0 key on the number pad was fully released). After participants completed the reproduction task for the considered trial, they were asked to press the space bar to begin the following trial.

The Operant Vocal trial procedure was identical to Operant Manual trial procedure, except for the action that participants were asked to perform. Participants were instructed that in the Operant Vocal trials, performing a specific action (i.e., pronouncing the word "go" in the microphone) would produce, after a variable time interval (randomized between 500 and 1500ms) a high-pitch tone (100ms, 1kHz). Participants then completed the reproduction task, pressing and holding down the central key on the response box for a duration equal to that of the time interval between their action and the tone. The target time interval was set to begin after the end of participants' action (i.e., when the microphone's noise threshold reached 1%).

After participants completed the reproduction task for the considered trial, they were asked to press the space bar to begin the following trial.

Using the same trial procedure of the Operant Manual condition, participants completed the Observed Manual condition where they observed the experimenter performing an action (i.e., pressing the 0 key on the number pad at any moment of his choosing). After a variable period of time (randomized between 500 and 1500ms) a high-pitch tone (100ms, 1kHz) was produced. Participants then completed the reproduction task and were asked to press the space bar to begin the following trial (See Fig. 7 for a graphical representation of the trial procedure). When participants were completing the Control, Operant Manual, and Operant Vocal conditions, the experimenter sat approximately 120cm distant from the participants' position, on the orthogonal side of the table. During the Observed Manual condition, the experimenter sat adjacently on the right of the participants, approximately 10cm distant from their position.

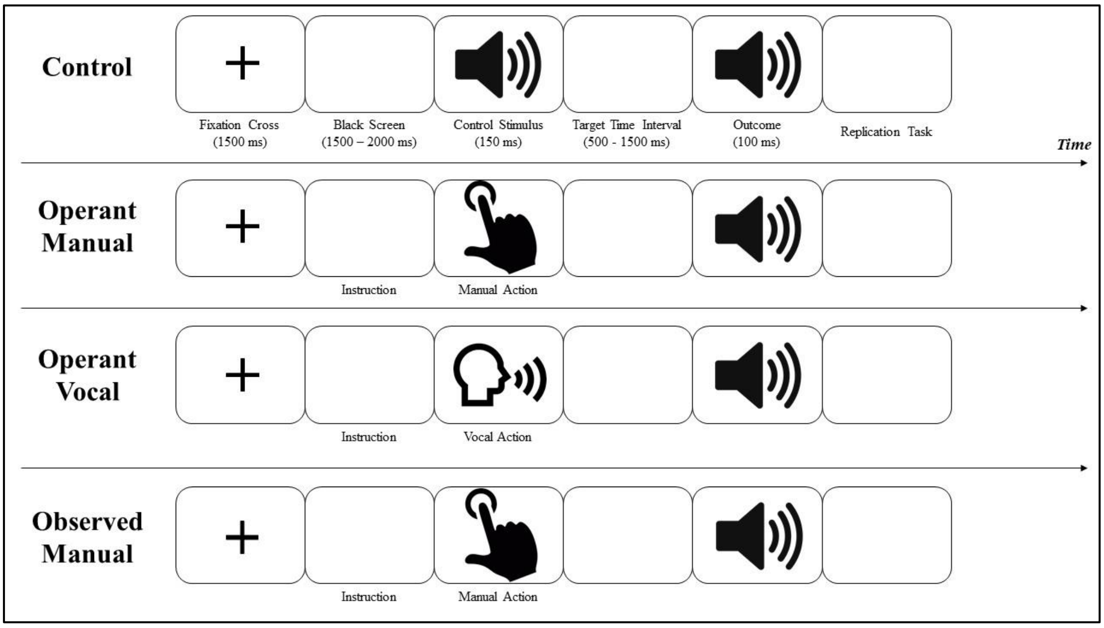


Figure 7. Trial procedure for the four conditions in Experiment 1.

Results

The calculated dependant measure was the inter-event proportional interval reproduction, derived by dividing the reproduced time interval by the actual time interval for the same trial (ms). Thus, scores equal to 1 represented perfect accuracy, while scores greater than 1 were over-reproductions, and scores lower than 1 were under-reproductions (that is, subjective temporal compression).

The following exclusion criteria were decided prior to the data collection: participants were excluded if they failed to produce temporal intervals covarying monotonically with actual action-tone intervals (average ρ across conditions lower than 0.4, meaning that participants were not continuously putting an effort in following the instructions; Caspar et al., 2016). Individual trial data were excluded based on failure to respond (trials where the reproduced time interval was lower than 100 ms were discarded) or extreme variability (trials falling outside +/- 3 SD from individual mean were removed). This resulted in 3 participants excluded and 22 trials deleted (0.57% of total trials).

For all Experiments reported in the Chapter, we followed the indications for stepwise multiple comparisons, as indicated in Welsch (1977) and in Howell (2002) with specific regard to repeated measures designs. Here the authors suggest to test as few comparisons as possible using only the data involved in those comparisons. Accordingly, we report the ANOVA, followed by the t-test between the Control condition and the predicted significantly different condition with the highest mean (with the assumption that comparisons with lower means in other predicted significantly different conditions would imply necessarily a lower p-value). As a practice of good and open science, we then report results of all predicted comparisons, that whilst uninformative about the significance of those differences, could still be of relevance to the scientific community for the differences in effect sizes between different conditions.

Proportional interval reproductions

Mean proportional interval reproductions were calculated for each participant and submitted for statistical analysis. Shapiro-Wilk tests were performed on each condition to test normality of each distribution, and all conditions reported p >.05. A one-way within-subjects ANOVA was conducted on proportional interval reproduction scores. There was a significant effect of the type of action: $F(3, 84) = 7.438$, $p < .001$, $\eta_p^2 = 0.210$. Planned paired-samples t-tests revealed that the mean proportional interval reproduction in the Control condition (M = 1.05, SD = 0.26, 95% CI [0.94, 1.15]) was significantly higher either than the Operant Vocal condition (M = 0.96, SD = 0.21, 95% CI [0.88, 1.05], $t(28) = 2.345$, $p = .026$, $d_z = 0.44$), the Operant Manual condition (M = 0.92, SD = 0.19, 95% CI [0.84, 0.99], $t(28) = 4.234$ $p < .001$, $d_z = 0.80$), and the Observed Manual condition (M = 0.92, SD = 0.20, 95% CI [0.85, 1.00], $t(28) = 3.318$, $p = .003$, $d_z = 0.63$). No significant differences were found between the Operant Manual, Operant Vocal, and Observed Manual conditions: $F(2, 56) = 1.815$, $p = .172$, $\eta_p^2 = 0.061$. Figure 8 shows a graphical representation of the data.

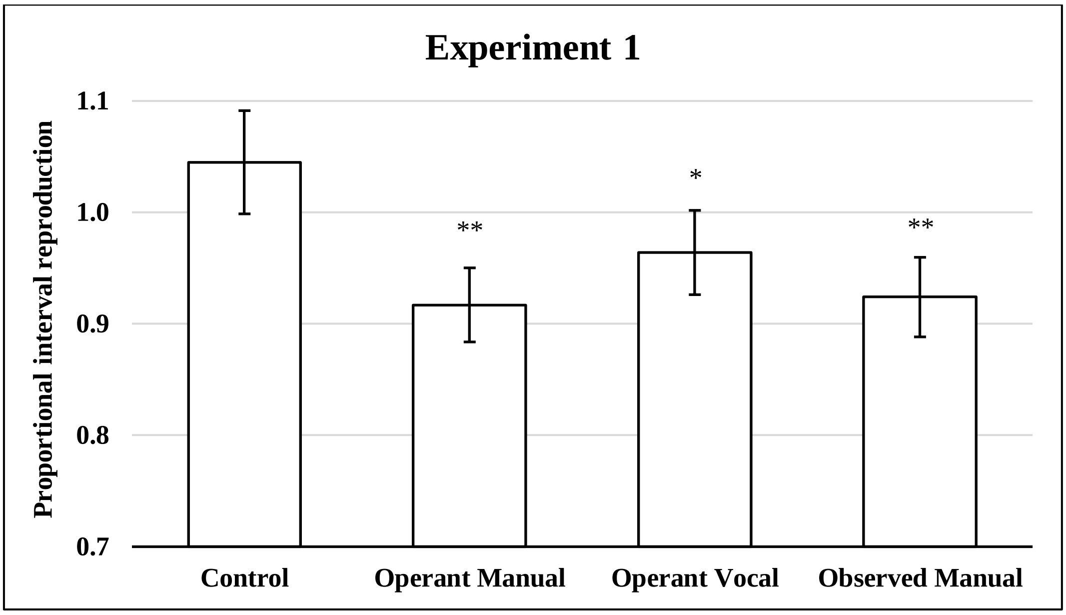


Figure 8. Mean proportional interval reproduction across the four conditions in Experiment 1. Error bars represent the standard error of the mean. Significant differences from the Control condition are highlighted for the .05 (*) and .01 (**) level.

Explicit agency ratings

When comparing self-reported explicit agency ratings, a one-way repeated measures ANOVA revealed a significant effect between the four conditions: $F(3, 84) = 23.146$, $p < .001$, $\eta_p^2 = 0.453$. Planned paired-samples t-tests revealed that the mean explicit agency rating in the Control condition (M = 2.66, SD = 2.04) did not differ from the Observed Manual condition (M = 2.97, SD = 2.01, $t(28) = -1.395$, $p = .174$), but was lower than both the Operant Manual condition (M = 5.35, SD = 1.99, $t(28) = -5.741$, $p < .001$) and the Operant Vocal condition (M = 5.28, SD = 2.03, $t(28) = -4.980$, $p < .001$) which did not differ from each other ($t(28) = .242$, $p = .810$).

Discussion

The data show that participants consistently under-reproduced the temporal interval between their own action and its subsequent outcome, both when they were executing a manual action and a vocal action. This provides new evidence that individuals' vocal actions may produce an implicit sense of agency, although their magnitude results reduced to manual actions that are capable of causing changes in the physical environment ($d_z = 0.44$ vs. $d_z = 0.80$). This finding is consistent with Limerick et al. (2015) who deployed a Libet clock method and reported reduced temporal compression for vocal actions, compared to manual actions. This congruency further corroborates the interval reproduction as a viable and effective method to investigate temporal binding. Our data also provide evidence in favour of the debated finding that observed (manual) actions also produce subjective temporal compression. In fact, these results highlight a binding effect over observed actions and consequent outcomes, to a similar extent as for self-produced actions.

Still, the mechanism underlying this effect for observed actions is not entirely clear, and further investigations should specifically address the motor simulation processes potentially involved. Explicit agency ratings reflected the pattern suggested by implicit measures, but only for self-generated actions. On the contrary, when considering observed actions participants reported low ratings of control towards the outcome, whilst still showing time compression between the events. This finding provides further evidence that while explicit and implicit dimensions of agency may be linked, they likely do not rely on the same mechanism (Moore et al., 2012; Synofzik et al., 2013).

Experiment 2

Experiment 2 aimed to replicate the novel temporal binding effect for self-generated vocal actions. Furthermore, building on the finding in Experiment 1 that time compression occurred between observed physical actions and their outcomes, it was tested whether *heard* vocal actions would lead to the same effect. In fact, if all actions are computed analogously (e.g., by ascertaining cause-effect relationships), we should also expect to find subjective time compression for heard vocal actions produced by another agent. To do this, the experimental design used in Experiment 1 was amended, while maintaining its structure. Thus, Experiment 2 featured the same

Control and Operant Vocal conditions from Experiment 1. Furthermore, in order to translate the observed action condition from the physical domain (Experiment 1) to the social domain (Experiment 2), a Heard Vocal condition was included, where the same social action was performed by the experimenter. To explore whether the capacity for intentional action was a necessary ingredient for generating binding over others' vocal actions, a Heard Artificial condition was also included, where the same utterance was provided by a computer.

This study further sought to contribute to the long-debated question of what elements generate binding effects. For example, according to Buehner and Humphreys (2009), a causal relationship between two events is sufficient to produce temporal compression (account referred as "causal binding"). On the other hand, Haggard and Chambon (2012) propose that causality is necessary but not sufficient to achieve temporal compression, while intentional action planning is needed (account referred as "intentional binding"). If causality is the only root to temporal compression, we should expect to observe binding effects in all conditions, except for the Control condition. Alternatively, if temporal compression for observed actions requires that we represent those actions in our own motor system, we should not necessarily expect to find binding effects in the Heard Vocal and Heard Artificial conditions, as participants would not be able to observe them (ad by consequence, to simulate them).

Method

Participants

To promote consistency between the experiments, we recruited a new sample of equal size to Experiment 1. Therefore, 32 participants aged 18-30 years (*M* = 19.90, *SD* = 2.12, 25 were females) recruited from the University of East Anglia, completed the experiment. Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported to have normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Apparatus and stimuli

The experimental setting and apparatus remained unaltered from Experiment 1. Two auditory start signals were created to be used in Experiment 2. First, in order to create comparable auditory stimuli across all conditions, 30 audio tracks of the experimenter saying the word ‘go’ were recorded. Durations and frequencies of these utterances were analysed. The average duration was 346.72ms (*SD* = 9.26), while the average frequency was 166.87Hz (*SD* = 3.41). Hence, for the Heard Artificial condition, a voice track was created using an on-line vocal synthesizer (www.cepstral.com), and then edited to match average duration and frequency of the human voice previously analysed (347ms, 167Hz). For the Control condition, a matching tone (347ms, 167Hz) was created as start stimulus. All auditory stimuli were created, analysed, or edited using MATLAB.

Design and procedure

The within-subjects design was identical to that of Experiment 1, with the exception that the start signal was replaced in two of the four blocked conditions. The four experimental conditions were Control, Operant Vocal, Heard Vocal, and Heard Artificial. Block order was pseudorandomized across participants. The experimental procedure remained unaltered from Experiment 1 for the Control and Operant Vocal condition. The Heard Vocal and Heard Artificial conditions were identical to the Operant Vocal condition, except for the source of the vocal input. In the Heard Vocal condition, the experimenter used the microphone to provide a vocal action (i.e., pronouncing the word “go” in the microphone), whilst remaining seated beside the participant (therefore, action execution information was not accessible). In the Heard Artificial condition, the microphone was positioned towards the left speaker, and the signal provided (347ms, 167Hz) served as vocal input. In this manner, the causal relationship between the events in the Heard Artificial condition was preserved. Across all conditions, the outcome tone (100ms, 1 kHz) was played by the right speaker only. A graphical representation of the trial procedure is shown in Figure 9.

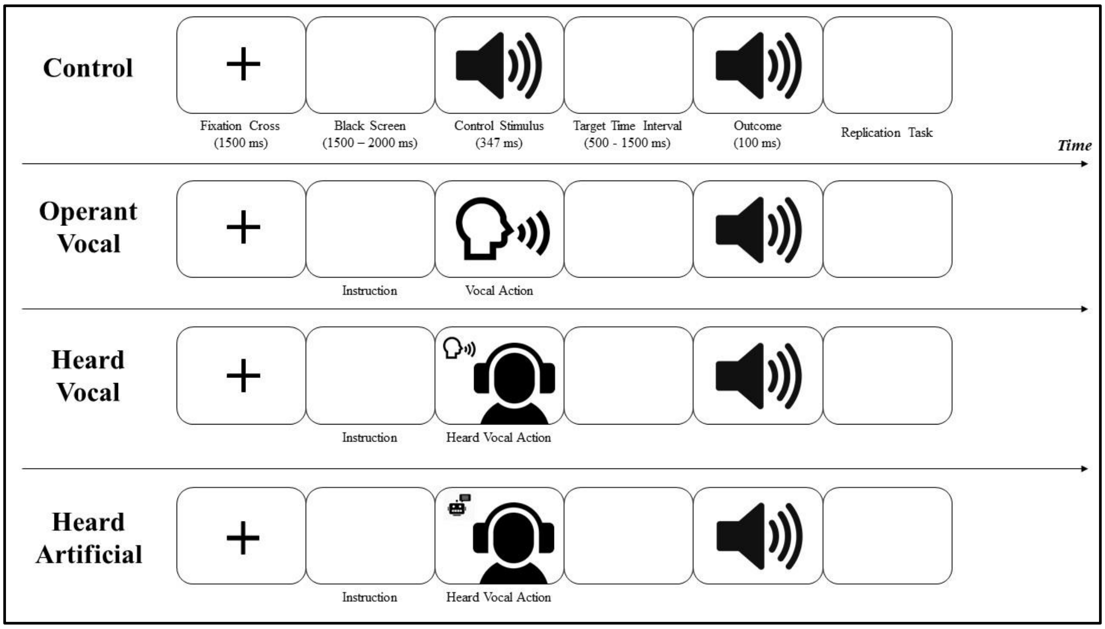


Figure 9. Trial procedure for the four condition in Experiment 2.

Results

The same exclusion criteria of Experiment 1 were adopted for Experiment 2. This resulted in 2 participants excluded for inaccurate reproductions and 29 trials deleted (0.76% of total trials).

Proportional interval reproductions

Mean proportional interval reproductions were calculated for each participant and submitted for statistical analysis. Shapiro-Wilk tests were performed on each condition to test normality of each distribution, and all conditions reported p >.05. A one-way within-subjects ANOVA was conducted on proportional interval reproduction scores. There was a significant effect of the type of action: $F(3, 87) = 7.757$, $p < .001$, $\eta_p^2 = 0.211$. Planned paired-samples t-tests revealed that the mean proportional interval reproduction in the Control condition (M = 0.95, SD = 0.19, 95% CI [0.88, 1.02]) was significantly higher than the Operant Vocal condition (M = 0.82, SD = 0.19, 95% CI [0.75, 0.89], $t(29) = 3.833$ $p = .001$, $d_z = 0.71$). No significant differences were found between the Control, Heard Vocal (M = 0.92, SD = 0.14, 95% CI [0.87, 0.97]) and Heard Artificial (M = 0.97, SD = 0.20, 95% CI [0.89, 1.04]) conditions: $F(2, 58) = 1.158$, $p = .321$, $\eta_p^2 = 0.085$. Figure 10 shows a graphical representation of the data.

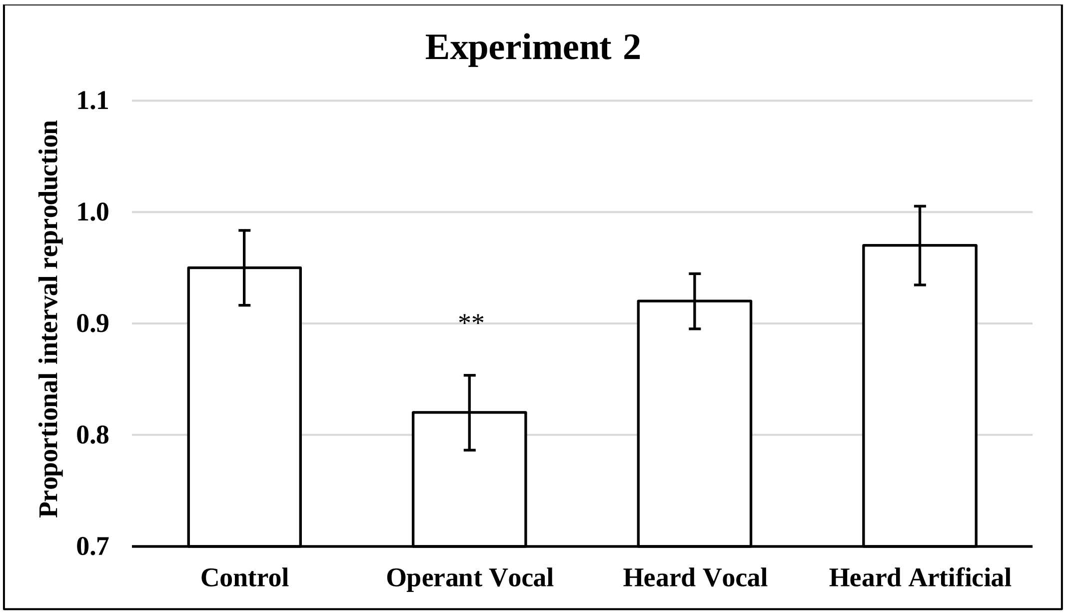


Figure 10. Mean proportional interval reproduction across the four conditions in Experiment 1. Error bars represent the standard error of the mean. Significant differences from the Control condition are highlighted for the .05 (*) and .01 (**) level.

Explicit agency ratings

When comparing self-reported explicit agency ratings, a one-way repeated measures ANOVA revealed a significant effect between the four conditions: $F(3, 87) = 21.230$, $p < .001$, $\eta_p^2 = 0.423$. Planned paired-samples t-tests revealed that the mean explicit agency rating in the Operant Vocal condition (M = 5.53, SD = 1.96) was higher than the Control condition (M = 3.53, SD = 1.78, $t(29) = 4.754$, $p < .001$) the Heard Vocal condition (M = 3.53, SD = 1.98, $t(29) = 4.882$, $p < .001$) and the Heard Artificial condition (M = 3.37, SD = 1.75, $t(29) = 5.559$, $p < .001$). No significant differences were found between the Control, the Heard Vocal, and the Heard Artificial conditions: $F(2, 58) = 0.535$, $p = .588$, $\eta_p^2 = 0.018$.

Discussion

In Experiment 2, participants again consistently under-reproduced the temporal interval between their self-generated vocal actions and their subsequent outcomes. This corroborates findings from Experiment 1, providing confidence about the reliability of this novel effect. Given that data from Experiment 1 showed binding for observed physical actions, it was expected that participants might analogously underestimate the temporal gap between heard vocal actions and their consequent outcomes. Importantly, no reliable time compression was observed when participants listened to a vocal action performed by the experimenter, nor when the same utterance was generated by the computer. This seems to contrast with Experiment 1, where participants showed a temporal binding effect over observed physical actions.

A possible explanation for this discrepancy is that merely hearing the experimenter's vocal action did not allow participants to simulate it in their own motor systems. This element drove the research interest into investigating the effects of direct action observation on the generation of temporal compression. Indeed, the Observed Manual (Experiment 1) and Heard Vocal (Experiment 2) conditions were not comparable in the extent to which they allowed visual access to action kinematic information. Kinematics in action observation has been shown to be of critical importance to understand the action itself, and it is also associated with higher-level mechanisms such as intention attribution and theory of mind (Aglioti et al., 2008; Cavallo et al., 2016). Thus, the open question become whether direct action observation (intended as observed muscular activation) could be a contributing feature to the generation of a temporal binding effect over observed actions performed by others, which is investigated in Experiment 3.

It is worth mentioning that the data reported in Experiment 2 provides insight into the fundamental elements that generate temporal compression. Some accounts claim that temporal compression is achieved whenever a causal relationship between events is detected, whereas others imply that intentional action planning is needed. Here, in both the Heard Vocal and Heard Artificial conditions, the events were not mutually independent (as opposed to the Control condition), but linked in a cause-effect relationship. As reliable time compression was not found in these two conditions, this argues against the view that mere causation between events is sufficient to generate a temporal binding effect. It is therefore possible to infer that, while causation between the events is indeed a necessary prerequisite to achieve a binding effect, it may not be sufficient by itself.

Experiment 3

Experiment 2 did not show reliable time compression between heard vocal actions performed by the experimenter and their outcomes, in contrast with Experiment 1 where binding did occur for observed manual actions. Considering that self-generated vocal actions appear to be consistently capable of producing a temporal binding effect (as reported in Experiment 1 and 2), Experiment 3 set out to investigate the discrepancy between manual and vocal actions performed by others. A critical difference between the observed conditions in Experiment 1 and 2 was that while manual actions performed by the experimenter were entirely visible to the participants, this was not true for heard vocal actions. In other words, while there were two major differences between the Observed Manual and the Heard Vocal conditions, only one of them was manipulated. Namely, while participants could see the hand moving in the Observed Manual condition (Experiment 1), they could not see the mouth moving in the Heard Vocal condition (Experiment 2). Crucially, having visual access to action kinematic information is widely considered to play a major role in action prediction (Cavallo et al., 2016). Thus, if visual information plays a role in the generation of the temporal binding effect for observed actions (regardless their physical or social nature), it could be argued that a visuomotor component (and as such an action prediction process) should be considered for extending the comparator model. To address this divergence, in Experiment 3 a visual component was included in every condition.

Method

Participants

To promote consistency between the experiments, we recruited a new sample of equal size to Experiment 1 and 2. Therefore, 32 participants aged 18-24 years (*M* = 19.35, *SD* = 1.41, 25 were females) recruited from the University of East Anglia, completed the experiment. Sample size was determined in the same way as described in Experiment 1 and 2. Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported to have normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Apparatus and stimuli

The experimental setting and apparatus remained unaltered from Experiment 1 and Experiment 2, except for the experimenter's position in the Observed Human condition, which was now facing the participants' position. The experimenter's position was located 60cm to the left of participants, 60cm from the edge of the table (thus creating a 45° angle facing North-West from participants' midline). Two audio-visual start signals were created to be used in experiment 3 (.avi format, 25 frames/s), each composed of two video clips. For the Control condition, the audio-visual start signal consisted of a first video clip depicting a white dot on a black screen (duration randomized between 1500 and 2000ms), and a second video clip showing the white dot enlarging at a constant speed while the same tone used in the Control condition in Experiment 2 was played (347ms, 167Hz). For the Observed Artificial condition, the audio-visual start signal consisted of a first video clip displaying a static frontal medium close-up of a human avatar, created using an online avatar generator (www.voki.com) with duration randomized between 1500-2000ms. A second video clip was created, displaying the avatar opening its mouth while the same artificial voice used in the Observed Artificial condition in Experiment 2 was played (347ms, 167Hz). All audio-visual stimuli were created and edited using Adobe Premiere Pro CS6 (Adobe Systems Software Ltd, Dublin, Ireland). Illustrations of the visual stimuli can be found in Figure 11.

Design and procedure

The design was identical to Experiment 2, with the exception that three of the blocks presented modified start signals. The four experimental conditions were: Control, Operant Vocal, Observed Vocal and Observed Artificial. Block order was pseudorandomized across participants. The experimental procedure for the Operant Vocal condition remained unaltered from Experiments 1 and 2.

Control trials began with a white fixation cross displayed in the centre of the screen (1500 ms) after which both video clips composing the control audio-visual start signal were presented in sequence. During the stimulus presentation, only the left speaker was operative. Then, after the target time interval (randomised between 500 and 1500ms) a high-pitch tone (100ms, 1kHz) was presented through the right speaker. Participants were then asked to complete the reproduction task.

The Observed Vocal trial procedure was identical to the Heard Vocal trial procedure in Experiment 2, but participants could now directly observe the experimenter as he was executing the vocal action, as he was sitting in front of them. Participants were instructed to wait for the fixation cross to disappear (1500ms) and then to move their gaze towards the centre of the experimenter's face, who would then execute the vocal action at a time of his choosing. Participants were instructed to move their gaze back to the monitor immediately after the end of the observed vocal action. After a variable time interval (randomized between 500 and 1500ms), a high-pitch tone (100ms, 1kHz) was released through the right speaker, and participants completed the reproduction task.

The Observed Artificial trial procedure was designed on the basis of the Heard Artificial condition in Experiment 2. Trials in this condition began with a white fixation cross (1500ms) after which both video clips composing the avatar audio-visual start signal were presented in sequence. The microphone was positioned toward the left speaker, and the signal provided (347ms, 167Hz) served as vocal input. In this manner, causal relationship between the events in the Observed Artificial condition was preserved. Then, after the target time interval (randomised between 500 and 1500ms) a high-pitch tone (100ms, 1kHz) was presented through the right speaker. Participants were then asked to complete the reproduction task. A graphical representation of the trial procedure is shown in Figure 11.

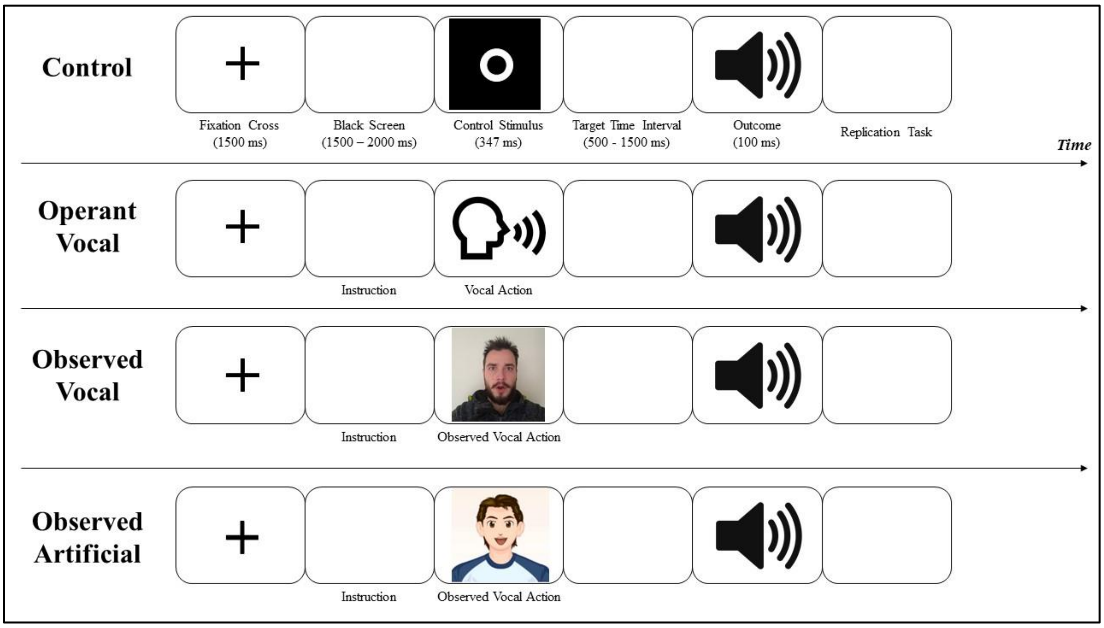


Figure 11. Trial procedure for the four conditions in Experiment 3.

Results

The same exclusion criteria of Experiments 1 and 2 were adopted for Experiment 3. This resulted in 1 participant being excluded for inaccurate reproductions and 33 trials deleted (0.86% of total trials).

Proportional interval reproductions

Mean proportional interval reproductions were calculated for each participant and submitted for statistical analysis. Shapiro-Wilk tests were performed on each condition to test normality of each distribution, and all conditions reported $p > .05$. A one-way within-subjects ANOVA was conducted on proportional interval reproduction scores. There was a significant effect of the type of action: $F(3, 90) = 6.783$, $p < .001$, $\eta_p^2 = 0.184$. Planned paired-samples t-tests revealed that the mean proportional interval reproduction in the Control condition (M = 0.91, SD = 0.24, 95% CI [0.82, 1.00]) was significantly higher either than the Observed Vocal condition (M = 0.86, SD = 0.24, 95% CI [0.76, 0.95, $t(30) = 2.190$, $p = .036$, $d_z = 0.40$), and the Operant Vocal condition (M = 0.81, SD = 0.27, 95% CI [0.71, 0.91], $t(30) = 3.096$, $p = .004$, $d_z = 0.56$). No significant difference was found between the Control condition and the Observed Artificial condition (M = 0.92, SD = 0.24, 95% CI [0.83, 1.01], $t(30) = -.313$, $p = .756$. Figure 12 shows a graphical representation of the data.

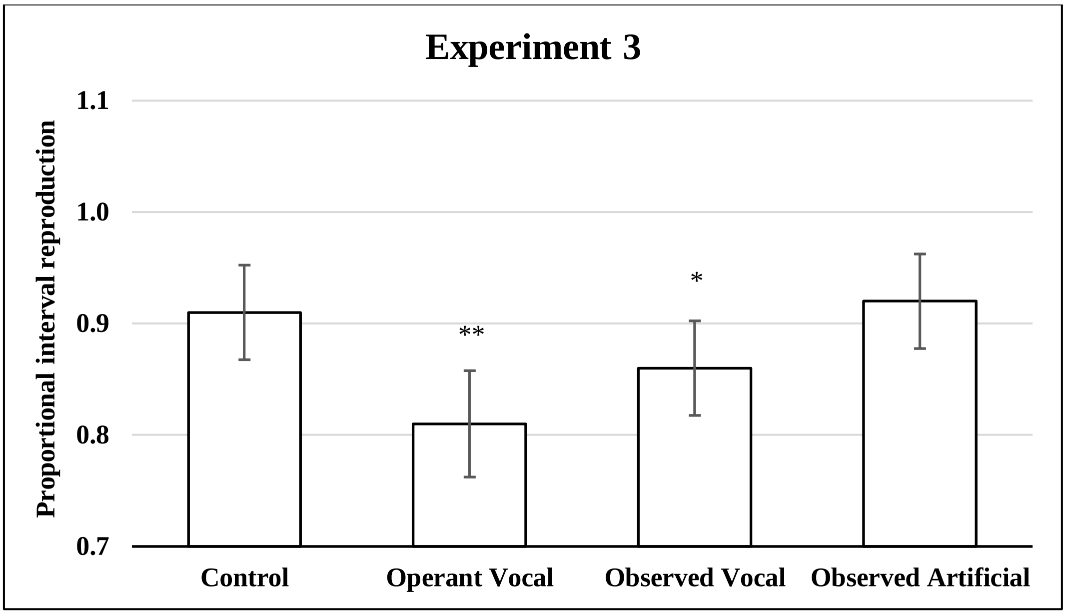


Figure 12. Mean proportional interval reproduction across the four conditions in Experiment 1. Error bars represent the standard error of the mean. Significant differences from the Control condition are highlighted for the .05 (*) and .01 (**) level.

Explicit agency ratings

When comparing self-reported explicit agency ratings, a one-way repeated measures ANOVA revealed a significant effect between the four conditions: $F(3, 90) = 20.082$, $p < .001$, $\eta_p^2 = 0.401$. Planned paired-samples t-tests revealed that the mean explicit agency rating in the Operant Vocal condition (M = 5.10, SD = 1.89) was higher either than the Control condition (M = 2.65, SD = 2.04, $t(30) = 5.517$, $p < .001$), the Observed Vocal condition (M = 3.16, SD = 2.01, $t(30) = 4.790$, $p < .001$), and the Observed Artificial condition (M = 2.84, SD = 1.97, $t(30) = 5.700$, $p < .001$). No significant differences were found between the Control, the Observed Human, and the Observed Artificial conditions: $F(2, 60) = 1.683$, $p = .194$, $\eta_p^2 = 0.053$. Explicit agency ratings across all experiments are summarised in Table 1.

Discussion

Replicating Experiments 1 and 2, participants again showed a temporal binding effect for self-generated vocal actions, providing further evidence that (like motor actions) vocal actions can generate a sense of agency towards their outcome. Also consistent with the previous experiments, binding was not observed when participants were shown a neutral stimulus on screen, nor when an artificial avatar was presented. Strikingly however, and in contrast to Experiment 2, participants showed a reliable time compression over vocal actions performed by a visible experimenter. This finding seems to indicate that direct observation of an actual action is necessary in order to experience a temporal binding effect for actions performed by others. Taken together, these findings are consistent with the possibility that the temporal binding effects for observed actions rely on participants' ability to represent, or embody, the actions of others.

The role of embodiment in producing temporal binding for observed actions received little investigation, and to the author's knowledge no previous study has directly compared binding for others' genuine actions under conditions where the action can or cannot be directly observed. Here, too, it is only possible to draw inferences from comparing binding effects across experiments. Thus, further research is needed to replicate and verify the importance of direct action observation. However, it is note-worthy that no binding emerged in the Observed Artificial condition (nor in the Heard Artificial condition of Experiment 2), providing further evidence that merely recognizing a causal relationship between events, while necessary, may not be sufficient to produce a binding effect.

Chapter discussion

The aim of this Chapter was to investigate how people detect agency for themselves, other humans, and artificial agents when they are performing a physical (e.g., manual) or social (e.g., vocal) action. To this end, subjective temporal compression was measured across three different experiments. This effect, known as temporal binding, has been interpreted by some (e.g., Haggard et al., 2002) as an index of sense of agency, and by others (e.g., Buehner, 2012; Buehner & Humphreys, 2009) as merely reflecting the recognition of cause-effect relations between events, even in the absence of intentionality. The data presented contribute to this debate by directly comparing conditions where causality occurred with or without intentionality. Throughout the three experiments reported, there was consistent evidence for the novel finding that vocal utterances produced a temporal binding effect, in a similar fashion as other motor actions (e.g., hand movements) do. This is an indication that we feel a sense of agency for vocal actions as well, despite their inability to directly affect the physical environment. However, it is worth noting that effect sizes reported for self-generated vocal and manual actions were of a different magnitude ($d_z = 0.44$ vs. $d_z = 0.80$), which is consistent with Limerick et al. (2015). One critical implication of this is that the social environment – like its physical counterpart – affords opportunities to modify our surroundings and to experience a sense of agency over the effects we produce in other people. These findings provide evidence that a common mechanism may be accounting for the sense of agency both for physical and communicative actions, as originally theorised by Stephenson et al. (2018). The fact that vocal actions can produce a binding effect meaningfully expands our understanding of how agency is experienced, as the majority of previous studies focused on outcomes produced by button presses (see Moore & Obhi, 2012 for a review).

Mechanisms of temporal binding

Experiment 1 reported results showing evidence for a temporal binding effect following the observation of others' physical actions, which is consistent with previous research (Poonian & Cunnington, 2013; Strother et al., 2010; Wohlschläger et al., 2003). Still, there is debate over why observed actions lead to temporal compression. Some authors favour the interpretation that this reflects a vicarious sense of agency, suggesting that such a mechanism could be an important feature of the empathic experience towards others' actions (Wegner et al., 2004). Their claim is that the emotional sensitivity gained for actions performed by others might be due to our ability to build foreknowledge of their actions, leading us to experience those actions as they belong to us and are under our personal control. Data from Experiments 1-3 contrast with this hypothesis: while participants did experience a binding effect over observed actions, they did not perceive personal control over the outcome produced by those actions (as measured through explicit agency ratings). This also suggests that implicit and explicit sense of agency may be driven by separate mechanisms (Moore et al., 2012).

An alternative explanation for temporal binding in action observation is that perceivers mentally represent others' actions within their own motor systems. Viewing actions performed by other individuals activates frontal and parietal cortical areas typically involved in action planning and execution (Di Pellegrino et al., 1992;

Fogassi et al., 2005; Gallese et al., 1996). This process may include not only a simulation of motor planning and execution, but also a prediction of the outcomes that would be generated by the observed action (Aglioti et al., 2008). In other words, when looking at someone else's action, the mirror neuron system may simulate that action as if it was our own, and generate a sensory prediction of the anticipated outcome. These predictions are then processed by the comparator and matched with actual sensory feedback. If the prediction matches the actual outcome, subjective time compression would be produced following the same mechanism as for self-generated actions (Blakemore et al., 2002; Haggard, 2017). This alternative explanation is corroborated by the discrepancy found between the heard and observed vocal actions in Experiments 2 and 3. In fact, in Experiment 2 vocal actions produced by an experimenter seated out of participants' sight did not generate a binding effect over their outcomes. However, when participants were granted visual access to the action kinematics of the experimenter's speech (Experiment 3), subjective time compression re-emerged.

Across the three experiments reported here, auditory signals alone (whether computer-, human- or avatar-generated) were not capable of producing a temporal binding effect. This suggests that explicit knowledge of causality (by itself) may be necessary (Buehner, 2012) but not sufficient to induce an implicit sense of agency. This evidence is in contrast with previous research which advanced the idea of a causal binding to occur between two events whenever the latter is thought to depend on the prior (Buehner & Humphreys, 2009). However, methodological differences must be noted: while our task asked participants to reproduce a time interval they just experienced, Buehner and Humphreys (2009) asked participants to indicate the time point at which they predicted an outcome that they had (or had not) caused to occur. In other words, our paradigm involved a retrospective assessment of time, while that of Buehner and Humphreys used a prospective assessment task. This element could lead to inappropriate comparisons between the studies, as while our method was focussed on both onset and offset of the target time interval (i.e., action and outcome), theirs specifically analysed the offset point (i.e., the outcome). Future research could tackle this controversy, by assessing causality and intentionality as features of the temporal binding effect using a consistent methodology to allow direct comparisons.

As we predicted, reported data showed no temporal binding effect for artificial vocal actions, both when they were heard and observed. This finding corroborates the idea that causality alone is not the only key to achieve temporal compression. In fact, the mere communicative act (pronouncing a word) was, when uttered by artificial agents, ineffective at generating binding effects. In other words, unlike observing another person producing a vocalization, viewing an artificial agent pronounce the same vocalization did not yield a temporal binding effect. One potential constraint on the interpretation of these results is that the artificial agent was a 2D avatar created through the animation of a drawn human face, and as such it was intrinsically different from the (3D, realistic) embodied agent that executed the vocal action in the Observed Human condition (i.e., the experimenter). Because the extent to which agents are embodied is an important aspect of action perception and intention attribution (Hostetter & Alibali, 2008; Niedenthal et al., 2005), Chapter 3 will investigate whether temporal binding is produced when observing actions carried out by embodied artificial agents (i.e., a social robot) other than human actors, or if in fact it is an exclusive feature of observed human actions.

Taken together, these findings suggest that visuomotor information plays a fundamental role in the generation of the temporal binding effect over observed actions. Still, this hypothesis remains speculative as based on preliminary findings yet to be replicated. Future research should directly focus on this topic, with the scope to understand specific contributions of visual information and accessibility in observed actions.

The role of agency in social interaction

The findings from Experiments 1-3 are also consistent with Kunde et al. (2018), who proposed a theoretical framework of sociomotor action control. According to this model, others' responses to our communicative actions (including our vocalizations) are used to plan subsequent communicative actions. Hence, being able to detect whether our actions were effective in producing the desired outcome acquires critical importance. For example, when a friend does not respond when we call their name, it may mean that our intention (to draw their attention) has not been achieved, and further actions are needed to achieve our goal. Indeed, experiencing agency over the outcomes we produce with our communicative actions is necessary for planning what to do next. Thus, the role of agency in social interaction is of central importance, and may also support higher-level mechanisms such as theory of mind. In these terms, agency may be a critical link in the chain connecting joint action and social cooperation, as ascribing agency to others (i.e., to perceive them as agentic, capable of producing changes in the environment) is crucial to develop expectations about their intentions and mental states (Sebanz & Knoblich, 2009).

A possible limitation to the current research relates to the nature of the outcome that followed the action. While this work was motivated by the goal of shedding light on the social aspect of the action (i.e., talking), that action was not followed by a meaningful social outcome. This was essential in order to isolate the phenomenon of temporal binding for vocal actions. However future research might fruitfully explore how the relationship between communicative actions and their outcomes (e.g., by matching the outcome to the vocal command) impacts the experience of agency.

These findings may be of direct interest for developers of vocal assistants like Siri and Alexa - those user-friendly devices of increasing prevalence worldwide. During the design process, they should take into account that both robotic and human voices may not be perceived as agentic per se. This should direct future research into exploring different interfaces that allow more intuitive and spontaneous interactions, focusing on other typical factors of human-human interaction such as eye-gaze (Bayliss & Tipper, 2006), embodiment (Niedenthal et al., 2005) and temporal coordination (Schmidt et al., 2011). New insights and innovation in the development of vocal artificial agents will enhance the quality of the social engagement we experience towards them, which is likely to be a central element of the social interactions we will build in the near future.

Here was reported consistent evidence of a novel temporal binding effect for vocal actions that produce a systematic outcome. Such effects occurred not only when participants produced the vocal action themselves, but also when they observed someone else doing it, as long as direct visual access to the other’s vocal action was possible. These findings make an important contribution to the growing literature concerned with how perceivers represent intentional actions and their consequences, demonstrating that action is not limited to movements by our hands and feet, but include a range of social behaviours as well.

Chapter 3

Agency detection for observed robotic actions

"We are fascinated with robots because they are reflections of ourselves."

Ken Goldberg, 2001

Throughout our daily lives, we commonly interact with others in different and complex ways. Such interaction has adapted to our social systems. However, while humans are experienced in interacting with other natural social agents, interaction with computerised and artificial systems which have recently become popular in our homes (e.g., vocal assistants) is still relatively novel. Humans are social animals and thus we commonly interact with one another in their everyday life. During such interactions, we attribute to ourselves and to others independent mental states, desires, and intentions (Leslie, 1995). Intuitively, these elements are not considered in relation to artefacts that we use to fulfil our goals. For example, when we use a food processor to chop some vegetables, there will be no attribution of intentions or desires to the machine. Nonetheless, the distinction between agents and artefacts can become blurred when considering artificially intelligent entities, including social robots. Consequently, investigating the process through which social robots may activate aspects of our cognitive systems that are typically reserved for other humans gains crucial importance. Within the next decades, the development of artificial agents will undergo a ground-breaking revolution, transforming robots into an innovative category of social entities among the human population. Currently, robots are intended as programmable and mechanical artefacts. Yet, the scope of this line of research is to undercover how they can become artificial agents, capable of modifying their surroundings through their physical (i.e., embodied) presence in the environment and the ability to spontaneously navigate it, executing the actions necessary to effect change.

As discussed in the previous chapters, agency detection for other agents affects the nature of our interaction. For instance, if we are looking at someone perceived as an agent, we will spontaneously try to predict the outcomes of their actions and infer their goals. Previous research showed that agency detection for others affects (and is affected by) crucial cognitive systems, including attention, perspective taking and spatial coding (Wiese et al., 2012; Wykowska et al., 2014; Stenzel et al., 2012; Zwickel, 2009;Ward et al., 2019). Taking into account robots as agents, the scientific community usually resist direct comparisons with human agents, conferring robots with an entirely different status (Perez-Osorio & Wykowska, 2019). Consequently, investigating the process through which social robots may activate aspects of our cognitive systems that are typically reserved for other humans gains crucial importance. Most importantly for the scope of this thesis, the research reported in this chapter aims to investigate whether robotic actions can be perceived as agentic, and which factors can modulate our agency detection in relation to them.

Cognitive systems in human-robot interaction

Previous research showed that the mere presence of a robot in the action context can affect participants' behaviour, inducing them to follow its suggestions while undertaking a decision-making task (Shinozawa et al., 2005), or to increase their peri-personal and reaching space (Bainbridge et al., 2008). These findings suggest that being in the presence of a robot can alter our decisions and perceptions, indicating that they are classified differently compared to other artefacts. Beyond the effects of their mere physical presence, robots can also impact human behaviour through their actions. For instance, Kompatisari et al. (2018) used a target discrimination task to investigate the effects of an anthropomorphic robot's (iCub) gaze on participants' attention. Their findings suggest that participants' own attention was biased towards the direction of the robot's gaze, in much the same way that happens in response to human gaze (Frischen et al., 2007). Similar findings have been reported in the domain of joint actions. It is widely established that people interacting together tend to synchronize their actions (Sebanz et al., 2003), which led Ciardo and Wykowska (2018) to wonder whether this effect may also emerge when interacting with artificial agents. Using a joint Simon task (Simon, 1990), participants showed the tendency to coordinate their action with a non-anthropomorphic robot. However, it is important to note that such effects may rely on participants' belief that the robot is controlled by another person (Stenzel et al., 2012). Taken together, these findings suggest that artificial agents can actually gain social resonance when interacting with humans, earning the name of social robots.

Sense of agency and artificial agents

With specific reference to human-robot interaction, Ciardo et al. (2018) showed that the sense of agency experienced towards our own actions can be reduced when such actions are executed together with a social robot. Specifically, participants reported lower levels of agency when they executed an action jointly with a social robot (compared to by themselves alone), but only if that action was associated with a negative cost (losing a variable amount of points). These findings are consistent with previous research that showed a strong self-serving bias in attributing actions with a positive outcome to ourselves, while ceding responsibility for negative actions to others (Wegner & Wheatley, 1999, Yoshie & Haggard, 2013). To avoid the confounding influence of such top-down processes, many studies started investigating implicit measures of the sense of agency (such as the temporal binding), which would better control for desirability effects and cognitive biases. Thus, adopting an implicit methodology may be useful to provide new insights into how people compute agency in relation to social robots.

Roselli et al. (2019) expanded on the work of Ciardo et al. (2018), comparing temporal binding effects when participants acted alone or interacted with an embodied non-humanoid robot. In this study, participants used a Libet clock method to report the timing of actions they executed alone (individual condition) or jointly with a robot (social condition). During the social trials, either participants or the robot could initiate the action. Their findings indicated that participants showed a weaker temporal binding effect towards their actions when they were executed in the presence of the robot, compared to when participants acted alone. In other words, their actions were less bound to the outcome when that outcome could also be attributed to the robot. These results are consistent with what was reported on an explicit level by Ciardo et al. (2018), suggesting that the temporal binding effect is indeed reflecting an implicit dimension of the sense of agency. Taken together, these findings suggest that responsibility for the outcomes of our actions can be displaced towards another agent, that is also capable of executing the action, regardless of it being human or robotic.

Action observation and artificial agents

The experiments reported in Chapter 2 showed that although participants did not experience a sense of agency towards observed actions (as measured through explicit agency ratings), temporal binding effects still emerged when they had visual access to kinematic information of the action execution. This evidence suggests that temporal binding effects can be produced for actions performed by others, perhaps with the contributing activation of the Action Observation Network (AON; Gallese & Goldman, 1998), a collection of brain regions comprising sections of the parietal, premotor, and occipitotemporal cortex, which respond when observing other humans performing an action. Interestingly, the AON has been shown to be sensitive to both human and non-human actors, as Gazzola et al. (2007) reported that the portions of the parietal, premotor, and middle temporal cortex ascribed to its activation respond both when observing humans grasping and manipulating objects, as well as a robot arm performing the same actions. These findings are consistent with an electroencephalography (EEG) study that showed a mu-suppression effect over sensorimotor areas when participants were observing either human or robotic actions (Oberman et al., 2007). Further investigations have implemented functional magnetic resonance (fMRI) techniques, and have reported that the AON was more strongly engaged during the observation of a mechanical movement, regardless of whether the action was performed by a robot or emulated by a human (Cross et al., 2012). These pioneering findings have been interpreted as a greater modulation of the AON in response to greater prediction errors that arise from unfamiliarity with robotic motions (see Press, 2011 for a review).

As discussed above, the observation of robotic movements typically engages brain areas related with action perception. Yet, the extent to which humans also attribute intentions and emotions to robotic entities remains unclear. Previous research featuring brain imaging methodologies suggested that the Person Perception Network (a group of brain regions comprising the fusiform face area and the extrastriate body area, which responds to the observation of other individuals, especially their faces and bodies; Quadflieg et al., 2011) was strongly engaged both when observing robots producing emotional expressions (Hortensius, et al., 2018), as well as when observing humans and robots interacting together (Wang & Quadflieg, 2015). Furthermore, an fMRI study (Özdem et al., 2017) adopted a robot gaze cuing paradigm to show behavioural and brain responses associated with the generation of a theory of mind (such as enhanced activation of the bilateral anterior temporoparietal junction) but only when participants believed that another person was controlling the robot (Henschel et al.,2020).

Overview of Experiments 4-5

Experiment 4 sought to investigate whether the observation of actions produced by a social robot is capable of eliciting temporal binding effects, and to what extent these effects are comparable to those associated with observing human actions. To this end, participants were asked to complete a time interval reproduction task (Humphreys & Buehner, 2010) on actions executed either by themselves (Operant condition), by a human (Observed Human condition), by a social robot (Observed Robot condition), or in the absence of any action at all (Control condition). We predicted that, if observing a social robot acting in the environment attunes our motor system as human actions do, participants will underestimate the time interval between actions and outcomes produced by the robot (to a similar extent of human actions), while no time underestimation will be achieved when witnessing two independent events. Experiment 5 aimed to investigate whether prior experience interacting with the social robot can modulate the binding effects, and how interactions of different types can affect participants' performance. In all experiments, details of how sample size was determined, as well as all data exclusions (if any), all manipulations, and all measures are reported.

Experiment 4

In Experiment 4, participants completed an interval reproduction task under four different conditions, manipulated within-subjects. Each condition featured a different start stimulus, each of which produced an identical outcome. The first aim of the experiment was to replicate previous findings from Experiment 1, including an Operant condition and an Observed Human condition where participants reproduced the time interval between their own manual action (pressing a key on the computer keyboard) or an observed manual action performed by the experimenter, and a tone produced in response to the action. To test whether observed robotic actions were also capable of producing a similar level of temporal compression, an Observed Robot condition was added, where participants reproduced the time interval between the same action performed by a robot (i.e., a Cozmo robot, distributed by Anki) and a tone produced in response to that action. As is typical for temporal binding paradigms, performance in the three experimental conditions (Operant, Observed Human, and Observed Robot) were compared with a Control condition in which no action was performed by any agent. Here, participants reproduced the time interval between two tones that were automatically produced by a computer.

Method

Participants

A power analysis (carried out with G*Power 3; Faul et al., 2007) indicated that to detect a medium-large effect size (as shown in Experiment 1 for the Observed Manual condition) $d_z = 0.69$, with $1 - \beta = 0.95$ at $\alpha = 0.05$, a minimum sample size of 30 would be required. Therefore, 32 participants (a sample size capable of detecting an effect size $d_z = 0.66$, with $1 - \beta = 0.95$ at $\alpha = 0.05$) aged 18-25 years (M = 20.03, SD = 1.60, 21 were females), recruited from the University of East Anglia, completed the experiment. Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported having normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Design

Subjective temporal compression was measured using an interval reproduction task where participants were asked to reproduce the duration of the time interval between two events by pressing and holding down the 0 key on the keyboard number pad for the same amount of time. In order to effectively manipulate participants' agency perceptions, the nature of the first event varied by block, while the second event was identical across blocks. Participants completed a repeated measures design experiment, having four blocks of trials consisting of 5 practice trials and 30 experimental trials, each with a different start stimulus, for a total of 120 experimental trials. Throughout the experiment, they were asked to replicate the time interval between an action and an outcome. Such actions could have been executed either by themselves (Operant condition), by the experimenter (Observed Human condition), or by Cozmo (Observed Robot condition). An additional condition was used as baseline, where no action was executed (Control condition). Block order was pseudorandomized across participants. The calculated dependant measure was the inter-event proportional interval reproduction, derived by dividing the reproduced time interval by the actual time interval for the same trial (ms). Thus, scores equal to 1 represented perfect time accuracy, while scores greater than 1 were over-reproductions, and scores lower than 1 were under-reproductions (that is, subjective temporal compression).

Stimuli

A first low-pitch tone (150 ms, 440 Hz sine wave, sample rate: 44100 Hz, bitrate 16: Poonian & Cunington, 2013) was created as start stimulus in the Control condition. A second high-pitch tone (100 ms, 1 KHz sine wave, sample rate: 44100 Hz, bitrate 16: Humphreys & Buehner, 2010) was created as end stimulus in the Control condition and as action's outcome in the Operant, Observed Human and Observed Robot conditions. The experiment was run using E-prime version 3.0 (Psychology Software Tools, Inc., Sharpsburg, PA, USA). All auditory stimuli were created using MATLAB (MathWorks, Natick, MA, USA).

Apparatus and materials

The experimental setting consisted of two adjacent chairs and a table in a dimly lit room. A computer monitor (BENQ XL2411: size: 24"; resolution: 1920x1080; refresh rate: 60 Hz) and a set of external speakers (Bose Companion 2 series III) were placed on the table and used to display/produce experimental stimuli. An external keyboard (Kensington KP400) was placed on the table and used to collect participants' responses. The height of the table was 80 cm. The position of the monitor was centred to participant's body midline, 60 cm from the edge of the table. The position of the speakers was respectively 30 cm on the right (right speaker) and 30 cm on the left (left speaker) of participants' body midline, 50 cm from the edge of the table. The position of the keyboard was centred to participants' body midline, 25 cm from the edge of the table. During the Observed Human trials, the experimenter sat adjacently on the left of the participants' position, with no further alteration of the experimental setting. In the Observed Robot trials, the action was executed by Cozmo. The robot had a non-anthropomorphic appearance: two sets of wheels allowed movement through track connection, while key press actions were enabled by a mobile joint on the front (see Fig. 13 for an image of Cozmo).

Figure 13. A Cozmo robot (distributed by Anki). Dimensions: 6 x 10 x 6 cm.

Procedure

Participants were invited to the laboratory and welcomed. They read an information sheet and provided informed consent to join the study. Prior to beginning the experiment, participants were introduced to the experimental equipment. They completed a sample of five Operant trials (with no reproduction task) to illustrate that given an action a high tone would be released, thus ensuring that they understood the causal relationship between the two events.

Control trials began with a white fixation cross displayed in the centre of the screen (1500 ms) after which, at a random interval (1500-2000 ms; Humphreys & Buehner, 2010) a low-pitch tone (150 ms, 440 Hz) was presented through the left speaker. Then, after the target interval of time (randomised between 500 and 1500 ms; Howard et al., 2016; Poonian & Cunnington, 2013) a high-pitch tone (100 ms, 1 kHz) was presented through the right speaker. Participants were then asked to reproduce the duration of the interval between the two tones by pressing and holding down the 0 key on the number pad for the same amount of time. After participants completed the reproduction task for each trial, they were asked to press the N key to begin the following trial.

In the Operant trials, participants were instructed that after the fixation cross disappeared (1500 ms) performing a specific action (i.e., press the space bar at any moment of their choosing) would release, after a variable time interval (randomized between 500 and 1500 ms) a high-pitch tone (100 ms, 1 kHz) through the right speaker only. Participants were free to perform the action whenever they chose to, but were encouraged to avoid rushing in giving the action, or to start the action at any predetermined fixed time (e.g., counting to 3 before performing the action). Following the tone participants were instructed to press and hold down the 0 key on the number pad for a duration equal to that of the time interval between their action and the tone. The target time interval was set to begin after the end of participants' action (i.e., when the space bar was fully released). After participants completed the reproduction task for each trial, they were asked to press the N key to begin the following trial.

Using the same trial procedure of the Operant condition, participants completed the Observed Human condition where they observed the experimenter performing an action (i.e., pressing the space bar on the keyboard at any moment of his choosing). After a variable period of time (randomized between 500 and 1500 ms) a high-pitch tone (100 ms, 1 kHz) was released through the left speaker. Participants then completed the reproduction task and were asked to press the N key to begin the following trial.

Analogously, in the Observed Robot condition Cozmo was placed at the centre of the keyboard, in reach of the space bar. The trial procedure was identical to the Observed Human condition, with the only difference being the agent executing the action. Cozmo's actions were remotely controlled by the experimenter positioned outside the room. Participants were given no specific information about how Cozmo was executing actions (i.e., by remote control, a programmed sequence, or spontaneously). In order to reduce possible interference, the experimenter waited outside the laboratory in all conditions, except for the Observed Human condition. A graphical representation of the trial procedure is shown in Fig. 14.

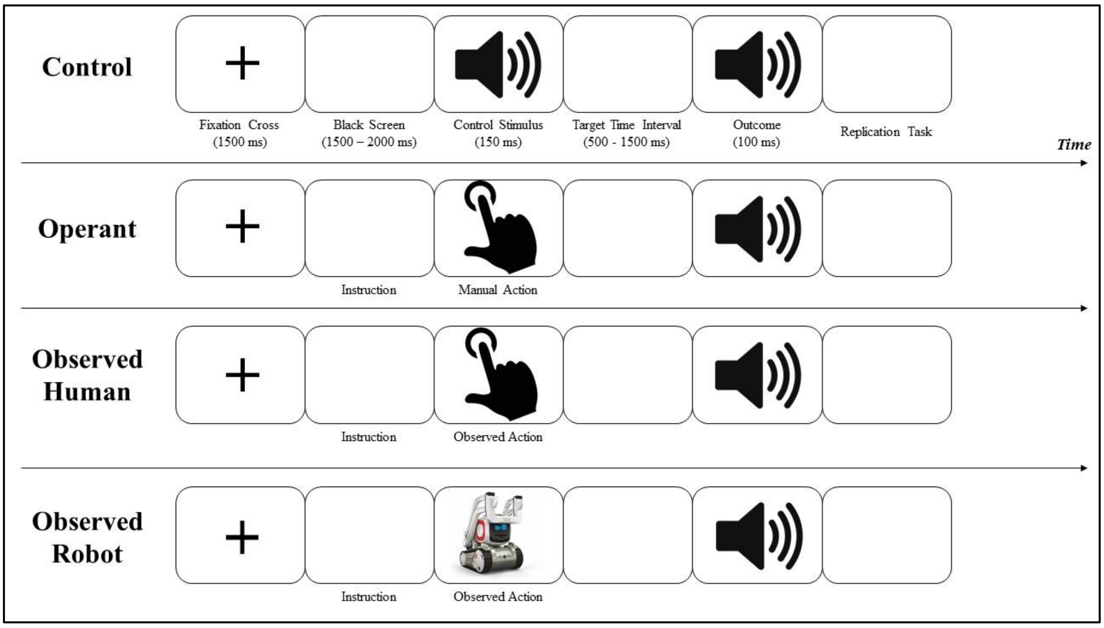


Figure 14. Trial procedure for the four conditions in Experiment 4.

Results

The following exclusion criteria were decided prior to the data collection: failure to produce temporal intervals covarying monotonically with actual action-tone intervals (average *r* across conditions lower than 0.4 would result in participant exclusion; Caspar et al., 2016), failure to respond (trials where the reproduced time interval was lower than 100 ms were discarded), and extreme variability (trials falling outside +/- 3 SD from individual mean were removed). This resulted in 1 participant excluded and 37 trials deleted (0.96% of total trials).

For all Experiments reported in the Chapter, we followed the indications for stepwise multiple comparisons, as indicated in Welsch (1977) and in Howell (2002) with specific regard to repeated measures designs. Here the authors suggest to test as few comparisons as possible using only the data involved in those comparisons. Accordingly, we report the ANOVA, followed by the t-test between the Control condition and the predicted significantly different condition with the highest mean (with the assumption that comparisons with lower means in other predicted significantly different conditions would imply necessarily a lower p-value). As a practice of good and open science, we then report results of all predicted comparisons, that whilst uninformative about the significance of those differences, could still be of relevance to the scientific community for the differences in effect sizes between different conditions.

Proportional interval reproductions

Mean proportional interval reproductions were calculated for each participant and submitted for statistical analysis. Shapiro-Wilk tests were performed on each condition to test normality of each distribution. The Observed Robot condition was the only one to reveal a non-normal distribution: $W(30) = 0.911$, $p = .016$, while all other conditions reported $p > .05$ (See Fig. 15).

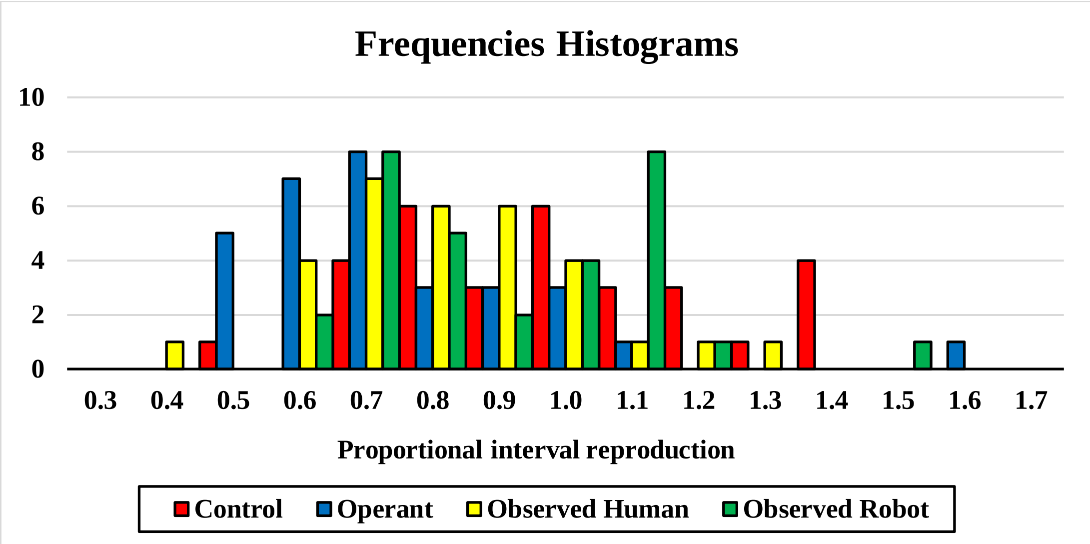


Figure 15. Frequencies histograms for the four conditions in Experiment 4.

Because of the non-normality of the distribution of scores in the Observed Robot condition (described above), a non-parametric analysis was performed. In order to run non-parametric analyses, the median (IQR) proportional interval reproduction was calculated for the Control (Mdn = 0.92, 0.75 to 1.09), the Operant (Mdn = 0.65, 0.52 to 0.81), the Observed Human (Mdn = 0.75, 0.62 to 0.84), and the Observed Robot (Mdn = 0.81, 0=65 to 1.05) conditions. A Friedman test showed a statistically significant effect on interval reproductions depending on which type of agent executed the action: $\chi^2(3) = 32.15$, $p < .001$. Planned comparisons were conducted with Wilcoxon signed-rank tests. Reported effect sizes (ES) are calculated dividing the Z statistics by the square root of the number of pairs. There were significant differences between the Control and the Observed Robot conditions ($Z = -2.088$, $p = .037$, ES = 0.37), between the Observed Robot and the Observed Human conditions (Z = -2.481, $p = 0.013$, ES = 0.45), and between the Observed Human and Operant conditions (Z = -2.825, $p = .005$, ES = 0.52). Fig. 16 shows a graphical representation of the data.

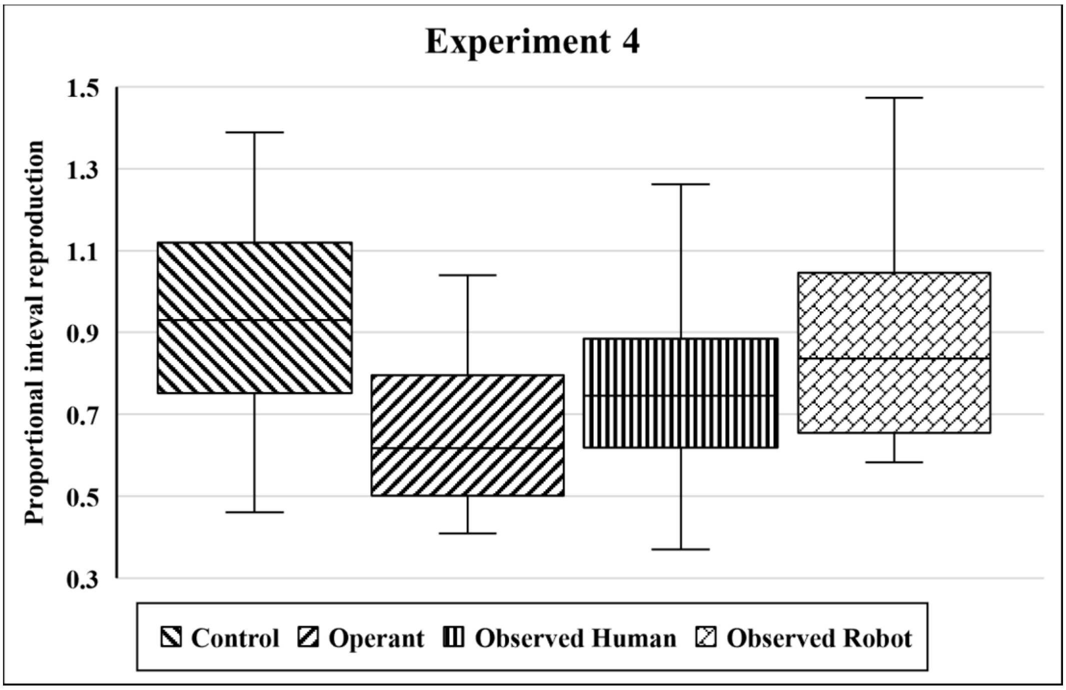


Figure 16. Boxplot comparing the mean proportional interval reproductions across the four conditions in Experiment 4.

Discussion

Experiment 4 aimed to investigate whether observing a human or robotic-generated action could induce a Temporal Binding effect. In order to do this, an interval reproduction task was adopted (Howard et al., 2016; Humphreys & Buehner, 2010; Poonian & Cunnington, 2013; Stephenson et al., 2018). Participants were requested to replicate the time interval between two events, the first of which varied by conditions (the source of the action) whereas the second remained constant (the action outcome). More specifically, participants were asked to replicate the time between an action performed either by themselves, the experimenter, or a robot, and a tone. In one additional control condition participants were asked to replicate the time between two independent events (two tones produced by the experimental apparatus).

Our findings show that when compared to the Control condition, participants consistently underestimated the time duration between actions performed by themselves and their outcomes, replicating the main binding effect known in the literature (see Haggard, 2017 for a review) and adding reliability to the interval reproduction task as an effective method to measure it. In a similar fashion, observing human-generated actions led to consistent time compression towards the outcomes, albeit with a smaller effect size compared to self-produced actions. This finding is not entirely consistent with previous research (Poonian & Cunnington, 2013; Strother et al., 2010; Wohlschläger et al., 2003) and with data reported in Experiment 1, but diverges only in the degree of binding rather that the presence or absence of it. This suggests that the temporal binding effect does not simply manifest on a dichotomous level (either it is achieved or not), but that it can be moderated on a continuous level, thus introducing a more flexible approach to its use in studies on the sense of agency.

Robotic actions also showed a significant time compression when compared to the control condition, though with a smaller effect size when compared to the human-generated actions. Interestingly, the extent to which participants experienced temporal compression in the Observed Robot condition proved to be distributed bimodally (see Fig. 15), which suggests that part of the experimental sample showed a strong temporal binding effect, whereas the rest had no differences between the Observed Robot and Control conditions. This discrepancy may be due to individual factors, such as prior belief about robots (Perez-Osorio & Wykowska, 2019). In Experiment 5 this element will be further explored, implementing a manipulation aimed at inducing different perceptions of Cozmo as either capable or incapable of intentional action.

Taken together, the findings reported in Experiment 4 do not fully support a pure motor simulation model for the explanation of temporal compression experienced over observed actions performed by another volitional agent, as some participants reported a temporal binding effect for actions performed by a robot. Yet, the data shows reliable temporal compression happening in every experimental condition when compared to the Control condition. This element suggests that while action observation and concurring motor simulation can indeed affect subjective time perception, they may not be the sole factors involved in the generation of temporal binding over observed actions. Experiment 5 will address new explicit components to include in the model (i.e., theory of mind, anthropomorphism), exploring whether explicit knowledge and attitudes can play a role in the generation of temporal compression.

Experiment 5

Experiment 4 showed that when observing a social robot acting, participants' performance in a time interval reproduction task was highly variable. The bimodal distribution observed in this condition may suggest that the generation of a binding effect is not only related to the action observation itself, but relies on the prior beliefs we have about the agent we are observing. More specifically, human and robotic actions may not differ on the sole base of the actual kinematics of the action. On the contrary, a second element possibly involved in the distinction between human and robotic actions could be individuated in the perceived volitional nature of the agent. This element could help to explain the two clusters of results in the Observed Robot condition: the temporal binding effect could have been influenced by participants' prior beliefs about the robot, as it was not specified whether it was an agent acting independently or as an artefact controlled by someone else.

Anthropomorphism models and tendencies

To further explore the pattern of results observed over the Observed Robot condition in Experiment 4, Experiment 5 aimed to manipulate participants' social perception of the robot, as well as their prior beliefs about its action dynamics. In other words, the intention was to prime participants in such a way to shift the extent to which they anthropomorphised Cozmo in polarized directions. Anthropomorphism has been described as the tendency to attribute human traits such as emotions and intentions to non-human entities (Waytz et al., 2010). Humans tend to anthropomorphise agents (and even events) that are perceived as displaying independent agency. This was true in ancient times when earthquakes were interpreted as a divine punishment, as well as today when our computers seem to freeze or run an update just in the most inappropriate moment. Previous research suggests that humans are easily triggered to provide anthropomorphic descriptions of other agents (Epley et al., 2007). As such, anthropomorphism is believed to reflect an implicit cognitive process (Mitchell et al., 1997), which is often deployed towards complex systems that are beyond our understanding, in order to render them more familiar and easily explainable (Waytz et al., 2010, Wiese at al., 2017). On the other hand, previous research has also shown that providing anthropomorphic features (e.g., name, gender, voice) to a mechanical artefact such as an autonomous vehicle, led participants to attribute it with a higher degree of mental states compared to when the same vehicle was not given such features. Additionally, participants reported higher trust scores in the anthropomorphised artefact, as well as reduced stress (measured through heart rate changes) during a simulated accident (Waytz et al., 2014). Taken together, the evidence reported above suggest anthropomorphism and the attribution of mental states are deeply interconnected, both capable of influencing and promoting the other.

Crucial to this thesis, previous research has shown that the mechanisms responsible for anthropomorphism rely on the same cognitive processes involved in the attribution of intentions to observed behaviour (Castelli et al., 2000; Iacoboni et al., 2004). Different degrees of said attribution include assumption of mental states (e.g., conscious experience, intentions, metacognition; Gray et al., 2007), emotional states, and human forms in non-human agents (Heider & Simmel, 1944; Epley et al., 2007). Given the evidence described above, Experiment 5 will directly address this topic, implementing a manipulation aimed to shift participants' perception of Cozmo (either as an artefact or as a spontaneous agent), and thus observing how prior beliefs and knowledge can affect anthropomorphism attitude towards social robots

The intentional stance

Dennett (1989) explored anthropomorphism with specific regards to the cognitive processes involved in the observation of other individuals. More specifically, Dennett hypothesised that when interacting with other people, we are likely to adopt a strategy to predict and explain their behaviour with reference to their mental states. This construct took the name of intentional stance, as opposed to the physical stance (explaining the natural events we observe through the laws of Physics), and the design stance (explaining more complex systems we observe through the experience and knowledge we have about how that system). As a unique trait, the intentional stance relies on the attribution of intentions and beliefs to a system, in order to explain and predict its behaviour.

Experiment 5 outline

Following this approach, Experiment 5 was designed to replicate the main structure of Experiment 4, but adding a between-subjects manipulation consisting in an interaction priming procedure during which participants experienced either a mechanistic or mentalistic interaction with Cozmo before the main study. The interaction priming procedure consisted of a brief interaction with the robot, in which participants could see that it was manoeuvred by the experimenter (Mechanistic condition), or acting spontaneously (Mentalistic condition). The idea beyond this manipulation was to shift participants' tendency to adopt either a design (mechanistic) stance or an intentional (mentalistic) stance, through a direct experience of the robot's behaviour. This manipulation was designed to minimise the appearance of Cozmo as autonomous in the Mechanistic condition and maximise it in the Mentalistic Condition. It was predicted that participants in the Mechanistic condition would produce a temporal compression for actions performed by themselves and the experimenter, while no differences will be observed between robotic actions and independent events. On the other hand, it was also predicted that participants in the Mentalistic condition would produce a temporal binding effect for robot-generated actions as well, to a similar extent as for self- and human-generated actions. Furthermore, given the deep connection between agency, intention attribution, and anthropomorphism (Castelli et al., 2000; Iacoboni et al., 2004), it was hypothesised that having participants who experienced Cozmo's behaviour as spontaneous (as opposed to mechanical) would also increase the extent to which they anthropomorphised and mentalised the robot.

Method

Participants

A power analysis (carried out with G*Power 3; Faul et al., 2007) indicated that to detect the same effect size observed between the Control and the Observed Human condition in Experiment 1 $d_z = 0.88$, with $1 – \beta = 0.95$ at $\alpha = 0.05$, a minimum sample size of 54 (27 per group) was required. Because the central hypothesis involved an interaction in which this target effect size would be achieved in the mentalistic condition but reduced or eliminated in the mechanistic condition, we followed Simonsohn's (2014; see also Blake & Gangestad, 2020) recommendation that sample size (per cell) should be doubled to ensure the same level of power. Therefore, 112 participants (a sample size capable of detecting an effect size $d_z = 0.69$ with $1 - \beta = 0.95$ at $\alpha = 0.05$) aged 18-39 years (M = 19.96, SD = 2.74, 91 were females), recruited from the University of East Anglia, completed the experiment. Participants were randomly allocated either to the Mechanistic or Mentalistic group. The Mechanistic group was composed by 56 participants aged 18-39 years (M = 20.14, SD = 3.28, 46 were females), whereas the Mentalistic group was composed by 56 participants aged 18 – 31 years (M = 19.79, SD = 2.05, 45 were females). Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported having normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Design

The mixed design featured both a within- and a between-subjects manipulation. While the within-subjects manipulation remained unaltered from Experiment 4, a between-subjects manipulation was added to randomly assign participants either to the Mechanistic or Mentalistic condition. The final design was a 4x2 factorial, with four within conditions (Control / Operant / Observed Human / Observed Robot) and two between conditions (Mechanistic / Mentalistic). Block order on the within manipulation was pseudorandomized across participants.

Apparatus and materials

The same apparatus and laboratory setting from Experiment 4 was retained. In order to provide an effective manipulation check, two questionnaires were developed to assess both participants' tendency to adopt either a design or intentional Stance, and their propensity to anthropomorphise Cozmo.

The first questionnaire was modelled on the InStance questionnaire (Marchesi et al., 2019), and each item depicted a scenario along with two interpretations of the scenario marking endpoints of a bipolar scale (see Fig. 17 for an example). A total of five scenarios were created, depicting Cozmo interacting with objects and/or humans, where each scenario was composed of three pictures (size 1080 x 920 pixels). Accompanying each scenario were two descriptions corresponding to different interpretations of the scenario. One of the interpretations explained Cozmo's actions by referring to the design stance (i.e., mechanistic explanation), whereas the other described Cozmo's actions by referring to the intentional stance (i.e., mentalistic explanation). Mechanistic and mentalistic interpretations were equally likely to appear either on the left or on the right side of the scale. As a further control procedure, Cozmo's emotional facial expression display was maintained neutral across the scenarios, in order to avoid biases towards mentalistic explanations. Participants' responses were recorded through the slider, which featured a hidden bipolar 0-100 scale. The value 0 corresponded to a complete mechanistic explanation (design stance) whereas 100 corresponded to a complete mentalistic explanation (intentional stance). The null value of the scale (i.e., the value in between the two) was set to 50 and as the initial position of the cursor. The complete InStance questionnaire is included in Appendix B.

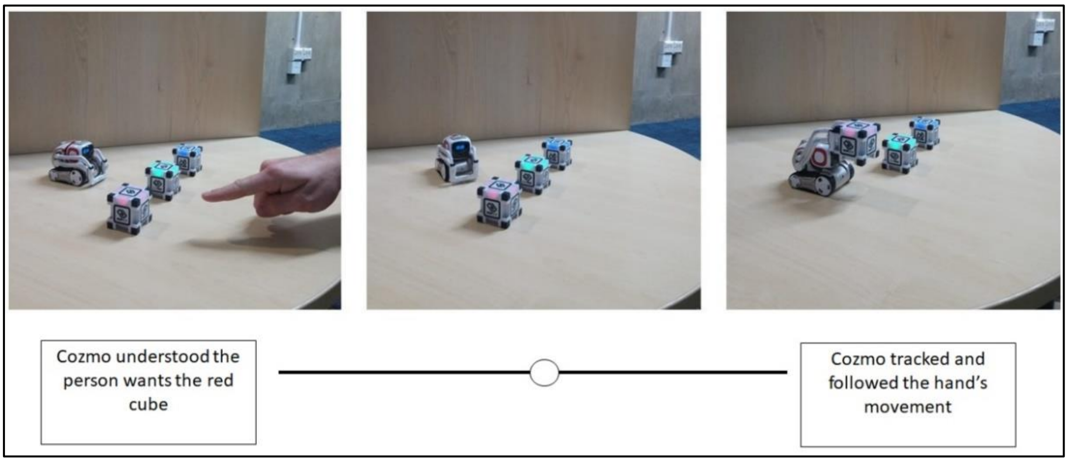


Figure 17. An example of an InStance questionnaire item. Participants were asked to move the slider towards the description they found more plausible.

A second questionnaire was created to assess participants' tendency towards anthropomorphism, that is the extent to which they referred to the robot's components by adopting humanoid words. This questionnaire depicted a digitalised sketch of Cozmo, with empty labels linked to four of its physical features (see Fig. 18).

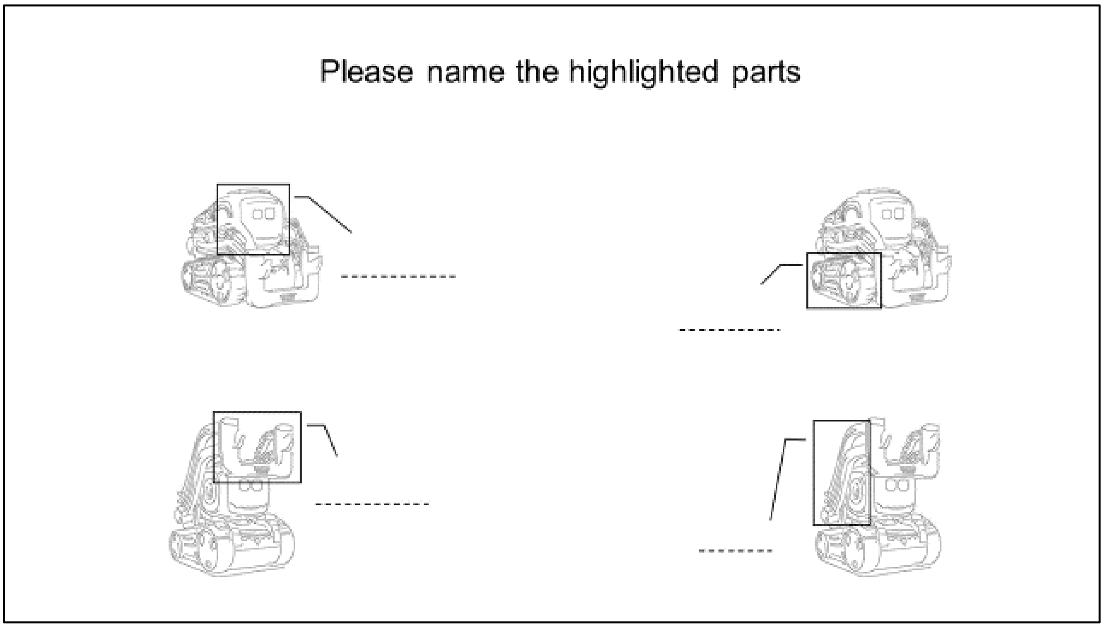


Figure 18. Anthropomorphism questionnaire developed to assess participants' anthropomorphism attitudes towards Cozmo.

Participants' mentalizing skills towards each acting agent were measured with the mind attribution scale (Kozak, et al., 2006), which featured a Likert measure ranging from 1 (strongly disagree) to 9 (strongly agree), composed by ten items that question about the explicit knowledge about others' mental states (i.e., emotion, intention, and cognition). The complete mind attribution scale can be found in Appendix A).

Procedure

Prior to the beginning of the experiment, participants were primed with an interaction with Cozmo based on the group they were assigned to. This activity was the same, but in one group Cozmo was visibly controlled by the experimenter in front of the participants (Mechanistic group) while in the other Cozmo was programmed such that it appeared to act spontaneously (Mentalistic group). The interaction between participants and Cozmo was structured as a brief session of a keep-away game. Participants were asked to slowly slide a cube towards Cozmo, which was ready to tap it from the top. Participants' task was to get the cube as close as possible to Cozmo, and quickly retreat when the robot attempted to tap it. Hence, a successful tap from Cozmo would have meant a point lost for the participants, while a miss would have meant a point gained. Points had no consequences for the compensation participants received. Each priming interaction consisted of ten trials. Notably, in the Mentalistic group Cozmo was programmed to display a brief emotional animation (cheerful or disappointment) according to the win/loss performance. After the priming procedure was over, participants completed the same experimental procedure as described in Experiment 4. The only difference was that in the Mechanistic condition, the experimenter remained in the laboratory during the Observed Robot condition, seated 160 cm 45° South-West away from the participants, as he was controlling the robot. In the Mentalistic condition, the experimenter also controlled Cozmo, but dis so from outside the laboratory in order to preserve participants' illusion of Cozmo acting spontaneously. At the end of each experimental block, participants completed a mind attribution scale (Kozak et al., 2006) which was adapted to refer to the specific agent that acted in that block (i.e., themselves, the experimenter, Cozmo, or the experimental apparatus). Before they were dismissed, participants further completed the InStance questionnaire and the anthropomorphism questionnaire.

Results

The same exclusion criteria of Experiment 4 were adopted for Experiment 5. This resulted in 2 participants excluded in the Mechanistic group and 5 in the Mentalistic group, with respectively 102 and 107 trials deleted (1.52% and 1.59% of total trials).

Proportional interval reproductions

Mean Proportional Interval Reproductions were calculated for each participant and submitted for statistical analysis. Shapiro-Wilk tests were performed on each condition to test normality of each distribution, and no condition violated normality assumptions ($p > .05$). A mixed ANOVA was conducted on proportional interval reproduction scores. As sphericity assumptions were not met, results are reported adopting a Greenhouse-Geisser correction. There was a significant main effect for the acting agent: $F(2.42, 244.47) = 62.06$, $p < .001$, $\eta_p^2 = .381$. A significant interaction effect between acting agent and interaction priming was also found: $F(2.42, 244.47) = 6.05$, $p < .001$, $\eta_p^2 = .057$.

Planned paired-samples t-tests revealed that in the Mechanistic group the mean proportional interval reproduction in the Observed Robot condition (M = 0.98, SD = 0.22, 95% CI [0.92, 1.04]) was significantly higher than the Observed Human condition (M = 0.78, SD = 0.20, 95% CI [0.72, 0.83], $t(53) = 7.76$, $p < .001$, $d_z = 1.045$), while no difference was found between the Observed Human and the Operant condition (M = 0.75, SD = 0.20, 95% CI [0.70, 0.81], $t(53) = 1.18$, $p = .242$). Similarly, no significant difference was found between the Control (M = 0.99, SD = 0.27, 95% CI [0.91, 1.06]) and the Observed Robot condition: $t(53) = 0.24$, $p = .808$.

On the other hand, in the Mentalistic group the mean proportional interval reproduction in the Control Condition (M = 1.01, SD = 0.25, 95% CI [0.94, 1.08]) was significantly higher than the Observed Robot condition (M = 0.85, SD = 0.20, 95% CI [0.80, 0.91], $t(48) = 3.93$, $p < .001$, $d_z = 0.57$), which was in turn higher than the Observed Human condition (M = 0.78, SD = 0.17, 95% CI [0.73, 0.83], $t(49) = 3.18$, $p = .003$, $d_z = 0.41$). No significant difference was found between the Observed Human and Operant condition (M = 0.77, SD = 0.19, 95% CI [0.72, 0.83], $t(50) = 0.35$, $p = .730$).

Planned independent-samples t-tests showed a significant difference only in the Observed Robot condition: $t(102) = 2.60$, $p = .011$, $d_z = 0.62$. Fig. 19 shows a graphical representation of the data.

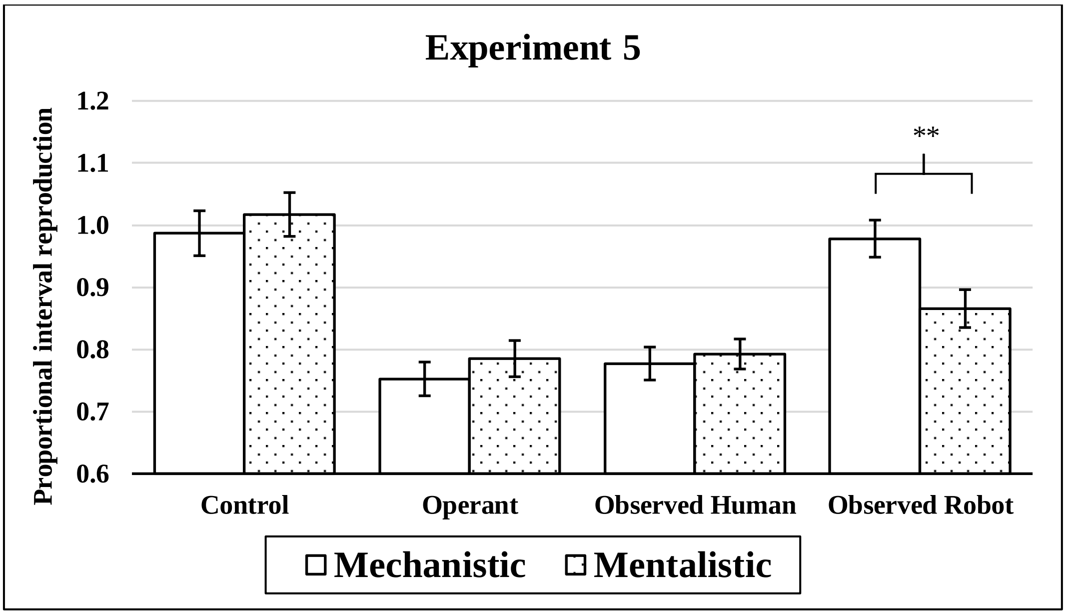


Figure 19. Mean proportional interval reproduction for the four conditions across the two groups in Experiment 5. Error bars represent the standard error of the mean. Significant differences are highlighted for the .05 (*) and .01 (**) level.

Mind attribution scale

The mind attribution scale (Kozak et al., 2006) score was computed by averaging the scores of all items for each participant. A mixed ANOVA was conducted on proportional interval reproduction scores. As sphericity assumptions were not met, results are reported adopting a Greenhouse-Geisser correction. There was a significant main effect for the acting agent: $F(1.65, 169.85) = 692.74$, $p < .001$, $\eta_p^2 = .871$. A significant interaction effect between acting agent and interaction priming was also found: $F(1.65, 169.85) = 17.79$, $p < .001$, $\eta_p^2 = .147$. An independent samples t-test was found significant only on the Observed Robot condition between the Mechanistic (M = 2.31, SD = 1.38, 95% CI = [1.93, 2.68]) and the Mentalistic group (M = 4.32, SD = 1.93, 95% CI [3.78, 4.87], $t(90.16) = 6.14$, $p < .001$, $d_z = 1.20$). See Fig. 20 for a graphical representation of the data.

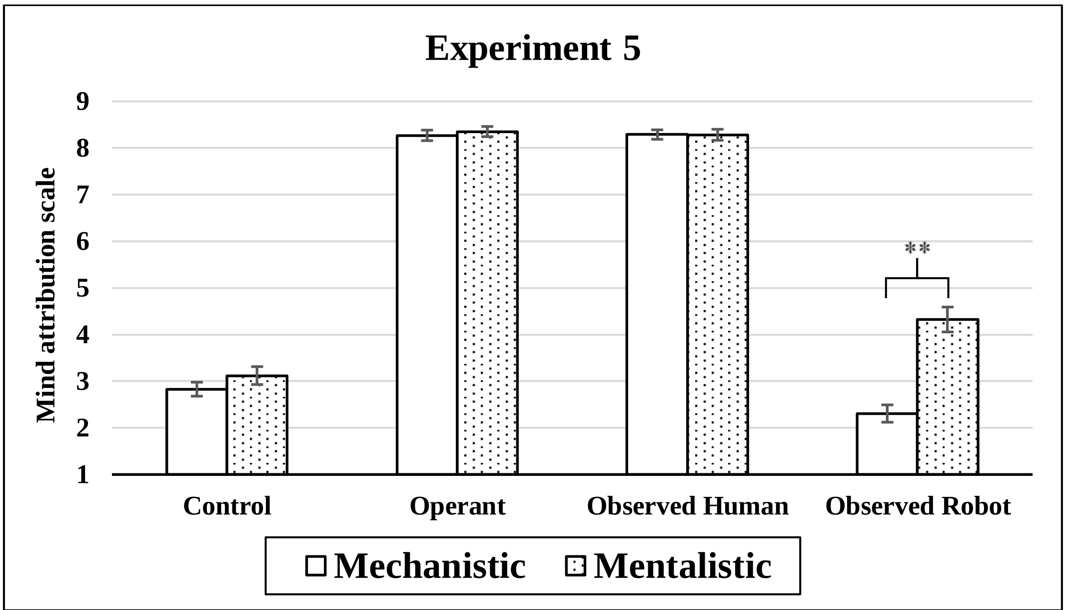


Figure 20. Mind attribution scale scores across the for the four conditions across the two groups in Experiment 5. Error bars represent the standard error of the mean. Significant differences are highlighted for the .05 (*) and .01 (**) level.

InStance questionnaire

The InStance questionnaire was scored converting the bipolar scale into a 0 – 100 scale, in which 0 corresponded to a complete mechanistic explanation (design stance) whereas 100 corresponded to a complete mentalistic explanation (intentional stance). The final InStance questionnaire score was computed as the mean of all items for each participant. Average scores below 50 indicated the propensity to adopt a design stance, whereas scores above 50 were in favour of an intentional stance. An independent samples t-test found a significant difference between the Mechanistic (M = 24.10, SD = 15.97, 95% CI = [19.74, 28.46]) and the Mentalistic group (M = 52.78, SD = 24.40, 95% CI [45.92, 59.64], $t(85.45) = 7.08$, $p < .001$, $d_z = 1.40$). See Fig. 21 for a graphical representation of the data.

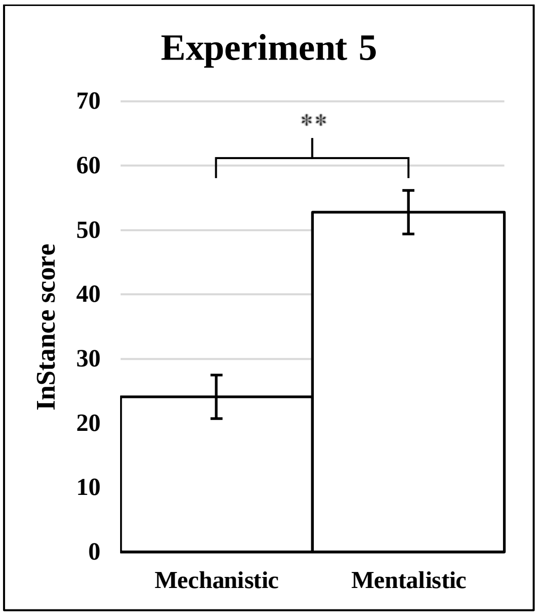


Figure 21. Average InStance score between the Mechanistic and Mentalistic groups. Error bars represent the standard error of the mean. Significant differences are highlighted for the .05 (*) and .01 (**) level.

Anthropomorphism questionnaire

The anthropomorphism score was calculated by first coding each label participants assigned to either a human or mechanical category. Words associated with human features (e.g., face, hands, legs) were operationalised as ‘1’, while words associated with mechanical features (e.g., screen, grabber, wheels) were operationalised as ‘0’. Ambiguous words, valid for both the human and mechanical domain (e.g., arm) were always considered as mechanical. The final anthropomorphism score was calculated by averaging values on all labels for each participant. The final score ranged from 0 to 1, with higher values representative of greater anthropomorphism tendency. An independent samples t-test found a significant difference between the Mechanistic (M = 0.35, SD = 0.22, 95% CI = [0.29, 0.41]) and the Mentalistic group (M = 0.65, SD = 0.20, 95% CI [0.60, 0.71], t(103) = 7.32, p < .001, dz = 1.44). See Fig. 22 for a graphical representation of the data.

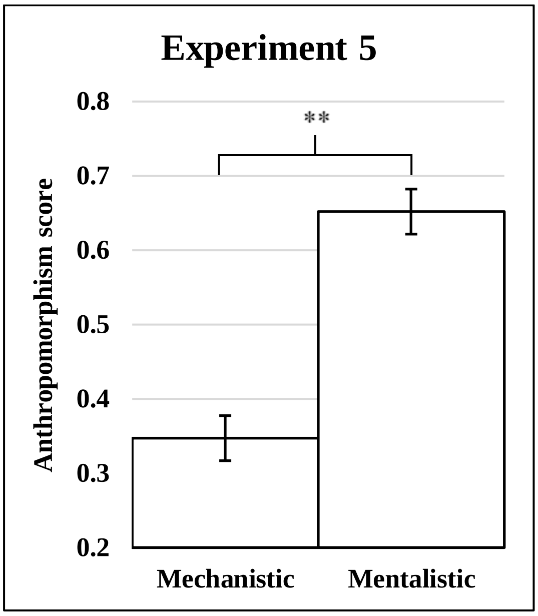


Figure 22. Average anthropomorphism score between the Mechanistic and Mentalistic groups. Error bars represent the standard error of the mean. Significant differences are highlighted for the .05 (*) and .01 (**) level.

As exploratory analysis, a correlation matrix was created between all explicit measures (i.e., mind attribution scale, InStance score, anthropomorphism score) and temporal binding instances in the Observed Robot condition. Results showed a significant weak correlation between the mind attribution scale and the anthropomorphism score, but only in the Mentalistic group: $r(49) = .34$, $p = .016$. Nevertheless, it is worth noting that since this exploratory analysis was not planned, it may be underpowered to detect any significant interaction. Additionally, results from this analysis may be biased by the direct comparison of measures obtained with different methods.

Discussion

Experiment 5 aimed to shed light on the findings reported in Experiment 4. Namely, to determine which features lead an observed robotic action to produce a temporal compression towards its outcome. Findings from this experiment show that when compared to the Control condition, participants reliably compressed the time duration between both self- and observed human-generated actions and their outcomes, producing binding effects of comparable sizes. This result is consistent with previous research that investigate observed human actions (Poonian & Cunnington, 2013; Strother et al., 2010; Wohlschläger et al., 2003) and reinforces the idea that directly observing other people acting in the environment generates a subjective time compression effect.

Even more interestingly, the nature of the social interaction that participants experienced with Cozmo did have an effect on the extent to which they experienced subjective temporal compression in relation to Cozmo's actions. More specifically, while participants in the Mechanical group showed no differences between the Control and the Observed Robot conditions, participants in the Mentalistic group reliably underestimated the time elapsed between Cozmo's actions and their outcomes. Notably, the achieved effect size was smaller when compared to self- and human-generated actions.

Mentalizing measures were also included, showing significant differences according to the nature of the interaction priming. Participants' explicit reports in the mind attribution scale remained unaltered between the Mechanistic and Mentalistic groups for the Control, Operant, and Observed Human conditions, but differed significantly in the Observed Robot condition. More specifically, participants in the Mentalistic group reported higher scores than those in the Mechanistic group, suggesting that an explicit belief about Cozmo being an autonomous being can increase the likeliness to attribute it mental states. Similarly, the InStance questionnaire results showed that participants in the Mentalistic group were more likely to explain Cozmo's behaviour adopting an intentional stance, thus ascribing its actions to a "mind" to a greater extent than did participants in the Mechanistic group. The anthropomorphism questionnaire showed analogous results, suggesting that the likelihood of attributing human physical features to robots is also influenced by the explicit understanding we have of them. These results fit well with previous literature (Castelli et al., 2000; Iacoboni et al., 2004; Waytz et al., 2014), providing further evidence that agency, intention attribution, and anthropomorphism are deeply interconnected.

Taken together, these findings indicate that the different experiences we have with robots can alter our social perception and engagement with them, both on an explicit and implicit level. Findings from Experiment 5 contribute to explaining the mixed results observed in Experiment 4 in the Observed Robot condition. In Experiment 4, participants were not given any information about the nature of the functioning of Cozmo, thus leaving them free to assume that information by themselves, possibly biasing it according to their prior beliefs and knowledge about robots. This factor resulted in participants showing a non-normal bimodal distribution, where part of the sample did experience time compression, whereas the other did not. In Experiment 5, inducing participants to adopt either a mechanistic or mentalistic approach towards the robot resulted in different performances in the interval reproduction task, as temporal binding was only observed when participants interacted with Cozmo as a spontaneous agent before the experiment. This evidence suggest that temporal compression does not solely rely on action observation as it is, but higher-level mechanisms do indeed modulate the temporal binding effect. Furthermore, different effect sizes reported in the different experimental conditions suggest again that temporal compression is not a dichotomous effect (i.e., either we experience it or not), but it can move on a continuous scale, thus providing a more sensitive measure for the studies on the sense of agency.

Chapter discussion

Chapter 3 investigated the cognitive mechanisms underlying the dynamics of human-robot interaction, and how they are related to the more common interaction between humans. More specifically, we posed the question of whether observing actions performed by a social robot (Cozmo by Anki) leads participants to experience a subjective time compression towards the outcomes of those actions. Throughout two experiments, data showed that observing robotic actions produced a temporal binding effect, and also provided evidence that higher-level mechanisms may be involved. These findings are consistent with previous research investigating the sense of agency in relationship to the human-robot interaction (Ciardo et al., 2018; Roselli et al, 2019), which suggested that both explicit agency perceptions and temporal compression for self-generated actions is weakened when a robot was also capable of acting in the environment compared to when participants were acting alone. Indeed, if participants felt less responsible for the action outcome when a robot was acting, it is implied that some of that responsibility is attributed to the robot. However, in the work presented here, when participants received no information about the robot, and thus they were free to interpret its actions according to their prior beliefs (Experiment 4), their performance in an interval reproduction task were highly variable, revealing a bimodal distribution. This result showed that some participants did experience temporal compression (just as they did for human-generated actions), while others were accurately replicating the time intervals between robotic actions and their outcomes (just as they replicated the interval between two computer-generated tones). Findings from Experiment 4 suggest that more than one mechanism may be involved in the generation of a binding effect for observed actions. Based on the evidence reported jointly with Chapter 2, it is possible to claim that action observation is a necessary (as suggested by Poonian & Cunnington, 2013; Poonian et al., 2015; Strother et al., 2010; Wohlschläger et al., 2003), but not sufficient condition to achieve a reliable time compression. Thereafter, the implications on the temporal binding effect and the wider sense of agency construct acquire a more complex dimension.

Explicit beliefs and agency detection

In Experiment 5, the prior belief participants had about the robot was systematically manipulated, priming participants with wither a mechanistic or mentalistic interaction with the robot. The aim of the study was to polarize the pre-existing attitude participants had towards the robot, inducing them to see it as an artefact controlled by the experimenter or more as a spontaneous agent. This manipulation proved to be effective, as participants in the Mentalistic group produced a reliable temporal binding effect, whereas those in the Mechanistic group showed no differences between the Observed Robot and Control condition. In other words, when Cozmo was seen an extension of the experimenter for the purpose of action production, it did not lead participants to experience time compression. This finding can be confronted to a fMRI study from Özdem et al. (2017), who reported increased brain activity in areas typically associated with the generation of a theory of mind (such as enhanced activation of the bilateral anterior temporoparietal junction), but only when participants believed that another person was controlling the robot. Taken together, these findings suggest that agency detection and mentalising processes are indeed deeply interconnected, but clearly dissociable. On the other hand, participants from Experiment 5 appeared to detect agency for Cozmo when it was experienced as a spontaneous agent. This innovative finding suggests that temporal compression is not generated for every action observed, but is rather influenced by top-down mechanisms operating an explicit filter of information we have about the agent who is performing the action. In addition, data from Experiment 4 suggests that if that information is not available, prior beliefs and knowledge may fill that gap, as suggested by the fact that some participants did spontaneously show temporal compression, while others did not.

Lastly, when addressing the core mechanisms driving temporal binding effects, findings from Chapter 3 corroborate results of Chapter 2, which was in contrast with previous research claiming that temporal compression is experienced when two events are linked by a cause-outcome relationship (Buehner, 2012; Buehner & Humphreys, 2009). In fact, as even though in Experiment 5 the causal properties of the action were preserved in the Observed Robot condition for the Mechanistic group, reliable time compression was not achieved.

Agency, intention attribution, and anthropomorphism

Additional findings from Experiment 5 showed that the two experimental groups revealed differences also in explicit theory of mind measures. More specifically, participants who experienced Cozmo as a spontaneous agent were more likely to attribute mental states to it, to explain its behaviour as a result of mental activities (i.e., adopting an intentional stance), and to refer to its components in human terms (e.g., face, hands). On the contrary, participants who saw Cozmo as an artefact tended to explain its behaviour in terms of its function (i.e., design stance), and referred to its components adopting more mechanical terms (e.g., screen, wheels). Taken together, this evidence seems to suggest that the temporal binding is not only deeply related to the sense of agency, but also with anthropomorphism and theory of mind skills. This is consistent with the models proposed by Castelli et al. (2000) and Iacoboni et al. (2004), who suggested that anthropomorphism mechanisms rely on the same cognitive processes involved in the attribution of intentions. To further investigate this line of research, more work could be done exploring this connection. For example, future work could compare subjective time compression for observed actions between specific populations known to be impaired in mentalizing skills (e.g., autistic individuals) and controls.

Critical to this study is the lack of a specific method to address metalizing skills for artificial agents. Although a validated questionnaire such as the mind attribution scale (Kozak et al., 2006) was adopted, this was not initially developed to address mentalising towards artificial agents. By consequence, a few items could have been misinterpreted by participants when referring to artificial agents. For example, an item stated "__________ has good memory", with the premise that greater memory is associated with greater cognition. However, while this is true for human agents, when referring to computerised and robotic entities, memory can be widely extended with little to no implication towards their cognition. To overcome this limitation, future research could focus to develop a method specifically aimed to assess perceived mental states and cognitive ability towards computerised and artificial agents.

Relatedly, the original InStance questionnaire (Marchesi et al., 2019) served the purpose of providing a method specifically dedicated to human-robot interaction, but depicted a different robot (i.e., iCub) and thus was not flexible enough to be adopted in its original form for the purpose of this thesis. To the author's knowledge, an explicit measure to assess mentalising for artificial agents has not yet been developed. Further research could focus on validating a flexible tool to assess this specific domain, as the emerging field of research on human-robot interaction would gain immense benefits from it.

Explicit and implicit cues in temporal compression

Taken together, the findings discussed above suggest that the temporal binding effect for observed actions is ascribable to more than one mechanism, including both implicit and explicit processes. Data from these studies provided evidence that mere action observation is not sufficient to achieve reliable time compression. Instead, explicit knowledge about the observed agent's mental states is also needed. These data would fit a theoretical model in which the Action Observation Network (Gallese & Goldman, 1998) and the Person Perception Network (Quadflieg et al., 2011) both concur to generate subjective time compression. Accordingly, the Action Observation Network could account for implicit and low-level mechanisms, while the Person Perception Network is dedicated to explicit and higher-level social cues. Further neuroimaging research could focus on these two networks, investigating their engagement and the involvement of the theory of mind network (Péron et al., 2010) on the observation of actions executed by humans and social robots, with specific attention to the effect produced by the display of social engagement and reward from those robots. This model describes a deeper connection between implicit and explicit cues involved in agency detection. In fact, while Synofzik et al. (2008) suggested a sharp distinction between judgements of agency (driven by explicit cues) and feelings of agency (driven by implicit cues), this thesis advances the idea that explicit cues may be relevant to the generation of feelings of agency as well.

Addendum to Chapter 3

Neural correlates of agency in action observation

As discussed in Chapter 1, previous research investigating the neural mechanisms associated with actions and their outcomes has mainly focused on two separate EEG components. The first component (pre-action readiness potential) inspects the neural activity which leads to voluntary actions, representing action planning and preparatory activity before the initiation of a movement. In EEG paradigms where self-generated actions are executed, a slow-wave negative potential arises approximately 2000 ms prior to the movement, reaching its peak on the movement onset (Libet et al., 1983; Deecke et al., 1969). The second component (N1) examines the neural processing of outcomes resulted from self-generated actions (Aliu et al., 2009; Martikainen et al., 2005). These components have been measured in action-outcome tasks both to investigate the discrete processing of actions and outcomes, and also to assess the contribution that predictive mechanisms exert during agency detection.

Previous studies investigating outcome prediction and sense of agency have mainly focused on the N1 component of the sensory ERP. The N1 component is a negative waveform, generally occurring around 80-120 ms after the presentation of an auditory stimulus. Sensory outcomes resulting from self-generated actions typically show decreased amplitude in the N1 component (Martikainen et al., 2005), both for auditory outcomes (Baess et al., 2011; Knolle et al., 2013; Lange, 2011) and visual outcomes (Hughes & Waszak, 2011; Gentsch & Schutz-Bosbach, 2011). Interestingly, N1 amplitudes have been shown to further decrease when the action outcome is predictable, as opposed to when it is not (Gentsch et al., 2012). On the other hand, N1 amplitudes are reduced to a lesser extent when the action is produced under coercion, suggesting that reduced experience of responsibility towards the outcome is associated with decreased N1 suppression (Caspar et al., 2016). Such decreases in N1 amplitudes are believed to occur as a result of top-down predictions from the motor system, which suppress the activity of the auditory cortex (Aliu et al., 2009; Blakemore et al., 2002). Several authors have therefore inferred that N1 suppression during action-outcome tasks reflects the activity of an internal model of action, which predicts the outcomes that result from our actions (Baess et al., 2008; Hughes et al., 2013).

Not many studies have investigated whether the N1 suppression that occurs when an action is self-produced also occurs when an action is observed as produced by another agent. Kuhn et al. (2011) had participants produce an action that in turn generated a tone, and in half of the trials participants were induced to believe that the tone was produced by someone else's action. Results showed no significant difference in N1 suppression over tones resulting from actions judged as self-produced, when compared to actions judged as other-produced. Observing a genuine action, namely lip movements releasing auditory speech sounds, resulted in suppression of the N1 component (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). This evidence is consistent with data shown in Chapter 2, providing further strength to the claim that indirect social actions (while not capable to directly affect the physical environment) are associated with agency markers, thus revealing a social domain of the sense of agency. Analogous results have been reported for observation of other classes of actions, such as clapping or tapping an object (Stekelenburg et al., 2013).

It has been suggested that N1 suppression in action observation occurs as a result of internal predictions over the sensory outcome of the observed action (Arnal et al., 2009; Stekelenburg & Vroomen, 2012). Poonian et al. (2015) investigated the connection between N1 suppression and temporal compression over self-produced and other-produced observed actions. Their results showed N1 suppression over auditory stimuli resulting from both classes of actions (self-executed as opposed to observed), with no significant differences in the N1 amplitude between the two. This effect was mirrored by the temporal compression that participants experienced, showing a temporal binding effect both for executed and observed actions. This finding supports the studies discussed above, and was interpreted as evidence that similar cognitive processes are involved in agency detection for both self-produced and observed actions. Poonian et al. (2015) have further suggested that N1 suppression occurs over outcomes produced by an action (self- or other-produced) as a result of an implicit sense of agency generated whenever that action (executed or observed) affects the environment.

N1 suppression and observed robotic actions

Given the premises discussed above, and considering the results from Chapter 3, the doctoral program initially included further explorations to investigate the neural mechanisms involved in agency detection in relation to observed robotic actions. More specifically, a study was designed to investigate whether observed robotic actions (when the robot is experienced as a spontaneous agent which synchronises its movements to those of the participant's) also lead to N1 suppression associated with the auditory outcomes produced by those actions. However, UK and UEA efforts to prevent further spread of the COVID-19 infection resulted in national lockdowns and laboratory closures, which made any EEG data impossible to collect.

To run this planned experiment, a set of three controlled video stimuli was created, each showing a different class of action. All videos depicted a white cube, which was approached either by a human hand or Cozmo. The agents performed an action (i.e., tapping the cube) and then exited the scene. For the control condition, no action was shown, but the cube flashed a green light. These three experimental conditions (i.e., observed human, observed robot, and control) would then be added to an operant condition where a key on the keyboard was pressed by participants, as they completed an interval reproduction task. It was predicted that participants primed with a mentalistic interaction (i.e., experienced motor synchronisation) with Cozmo would have shown significant N1 suppression, compared to participants who experienced the robot as an artefact non-responsive over their movements.

Method

Participants

A power analysis (carried out with G*Power 3; Faul et al., 2007) indicated that to detect the same a large effect size (as reported by Poonian et al., 2015) d_z = 0.96, with $1 - \beta = 0.95$ at $\alpha = 0.05$, a minimum sample size of 50 (25 per group) was required. Therefore, it was planned to recruit a total of 50 participants (a sample size capable of detecting an effect size $d_z = 1.04$ with $1 - \beta = 0.95$ at $\alpha = 0.05$) from the University of East Anglia.

Design

The mixed design featured both a within- and a between-subjects manipulation, keeping unaltered the design from experiment 5. This resulted in a 4x2 factorial, with four within conditions (Control / Operant / Observed Human / Observed Robot) and two between conditions (Mechanistic / Mentalistic). Block order on the within manipulation was set to be pseudorandomized across participants.

Apparatus and Stimuli

For all conditions, except the Operant condition, a set of three video stimuli was created using Adobe Premiere Pro CS6 (Adobe Systems Software Ltd, Dublin, Ireland). The video stimuli were composed by 41 bitmap images presented in rapid succession (25 frames/s), for a total duration of 164 ms. Throughout the videos, a white cube was shown on a plane surface (Frame 1 and 41). For the Observed Human and Observed Robot conditions, the video sequence involved a right hand or Cozmo appearing from the top-right corner of the screen (Frames 2-22), reaching over and tap the cube (Frames 23-25), then retreating backwards and exiting from the top-right corner of the screen (Frames 26-40). For the Control condition, the cube was presented static for the whole duration, except for the action sequence (Frames 23-25) where the cube was flashing a green light (See Fig. 23 for a graphical representation of the video sequence).

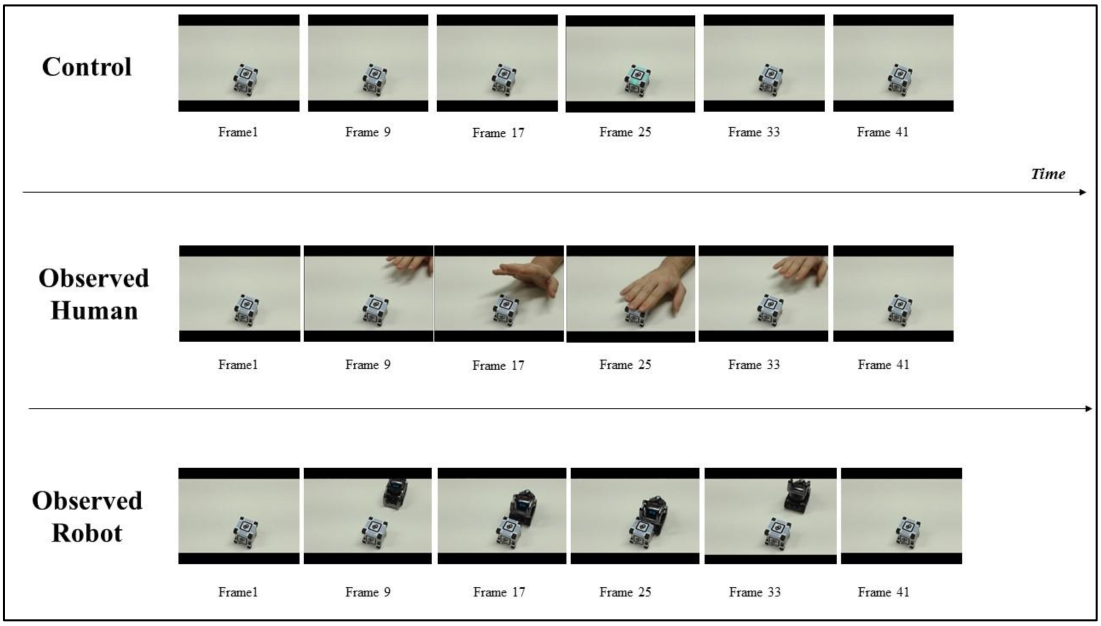


Figure 23. Graphical representation of the video sequence used in the stimuli created for each of the observed conditions.

In order to provide increased control on the stimuli presentation, participants would have rested their head on a chinrest placed approximately 60 cm away from a 24’’ computer monitor (resolution 1920x1080 pixels). To record participants behavioural responses, a standard computer keyboard was available.

EEG acquisition

A 64-channel active electrode system (Brain Products GMbH) with a cap (BrainCap-64 channels) and an amplifier (BrainAmp MR 64 PLUS) was available for EEG acquisition. EOG recordings were planned to be measured using an electrode positioned below the left eye for vertical EOG. Recording of the reference mastoid would also have been taken.

Procedure

Prior to the beginning of the experiment, participants would have been primed with an interaction with Cozmo based on the group they were assigned to. This activity would have been the same, but in one group Cozmo would have been responsive to participants’ movements following them (Mentalistic group) while in the other Cozmo would have not follow participants’ movements (Mechanistic group). The interaction between participants and Cozmo was structured as a brief session of a memory game. Cozmo would have been positioned in front of the participant, and three cubes (displaying red, blue, and green lights) would have been orthogonally aligned between the robot and the participant. As the memory game started, the cubes would flash their lights in a specific order (e.g., red-green-red-blue-green-blue). Participants' task would have been to tap the cubes in the same order they flashed they lights. For the Mentalistic group, Cozmo would have followed each tap participants executed, looking at each cube that was tapped just after the movement was executed. For the Mechanistic group, Cozmo would have not follow participant's actions, but just flash a led light on the top of its screen on the same colour of each cube tapped. After the priming procedure, participants would have completed the experimental procedure.

In the Operant condition, the trial procedure was planned to remain unaltered from Experiment 4 and 5. Throughout all three observe conditions (Control, Observed Human, and Observed Robot), the respective video stimulus would have been played, followed by a prolonged presentation of Frame 1 (randomised between 1500 and 2000 ms). The target time interval (randomised between 500 and 1500 ms) would have started at the end of the action sequence (Frame 25), after which an auditory tone (100 ms, 1 KHz sine wave, sample rate: 44100 Hz, bitrate 16) would have been played. Participants would then be asked to reproduce the duration of the interval between the action (or the flash in the Control condition) and the tone by pressing and holding down the 0 key on the number pad for the same amount of time. Each block would have consisted of 60 trials.

Conclusions

Chapter 3 aimed to investigate the cognitive mechanisms underlying agency detection during human-robot interaction, and how they are related to the more common interaction between humans. The findings discussed above are consistent with previous literature, suggesting that our cognitive abilities (e.g., agency detection, action perception, decision-making, spatial mapping, joint attention) can be affected by the presence or the interaction with a social robot (Bainbridge et al., 2008; Kompatisari et al., 2018; Shinozawa et al., 2005). Innovative insight from this thesis suggests that the mere presence of the robot could not be sufficient by itself to alter our cognition. On the contrary, explicit belief and expectations about the robot may play a critical role. Chapter 4 will expand the domain of this investigation, addressing agency detection in function of action representation processes. The aim is to provide a more comprehensive theoretical framework to the understanding of human-robot interaction, which includes both cognitive and motor effects.

Chapter 4

Action representation, social contexts, and embodiment

"Embodiment means we no longer say, I had this experience.

We say, I am this experience."

Sue Monk Kidd, 1987

Producing cognitive activity requires a cost, both in terms of physical energy (i.e., increased neural metabolism) and evolutionary effort. In a phylogenetics view, these costs are justified for the purpose of acting on (or interacting with) the environment. Evidence to support this claim can be found in the *megalodicopia hians*, a tunicate also known as “ghostfish” thanks to its pale white and semi-transparent texture. In the first phase of its life cycle, the ghostfish looks and behaves much like a tadpole. It has indeed a very simple nervous system, yet sufficient to grant survivability by the means of perception and action. However, when reaching the second phase of its life cycle, the ghostfish will drop all its limbs and appendices, to anchor itself to an underwater rock for the rest of its life. Interestingly, following this anchoring, its nervous system shrinks and eventually disappears, turning the ghostfish into an underwater venus flytrap. While still capable of surviving by digesting naïve prey, the ghostfish loses ability to produce cognitive activity: without action, it is not worth it. The ghostfish example suggests that our nervous system evolved to grant action possibilities within our environment.

Indeed, some authors agree that if something is capable of spontaneous movement, they will necessarily be able to produce cognitive activity (Llinás, 2002). Thus, our complex cognitive systems evolved to allow predictive interactions between mobile creatures and the environment. To securely navigate the environment, we must anticipate the outcome of each action on the basis of incoming sensory information. As discussed in Chapter 1, agency detection is not the only cognitive mechanism that can be used to assess intentionality and volition of our actions. In fact, the relationship between cognition, action, and the environment is believed to be driven by cognitive constructs known as affordances (Gibson, 1979). Affordances are mental representations of motor acts associated with specific objects (Tucker & Ellis, 1998). For example, a pen affords a specific action, that is a precision grip grasping. Affordances were initially theorised to reflect environmental offerings as perceived by a sentient being, but in more recent times the scientific community has adopted the affordance construct to indicate action-oriented representations and dispositions (Sakreida et al., 2016). Consequently, research on affordances investigating the relationship between cognition and environment has gained crucial relevance to understand how actions are computed and represented.

Action representation and motor facilitation

Recent evidence has shown that perceived affordances can affect our motor system, indicating that seeing an object directly triggers the motor representation of appropriate action possibilities (Tucker & Ellis, 2004; Vainio et al., 2007). Furthermore, research findings have shown that task-irrelevant information (e.g., the right-left orientation of a handled mug) may potentiate (i.e., facilitate) the execution of right-left actions, if the orientation of the affording component of the object (e.g., the handle) is spatially aligned with the hand participants moved (Tucker & Ellis, 1998). This motor facilitation effect, also known as spatial alignment effect, is intended as a decrease of RTs when participants perform an action that is congruent with the action afforded by the object seen (Bub & Masson, 2010). These behavioural findings were later supported by neurophysiological data that showed how specific parieto-frontal networks are dedicated to encode observed objects in relation to action possibilities (Buccino et al., 2009; Chao & Martin, 2000; Grafton et al., 1997; Grèzes et al., 2003).

In a more recent study, Costantini et al. (2010) used the spatial alignment paradigm to investigate the possibility of an object (i.e., a handled mug) to afford a congruent action (i.e., a hand grasp with a precision grip) in relation to its spatial position from the participants. They instructed participants to execute a precision grasping action with their right or left hand as soon as a task-irrelevant stimulus (i.e., a handled mug) was presented. The affording component of the object (i.e., the handle) could trigger an action representation either congruent with the executed action (e.g., right handle grasped with the right hand) or incongruent (e.g., right handle grasped with the left hand). Crucial to this study, the mug could have been presented either within or outside participants' reachable space (See Fig. 24 for an example of the task-irrelevant stimuli). Results showed that a motor facilitation effect occurred in congruent actions only when the mug was presented within participants' reachable space, and interpreted this finding as evidence that affordances are not intrinsic to the object, but are rather generated whenever an actual action opportunity arises. This claim was supported by a successive experiment (Costantini et al., 2010) in which a mug was always presented in participants' reachable space, but in half of the trials a transparent barrier was placed between participants and the object. Results showed that a motor facilitation effect occurred in congruent actions only when there was no barrier to prevent action opportunity.

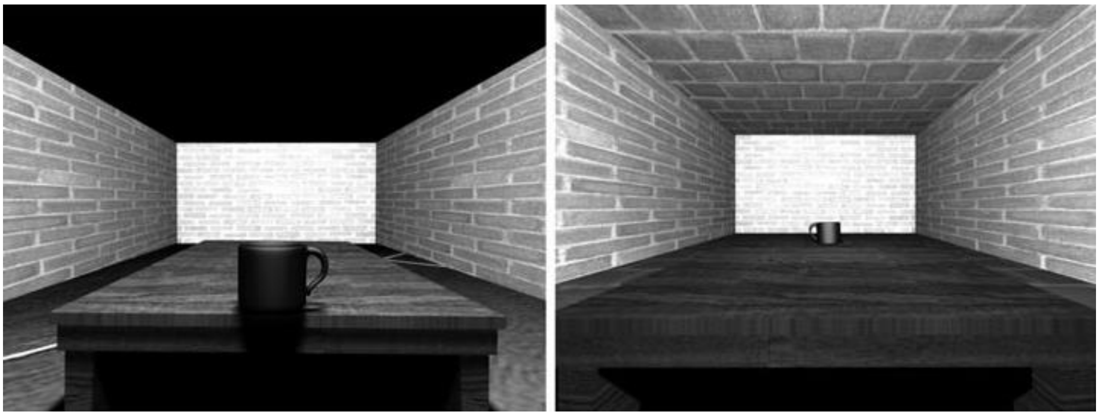


Figure 24. Examples of task-irrelevant stimuli used in Costantini et al. (2010). Assuming participants moved their right hand, the two stimuli correspond to a reachable congruent affording action (left figure) and an unreachable incongruent affording action (right figure).

However, given the complexity of the environment, we don’t always perceive or grasp isolated objects by ourselves. Therefore, Costantini et al. (2011) investigated whether perception of affording objects could be influenced by the presence of other agents. To answer this question, they expanded their previous study, introducing a human avatar placed on the side of the table. As such, the mug that was outside participants’ reachable space, was now within the avatar’s reachable space. They compared this condition to a control condition in which the avatar was substituted with a non-corporeal object (i.e., a cylinder; see Fig. 25 for an example of the task-irrelevant stimuli).

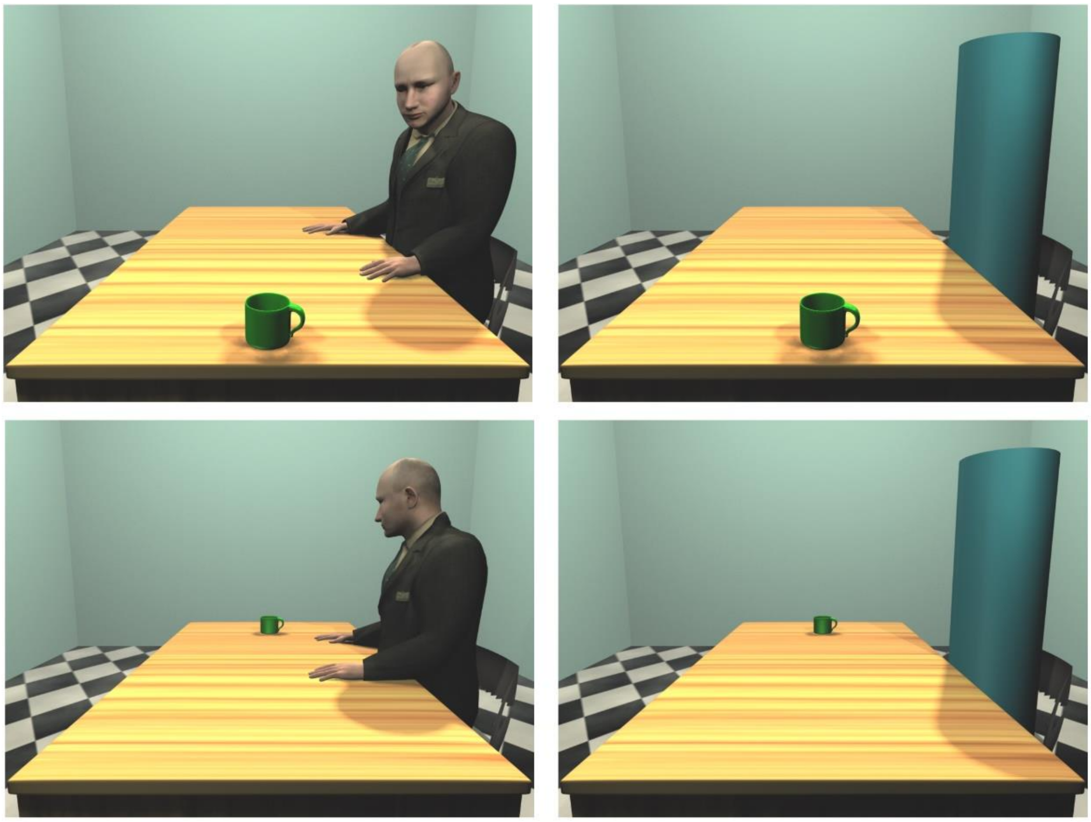


Figure 25. Examples of task-irrelevant stimuli used in Costantini et al. (2011).

Results showed that motor facilitation was achieved when the mug was presented within participants' reachable space, thus corroborating previous findings. In addition, they also showed a motor facilitation effect to occur when the mug was placed outside participants' reachable space, but only when the human avatar was present in the scene. On the other hand, no motor facilitation effect was achieved for mugs placed outside participants' reachable space when the non-corporeal object was on the side of the table. These findings suggest that seeing an affording object could invite (or even demand) a specific action to the observer, not only when it is located within their own reachable space, but also when placed in the reachable space of someone else. Such evidence could be interpreted as a motor reflection of agency detection mechanisms for other agents, which were discussed in the previous chapters. Costantini et al. (2011) theorised that this effect is due to a re-mapping of our own and others' arm reaching space (and consequently, action possibilities within that space). This does not imply that our own reachable space is extended by the presence of others, but rather that we map action opportunities available to others as if they were available to us. Consequently, a seen object may afford an action either directly (when it is within our reachable space) or indirectly (when it falls within someone else's reachable space).

Action representation and body representation

While this model of action representation in social contexts still needs further corroboration, a few studies have reported behavioural and neurophysiological evidence suggesting that others' bodily space may be mapped within our own body representation (Maravita et al., 2002; Reed & Farah, 1995; Sirigu et al., 1991). Thomas et al. (2006) deployed a cueing paradigm to investigate the contribution of spatial mapping during the processing of sensory events inflicted on our own or others' body. Cues consisted of rapid light flashes pointed at different locations of an observed agent's body, while the target was a haptic stimulus (i.e., a vibration) produced either on participants' analogous anatomical location (congruent) or on a different one (incongruent). Participants' task was to report when they could feel the haptic stimulation, while their RTs were measured. Results showed that a significant congruency effect was achieved, as participants were significantly faster to detect haptic stimulation on their bodies when a visual stimulus was projected on the same location on someone else's body. Interestingly, this congruency effect was shown to be body-specific, as it did not occur when the visual cues were projected at a non-bodily object (e.g., a house). Authors interpreted these findings as evidence that visuo-tactile processes relevant to mapping our bodily space could also be used to map others' bodily space, therefore providing an interpersonal bodily space representation.

In relation to others' bodily representation, results reported in Costantini et al. (2011) were interpreted by the authors as a spontaneous re-mapping of environmental affordances. Another possible model to account for their findings could derive from spatial perspective-taking processes, which enable own-body transformation effects (Ward et al., 2019). According to this model, participants would not necessitate to represent action possibilities of another person, but rather their own action possibilities if they were on the same spatial location of that person. This element posed the question of whether the other's body that we represent actually needs to be capable of action at all, and how spatial alignment effects could result affected when the agent in the scene does not have any potential for action. To tackle this question, perceived agency was investigated for actions available to participants only, to other humans, and to non-human (but humanoid) agents.

Overview of Experiments 6-7

Given the premises discussed above, Experiment 6 aimed to investigate the role of bodily representation of others during action representation. To this end, a spatial alignment paradigm (Bub & Masson, 2010) was deployed, where participants were asked to perform a reach to grasp action with a precision grip as soon as a task-irrelevant stimulus (i.e., a handled mug) was displayed. Crucial to the manipulation of this experiment, we expanded on Costantini et al (2011) design, implementing a condition where the human avatar was replaced by a humanoid mannequin. They reported a motor facilitation effect for objects that were unreachable to participants, but reachable to the human avatar. We predicted that, if action representation for others is driven by the bodily representation we have of them, the same effect would have been reported also when a humanoid object is presented in the scene. On the other hand, if action representation for others is driven by the representation we have of their mental states, no motor facilitation effect would have emerged for objects falling outside participants' reachable pace, but within the mannequin's. Experiment 7 expanded this investigation, examining action representation mechanisms in relation to representation of others' mental states. To this end, we amended the experimental design to include a joint attention manipulation, in which the entity on the side of the table could either face or not the mug available to its reach. In all experiments, details of how sample size was determined, as well as all data exclusions (if any), all manipulations, and all measures are reported.

Experiment 6

In Experiment 6, participants completed a spatial alignment paradigm under a total of twelve different conditions, manipulated within-subjects. We sought to replicate previous findings from Costantini et al. (2011), including a human avatar and a non-bodily object (i.e., a cylinder). To test whether bodily representation of others is a crucial feature to trigger action representation of their action, a third condition was added, in which the human avatar was substituted with a humanoid mannequin. Participants were asked to execute a grasping action with a precision grip as soon as the task-irrelevant stimulus was displayed, while their RTs were measured using a dead-man switch. According to previous research, we predicted that a motor facilitation effect should emerge across conditions when the mug is placed within participant's reach, regardless of the entity seated on the side of the table. On the other hand, when the mug is placed outside participants' reach, a motor facilitation effect should emerge if either the human avatar or the humanoid mannequin is seated on the side of the table.

Method

Participants

Since the data collection was executed during a national lockdown resulting from UK efforts to prevent further spreading of the COVID-19 pandemic, rigorous testing in controlled environments was unachievable. To counteract possible data loss and increased error variability (and hence smaller effect sizes) associated with on-line testing, all original sample sizes were doubled.

As no effect size was reported in previous studies using this paradigm, estimation of an appropriate sample size was not possible. To ensure consistency within the scientific literature, the same sample size adopted in Costantini et al. (2011) was adopted. Therefore, 40 participants (a sample size capable of detecting an effect size $d_z = 0.66$, with $1 - \beta = 0.95$ at $\alpha = 0.05$) aged 19-40 (M = 21.54, SD = 4.36, 32 were females) were recruited from the University of East Anglia to complete the experiment. Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported having normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Design

Motor facilitation effects were measured using a spatial alignment paradigm, where participants were asked to perform a grasping action with a precision grip as soon as a task-irrelevant stimulus was presented. There were three independent variables, the first being the entity displayed on the side of the table (i.e., a human avatar, a non-corporeal cylinder, or a humanoid mannequin), the second being the spatial position where the affording object (i.e., a mug) was placed (reachable or unreachable from participants' point of view), and the third being the congruency between the spatial direction of the mug handle and the hand participants used to respond (congruent or incongruent). This ended in a 3x2x2 within-subjects factorial design, with 32 randomised trials for condition, for a total of 384 trials per participant, divided in four identical blocks of 96 trials each. The dependant variable was the action onset time, operationalised as the time elapsed between the presentation of the task-irrelevant stimulus and the onset of participants' grasping action.

Stimuli

Two sets of stimuli were created. The first set consisted in coloured images depicting either a right or a left hand from a first-person point of view, pantomiming a precision grip action, which served as instruction stimuli to tell participants which hand to use in that trial (See Fig. 27). The second set consisted of 3D scenes, created with Autodesk 3ds Max, which were used as go-stimuli to prompt participants' responses. The scenes depicted 3D rooms, containing a table with a mug on top of it. The handle of the mug was oriented either towards the right or the left, resulting in a congruent or incongruent alignment with the hand participants used to respond. In half of the trials the mug was placed within participants' reachable space, while in the other half it was placed outside their reachable space (but within that of the other entity). For each spatial sector (i.e., reachable or unreachable), 33% of the trials shown a human avatar seated on the side of the table, another 33% shown the same chair occupied by a non-corporeal cylinder, and the last 33% showed a humanoid mannequin on the same chair. All three entities were always displayed on the same side where the mug handle was oriented, thus being seated either on the right or on the left side of the table (See Fig. 26 for an example of the stimuli used). The experiment was run using an online extension of E-prime 3.0 (E-Prime Go; Psychology Software Tools, Inc., Sharpsburg, PA, USA).

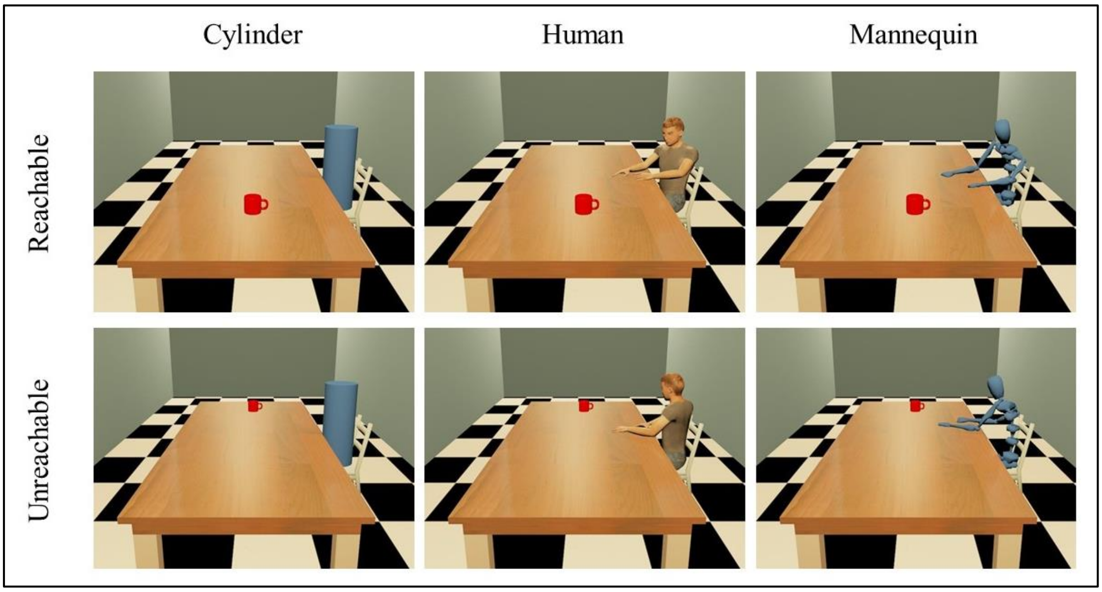


Figure 26. Example of the experimental stimuli used in Experiment 6.

Procedure

After being contacted by the experimenter, participants agreed a date and time to perform the on-line experiment remotely. Before the agreed time, participants read an information sheet and provided informed consent. At the agreed time, the experimenter opened a video call with the participants, sent all the experimental program via e-mail, and guided participants through the installation. Once the experimental procedure started, the experimenter remained available to answer question as participants read the instructions. Once participants had no doubt about their task, the experimenter ended the call to avoid further interference.

At the beginning of each trial, participants were requested to rest their index fingers on two specific keys on the keyboard (i.e., X and M), and hold them down. Each trial started with a white fixation cross (2000 ms), followed by the presentation of the instruction stimulus (i.e., a right or left hand performing a pantomimed grasping action) for 150 ms. After a variable interval (150-450 ms), the go stimulus was presented (500 ms). Participants were instructed to perform a grasping action as soon as the go stimulus appeared on the screen, thus replicating the movement they saw in the instruction stimulus (see Fig. 27). Consequently, congruent trials are referred to as the condition in which participants performed the grasping action with either the right or the left hand, while the mug handle was shown ipsilaterally. On the other hand, incongruent trials are referred to as the condition in which participants performed the grasping action with either the right or the left hand, while the mug handle was shown contralaterally. Participants' responses were made lifting the index finger of the instructed hand, and then performing the grasping action. The experimental software measured the action onset time (i.e., the time elapsed between the onset of the go-stimulus and the release of the appropriate key). As a manipulation check, at the end of the experiment, participants were asked to estimate the distance of the mug in both the reachable and unreachable position, with respect to their bodies. Reachable and unreachable mugs were, on average, estimated as being respectively 29.89 cm (SD = 14.98) and 103.30 cm (SD = 37.32) away, with all participants reporting distance to the reachable mug lower than distance to the unreachable mug.

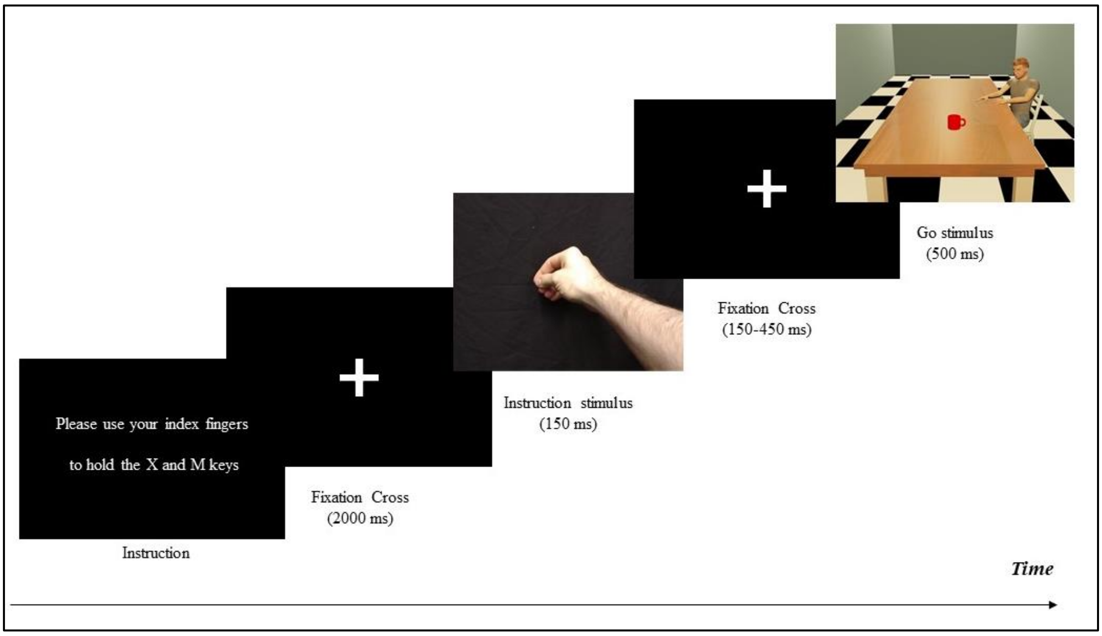


Figure 27. Trial procedure in Experiment 6.

Results

Due to technical errors during the on-line testing procedure, data from five participants were lost. The following exclusion criteria were decided prior to the data collection: action onset time lower than 100 ms (which would imply participants started to respond before the go stimulus was presented) or response given with the incorrect hand resulted in trial deletion, poor performance (participants with less than 80% of correct responses were removed), and extreme variability (trials falling outside +/- 2 SD from individual mean were removed). This resulted in 7 participants excluded (mean accuracy = 0.49, SD = 0.25), 208 trials removed for extreme variability (1.9% of total trials), and 830 trials removed for incorrect responses (7.7% of total trials). The remaining sample size consisted of 28 participants (a sample size capable of detecting an effect size $d_z = 0.71$, with $1 - \beta = 0.95$ at $\alpha = 0.05$).

Action onset time

Mean action onset time was calculated on each condition for each participant and submitted for statistical analysis. Shapiro-Wilk tests were performed on each condition to test normality of each distribution, and all conditions reported $p > .05$. A three-way within-subjects ANOVA was conducted with entity (cylinder vs. human vs. mannequin), mug position (reachable vs. unreachable), and handle position (congruent vs. incongruent) as main factors.

RTs analysis revealed a main effect of handle position: $F(1,27) = 5.18$, $p = .031$, $\eta_p^2 = .161$, and a three-way interaction effect: $F(2,54) = 4.32$, $p = .018$, $\eta_p^2 = .138$. Planned paired-samples t-tests revealed that when the cylinder was placed on the side of the table a significant motor facilitation effect was achieved in the reachable condition (congruent M = 361.62, SD = 79.01, incongruent M = 379.82, SD = 95.37, $t(27) = 2.40$, $p = .024$, $d_z = 0.45$), but not in the unreachable condition (congruent M = 367.34, SD = 85.36, incongruent M = 365.34, SD = 76.74, $p = .744$, $d_z = 0.06$). A motor facilitation effect was also found when the human avatar was shown when the mug was unreachable to participants (congruent M = 361.47, SD = 73.27, incongruent M = 376.30, SD = 84.39, $t(27) = 2.12$, $p = .044$, $d_z = 0.40$), but not when it was reachable (congruent M = 366.43, SD = 92.71, incongruent M = 367.85, SD = 96.46, $p = .803$, $d_z = 0.05$). When the mannequin was shown on the side of the table, no motor facilitation effect emerged, neither in the reachable condition (congruent M = 366.98, SD = 78.55, incongruent M = 374.55, SD = 94.09, $p = .284$, $d_z = 0.21$) nor in the unreachable condition (congruent M = 372.74, SD = 88.12, incongruent M = 370.73, SD = 87.60, $p = .622$, $d_z = 0.10$). Figure 28 shows a graphical representation of the data.

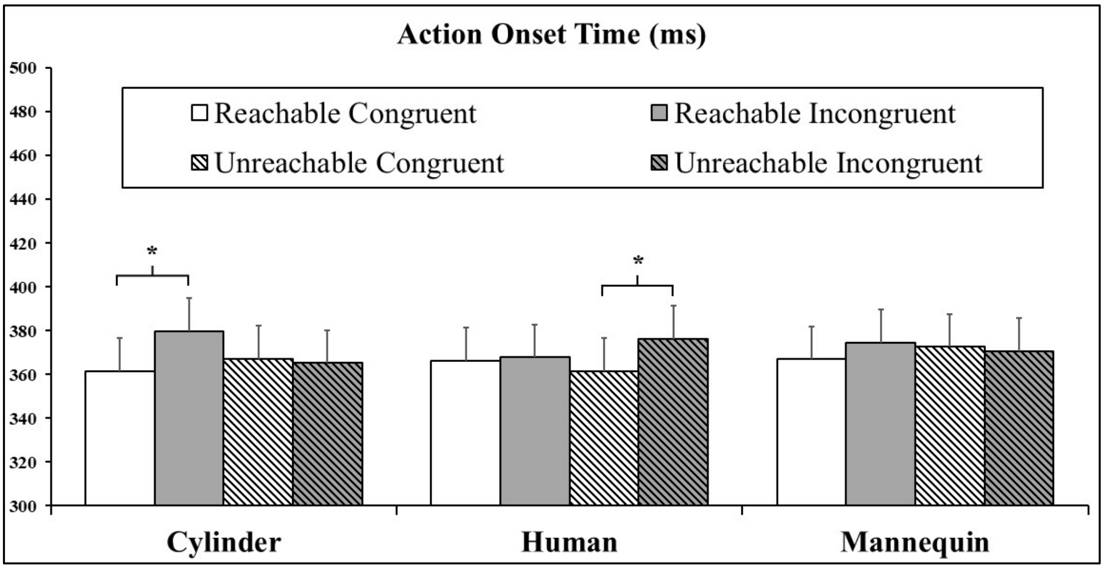


Figure 28. Mean action onset time across conditions in Experiment 6. Error bars represent the standard error of the mean. Significant differences are highlighted for the .05 (*) level.

Discussion

Experiment 6 aimed to investigate whether body representation for other agents could influence action representation mechanisms we have for them. To this end, a spatial alignment paradigm was deployed (Bub & Masson, 2010; Costantini et al., 2010; Costantini et al., 2011). Participants were requested to perform a grasping action with a precision grip, as soon as a task-irrelevant go stimulus was presented. Notably, in the stimuli the position of the affording object (i.e., a mug) was manipulated, placing it either within or outside participants' reachable space. Across three conditions, the action onset time (i.e., the time elapsed between the stimulus presentation and the beginning of participants' action) was measured, while on the side of the table either a human avatar, a non-corporeal object (i.e., a cylinder) or a humanoid mannequin were seated.

Our findings show that when the cylinder was displayed on the side of the table, participants reported a motor facilitation effect for the reachable mug, but not for the unreachable mug. In other words, participants were faster to act with the congruent hand (e.g., right hand for a right-handled mug) compared to the incongruent hand (i.e., right-hand for a left-handled mug), but only when the mug was placed in their reachable space of action. On the contrary, no motor facilitation effect (i.e., no difference between congruent and incongruent actions) was found when the mug was placed outside of participants' reach. This result corroborates previous findings (Cardellicchio et al., 2011; Costantini et al., 2010; Costantini et al., 2011), suggesting that the capability of objects to afford a specific action is not intrinsic of that object, but is rather influenced by the actual possibility we have to act towards the object.

When a human avatar was displayed in the scenario, our data show a motor facilitation effect to occur when the mug was placed outside participants' reachable space (but within the avatar's). On the contrary, no difference was found between congruent and incongruent actions directed at the reachable mug. This finding only partially fits with previous research (Costantini et al., 2011; Cardellicchio et al., 2013), which showed evidence for a motor facilitation effect to occur both for reachable and unreachable objects when a human avatar is present in the scenario. This inconsistency can be explained with reference to the reachable mug position between the agents. While Costantini et al. (2011) placed the mug to the very end of the table, at the nearest point with the participants' point of view, we placed the mug half the distance between participants' and the human avatar. Our data show no difference between congruent and incongruent actions performed to the reachable mug. However, rather that no effect at all, we interpret this result as evidence of a joint action representation. In fact, if participants represented the avatar's action for the unreachable mug, they could have done the same for the reachable mug. For example, when participants were shown a right-handled reachable mug, they would have been faster to grasp it with their right hand, but the avatar seated in front of them would have been faster with his left hand. Consequently, participants could have experienced a motor facilitation effect driven by the representation of their action (congruent grasping) and the avatar's action (incongruent grasping), leading to reduced RTs in both conditions. To provide more insightful data, further research could focus on the relationship between action representation in social contexts and the spatial positioning of affording objects in relation to the agents in the scene.

Crucial to the novelty of this study, when a humanoid mannequin was shown on the side of the table, no significant motor facilitation effect emerged, neither for the reachable mug, nor for the unreachable mug. This element is seemingly in contrast with previous research, as even if participants did not represent the mannequin's action opportunities, they should still represent their own and thus report lower RTs for congruent actions directed to reachable objects (analogous to the cylinder condition). Our data show that while not significant, there was a difference between congruent and incongruent actions for the reachable mug (p = .284, d_z = 0.21). However, because the mannequin is a more complex and less familiar object to process, higher statistical power may be needed to detect a significant difference. The absence of a significant effect for the unreachable mug can be interpreted as evidence that participants could not represent actions available to the mannequin, despite its humanoid form. This poses the question of whether mental features could be more important than bodily features to enable action representation of others, which will be explored in Experiment 7.

It is important to remember that this study was performed optimising on-line testing procedures, where keeping a high degree of control over the experimental environment and external factors is a delicate challenge. To try to overcome these intrinsic limitations, methodological and statistical strategies were adopted (i.e., doubling sample size and number of trials per condition). Yet, future research could replicate this study in a more controlled experimental setting, in order to provide more reliable data to answer the research question.

Experiment 7

Experiment 6 showed that when a humanoid mannequin was displayed on the side of the table, participants did not experience a motor facilitation effect for a mug placed outside their reachable space, but within the mannequin's. On the contrary, when a human avatar was present in the scene, participants experienced a motor facilitation effect for objects the avatar only could reach, but not for object they could reach (thus leading the hypothesis for the occurrence of a joint action representation). As previously discussed, this evidence suggests that bodily representation of other agents is not a sufficient condition to trigger action representation for their action possibilities. However, the question arises of whether mental (versus body) representation of the other agent may play a critical role. To answer this question, Experiment 7 will investigate the influence of joint attention on action representation mechanisms.

Joint attention and action representation

The gaze and eye movements of other agents are a fundamental source of information about what they can see, what they are aware of, and their internal states (Tomasello et al., 2005). This information can then be used to plan and execute actions together. For instance, when passing an object to another agent, we may use mutual gaze to predict whether the other is aware of an obstacle (e.g., a barrier). Joint attention allows both agents to monitor the other's gaze and attentional states (Emery, 2000). For example, if two agents need to synchronise their actions, they will divide attention between spatial locations which are relevant both to their own and the other's goal (Böckler et al., 2012; Ciardo et al., 2016; Kourtis et al., 2014). In addition, previous research showed that sharing gaze influences object and action processing, increasing the motor and emotional relevance conferred to attended objects (Becchio et al., 2008; Innocenti et al., 2012; Scorolli et al., 2014). Further investigations reported that in a joint attention search task, co-agents who received mutual information regarding the other's gaze direction searched faster than those who had no access to such information (Brennan et al., 2008; Wahn et al., 2015). Taken together, these findings illustrate that gaze information holds a crucial role in others' mental representation, especially in joint action settings where representing others' actions could be a critical ability to succeed (Vesper et al., 2017).

Given the premises discussed above, Experiment 7 will investigate whether joint attention can affect the action representation we have for other agents. To do this, the same procedure of Experiment 6 was deployed, but the experimental stimuli were amended to always show a human avatar on the side of the table, which could be either looking at or away to a mug placed within or outside participants' reachable space. We predicted that if the avatar is looking at the mug, a motor facilitation effect should emerge only when the object is placed outside participants' reachable space (but within the avatar's). On the other hand, if the avatar is not looking at the mug, a motor facilitation effect should only emerge when the object is placed within participant's reach.

Method

Participants

Analogously to Experiment 6, calculated sample sizes were doubled in order to adjust for data loss resulting from on-line testing procedures.

As no effect size was reported in previous studies using this paradigm, estimation of an appropriate sample size was not possible. To ensure consistency within the scientific literature and with Experiment 6, the same sample size adopted in Costantini et al. (2011) was adopted. Therefore, 40 participants (a sample size capable of detecting an effect size $d_z = 0.66$, with $1 - \beta = 0.95$ at $\alpha = 0.05$) aged 19-31 (M = 22.85, SD = 4.45, 31 were females) were recruited from the University of East Anglia to complete the experiment. Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported having normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Design

Motor facilitation effects were measured using a spatial alignment paradigm, where participants were asked to perform to perform a grasping action with a precision grip as soon as a task-irrelevant stimulus was presented. There were three independent variables, the first being the gaze orientation of the human avatar (joint or averted), the second being the mug's spatial position (reachable or unreachable from participants' point of view), and the third being the congruency between the spatial direction of the mug handle and the hand participants used to respond (congruent or incongruent). This ended in a 2x2x2 within-subjects factorial design, with 32 randomised trials for condition, for a total of 256 trials per participant, divided in four identical blocks of 64 trials each. The dependant variable was the action onset time, operationalised as the time elapsed between the presentation of the task-irrelevant stimulus and the onset of participants' grasping action.

Stimuli

Instruction stimuli showing either a right or a left hand from a first-person point of view, pantomiming a precision grip action were kept unaltered from

Experiment 6. 3D experimental stimuli were amended to always show a human avatar on the side of the table, which could either face the mug or looking away. (See Fig. 29 for an example of the stimuli used). The experiment was run using an online extension of E-prime 3.0 (E-Prime Go; Psychology Software Tools, Inc., Sharpsburg, PA, USA).

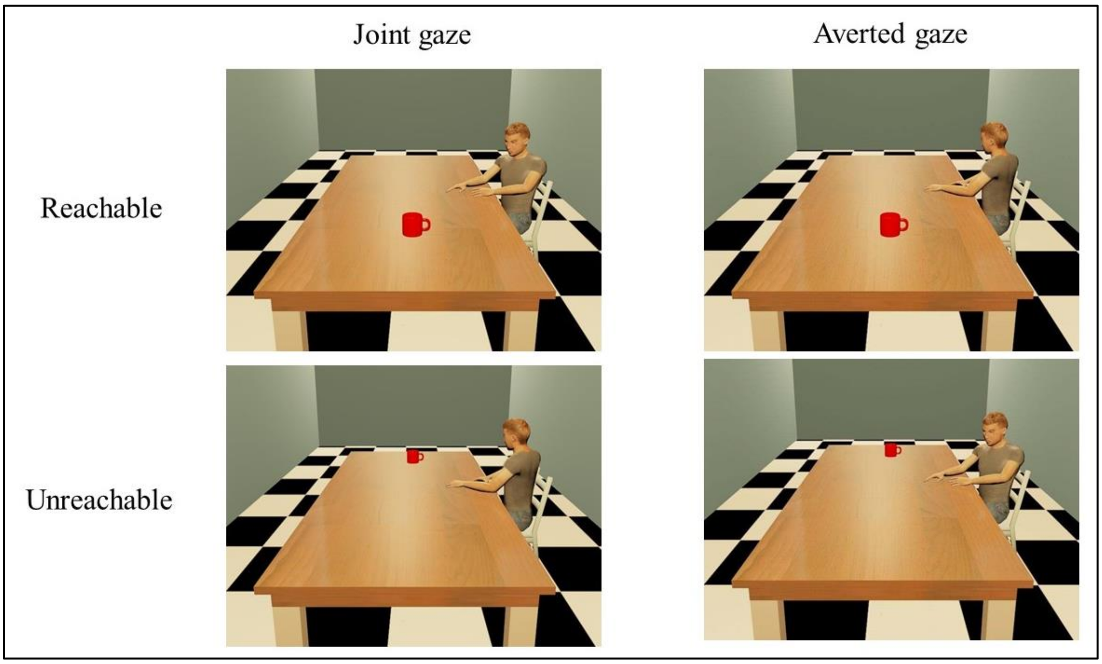


Figure 29. Example of the experimental stimuli used in Experiment 7.

Procedure

After being contacted by the experimenter, participants agreed a date and time to perform the on-line experiment remotely. Before the agreed time, participants read an information sheet and provided informed consent. At the agreed time, the experimenter opened a video call with the participants, sent all the experimental program via e-mail, and guided participants through the installation. Once the experimental procedure started, the experimenter remained available to answer question as participants read the instructions. Once participants had no doubt about their task, the experimenter ended the call to avoid further interferences.

At the beginning of each trial, participants were requested to rest their index fingers on two specific keys on the keyboard (i.e., X and M), and hold them down. Each trial started with a white fixation cross (2000 ms), followed by the presentation of the instruction stimulus (i.e., a right or left hand performing a pantomimed grasping action) for 150 ms. After a variable interval (150-450 ms), the go stimulus was presented (500 ms). Participants were instructed to perform a grasping action as soon as the go stimulus appeared on the screen, thus replicating the movement they saw in the instruction stimulus. Consequently, congruent trials are referred to as the condition in which participants performed the grasping action with either the right or the left hand, while the mug handle was shown ipsilaterally. On the other hand, incongruent trials are referred to as the condition in which participants performed the grasping action with either the right or the left hand, while the mug handle was shown contralaterally. Participants' responses were made lifting the index finger of the instructed hand, and then performing the grasping action. The experimental software measured the action onset time (i.e., the time elapsed between the onset of the go-stimulus and the release of the appropriate key). As manipulation check, at the end of the experiment, participants were asked to estimate the distance of the mug in both the reachable and unreachable position, with respect to their bodies. Reachable and unreachable mugs were averagely estimated as being respectively 34.61 cm (SD = 27.76) and 110.00cm (SD = 62.47) away, with all participants reporting distance to the reachable mug lower than distance to the unreachable mug.

Results

Due to technical errors during the on-line testing procedure, data from 3 participants were lost. The following exclusion criteria were decided prior to the data collection: action onset time lower than 100 ms (which would imply participants started to respond before the go stimulus was presented) or response given with the incorrect hand resulted in trial deletion, poor performance (participants with less than 80% of correct responses were removed), and extreme variability (trials falling outside +/- 2 SD from individual mean were removed). This resulted in 4 participants excluded (mean accuracy = 0.57, SD = 0.20), 385 trials removed for extreme variability (4.6% of total trials), and 342 trials removed for incorrect responses (4.0% of total trials). The remaining sample size consisted of 33 participants (a sample size capable of detecting an effect size $d_z = 0.65$, with $1 - \beta = 0.95$ at $\alpha = 0.05$).

Action onset time

Mean action onset time was calculated on each condition for each participant and submitted for statistical analysis. Shapiro-Wilk tests were performed on each condition to test normality of each distribution, and all conditions reported p >.05. A three-way within-subjects ANOVA was conducted with gaze (joint vs. averted), mug position (reachable vs. unreachable), and handle position (congruent vs. incongruent) as main factors.

RTs analysis revealed a main effect both on mug position ($F(1,32) = 6.22$, $p = .018$, $\eta_p^2 = .163$) and handle position ($F(1,32) = 5.68$, $p = .023$, $\eta_p^2 = .151$). A significant interaction between gaze and mug position was also found: $F(1,32) = 7.59$, $p = .010$, $\eta_p^2 = .192$. Planned paired-samples t-tests revealed that when the avatar was facing the mug, no motor facilitation effect emerged either in the reachable condition (congruent M = 418.08, SD = 93.53, incongruent M = 429.64, SD = 106.74, $p = .091$, $d_z = 0.31$), nor in the unreachable condition (congruent M = 432.34, SD = 106.20, incongruent M = 436.36, SD = 104.39, $p = .276$, $d_z = 0.19$). When the avatar was looking away from the mug, no motor facilitation effect emerged either in the reachable condition (congruent M = 426.55, SD = 96.81, incongruent M = 433.59, SD = 106.01 $p = .180$, $d_z = 0.23$), nor in the unreachable condition (congruent M = 430.36, SD = 104.41, incongruent M = 430.10, SD = 104.39, $p = .957$, $d_z = 0.01$). Figure 30 shows a graphical representation of the data.

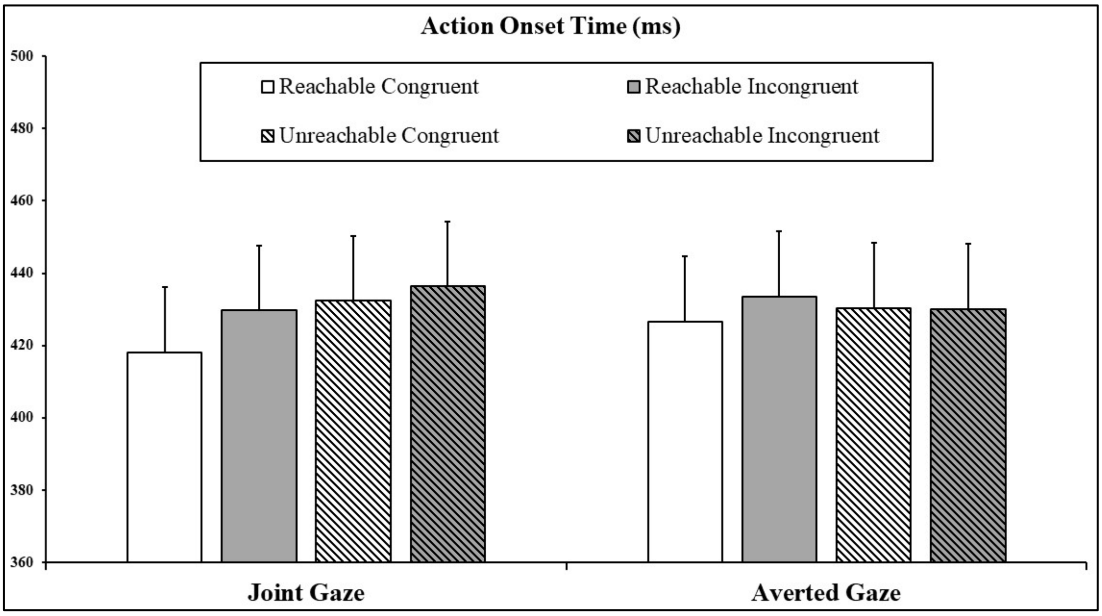


Figure 30. Mean action onset time across conditions in Experiment 7. Error bars represent the standard error of the mean.

Discussion

Experiment 7 aimed to investigate whether joint attention could influence action representation mechanisms we have for other agents. To this end, a spatial alignment paradigm was deployed (Bub & Masson, 2010; Costantini et al., 2010; Costantini et al., 2011). Participants were requested to perform a grasping action with a precision grip, as soon as a task-irrelevant go stimulus was presented. Notably, in the stimuli the position of the affording object (i.e., a mug) was manipulated, placing it either within or outside participants' reachable space. Across two conditions, the action onset time (i.e., the time elapsed between the stimulus presentation and the beginning of participants' action) was measured, while the human avatar seated by the side of the table was either looking at or away from the mug.

Our findings show that despite two significant main effects respectively on mug position (reachable vs. unreachable) and handle position (congruent vs. incongruent), and a significant interaction effect between gaze (joint vs. averted) and mug position, no motor facilitation effect emerged from direct comparisons. These results are not consistent with previous research or with Experiment 6, which detected motor facilitation effects for objects presented within someone else's reachable space. Our data show that while not significant, there was a difference in the joint gaze condition between congruent and incongruent actions both for the reachable (p = .091, d_z = 0.31) and unreachable mug (p = .276, d_z = 0.19), as well as for the averted gaze condition for the reachable mug (p = .180, d_z = 0.23). Given the wide consistency of the effects discussed in the literature, shown both on a behavioural (Costantini et al., 2011) and neurophysiological (Cardellicchio et al., 2013) level, we speculate that increased statistical power and a more controlled experimental environment could lead to clearer results. If that would be the case, it could suggest that in order to experience motor facilitation effects towards other agents, we need to reconstruct a representation of their mental states (such as their attention) first. However, data from Experiment 7 is not strong enough to fully support this claim, and future research should address this topic to provide further evidence.

Chapter discussion

Chapter 4 investigated action representation processes in social context, and how these mechanisms are influenced by bodily and mental representation of other agents. More specifically, across two experiments, participants' readiness to act (i.e., action onset time) towards an affording object was measured. This object (i.e., a handled mug) featured a lateralised affordance, as the action it suggests was directed preferentially to a specific hand (i.e., a right-handled mug would invite an action performed with the right hand). By consequence, participants could produce an action that was either congruent or incongruent with the stimulus presented. This paradigm, known as spatial alignment effect, has been widely adopted in previous studies to investigate action representation and object perception (e.g., Bub & Masson, 2010; Costantini et al., 2010; Costantini et al., 2011; Tucker & Ellis, 1998). By these means, a motor facilitation effect is detected when there is a significant difference between congruent and incongruent actions, evidence that the lateralised affordance is inviting a specific motor act.

In Experiment 6, a spatial alignment paradigm was adopted to investigate whether bodily representation of another agent could play a central role to enable action representation for them. To this end, 3D stimuli were created showing the mug placed on a table, either within or outside participants' reachable or unreachable space. In addition, three different classes of entities were placed by the side of the table, in such a way that a mug that was unreachable to participants fell within the entity reachable space. These three entities varied on the degree of bodily representation: a non-corporeal cylinder, a humanoid mannequin, and a human avatar. Results showed that participants experienced a motor facilitation effect for reachable objects, but not for unreachable objects, when they were acting alone (i.e., the cylinder and, to a lesser extent, the mannequin condition). On the contrary, participants showed a motor facilitation effect when the mug was placed outside their reachable space, but within another human's. This evidence replicates previous findings (Costantini et al., 2011; Cardellicchio et al., 2013), suggesting that when another agent is present in the scene, we may represent which actions are available to them.

Joint action representation

Results reported for the human avatar condition in Experiment 6 pose the question of which action representation is prioritised when an object is graspable by more than one agent. Costantini et al. (2011) found a significant motor facilitation effect when the mug was reachable to participants, hence suggesting that egocentric action representation is prioritised compared to allocentric action representation. However, it is worth noting that they displayed the reachable mug on the very end of the table (from participants' point of view), thus hardly reachable by the human avatar. On the other hand, Experiment 6 in the analogous condition showed the mug exactly in between participants' and the human avatar, reporting no difference between congruent and incongruent actions. Rather than no action representation at all, it is possible to interpret this finding as evidence that when an affording object is available to more than one agent, both actions are represented altogether. For example, a right-handled mug would invite a right-hand action to us, but a left-hand action to someone seated in front of us, leading to decreased action onset time both for participants' right hand (egocentric action) and left hand (allocentric action). This model would be consistent with the literature on joint action, which consistently agrees that when performing a motor task with other agents, we have a tendency to represent each other's task, even in the absence of a shared goal (Atmaca et al., 2008; Böckler et al., 2012; Sebanz et al., 2003; Sebanz et al., 2005; Schmitz et al., 2017; Tsai et al., 2006; Welsh, 2009). However, when replicating this condition in Experiment 7, a nearly significant motor facilitation effect was found when the mug was available to both participants and the human avatar (p = .091). This may suggest that individual differences could bias statistical findings, as some participants could differ on their tendency to represent other's actions (at an increased cognitive cost). Further research should address this controversy, exploring how action representation mechanisms in social contexts are influenced by the spatial position of an affording object available to two agents (e.g., egocentric space, neutral, allocentric space).

Bodily and mental representation

Experiment 6 and 7 respectively investigated the contribution of bodily and mental representations in the formation of action representation for others' actions. In Experiment 6, a humanoid mannequin condition was used to test whether the sole representation of a human-like shape could be a decisive feature to produce a motor facilitation effect for objects placed outside of our reach. Results revealed no difference between congruent and incongruent actions, suggesting that we can't represent actions for a mindless entity, despite its humanoid appearance. These findings build on the previous literature on other's bodily representation (e.g., Maravita et al., 2002; Reed & Farah, 1995; Sirigu et al., 1991; Thomas et al., 2006), which consistently agree that others' bodily space could be mapped within our own body representation. Work from this thesis expands this model, suggesting that in order to trigger body representation mechanisms, a mind must be attributed to the other agent first.

In Experiment 7, representation of the other's mental states was manipulated. Namely, this experiment investigated whether joint attention could influence representation of others' actions. Results revealed that although not significant, minor motor facilitation effects were detected, hence indicating that this study could be underpowered. A speculative interpretation of our data, based on increased power, would suggest that action representation for others is only possible when we represent their mental states. In this case, attention is a prerequisite to infer whether others are aware of the mug available to their reach. This interpretation would fit nicely with previous literature on joint action, which suggest that gaze information gains a crucial role in others' mental representation, especially in joint action settings where representing others' actions could be a critical ability to succeed (e.g., Brennan et al., 2008; Vesper et al., 2017; Wahn et al., 2015). However, extreme caution should be taken interpreting these results. Given the intrinsic limitations over the experimental setting resulting from on-line testing procedures, it is advisable for future research to replicate these studies in controlled environments, in order to provide more reliable data.

A limitation to be addressed is the extent to which we can represent actions of a virtual individual (such as the human avatar displayed in the stimuli). Indeed, the experimental set-up differs from ecological environments, especially for the fact that the avatar was always presented in the same, static posture. Nonetheless, previous studies adopted similar stimuli to investigate mental representation mechanisms (e.g., spatial perspective-taking; Amorim, 2003; Lambrey et al., 2008; Vogeley, 2008). Notably, in the work from Amorim (2003) and Lambrey et al. (2008), the presence of a static human avatar was a sufficient condition to enable a perspective futuristic point of view on the scenario. In addition, it is worth noting that both in Experiment 6 and 7, the presence of the human avatar in the context was not relevant to the task participants performed. While it is not advisable to dismiss possible differences between a human avatar and a real person, these studies show that the object-avatar relationship was sufficient to invite a motor act on the observer (i.e., participants), granted that the object was available to the avatar's reach. On a side note, according to our findings and the models described, more human-realistic stimuli (e.g., photographs of real people acting) could lead to increased effect sizes, thus requiring lower statistical power to be detected.

One point that needs further clarification is the separated contributions of action representation and spatial perspective taking to the generation of a motor facilitation effect for objects only reachable by another person. According to Costantini et al. (2011), such effects are due to a re-mapping process which allows us to represent others' action opportunities as if they were ours own. On the contrary, spatial perspective-taking models (e.g., Ward et al., 2019) could account for this evidence in terms of own-body transformation effects, in which participants would not necessitate to represent action possibilities of another person, but rather their own action possibilities if they were on the same spatial location of that person. A possibility to explore separate contributions of these mechanisms could be to manipulate the avatar's awareness of action, while keeping its spatial location unaltered (for instance, using a blindfolded avatar in half of the trials).

Taken together, findings from Chapter 4 illustrate that agency detection and attribution to others can affect our motor system. These mechanisms seem to operate on an automatic and implicit level, but may be influenced by higher-level mechanisms (such as theory of mind). If considered along with data from Chapter 3, it is possible to hypothesise that action representation for non-human agents could also be affected by explicit belief. Further research could explore this line of research, deploying a method similar to Experiments 6 and 7, whilst retaining the priming manipulation of Experiment 5. This evidence could be of extreme relevance to robot designers looking to develop social embodied artificial agents. Accordingly, rather than focussing on their physical and humanoid appearance, mental ergonomics should be prioritised. In fact, all the features capable of leading an inference towards the other's mental states (e.g., gaze direction) can become crucial elements to facilitate human-robot interaction, both on a cognitive and motor level.

SECTION 3

-

General Discussion

Chapter 5

General Discussion and Conclusions

"We can only see a short distance ahead,

but we can see plenty there that needs to be done."

Alan Turing, 1948

Existing literature on the sense of agency has mainly focussed on elements that help (or hinder) the experience of agency over physical, self-produced, discrete actions. However, our action domain is not limited to the one mentioned above. Every day, we execute social actions directed to other people, and observe their actions as well. Moreover, our social environment will soon not be limited to human agents, but will include artificial agents as well. It is therefore essential to understand the cognitive processes involved in the interaction between humans and with artificial agents, in order to promote a higher degree of social perception towards them. By these means, the current gap experienced between the interaction with humans and with artificial agents may be reduced, leading us to a more familiar user experience which could transform what today are tools and artifacts into actual social companions. Yet, in the past these important topics have received little attention by the scientific community. Therefore, the aim of this dissertation was to expand existing knowledge about sense of agency and agency detection mechanisms, with specific focus on observed actions executed by human and artificial agents. Namely, this work consisted of three empirical aims: to investigate whether agency can be detected for indirect social actions executed by humans or artificial agents, to explore agency detection on observed robotic actions, and to examine motor facilitation effects resulting from representing the action opportunities of other agents, whether they be human or artificial. This chapter will provide an extensive overview of findings across all experimental chapters, also discussing limitations, future directions, and possible integrations with existing cognitive models.

Results Overview

Social sense of agency and vocal actions

To explore how we detect agency for ourselves, other humans, and artificial agents when we are performing or observing a physical or social action, Chapter 2 presents three experiments where the agent's nature (i.e., self, human, or artificial) and the action domain (i.e., manual or vocal) were manipulated. To this end, subjective temporal compression was measured as main dependant variable, an effect known as temporal binding (Haggard et al., 2002). More specifically, Experiment 1 investigated whether performing vocal actions leads to perceived time compression, and to what extent this effect would be similar to that experienced when performing or observing someone else executing manual actions. Experiment 2 further expanded the topic, exploring whether hearing someone else performing a vocal action would result in a temporal binding effect. Experiment 3 investigated the role of the agent's nature, comparing temporal compression experienced for observed vocal actions performed by human and artificial agents.

Taken together, data from Chapter 2 showed consistent evidence of a temporal binding effect for self-generated vocal actions that produce a contingent outcome, the extent of which appears to be reduced compared to manual actions, as consistent to Limerick et al. (2015). These findings make an important contribution to the growing literature concerned with how we represent intentional actions and their consequences, demonstrating that action is not limited to our physical movements, but includes a range of social behaviours as well (as initially theorised by Stephenson et al., 2018). In addition, temporal binding was experienced by participants for observed manual and vocal actions, as long as direct visual access to the other's action was possible. This finding is consistent with previous research (Poonian & Cunnington, 2013; Strother et al., 2010; Wohlschläger et al., 2003), and we interpreted it as evidence of engagement of the observer's motor system, possibly related to the mirror neuron system (Di Pellegrino et al., 1992; Gallese et al., 1996) and action representation processes according to which observing an action results in simulating it as if it was our own (Fogassi et al., 2005). These processes may include not only a simulation of motor planning and execution, but also a prediction of the outcomes that would be generated by the observed action (Aglioti et al., 2008). As a result, simulating others' actions as if they our own may lead to outcome prediction, and consequently to agency detection.

According to our data, existing models of agency detection (e.g., Haggard et al., 2017, Kunde et al., 2018) should be expanded and integrated with a dedicated mechanism accounting for action observation. In fact, results from Chapter 2 illustrate that agency is not solely detected as a result of predictions on self-generated actions, but can be experienced for observed actions as well. Furthermore, having a specific system dedicated to detect agency for observed actions would have proven useful in terms of adaptability and survivability, helping us not only to distinguish *what we are doing* from *what is happening*, but also from *what others are doing*. Interestingly, reported data showed no temporal binding effect for artificial vocal actions, both when they were heard and observed. This finding corroborates the idea that causality alone is not the only key to achieve temporal compression. In fact, the mere communicative act (pronouncing a word) was, when uttered by artificial agents, ineffective at generating binding effects. This evidence is in contrast with previous research which advanced the idea of binding effects to occur between two events whenever the latter is thought to depend on the prior (Buehner & Humphreys, 2009). However, as mentioned in Chapter 2, while our method involved a retrospective assessment of time, that of Buehner and Humphreys used a prospective assessment task. This element could lead to inappropriate comparisons between the studies, as while our method was focussed on both onset and offset of the target time interval (i.e., action and outcome), theirs specifically analysed the offset point (i.e., the outcome). Future research could tackle this controversy, by assessing causality and intentionality as features of the temporal binding effect using a consistent methodology to allow direct comparisons.

Sense of agency and embodied robots

Nowadays, most of the public has access to artificial agents in the form of vocal assistants (e.g., Siri, Alexa). However, a new class of artificial agents will soon start to populate our environments: social robots. Dealing with embodied entities, which appear to modify their surroundings according to inner states (i.e., being agentic), poses the question of whether we can efficiently interact with them, and perceiving their action as hauling intentionality. Chapter 3 directly examined whether observing actions performed by a social robot produced subjective temporal compression, and if this effect is mediated by higher-level mechanisms, such as explicit belief and theory of mind. Experiment 4 compared temporal binding effects reported by participants when they were executing an action, or observing it performed either by a human or a robot (i.e., Cozmo). Experiment 5 built on data reported in Experiment 4, investigating whether different prior experience of Cozmo's behaviour could affect participants' time perception.

Taken together, data from Chapter 3 shows that across all conditions, participants experienced binding effects when they were executing an action, or observing a human performing it. This is consistent with data reported in Chapter 2 and previous literature (Poonian & Cunnington, 2013; Strother et al., 2010;

Wohlschläger et al., 2003), suggesting that a specific agency detection mechanism may be employed to assess intentionality for actions performed by others. When participants were observing actions performed by a robot without specific knowledge of how it functioned, Experiment 4 showed that interval reproduction performance varied greatly between participants, with some experiencing temporal compression, and some not. This finding, although requiring further investigation, is consistent with previous research, which suggests that being in the presence of a robot can alter our decisions and perceptions (Bainbridge et al., 2008; Frischen et al., 2007; Kompatisari et al., 2018; Shinozawa et al., 2005), indicating that they are classified differently from other artefacts. In Experiment 5 participants completed a brief interaction activity with Cozmo before the experiment. Half of the sample experienced Cozmo as being manually controlled by the experimenter (mechanistic group) whereas the other half experienced Cozmo as a spontaneous agent (mentalistic group), capable of movement and of displaying outcome-congruent emotional responses (i.e., happiness when winning, frustration when losing). Results showed that our manipulation was effective as participants in the mechanistic group did not experience significant binding effects. On the other hand, participants in the mentalistic group reported increased temporal compression, yet not to the same degree of self- and human-produced actions. This element fits well with the model of agency detection for observed action described above. In fact, observing robotic actions were shown to recruit the mirror neuron system (Gazzola et al., 2007), analogously to observed human actions.

Data from Experiment 5 further expands existing models (e.g., Haggard et al., 2017, Kunde et al., 2018), showing that action observation is indeed necessary to agency detection, but not sufficient by itself. On the contrary, explicit information about the agent's behaviour is needed to generate temporal compression. Interestingly, participants in the mentalistic group tended to ascribe Cozmo with a higher degree of mental states, explain its behaviour referring to intentions and desires, and refer to its components with human terms. Previous research suggests that mentalising and anthropomorphism processes can influence action perception (Castelli et al., 2000; Iacoboni et al., 2004; Mitchell et al., 1997; Waytz et al., 2014). Our data expand on this finding, suggesting that action perception is not only affected by mentalising and anthropomorphism mechanisms, but can affect them as well.

Social affordances and action representation

Having explored how our cognitive systems can be affected by agency detection processes, Chapter 4 expanded on previous findings to explore whether any effect may be exerted on our motor system as well. To do this, we took advantage of a specific cognitive process known as affordance (Gibson, 1979). According to existing literature, seeing an object triggers the representation of the congruent motor act needed to interact with it, which results in reduced RTs to initiate that action, compared to an incongruent action (Ellis & Tucker, 2001). Interestingly, this motor facilitation effect is not generated just by intrinsic visual features of the object, but it also depends on the contextual information related to action opportunities with that object (Cardellicchio et al., 2011; Costantini et al., 2010). By these means, action representation can be achieved when an affording object is presented within an action opportunity (e.g., inside one's own reachable space). Interestingly, previous research has shown that action representation could be sensible to social context, as perceiving objects available to other agents still resulted in a motor facilitation effect (Cardellicchio et al., 2013; Costantini et al., 2011).

Chapter 4 adopted this framework to investigate whether action representation is dependent on the representation we have of the observed agent's body (Experiment 6) or mental states (Experiment 7). Experiment 6 explored action representation for observed agents in relation to the agent's nature, manipulating it between non-corporeal (i.e., a cylinder), humanoid (i.e., a mannequin), or human (a 3D avatar). The object location could have been either available to participants and the displayed agent (i.e., reachable condition), or only to the displayed agent (unreachable condition). Results showed that a motor facilitation effect was only achieved for unreachable objects when the human avatar was displayed (consistently with Costantini et al., 2011), which suggests that an even though we represent a human body, action representation is not achieved without mental states attribution.

When the object was available to both participants and the displayed agent (reachable condition), a motor facilitation effect was achieved when the cylinder and the mannequin were displayed, suggesting that participants could represent their own actions towards an object with which they could interact. On the other hand, no difference in RTs between congruent and incongruent actions was reported for reachable objects when the human avatar was displayed, evidence that we interpreted in favour of a joint action representation mechanism. Accordingly, displaying a right-handled mug would invite a right-hand action to us, but a left-hand action to someone seated in front of us, leading to decreased action onset time both for participants' right hand (egocentric action) and left hand (allocentric action). This model would be consistent with the literature on joint action, which consistently agrees that when performing a motor task with other agents, we have a tendency to represent each other's task, even in the absence of a shared goal (Atmaca et al., 2008; Böckler et al., 2012; Sebanz et al., 2003; Sebanz et al., 2005; Schmitz et al., 2017; Tsai et al., 2006; Welsh, 2009).

Experiment 7 investigated whether representing others' mental states could influence action representation mechanisms. To do this, participants' perception of the avatar's attention was manipulated directing its gaze either towards or away from the object. The experiment resulted to be underpowered to detect any significant effect. However, preliminary data showed minor effects for the reachable object regardless of the avatar's gaze direction, whereas for the unreachable object the same was true only in the joint gaze condition. Taken together, findings from Chapter 4 suggest that action representation for others does not depend only on their bodily representation, but is only possible when we represent their mental states. In this case, attention is a prerequisite to infer whether others are aware of the mug available to their reach. This interpretation would fit nicely with previous literature on joint action, which suggest that gaze information gains a crucial role in others' mental representation, especially in joint action settings where representing others' actions could be a critical ability to succeed (e.g., Brennan et al., 2008; Vesper et al., 2017; Wahn et al., 2015).

Limitations

This dissertation explored agency detection through temporal binding and motor facilitation effects. However, it is worth noting that subjective temporal compression measures only one component of the sense of agency. As previously discussed in Chapter 1, the sense of agency is composed by two sub-constructs.

Feelings of agency are known to be measured using temporal binding effects. On the contrary, *judgements* of agency reflect the explicit awareness of ourselves as intentional agents. Typically, studies investigating sense of agency using temporal binding paradigms only focus on the implicit measure of agency. However, more recently the scientific community has focussed on how implicit agency measures (i.e., temporal binding effects) relate to explicit agency measures (i.e., judgements of agency). In typical instances, where there is no ambiguity over agency detection, feelings of agency and judgements of agency tend to reflect each other (Haggard & Tsakiris, 2009; Moore & Haggard, 2010; Moore et al., 2012, see Haggard, 2017 for a review). Results showed in Chapter 2 are consistent with this evidence. However, although deeply interconnected, judgements and feelings of agency have been shown to be separate and dissociable. For example, a study investigating the relationship between neural sensory attenuation, explicit ratings of agency, and temporal binding (Dewey & Knoblich, 2014) reported no significant correlations between these measures. In addition, judgements of agency were found not to be reflected by the neural N1 component, but by the successive P3 component, which resulted attenuated for predictable self-generated outcomes (Kühn et al., 2011). On the other hand, feelings of agency have been shown to be reflected by the earlier N1 component (Caspar et al., 2016; Poonian et al., 2015). This seems to suggest that feelings of agency and judgements of agency are not only computed separately, but also arise at different time points. Therefore, previous research has suggested that explicit and implicit measures of agency rely on separate mechanisms, which concur together in the generation of a complete sense of agency.

As previously discussed in Chapter 1, measuring subjective temporal compression with an interval reproduction task confers some advantages over the typical Libet clock method, especially when investigating observed actions. For example, interval reproductions are directly informative of the perceived time interval between two events, and take advantage of experiential reproduction of perceived time intervals. On the other hand, the Libet clock method consists of converting positional judgments into temporal measurements. Hence, adopting a temporal reproduction task reduces the risk to bias the interval estimations. However, interval reproduction tasks may suffer increased variability in the perception of the onset of the target time interval. In the control condition, the target time interval begins at the offset of the first tone, but participants may associate the onset of the first tone with the onset of the target time interval. In such situation, the perceived time interval may result longer, and this bias could account for many of the differences detected between the control condition and agency conditions. However, in Experiment 2 and 3 the modality of presentation of the first event (interval onset) was maintained unaltered. Comparing computer-generated tones with verbal utterances produced wither by human or artificial agents allows to overcome this limitation, thus supporting the interval reproduction task as a reliable method to measure subjective time compression.

Future Directions

Findings from Chapter 2 suggest that agency detection is not limited to the physical domain of action, but includes social actions as well. The data presented show how vocal actions were computed by participants when they were self-, human-, or artificial-generated. Given the intrinsic mutual relationship of action (and interaction) in social context, future research should address how vocal actions are computed when directed to other agents, human or artificial. For example, temporal binding could be measured for self-produced vocal actions (pronouncing the word “go”) directed to a standing agent (human vs. artificial vs. inanimate), which would start to move after a variable time interval.

Experiment 4 and 5 showed that it is possible to detect agency for actions performed by embodied artificial agents (i.e., social robots), but this process is influenced by the explicit and contextual information we have about them. In particular, theory of mind and anthropomorphism mechanisms seem to both affect and be affected by action observation. Given the result reported in Chapter 3, future research should investigate agency detection in action observation for social robots with more advanced specimens, capable of more refined features. For example, it could be investigated the role of human action kinematics, having a robot that could act imitating human actions, or in a more mechanical way.

Data from Chapter 4 posed the basis for a joint action representation mechanism, in which actions towards an affording object available to multiple agents are represented simultaneously. Previous findings showed that action representation is sensible to the spatial location of the perceived object (Cardellicchio et al., 2011;

Costantini et al., 2010), but data from Chapter 3 suggest that explicit information about the object (e.g., ownership) could be a relevant element as well. Future research should investigate action representation in joint action, and shed light on whose action we represent (or which one is prioritised) in different contexts.

Thesis Summary

This thesis investigated the dynamics beyond agency detection in social interaction with human and artificial agents, both on a cognitive and motor level. For the first time, a novel temporal binding effect has been reported to arise for social actions (i.e., vocalisations), which although incapable of directly producing changes in the environment, can still generate predictable outcomes in someone else's behaviour (Chapter 2). Interestingly, agency detection for observed actions seems to depend on access to visuomotor information about the agent's movement, which suggest a potential involvement of the mirror neuron system. A direct investigation into agency detection for observed actions performed by a robot (Chapter 3) revealed, again for the first time, that it is possible to experience subjective temporal compression for embodied artificial agents. However, this implicit mechanism may be affected by explicit and contextual information, such as prior experience and belief about robots. In addition, explicit and contextual information about the robot's behaviour showed a significant impact over participants' mentalising and anthropomorphism instances, evidence that intention ascription, theory of mind, and anthropomorphism may be deeply interconnected. Finally, this thesis explored the role of bodily and mental representation in action representation in social contexts of action (Chapter 4). This study revealed that when representing other's actions, their physical appearance is not crucial to detect agency. On the other hand, representation of the agent's mental states and availability of action opportunity seem to play a more decisive role.

This thesis explored multiple elements relevant to agency detection in social interaction contexts, both between humans and with artificial agents. Ascribing agency to artificial agents may seem a contradiction, given that at their current stage of development, none of them truly has intentions, desires, or goals. However, investigating the circumstances and the features that can make artificial agents resulting as more agentic, can be an invaluable resource to achieve a smoother interaction. Innovative work presented in this thesis increases our insights into this field of research, which should further aim to understand cognitive ergonomics in social interaction. This will provide concrete assets to facilitate the acceptance of social robots in our daily environments, having them as friends rather than tools.

References

Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and cognition*, *14*(3), 439-458.

Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends in neurosciences*, *31*(6), 265-272.

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annual review of psychology*, *60*, 693.

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

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of cognitive neuroscience*, *21*(4), 791-802.

Amorim, M. A. (2003). " What is my avatar seeing?": The coordination of" out-of-body" and" embodied" perspectives for scene recognition across views. *Visual Cognition*, *10*(2), 157-199.

Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., & Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia*, *51*(8), 1619-1629.

Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: Sharing tasks with real and imagined co-actors. *Experimental Brain Research*, *211*, 371–385.

Aziz-Zadeh, L., & Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology-Paris*, *102*(1-3), 35-39.

Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology*, *48*(9), 1276-1283.

Baess, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology*, *70*(2), 137-143.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American psychologist*, *37*(2), 122.

Bainbridge, W. A., Hart, J., Kim, E. S., & Scassellati, B. (2008, August). The effect of presence on human-robot interaction. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 701-706). IEEE.

Bak, T. H., & Hodges, J. R. (1999). Cognition, language and behaviour in motor neurone disease: evidence of frontotemporal dysfunction. *Dementia and Geriatric Cognitive Disorders*, *10*(Suppl. 1), 29-32.

Bar-Haim, Y., Kerem, A., Lamy, D., & Zakay, D. (2010). When time slows down: The influence of threat on time perception in anxiety. *Cognition and Emotion*, *24*(2), 255-263.

Barlas, Z., & Obhi, S. S. (2013). Freedom, choice, and the sense of agency. *Frontiers in human neuroscience*, *7*, 514-514.

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition*, *104*(3), 644-653.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic bulletin & review*, *13*(6), 1061-1066.

Bayliss, A. P., & Tipper, S. P. (2006). Predictive gaze cues and personality judgments: Should eye trust you?. *Psychological Science*, *17*(6), 514-520

Bays, P. M., Flanagan, J. R., & Wolpert, D. M. (2006). Attenuation of self-generated tactile sensations is predictive, not postdictive. *PLoS Biol*, *4*(2), e28.

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in cognitive sciences*, *12*(7), 254-258.

Berberian, B., & Cleeremans, A. (2010). Endogenous versus exogenous change: Change detection, self and agency. *Consciousness and cognition*, *19*(1), 198-214.

Beyer, F., Sidarus, N., Fleming, S., & Haggard, P. (2018). Losing Control in Social Situations: How the Presence of Others Affects Neural Processes Related to Sense of Agency. *ENeuro*, *5*(1), ENEURO-0336.

Birks, M., Bodak, M., Barlas, J., Harwood, J., and Pether, M. (2016). Robotic seals as therapeutic tools in an aged care facility: a qualitative study. *J. Aging Res.* 2016, 1–7. doi: 10.1155/2016/8569602

Blake, K. R., & Gangestad, S. (2020). On attenuated interactions, measurement error, and statistical power: guidelines for social and personality psychologists. *Personality and Social Psychology Bulletin*, *46*(12), 1702-1711.

Blakemore, S. J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*(4), 267-277.

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.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport*, *12*(9), 1879-1884.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience*, *1*(7), 635-640.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2000). Why can't you tickle yourself?. *Neuroreport*, *11*(11), R11-R16.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in cognitive sciences*, *6*(6), 237-242.

Böckler, A., Knoblich, G., & Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1404.

Booth, J. R., Wood, L., Lu, D., Houk, J. C., & Bitan, T. (2007). The role of the basal ganglia and cerebellum in language processing. *Brain research*, *1133*, 136-144.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465-1477.

Bub, D. N., & Masson, M. E. (2010). Grasping beer mugs: on the dynamics of alignment effects induced by handled objects. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 341.

Buccino, G., Sato, M., Cattaneo, L., Rodà, F., & Riggio, L. (2009). Broken affordances, broken objects: a TMS study. *Neuropsychologia*, *47*(14), 3074-3078.

Buehner, M. J. (2012). Understanding the past, predicting the future: causation, not intentional action, is the root of temporal binding. *Psychological Science*, *23*(12), 1490-1497.

Buehner, M. J., & Humphreys, G. R. (2009). Causal binding of actions to their effects. *Psychological Science*, *20*(10), 1221-1228.

Capozzi, F., Becchio, C., Garbarini, F., Savazzi, S., & Pia, L. (2016). Temporal perception in joint action: This is MY action. *Consciousness and Cognition*, *40*, 26-33.

Cardellicchio, P., Sinigaglia, C., & Costantini, M. (2011). The space of affordances: a TMS study. *Neuropsychologia*, *49*(5), 1369-1372.

Cardellicchio, P., Sinigaglia, C., & Costantini, M. (2013). Grasping affordances with the other's hand: a TMS study. *Social cognitive and affective neuroscience*, *8*(4), 455-459.

Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion changes the sense of agency in the human brain. *Current biology*, *26*(5), 585-592.

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, *12*(3), 314-325.

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports*, *6*(1), 1-8.

Chambon, V., Wenke, D., Fleming, S. M., Prinz, W., & Haggard, P. (2012). An online neural substrate for sense of agency. *Cerebral Cortex*, *23*(5), 1031–1037.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, *12*(4), 478-484.

Christensen, J. F., Yoshie, M., Di Costa, S., & Haggard, P. (2016). Emotional valence, sense of agency and responsibility: A study using intentional binding. *Consciousness and cognition*, *43*, 1-10.

Church, W., Ford, T., Perova, N., and Rogers, C. (2010). "Physics with robotics: using Lego Mindstorms in high school education," in *Proceedings of Advancement of Artificial Intelligence Spring Symposium*, Stanford, CA, 47–49.

Ciardo, F., Beyer, F., De Tommaso, D., & Wykowska, A. (2020). Attribution of intentional agency towards robots reduces one's own sense of agency. *Cognition*, *194*, 104109.

Ciardo, F., De Tommaso, D., Beyer, F., & Wykowska, A. (2018, November). Reduced sense of agency in human-robot interaction. In *International conference on social robotics* (pp. 441-450). Springer, Cham.

Ciardo, F., Lugli, L., Nicoletti, R., Rubichi, S., & Iani, C. (2016). Action-space coding in social contexts. *Scientific reports*, *6*(1), 1-8.

Ciardo, F., & Wykowska, A. (2018). Response coordination emerges in cooperative but not competitive joint task. *Frontiers in psychology*, *9*, 1919.

Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., & Committeri, G. (2010). Where does an object trigger an action? An investigation about affordances in space. *Experimental brain research*, *207*(1-2), 95-103.

Costantini, M., Committeri, G., & Sinigaglia, C. (2011). Ready both to your and to my hands: mapping the action space of others. *PloS one*, *6*(4), e17923.

Cross, E. S., Liepelt, R., de C. Hamilton, A. F., Parkinson, J., Ramsey, R., Stadler, W., & Prinz, W. (2012). Robotic movement preferentially engages the action observation network. *Human brain mapping*, *33*(9), 2238-2254.

David, N., Newen, A., & Vogeley, K. (2008). The sense of agency and its underlying cognitive and neural mechanisms. *Consciousness and cognition*, *17*(2), 523-534.

Davidson, P. R., & Wolpert, D. M. (2005). Widespread access to predictive models in the motor system: a short review. *Journal of Neural Engineering*, *2*(3), S313.

Deecke, L., Scheid, P., & Kornhuber, H. H. (1969). Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements. *Experimental brain research*, *7*(2), 158-168.

Dennett, D. C. (1989). *The intentional stance*. MIT press.

Desantis, A., Roussel, C., & Waszak, F. (2011). On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition*, *20*(4), 1211-1220.

Dewey, J. A., & Carr, T. H. (2013). When dyads act in parallel, a sense of agency for the auditory consequences depends on the order of the actions. *Consciousness and cognition*, *22*(1), 155-166.

Dewey, J. A., & Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing?. *PloS one*, *9*(10), e110118.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176-180.

Dirnberger, G., Fickel, U., Lindinger, G., Lang, W., & Jahanshahi, M. (1998). The mode of movement selection Movement-related cortical potentials prior to freely selected and repetitive movements. *Experimental Brain Research*, *120*(2), 263-272.

Doi, H., & Shinohara, K. (2009). The perceived duration of emotional face is influenced by the gaze direction. *Neuroscience letters*, *457*(2), 97-100.

Droit-Volet, S., Brunot, S., & Niedenthal, P. (2004). BRIEF REPORT Perception of the duration of emotional events. *Cognition and Emotion*, *18*(6), 849-858.

Droit-Volet, S., Fayolle, S., Lamotte, M., & Gil, S. (2013). Time, emotion and the embodiment of timing. *Timing & Time Perception*, *1*(1), 99-126.

Droit-Volet, S., & Meck, W. H. (2007). How emotions colour our perception of time. *Trends in cognitive sciences*, *11*(12), 504-513.

Eliades, S. J., & Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of neurophysiology*, *89*(4), 2194-2207.

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British journal of psychology*, *91*(4), 451-471.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews*, *24*(6), 581-604.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, *114*(4), 864.

Engbert, K., & Wohlschläger, A. (2007). Intentions and expectations in temporal binding. *Consciousness and Cognition, 16*(2), 255-264. Doi: 10.1016/j.concog.2006.09.010

Engbert, K., Wohlschläger, A., & Haggard, P. (2008). Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition, 107*(2), 693-704. Doi: 10.1016/j.cognition.2007.07.021

Engbert, K., Wohlschläger, A., Thomas, R., & Haggard, P. (2007). Agency, subjective time, and other minds. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1261-1268. Doi: 10.1037/0096-1523.33.6.1261

Farrer, C., Frey, S. H., Van Horn, J. D., Tunik, E., Turk, D., Inati, S., & Grafton, S. T. (2008). The angular gyrus computes action awareness representations. *Cerebral Cortex, 18*(2), 254–261.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, *308*(5722), 662-667.

Ford, J. M., Palzes, V. A., Roach, B. J., & Mathalon, D. H. (2013). Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. *Schizophrenia bulletin*, sbt072.

Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: auditory hallucinations in schizophrenia. *American Journal of Psychiatry*, *164*(3), 458-466.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, *133*(4), 694.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, *4*(1), 14-21.

Gallagher, S. (2012). Multiple aspects in the sense of agency. *New Ideas in Psychology*, *30*(1), 15-31.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage*, *35*(4), 1674-1684.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593-609.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Gentsch, A., Kathmann, N., & Schütz-Bosbach, S. (2012). Reliability of sensory predictions determines the experience of self-agency. *Behavioural brain research*, *228*(2), 415-422.

Gentsch, A., & Schütz-Bosbach, S. (2011). I did it: unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. *Journal of cognitive neuroscience*, *23*(12), 3817-3828.

Gentsch, A., & Synofzik, M. (2014). Affective coding: the emotional dimension of agency. doi.org/10.3389/fnhum.2014.00608

Gil, S., & Droit-Volet, S. (2012). Emotional time distortions: the fundamental role of arousal. *Cognition & emotion*, *26*(5), 847-862.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of sciences*, 423(1), 52-77.

Gibson, James J. (1979), *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Glasauer, S., Huber, M., Basili, P., Knoll, A., & Brandt, T. (2010, September). Interacting in time and space: Investigating human-human and human-robot joint action. In *19th International Symposium in Robot and Human Interactive Communication* (pp. 252-257). IEEE.

Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage*, *6*(4), 231-236.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *science*, *315*(5812), 619-619.

Greenberg, J., Pyszczynski, T., & Solomon, S. (1982). The self-serving attributional bias: Beyond self-presentation. *Journal of Experimental Social Psychology*, *18*(1), 56-67.

Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: an fMRI study of implicit processing. *European Journal of Neuroscience*, *17*(12), 2735-2740.

Grommet, E. K., Droit-Volet, S., Gil, S., Hemmes, N. S., Baker, A. H., & Brown, B. L. (2011). Time estimation of fear cues in human observers. *Behavioural Processes*, *86*(1), 88-93.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience, 9*(12), 934-946.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn.2017.14

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology*, *22*(10), R390-R392.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*(4), 382-385.

Haggard, P., & Tsakiris, M. (2009). The experience of agency feelings, judgments, and responsibility. *Current Directions in Psychological Science, 18*(4), 242-246.

Haggard, P., & Whitford, B. (2004). Supplementary motor area provides an efferent signal for sensory suppression. *Cognitive Brain Research, 19*(1), 52-58.

Hagura, N., Kanai, R., Orgs, G., & Haggard, P. (2012). Ready steady slow: action preparation slows the subjective passage of time. *Proceedings of the Royal Society B: Biological Sciences, 279*(1746), 4399-4406.

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. *Annual review of psychology*, *68*, 243-267.

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

Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social Cognition in the Age of Human–Robot Interaction. *Trends in Neurosciences*

Hesse, M. D., Nishitani, N., Fink, G. R., Jousmäki, V., & Hari, R. (2010). Attenuation of somatosensory responses to self-produced tactile stimulation. *Cerebral Cortex*, *20*(2), 425-432.

Hon, N., Poh, J. H., & Soon, C. S. (2013). Preoccupied minds feel less control: Sense of agency is modulated by cognitive load. *Consciousness and cognition*, *22*(2), 556-561.

Hornik, J. (1981). Time cue and time perception effect on response to mail surveys. *Journal of Marketing Research*, 243-248.

Hortensius, R., Hekele, F., & Cross, E. S. (2018). The perception of emotion in artificial agents. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 852-864.

Horváth, J., Maess, B., Baess, P., & Tóth, A. (2012). Action–sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. *Journal of cognitive neuroscience*, *24*(9), 1919-1931.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495-514.

Howard, E. E., Edwards, S. G., & Bayliss, A. P. (2016). Physical and mental effort disrupts the implicit sense of agency. *Cognition*, *157*, 114-125.

Howell, D. C. (2002) *Statistical Methods for Psychology, 5th ed.*. Belmont, CA: Duxbury Press.

Hughes, G., Desantis, A., & Waszak, F. (2013). Attenuation of auditory N 1 results from identity-specific action-effect prediction. *European Journal of Neuroscience*, *37*(7), 1152-1158.

Hughes, G., & Waszak, F. (2011). ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. *Neuroimage*, *56*(3), 1632-1640.

Humphreys, G. R., & Buehner, M. J. (2010). Temporal binding of action and effect in interval reproduction. *Experimental Brain Research*, *203*(2), 465-470.

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *Neuroimage*, *21*(3), 1167-1173.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1-2), 101-144.

Innocenti, A., De Stefani, E., Bernardi, N. F., Campione, G. C., & Gentilucci, M. (2012). Gaze direction and request gesture in social interactions. *PLoS one*, *7*(5), e36390.

Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioural brain research*, *142*(1-2), 1-15.

Jo, H. G., Wittmann, M., Borghardt, T. L., Hinterberger, T., & Schmidt, S. (2014). First-person approaches in neuroscience of consciousness: brain dynamics correlate with the intention to act. *Consciousness and cognition*, *26*, 105-116.

Kajopoulos, J., Wong, A. H. Y., Yuen, A. W. C., Dung, T. A., Tan, Y. K., and Wykowska, A. (2015). "Robot-assisted training of joint attention skills in children diagnosed with autism," in *Lecture Notes in Artificial Intelligence*, eds G. Randy, T. Yuzuru, and W. Wolfgang (Berlin: Springer), 296–305. doi: 10.1007/978-3-319-25554-5-30

Khalighinejad, N., Di Costa, S., & Haggard, P. (2016). Endogenous Action Selection Processes in Dorsolateral Prefrontal Cortex Contribute to Sense of Agency: A Meta-Analysis of tDCS Studies of "Intentional Binding." *Brain Stimulation*. https://doi.org/10.1016/j.brs.2016.01.005

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

Knolle, F., Schröger, E., Baess, P., & Kotz, S. A. (2012). The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *Journal of Cognitive Neuroscience*, *24*(3), 698-706.

Knolle, F., Schröger, E., & Kotz, S. A. (2013). Prediction errors in self-and externally-generated deviants. *Biological psychology*, *92*(2), 410-416.

Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, G., & Wykowska, A. (2018). On the role of eye contact in gaze cueing. *Scientific reports*, *8*(1), 1-10.

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

Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of personality and social psychology*, *90*(4), 543.

Kühn, S., Brass, M., & Haggard, P. (2013). Feeling in control: Neural correlates of experience of agency. *Cortex*, *49*(7), 1935–1942.

Kühn, S., Nenchev, I., Haggard, P., Brass, M., Gallinat, J., & Voss, M. (2011). Whodunnit? Electrophysiological correlates of agency judgements. PLoS One, 6(12), e28657.

Kunde, W., Weller, L., & Pfister, R. (2018). Sociomotor action control. Psychonomic Bulletin & Review, 25(3), 917-931.

Lambrey, S., Amorim, M. A., Samson, S., Noulhiane, M., Hasboun, D., Dupont, S., ... & Berthoz, A. (2008). Distinct visual perspective-taking strategies involve the left and right medial temporal lobe structures differently. *Brain*, *131*(2), 523-534.

Lange, K. (2011). The reduced N1 to self-generated tones: An effect of temporal predictability?. *Psychophysiology*, *48*(8), 1088-1095.

Leslie, A. M. (1995). Pretending and believing: Issues in the theory of ToMM. *Cognition on cognition*, 193-220.

Levelt, W. J. (1999). Models of word production. *Trends in cognitive sciences*, *3*(6), 223-232.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). Brain, 106(3), 623-642.

Limerick, H., Moore, J. W., & Coyle, D. (2015). Empirical evidence for a diminished sense of agency in speech interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3967-3970).

Llinás, R. R. (2002). *I of the vortex: From neurons to self*. MIT press.

Luck, S.J.; Kappenman, E.S., eds. (2012). The Oxford Handbook of Event-Related Potential Components. Oxford: Oxford University Press.

Luria, A. R., & Yudovich, F. I. (1971). Speech and the Development of Mental Processes in the Child.

Maravita, A., Spence, C., Kennett, S., & Driver, J. (2002). Tool-use changes multimodal spatial interactions between vision and touch in normal humans. *Cognition*, *83*(2), B25-B34.

Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., & Wykowska, A. (2019). Do we adopt the intentional stance toward humanoid robots?. *Frontiers in psychology*, *10*, 450.

Martikainen, M. H., Kaneko, K. I., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral cortex*, *15*(3), 299-302.

Martin, R. F., Carlos, A. D., Jose Maria, C. P., Gonzalo, A. D., Raul, B. M., Rivero, S., et al. (2013). Robots in therapy for dementia patients. *J. Phys. Agents* 7, 49–56.

Mella, N., Conty, L., & Pouthas, V. (2011). The role of physiological arousal in time perception: psychophysiological evidence from an emotion regulation paradigm. *Brain and cognition*, *75*(2), 182-187.

Merchant, H., & Yarrow, K. (2016). How the motor system both encodes and influences our sense of time. *Current Opinion in Behavioral Sciences*, *8*, 22–27.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological bulletin*, *130*(5), 711.

Mitchell, R. W., Thompson, N. S., & Miles, H. L. (Eds.). (1997). *Anthropomorphism, anecdotes, and animals*. Suny Press.

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: a review of cue integration approaches. *Consciousness and cognition*, *21*(1), 59-68.

Moore, J., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and cognition*, *17*(1), 136-144.

Moore, J. W., Middleton, D., Haggard, P., & Fletcher, P. C. (2012). Exploring implicit and explicit aspects of sense of agency. *Consciousness and cognition*, *21*(4), 1748-1753.

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: a review. *Consciousness and Cognition*, *21*(1), 546-561.

Moore, J. W., Ruge, D., Wenke, D., Rothwell, J., & Haggard, P. (2010). Disrupting the experience of control in the human brain: pre-supplementary motor area contributes to the sense of agency. *Proceedings of the Royal Society of London B: Biological Sciences*. doi. 10.1098/rspb.2010.0404

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and cognition*, *18*(4), 1056-1064.

Moretto, G., Walsh, E., & Haggard, P. (2011). Experience of agency and sense of responsibility. *Consciousness and cognition*, *20*(4), 1847-1854.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9*(3), 184-211.

Norman, D. A. (1996). *La caffettiera del masochista. Psicopatologia degli oggetti quotidiani*. Giunti editore.

Oberman, L. M., McCleery, J. P., Ramachandran, V. S., & Pineda, J. A. (2007). EEG evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots. *Neurocomputing*, *70*(13-15), 2194-2203.

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) (2012), 587-588.

Obhi, S. S., & Hall, P. (2011a). Sense of agency and intentional binding in joint action. *Experimental Brain Research*, *211*(3-4), 655.

Obhi, S. S., & Hall, P. (2011b). Sense of agency in joint action: Influence of human and computer co-actors. *Experimental Brain Research*, *211*(3-4), 663-670.

Obhi, S. S., Swiderski, K. M., & Farquhar, R. (2013). Activating memories of depression alters the experience of voluntary action. *Experimental brain research*, *229*(3), 497-506.

Özdem, C., Wiese, E., Wykowska, A., Müller, H., Brass, M., & Van Overwalle, F. (2017). Believing androids–fMRI activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents. *Social Neuroscience*, *12*(5), 582-593.

Pacherie, E. (2014). How does it feel to act together?. *Phenomenology and the cognitive sciences*, *13*(1), 25-46.

Perez-Osorio, J., & Wykowska, A. (2019). Adopting the intentional stance towards humanoid robots. In *Wording Robotics* (pp. 119-136). Springer, Cham.

Péron, J., Le Jeune, F., Haegelen, C., Dondaine, T., Drapier, D., Sauleau, P., ... & Vérin, M. (2010). Subthalamic nucleus stimulation affects theory of mind network: a PET study in Parkinson's disease. *PLoS One*, *5*(3), e9919.

Pfeiffer, U., Schilbach, L., Timmermans, B., Jording, M., Bente, G., & Vogeley, K. (2012). Eyes on the mind: investigating the influence of gaze dynamics on the perception of others in real-time social interaction. *Frontiers in Psychology*, *3*, 537.

Pfister, R., Obhi, S. S., Rieger, M., & Wenke, D. (2014). Action and perception in social contexts: intentional binding for social action effects. *Frontiers in human neuroscience*, *8*, 667.

Poonian, S. K., & Cunnington, R. (2013). Intentional binding in self-made and observed actions. *Experimental Brain Research*, *229*(3), 419-427.

Poonian, S. K., Mcfadyen, J., Ogden, J., & Cunnington, R. (2015). Implicit agency in observed actions: evidence for N1 suppression of tones caused by self-made and observed actions. *Journal of cognitive neuroscience*, *27*(4), 752-764.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(4), 515-526.

Press, C. (2011). Action observation and robotic agents: learning and anthropomorphism. *Neuroscience & Biobehavioral Reviews*, *35*(6), 1410-1418.

Quadflieg, S., Flannigan, N., Waiter, G. D., Rossion, B., Wig, G. S., Turk, D. J., & Macrae, C. N. (2011). Stereotype-based modulation of person perception. *Neuroimage*, *57*(2), 549-557.

Ratcliffe, M. (2013). Depression and the phenomenology of free will. *The oxford handbook of philosophy and psychiatry*, 574-591.

Reed, C. L., & Farah, M. J. (1995). The psychological reality of the body schema: a test with normal participants. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(2), 334.

Ridderinkhof, K. R., Van Den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and cognition*, *56*(2), 129-140.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169-192.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., & Fazio, F. (1996). Localization of grasp representations in humans by PET: 1. Observation versus execution. *Experimental brain research*, *111*(2), 246-252.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron*, *31*(6), 889-901.

Roselli, C., Ciardo, F., & Wykowska, A. (2019, November). Robots improve judgments on self-generated actions: an Intentional Binding Study. In *International conference on social robotics* (pp. 88-97). Springer, Cham.

Sakreida, K., Effnert, I., Thill, S., Menz, M. M., Jirak, D., Eickhoff, C. R., ... & Binkofski, F. (2016). Affordance processing in segregated parieto-frontal dorsal stream sub-pathways. *Neuroscience & Biobehavioral Reviews*, *69*, 89-112.

Saito, N., Takahata, K., Murai, T., & Takahashi, H. (2015). Discrepancy between explicit judgement of agency and implicit feeling of agency: Implications for sense of agency and its disorders. *Consciousness and cognition*, *37*, 1-7.

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and cognition*, *17*(4), 1219-1227.

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, *94*(3), 241-255.

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human Movement Science*, *30*(5), 834-845.

Scorolli, C., Miatton, M., Wheaton, L. A., & Borghi, A. M. (2014). I give you a cup, I get a cup: a kinematic study on social intention. *Neuropsychologia*, *57*, 196-204.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*, 70–76.

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*, B11–B21.

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Co-representing stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1234–1246.

Sebanz, N., & Lackner, U. (2007). Who's calling the shots? Intentional content and feelings of control. *Consciousness and Cognition*, *16*(4), 859-876.

Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential?. *Clinical neurophysiology*, *117*(11), 2341-2356.

Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *International journal of human-computer studies*, *62*(2), 267-279.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In *Advances in psychology* (Vol. 65, pp. 31-86). North-Holland.

Simonsohn, U. (2014, March 12). No-way interactions [Data colada]. http://doi.org/10.15200/winn.142559.90552

Sirigu, A., Duhamel, J. R., & Poncet, M. (1991). The role of sensorimotor experience in object recognition: A case of multimodal agnosia. *Brain*, *114*(6), 2555-2573.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, *25*(1), 76-89.

Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, *11*(5), 543-545.

Sowman, P. F., Kuusik, A., & Johnson, B. W. (2012). Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Experimental brain research*, *222*(1-2), 149-157.

Stenzel, A., Chinellato, E., Bou, M. A. T., Del Pobil, Á. P., Lappe, M., & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1073.

Stephenson, L. J., Edwards, S. G., Howard, E. E., & Bayliss, A. P. (2018). Eyes that bind us: Gaze leading induces an implicit sense of agency. *Cognition*, *172*, 124-133.

Strother, L., House, K. A., & Obhi, S. S. (2010). Subjective agency and awareness of shared actions. *Consciousness and cognition*, *19*(1), 12-20.

Symes, E., Ellis, R., & Tucker, M. (2007). Visual object affordances: Object orientation. *Acta psychologica*, *124*(2), 238-255.

Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. (2010). Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain*, *133*(Pt 1), 262-271. Doi: 10.1093/brain/awp291

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and cognition*, *17*(1), 219-239.

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: an interplay between prediction and postdiction. *Frontiers in Psychology, 4*, 127. Doi: 10.3389/fpsyg.2013.00127

Tapus, A., Mataric, M. J., and Scasselati, B. (2007). Socially assistive robotics [Grand challenges of robotics]. *IEEE Robot. Autom. Mag.* 14, 35–42. doi: 10.1109/MRA.2007.339605

Tapus, A., Peca, A., Aly, A., Pop, C. A., Jisa, L., Pintea, S., et al. (2012). Children with autism social engagement in interaction with Nao, an imitative robot–A series of single case experiments. *Interact. Stud.* 13, 315–347. doi: 10.1075/is.13.3.01tap

Takahata, K., Takahashi, H., Maeda, T., Umeda, S., Suhara, T., Mimura, M., & Kato, M. (2012). It's not my fault: Postdictive modulation of intentional binding by monetary gains and losses. *PLoS One*, *7*(12), e53421.

Takayama, L., Ju, W., and Nass, C. (2008). "Beyond dirty, dangerous and dull: what everyday people think robots should do," in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, Amsterdam, 25–32.

Thomas, R., Press, C., & Haggard, P. (2006). Shared representations in body perception. *Acta psychologica*, *121*(3), 317-330.

Timm, J., SanMiguel, I., Keil, J., Schröger, E., & Schönwiesner, M. (2014). Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *Journal of cognitive neuroscience*, *26*(7), 1481-1489.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, *28*(5), 675-691.

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*, 353–362.

Tsakiris, M., & Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. *Experimental Brain Research*, *149*(4), 439-446.

Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition*, *96*(3), 215-231.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human perception and performance*, *24*(3), 830.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual cognition*, *8*(6), 769-800.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta psychologica*, *116*(2), 185-203.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124-129.

Tversky, A., & Kahneman, D. (1974). Heuristics and biases: Judgement under uncertainty. *Science*, *185*, 1124-1130.

Vainio, L., Tucker, M., & Ellis, R. (2007). Precision and power grip priming by observed grasping. *Brain and Cognition*, *65*(2), 195-207.

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., Hristova, D., Karlinsky, A., McEllin, L., Nijssen, S. R., Schmitz, L., & Wahn, B. (2017). Joint Action: Mental Representations, Shared Information and General Mechanisms for Coordinating with Others. *Frontiers in psychology*, *7*, 2039.

Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., & Fink, G. R. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of cognitive neuroscience*, *16*(5), 817-827.

Vosgerau, G., & Synofzik, M. (2012). Weighting models and weighting factors. *Consciousness and cognition*, *21*(1), 55-58.

Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain*, *133*(10), 3104-3112.

Wahn, B., Schwandt, J., Krüger, M., Crafa, D., Nunnendorf, V., & König, P. (2016). Multisensory teamwork: using a tactile or an auditory display to exchange gaze information improves performance in joint visual search. *Ergonomics*, *59*(6), 781-795.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219-232.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113-117.

Wang, Y., & Quadflieg, S. (2015). In our own image? Emotional and neural processing differences when observing human–human vs human–robot interactions. *Social cognitive and affective neuroscience*, *10*(11), 1515-1524.

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, *29*(5), 874-880.

Warren, Z. E., Zheng, Z., Swanson, A. R., Bekele, E., Zhang, L., Crittendon, J. A., et al. (2015). Can robotic interaction improve joint attention skills? *J. Autism Dev. Dis.* 45, 1–9. doi: 10.1007/s10803-013-1918-4

Wearden, J. H., Pilkington, R., & Carter, E. (1999). 'Subjective lengthening' during repeated testing of a simple temporal discrimination. *Behavioural Processes, 46*(1), 25-38.

Wegner, D. M. (2002). *The illusion of conscious will.* MIT Press.

Wegner, D, M. (2003). The mind's best trick: how we experience conscious will. *Trends in Cognitive Sciences, 7*(2), 65 – 69.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology, 86*(6), 838-848. Doi: 10.1037/0022-3514.86.6.838

Wegner, D, M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American psychologist, 54*(7), 480.

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011). The self in action effects: Selective attenuation of self-generated sounds. *Cognition*, *121*(2), 207-218.

Welsch, R. E. (1977). Stepwise multiple comparison procedures. *Journal of the American Statistical Association*, *72*(359), 566-575.

Welsh, T. N. (2009). When 1 1 1: The unification of independent actors revealed through joint Simon effects in crossed and uncrossed effector conditions. *Human Movement Science*, *28*, 726–737.

Wenke, D., & Haggard, P. (2009). How voluntary actions modulate time perception. Experimental brain research, 196(3), 311-318.

Wenke, D., Atmaca, S., Holländer, A., Liepelt, R., Baess, P., & Wolfgang, P. (2011). What is shared in joint action? Issues of co-representation, response conflict, and agent identification. *Review of Philosophy and Psychology*, *2*, 147–172.

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Frontiers in psychology*, *8*, 1663.

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I see what you mean: how attentional selection is shaped by ascribing intentions to others. *PloS one*, *7*(9), e45391.

Wilke, C., Synofzik, M., & Lindner, A. (2012). The valence of action outcomes modulates the perception of one's actions. *Consciousness and cognition*, *21*(1), 18-29.

Wittmann, M. (2013). The inner sense of time: how the brain creates a representation of duration. *Nature Reviews Neuroscience*, *14*(3), 217-223.

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in cognitive sciences*, *1*(6), 209-216.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. N*ature Neuroscience*, *3*, 1212-1217.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269*(5232), 1880.

Wohlschläger, A., Haggard, P., Gesierich, B., & Prinz, W. (2003). The perceived onset time of self-and other-generated actions. *Psychological Science*, *14*(6), 586-591.

Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150375.

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PloS one*, *9*(4), e94339.

Yamada, Y., & Kawabe, T. (2011). Emotion colors time perception unconsciously. *Consciousness and cognition*, *20*(4), 1835-1841.

Yamazaki, R., Christensen, L., Skov, K., Chang, C., Damholdt, M., Sumioka, H., et al. (2016). Intimacy in phone conversations: anxiety reduction for danish seniors with hugvie. *Front. Psychol.* 7:537. doi: 10.3389/fpsyg.2016.00537

Yoshie, M., & Haggard, P. (2013). Negative emotional outcomes attenuate sense of agency over voluntary actions. *Current Biology, 23*(20), 2028-2032. Doi: 10.1016/j.cub.2013.08.034

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychonomic Bulletin & Review*, *16*(6), 1089-1093.

Appendices

Appendix A

Mind attribution scale (adapted from Kozak et al., 2006)

After each block in Experiment 5, participants completed the scale referring to the agent who executed the action in that block. Thus, four different version of the scale was created, one for the Operant condition (i.e., "you"), one for the Observed Human condition (i.e., "the experimenter"), one for the Observed Robot condition (i.e., "Cozmo"), and one for the Control condition (i.e., "the computer").

"*Please rate from 1 (strongly disagree) to 9 (strongly agree) the extent to which you agree with the following statements.*"

Emotion

a.___________ has complex feelings.

b.___________ can experience pain.

c.___________ is capable of emotion.

d.___________ can experience pleasure.

Intention

e.___________ is capable of doing things on purpose.

f.___________ is capable of planned actions.

g.___________ has goals.

Cognition

h.___________ is highly conscious.

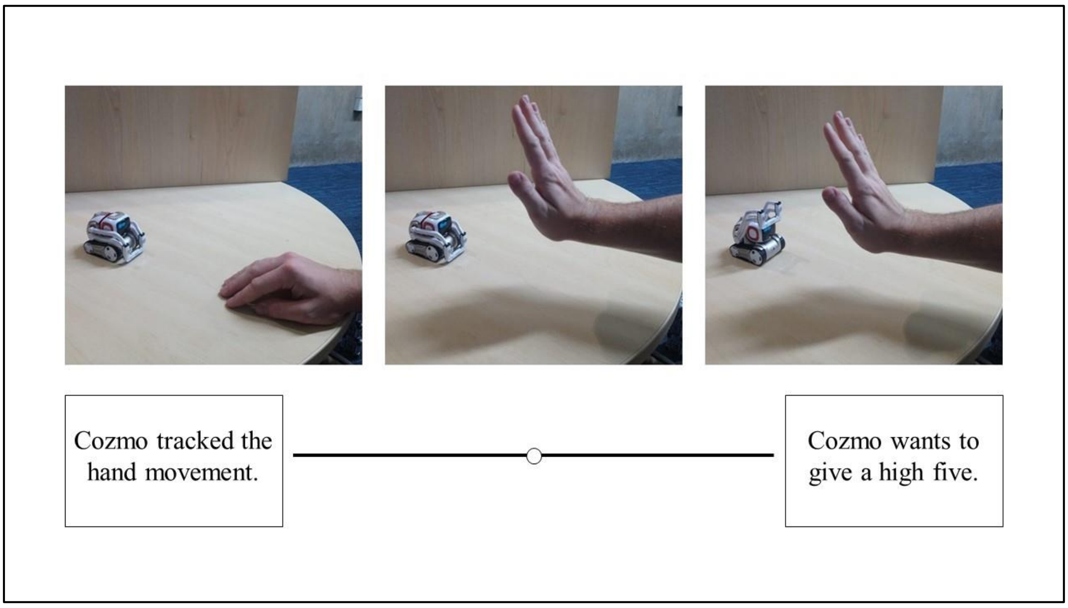
i.___________ has a good memory.

j.___________ can engage in a great deal of thought.

Appendix B

InStance questionnaire (adapted from Marchesi et al., 2019)

InStance questionnaire adopted in Experiment 5. Each of the five items featured a scenario composed of three pictures, depicting Cozmo interacting with objects and/or humans. Below the three images, each item included two judgements on the polar sides of a slider. One of the judgements explained Cozmo's actions referring to the design stance (i.e., mechanistic explanation), whereas the other described Cozmo's actions by referring to the intentional stance (i.e., mentalistic explanation). Participants' responses were recorded through the slider, which featured a hidden bipolar 0-100 scale. The value 0 corresponded to a complete mechanistic explanation (design stance) whereas 100 corresponded to a complete mentalistic explanation (intentional stance). The null value of the scale (i.e., the value in between the two) was set to 50 and as the initial position of the cursor.

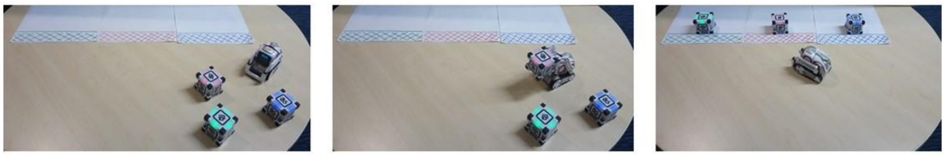


Cozmo categorises cubes by their color.

Cozmo likes the green cube.

Cozmo tracked the cube position and moves it to the assigned position.

Cozmo wants to keep its cube and not let the hand steal it.



Cozmo classifies cubes by colour.

Cozmo knows it is better to clean the table after playing.

Cozmo tracked the finger movement.

Cozmo understood that the person wants the red cube.