The sense of agency

when working with human and robotic partners

Thomas Peter Burke

100017801

School of Psychology

May 2024

Foreword and Acknowledgments

When first studying the sense of agency, it was something that immediately captivated me. Without understanding why, we can recognise our own actions and their effect on our environment. This seemingly trivial phenomenon is one of the fundamental parts of our conscious experience. It is an aspect of our mind which is so unnoticeable that you might go your whole life without considering its existence. Yet, it gives us the ability to feel freedom and experience our effect on the world. However, it also burdens us with knowing what it is like to be caged. Many of us might have felt times in our life where we feel we have been stripped of our agency. Whether chronic or acute, in those times we might describe ourselves as helpless, powerless, or useless. Yet other times, we might have also felt that we could accomplish anything. This dichotomy illustrates something. It shows us that our sense of agency is not fixed. We are not born to be useless, helpless, or powerless, and are not determined to be that way. We are born to act on the world and feel its reaction. And so, in our darkest moments, we will always have the capacity within ourselves to turn on the lights. From the research of many talented people, it seems that our agency is susceptible to how we are treated by others. Meaning, we seem to be able to affect how much agency others might feel in the world. This is both extraordinary and terrifying. If agency is tantamount to freedom, being able to alter another's experience of this is a responsibility that I don't think we take very seriously. We can imprison or liberate people, simply by our words. As Nelson Mandela put it, "For to be free is not merely to cast off one's chains, but to live in a way that respects and enhances the freedom of others". So, in the knowledge that I have gained while writing this thesis, I will endeavour to enable others better, even if in just small amounts.

On that note, I would like to take the time to acknowledge the people around me, who gave me the agency and purpose which enabled me to complete this thesis. Undoubtably without them what you are reading here wouldn't exist. First, I would like to thank my partner, Christiana Markou, for her enduring love and support throughout the years of my PhD and for making my waking moments better every day. I would also like to thank her family and my own family. I would like to especially thank my mum, who has supported me through this PhD and through my whole life. Finally, I would like to thank my supervisory team, particularly Andrew Bayliss, who has been a source of inspiration, patience and understanding throughout the years. I sincerely thank them for everything they have done for me.

Abstract

When we perform an action, we are aware that we are the one acting on our environment. This experience is known as the “sense of agency”. Research has shown that the sense of agency is not an all-or-nothing construct but instead is malleable and graded. As such, it can be increased and decreased by internal and external factors. More recently, researchers have sought to understand how the sense of agency can be manipulated through social context. This thesis was written to expand on previous research within this area.

Experiments 1-3 looked to understand how explicit agency can be altered by working with a partner. In each experiment, we manipulated whether the participants would be told that they were working with a human, a robotic avatar, or a preprogrammed computer. We asked participants how much control they felt over their action on each trial. We found that consistently people report control over their actions as lower when working with a partner, irrespective of the type of partner that they are working with.

Due to the COVID-19 outbreak, experiments in this thesis were all conducted online, The sense of agency can be measured through both implicit and explicit measures. Experiment 4 looks to understand whether temporal binding can occur through the interval estimation methodology, an implicit measure of the sense of agency, in an online environment. The results show that this is indeed the case.

Experiments 5-8 investigated how working with a partner could affect temporal binding, along with explicit agency. In a paradigm analogous to Experiment 1-3, we found that working with a partner reduced explicit agency compared with working alone, implicit agency (temporal binding) increased. We found that this temporal binding increase seems to be specific to the partner being able to act, rather than a mere presence effect.

Experiments 9-10 were further adaptations of Experiments 1-3, looking specifically at explicit agency. Experiment 9 investigated whether explicit agency when working with a partner can be affected by whether the participants are playing for their own points or not. We found that explicit agency was lowered by working for points that were not their own, along with the reduction in agency from working with a partner. Finally, Experiment 10 investigated how explicit agency when working with a partner is formed in moral scenarios. We found that the partner effect persists in paradigms with moral contexts. Furthermore, people still report agency over non-actions that have moral consequence. The results in this thesis expand the literature on the sense of agency and will contribute to a better understanding of our own experience when working with others.

Contents

List of Figures

Thesis Outline

Aims and Objectives

The aim of this thesis is to understand how the feeling of control over our own actions (“the sense of agency”) is affected by working with others. Specifically, the research contained here explores how working with both human partners and artificial agents can affect how we perceive and make judgments of our own agency under various conditions. First, we investigate how the judgments of our own agency are affected by working with a partner, with varying degrees of intentionality. Since an online environment was used to conduct the research, we next explore whether our implicit measure of the sense of agency (“temporal binding”) is possible to be produced online. After this, we use temporal binding to assess how working with partners marked as computers or robotic agents affects our implicit and explicit sense of agency. Finally, we explore how manipulating the relevance and the moral context of an outcome affects how we make judgments of agency.

Purpose of Investigation

The primary purpose of this investigation is to understand how our experience of control over our own actions is affected by working with others. In everyday life, we often experience performing tasks in groups. Such tasks can be as benign as choosing what you and your partner are going to eat for dinner, or as critical as performing surgery. Research into the sense of agency when working with others has suggested that our own feeling of control is modulated by performing tasks with others around us. Understanding how, when, and why this modulation occurs will be helpful in producing positive outcomes in shared tasks when it matters most.

In addition, this thesis also looks at how performing tasks with a computer and artificial agents can modulate agency over our own actions. As technology rapidly develops, we are seeing the presence of machines assisting with our daily activities. We can see this increasingly in cars, navigation, smart home devices and many other aspects of our day to day lives. Artificial intelligence has made massive strides over the last decade and robots are being developed which 30 years ago would be considered characters in a science fiction novel. It is an inescapable reality that our lives are going to be dramatically impacted by such developments in technology. The current understanding of the processes underpinning our interactions with machines is currently quite poor. The thesis in this research helps to contribute a small amount to the development of that knowledge. The following chapters first review current research on the sense of agency and its role in social interaction before outlining and discussing the research conducted for this thesis.

COVID-19 Impact Statement

As this doctoral programme started in the autumn preceding the COVID-19 outbreak, it was significantly impacted by the pandemic. The original intention of the programme was to conduct eye tracking experiments in the laboratories at UEA. However, mandated lockdowns by the government to mitigate spread of the virus meant that access to laboratories was not possible until May 2021 and even after opening they were severely restricted. Not knowing when the laboratories were going to open again (and if they did, whether they would stay open), this meant that the initial plan for the PhD had to change and instead, a new programme was created with the intention of conducting experiments online. The significant time lost as a result of the pandemic meant that there was not time to conduct some of the experiments that were planned.

Chapter 1 – The of Sense of Agency

Introduction

When a healthy individual acts, they generally feel that they are the instigator of that action and the cause of its effect. When I pick up a stone and throw it, I am aware that I was the author and the cause of the stone's trajectory. I do not attribute the stone moving to anything (or anyone) else. This phenomenological experience is known as the sense of agency (Gallagher, 2000). The sense of agency is thought to be integral in how we experience consciousness, through mechanisms that facilitate the distinction between the self and the other – the so-called 'Who' system (Georgieff & Jeannerod, 1998; Vignemont & Fourneret, 2004). Researchers have discussed the role of the sense of agency in formulating the 'Minimal self'. This is the immediate, pre-reflective sense of self that is rooted in embodied experience and agency (Gallagher, 2000). The sense of agency is automatic and is pervasive in almost all the actions we perform. It allows us to consciously detect the efficacy that we have over our environment and is sufficiently accurate for self-recognition of our actions (Jeannerod, 2003). This chapter will review literature on the sense of agency, looking at how researchers might believe it is formulated, how it is measured and how it might be manipulated.

Accounts of the Sense of Agency

It is currently thought that the sense of agency arises from two components (Figure 1), one being the feeling of agency and the second being judgments of agency (Synofzik et al., 2008; Synofzik, Vosgerau & Newen, 2012). The feeling of agency is thought to be the low-level feeling of being the author of one's actions and based on sensorimotor feelings. It is also thought to be pre-reflective (Gallagher, 2000) and non-conceptual. Alternatively, the judgments of agency are the conceptual judgments that one is an agent. For example, if you picked up a cup and were asked if you picked up the cup, your attribution that you indeed picked it up would be a judgment of agency. While the feeling of agency is automatic and experienced coherently in the flow of consciousness, judgments of agency are only realised when an agent becomes explicitly aware of their own volition over their action (Pacherie, 2001). Both the feeling of agency and judgments of agency appear to be open to modulation. For example, Desantis, Roussel & Wazak (2011) found that the feeling of agency was modulated by prior beliefs of causality. This suggests that the feeling of agency can be affected by top-down cognition and higher order cognitive processes. Likewise, judgments of agency can be affected by beliefs and distortions to sensory feedback such as temporal delays or low-resolution spatial feedback (Farrer et al., 2003; Wegner 2003;

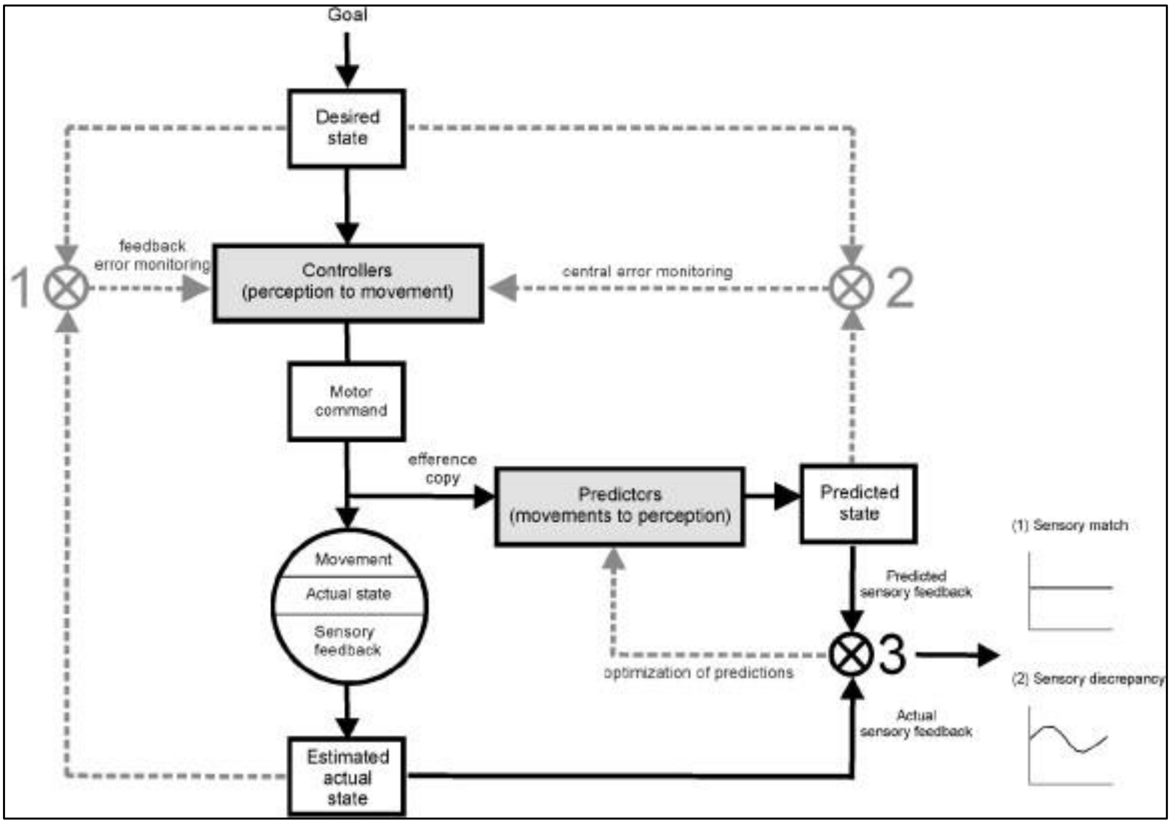

Figure 1- The comparator model for action monitoring from Synofzik, Vosgerau and Newen (2008). Comparisons between the desired state and actual state provides an estimation of motor error as shown in comparator 1. In comparator 2, a prediction of one's own movement or behaviour is sent to the forward model, before any sensory feedback is received. In comparator 3, when motor command is submitted, an efference copy is sent to a forward model which is used to predict sensory outcomes from an action.

Metcalfe & Greene, 2007). It is thought that the feeling of agency and judgments of agency are deeply interconnected, but it seems that the feeling of agency alone is not sufficient in producing judgments of agency (Haggard & Tsakiris, 2009). Furthermore, evidence suggests that the two can dissociate (Moore et al., 2012). Such evidence suggests that while feelings of agency and judgments of agency normally arise together, they are in fact separate systems that require different conditions to compute.

Synofzik et al. (2008) suggest that the feeling of agency is affected by the mismatch between predicted actions and their actual sensory feedback. Such an account is derived from the comparator model (Blakemore, Wolpert & Frith, 1998; Frith et al., 2000; Blakemore et al., 2002). Here, the sense of agency arises when actual sensory feedback is matched by the predicted outcome (*Figure* 2). According to the comparator model, when an action is to be performed, a motor intention is produced, which represents the target position of the body. Two motor commands are then produced by what is known as the inverse model. One will be sent to perform the action, the other, known as an "efference copy", will be sent to the forward model at the same time as the other motor command, which will predict the

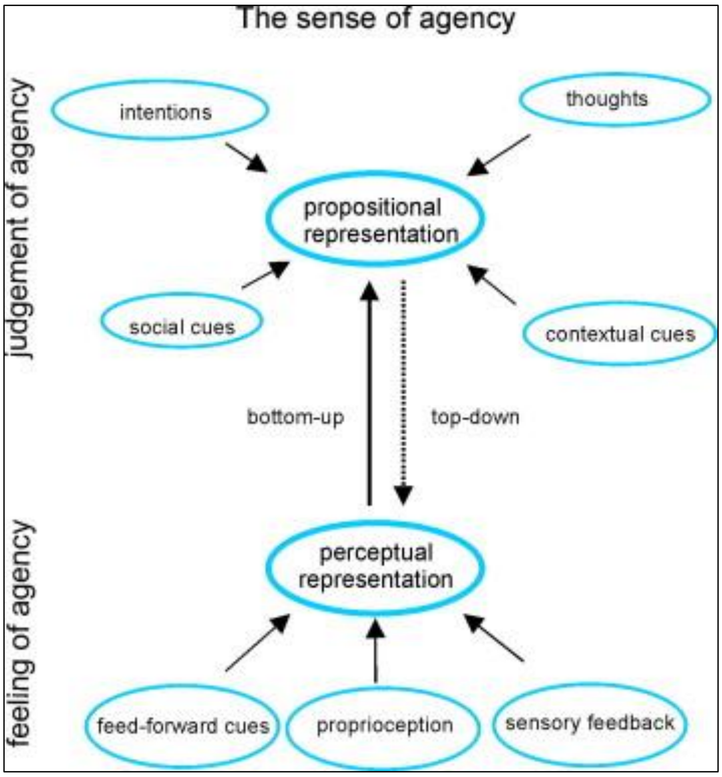

Figure 2 – Two step account of agency from Synofzik, Vosgerau and Newen (2008).

expected sensory outcomes of the action. Once the action is complete, sensory feedback from the motor action will be compared to the predicted sensory outcome and if the two are congruent, then the sense of agency is produced. If the efference copy does not match actual sensory feedback, then the sense of agency is not produced. In fact, on this account the sense of agency is *only ever* produced if sensory feedback is identical to the efference copy (Carruthers, 2012). The comparator model is thought to contribute to explaining the feeling of agency but is not sufficient or necessary to explain judgments of agency (Synofzik et al., 2008; Haggard & Tsakiris, 2009).

A contrasting theory to the comparator model that may help explain judgments of agency is the 'Theory of apparent mental causation' (*Figure 3*), which was first developed by Wegner and Wheatley (1999). This account suggests that the sense of agency arises out of a post-stimulus reconstruction of an action. In other words, the outcome of an action is perceived first, then ownership of the action is retrospectively administered. Unlike the comparator model, Wegner (2003) suggests that the sense of agency is conceptual and is formed from high level interpretative processes. Wegner (2002) outline three key components of action-outcome contingencies that together form the sense of agency: consistency, priority and exclusivity. Consistency means that there needs to be an intention to explain an action. We need to have an intention to perform an action to feel authorship over it. Priority means that intention has to occur before the action is performed. Finally, exclusivity means that the actor needs to be the only plausible cause of the action and its consequence. So, we first need to develop an intention before an action occurs and then there be little or no ambiguity that our action caused the desired change on the environment. This account of the sense of agency is reconstructionist and postdictive, meaning that the sense of agency is developed after the action occurs and was in part to argue against the idea of conscious free will (Wegner, 2002). Support for this theory comes from research showing that priming thoughts about an action generate a sense of agency even when the participant did not perform the action themselves (Wegner & Wheatley, 1999). Extending this research, Wegner, Sparrow & Winerman (2004) showed that internal motor contributions were not necessary to generate self-attributions of agency. The authors designed an experiment where participants were paired with another participant (called "hand helpers") whose arms were placed in a position to make it seem like they were the participants arms. The participant was directed to look at a mirror while the hand helper's arms followed the instructions from headphones. Participants were told that they may or may not receive audio that is related to the instruction of the hand helper. The results showed that participants who received the instructions were significantly more likely to report that they felt control over the arms than those who did not. Furthermore, in follow-up studies they also found that consistent instructions to the participant led to a greater sense of agency than inconsistent instructions and also that galvanic skin responses were greater when people heard the primes than when they did not. The authors conclude that the sense of agency can be produced without internal motor contributions if two of the three components outlined above are satisfied (consistency and priority). So according to this account, judgments of agency are produced by the retrospective inference of causality over our actions, not from motor contributions.

The two accounts of the sense of agency outlined above both have empirical research to support them. However, taken in isolation the research seems to be conflicting. The comparator model alone cannot explain why we feel a sense of agency without contributions from motor regions (Wegner and Wheatley 1999; Wegner 2004) yet there is evidence to suggest that sensorimotor regions play a role. Furthermore, Wegner's account cannot explain how the feeling of agency arises, nor can it account for a number of biological and cognitive disadvantages of solely relying on this system. Synofzik, Vosgerau & Voss (2013) discuss how this system would produce the sense of agency late in the action process, would be prone to error, would solely rely on external information and runs exclusively on a conceptual level. Furthermore, while most studies into the sense of agency fail to separate these two accounts of the sense of agency (Haggard, 2017), Moore and Haggard (2008) found that both sensorimotor prediction and retrospective inference contribute to the sense of agency. In their study, they found that when action-outcome contingency had a high level of predictability, the participants produced markers of the sense of agency even when they did not produce an outcome from their button press. However, when these contingencies were not predictable, participants only produced such markers of the sense of agency when their action caused the outcome. Such research led Moore, Wegner and Haggard (2009) to propose a optimal cue integration model. This model posits that the sense of agency is derived from both predictive (e.g., the comparator model) and postdictive (e.g., theory of mental causation) cues. The reliance on each of these cues is determined by the reliability and availability of each (Synofzik, Vosgerau & Voss 2013). Such a theory accounts for the discrepancies between the comparator model and the theory of mental causation and suggests that both predictive and postdictive processes play a part in generating the sense of agency.

Measuring the Sense of Agency

There has been an explosion of research into the sense of agency over the last 20 years. At least part of this increase in number of publications is due to the development of measurements used to assess the conscious experience of volitional action. Research into the sense of agency previously relied on explicit self-reports, where researchers would typically ask participants to make judgments about how much control they felt over an action or whether they were the cause of the outcome (Wegner et al., 2004; Sato & Yasuda, 2005). Researchers would try to manipulate agency by creating ambiguity as to who is causing the outcomes (Farrer et al., 2008), or by creating noise in between visual and motor feedback

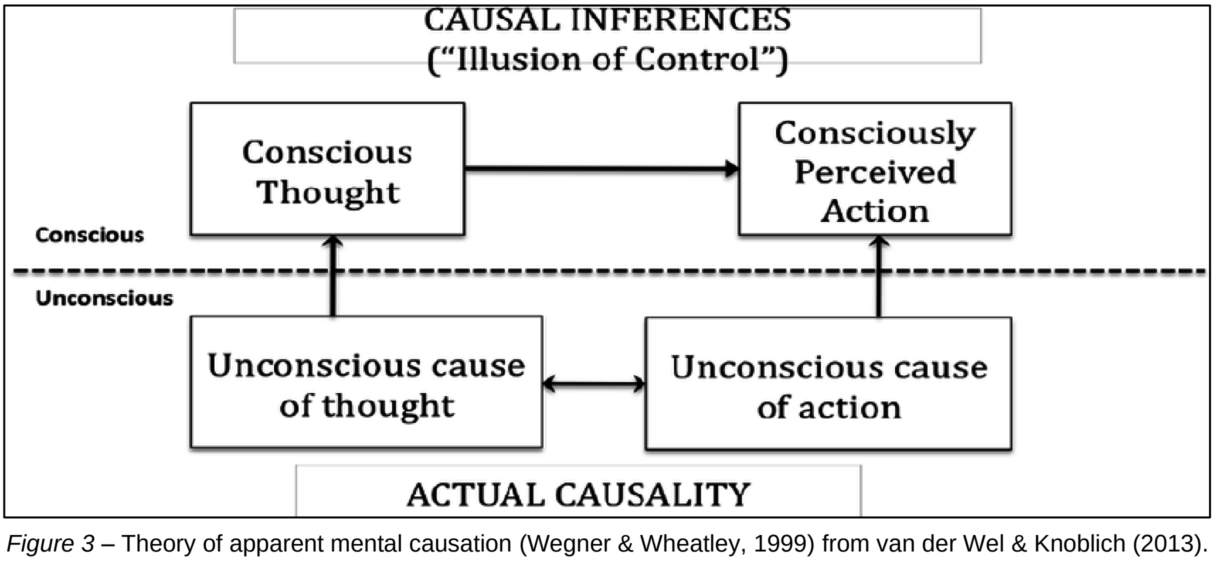

Figure 3 – Theory of apparent mental causation (Wegner & Wheatley, 1999) from van der Wel & Knoblich (2013).

(Metcalfe & Greene, 2007; Synofzik et al. 2010). It is also worth noting that how explicit agency is inquired can have qualitative differences over what aspect of the sense of agency we are capturing. For example, some experiments ask participants to rate how much control they felt over the *outcome* of an action (Chambon et al., 2014; Beyer et al., 2017; Beyer et al., 2018; Ciardo et al., 2020) and there are experiments which are related to control felt over their *action* (Wegner & Wheatley, 1999). These are subtle differences but are asking two different questions about the experience of volitional action. To date, these distinctions are not typically discussed and are poorly understood (Moore, 2016). There are benefits to using such a measure as they easily fit into experimental paradigms and are relatively straightforward and intuitive for experimenters. However, there are some limitations. Explicit agency can only ever measure judgments of agency as we are asking participants on a conceptual level to rate their conscious experience of their action. Therefore, if only using explicit measures, researchers do not have a measure of the feeling of agency in isolation. Furthermore, explicit agency ratings are subjective in nature. This leaves them prone to demand characteristics, such as implicit cues within the experiment or social desirability bias. It has also been found that participants often overattribute their own agency (Langer, 1975) and are open to the self-serving bias (Haggard, 2017). The self-serving bias is the tendency to attribute positive outcomes to oneself and to distance themselves from negative events (Bandura, 1982; Bandura, 1991). Finally, it is important to note that the experience of agency is phenomenologically thin (Haggard, 2005; Moore, 2016). That is, it occupies a space in consciousness that is not easily accessible. We might question how participants are supposed to rate their feeling of control over something that they might not be aware of. This is especially important considering that self-reports of the sense of agency are dependent on an individual’s level of introspection (Sebanz & Lackner, 2007). However, it has also been shown that the judgments of our own agency are reflective of the actual amount of control we have over our environment. Metcalfe and Greene (2007) asked participants to perform a task where they were required to click on targets as they moved down a screen, while avoiding others. Critically, the amount of control the participants had over the cursor varied, so that the cursor became less accurate and with more turbulence. Also, at times the cursor clicked on the target without input from the participant. Participants were then asked to rate their judgment of performance and their judgment of agency on each trial. They found that participants were able to successfully monitor their experience of agency independent of their performance. That is, they were able to monitor when they were less in control when the mouse was more turbulent and also able to not attribute causality to them themselves when they did not cause the outcome. This study suggests that we are consciously sensitive to, and able to reflect on, the factors which would be involved when monitoring our sense of agency.

The last point above notwithstanding, there is still a need for an objective measure to evaluate people's experience over volitional action. Such measures would focus on implicit agency, the non-conceptual aspect of our experience of our actions that the feeling of agency describes. One such measure was first discovered by Haggard, Clark & Kalogeras (2002). In their seminal paper, they found that both intentional actions and their resulting consequence were subjectively compressed in time by the participants. The experiment used an existing methodology from Libet et al. (1983), which was investigating something known as the readiness potential, a slow build-up of electrical potential in the brain that precedes the intention to make a movement (Schurger et al., 2021). Participants were asked to view a clock on a screen with a rotating hand. In the operant condition, participants were asked to press a button to generate a tone which occurred after a short delay. In two separate baseline conditions the computer produced the tone or their action led to no outcome. In each condition participants were asked to report where the clock hand was when the event occurred. They found that while in the baseline condition the events were

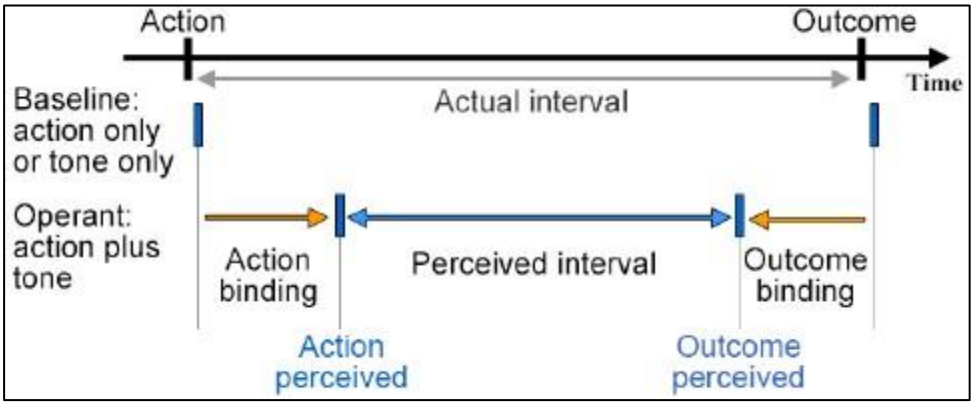

Figure 4 – Schematic of intentional binding using the Libet Clock methodology (Haggard, Clark & Kalogeras, 2002) from Limerick, Moore & Coyle (2015). In the Baseline condition, actions are produced without tones and tones are produced without actions. Participants are asked to time the onset of these events on a clock. In the Operant condition, the full causal chain of action and effect is present. Action binding represents a perceptual temporal shift towards the action's outcome compared with the veridical time interval. Outcome binding is the perception that the outcome of an action is drawn closer to the action in time.

reported accurately, the time was compressed in the operant condition when the action and effect were both present. Specifically, the action was perceived as delayed in time and the tone was rushed in time. Importantly, this compression did not occur from TMS stimulation causing the same action. Labelling this phenomenon as temporal binding, the authors concluded that volitional action and its effect are bound together by the brain to create a coherent experience of agency.

Since then, a vast amount of research has been conducted investigating this supposed marker of the sense of agency. Part of this research has questioned the nature of intentional binding, asking whether intentionality is necessary for the effect to occur. While it has been found that intentional binding is dependent on explicit intentional attributions (Engbert and Wohlschläger, 2007), other research suggests that intention is unnecessary to produce the effect (Kirsch, Kunde, and Herbort. 2019). Moreover, other researchers have proposed that this binding simply reflects the mental constructs of causality, suggesting that the term causal binding would be more appropriate (Buehner & Humphreys, 2009). Since it is up for debate on the necessary conditions to elicit binding, many authors use the term *temporal binding* (Engel & Singer, 2001). As this is a more conservative term that does not make claims about intentionality or what causes the temporal compression between action and effect, we will use this term henceforth. Since its inception, much of the research has used the Libet clock methodology to generate temporal binding. However, there are certain limitations to this methodology. For example, participants are required to divide their attention between the events and the clock, which could possibly create artifacts in the data (Engbert et al., 2007). Next, as the clock is required to be on screen during the trial, it limits experiments which might want to use that screen space for other stimuli (Silver, Tatler, Chakravarthi & Timmermans, 2020). Finally, Buehner and Humphreys (2009), pointed out that the Libet clock methodology was in an indirect measure of temporal binding, as it examines the timing of events relative to baseline, rather than the veridical time between events. This led the authors to develop an alternative measure for the sense of agency, which asks participants to replicate the interval between two events. In one condition, participants are required to perform an action which will lead to an event after a brief duration. In another condition, these two events occur independently of the participant's action. In both conditions, participants are required to estimate the interval between the two events occurring (either their action and its outcome or the two independent events). In this study however, participants were required to use an on-screen scale to estimate the time interval. As pointed out by Buehner and Humphreys (2010), this might lead to postdictive rationalisation or response bias. To remedy this, Buehner and Humphreys (2010) adapted the task so that participants were required to hold down a key to replicate the time interval, finding similar effects to their initial experiment. One criticism of this methodology is that participants will begin timing the interval after the event ends in the control condition. This lag in time might be sufficient to explain the temporal binding effect in this case. However, Humphreys and Buehner (2010) also demonstrated that operant intervals were still estimated as shorter even when taking this potential lag into account.

The Neural Basis of the Sense of Agency

The neural underpinnings of the sense of agency have been the focus of numerous studies to better understand the phenomenon from a biological standpoint. While the neural mechanisms are not well understood, evidence consistently seems to indicate the role of parietal brain regions in the sense of agency. One research method has involved studying individuals with damage to these regions. For instance, Sirigu et al. (2004) observed that patients with parietal lesions could detect the onset of movement but lacked awareness of their intention to act initially. Additionally, Daprati et al. (2000) found that damage to the right posterior parietal cortex led to difficulties in attributing actions to oneself and recognising one's own limbs. This suggests that parietal regions play a crucial role not only in initiating intentional actions but also in maintaining a sense of body ownership. Moreover, Desmurget et al. (2009) demonstrated that stimulating the inferior parietal cortex could evoke a desire to move among participants, whereas stimulation of the premotor cortex resulted in actual movement that participants denied initiating. These findings imply that the intention and awareness of movement originate in parietal areas, and that the sense of agency is constructed from both the intention to act and the anticipation of the sensory outcomes of the action. Research using transcranial direct current stimulation (TDCS) also supports this, showing that stimulating the angular gyrus and the right temporal parietal junction (TPJ) can diminish the sense of agency (Khalighhinejad & Haggard, 2015).) and the right temporal parietal junction (TPJ) could also attenuate agency. Furthermore, studies have examined the differences in neural responses between schizophrenia patients and healthy controls. For example, Whalley et al. (2004) noted increased resting activity in the inferior parietal lobe in patients with schizophrenia, which did not properly adjust to movement signals. This was also suggested as an early indicator of schizophrenia (Whalley et al., 2006).

On the other hand, there is substantial evidence indicating that these regions are not just associated with the positive aspects of agency but also with the sensation of losing control. A PET study by Ruby and Decety (2001) which analysed actions from both first- and third-person perspectives found that the right inferior parietal and somatosensory cortex play key roles in differentiating self-initiated actions from externally caused actions. Brain imaging research consistently shows activation in the right angular gyrus when participants experience a loss of agency (Farrer et al., 2003; Farrer et al., 2008). The TPJ, particularly, has been implicated more with external rather than self-agency, as highlighted in a meta-analysis by Sperduti et al. (2011) and in a review by Eddy (2016), which emphasised its role in self-other attributions.

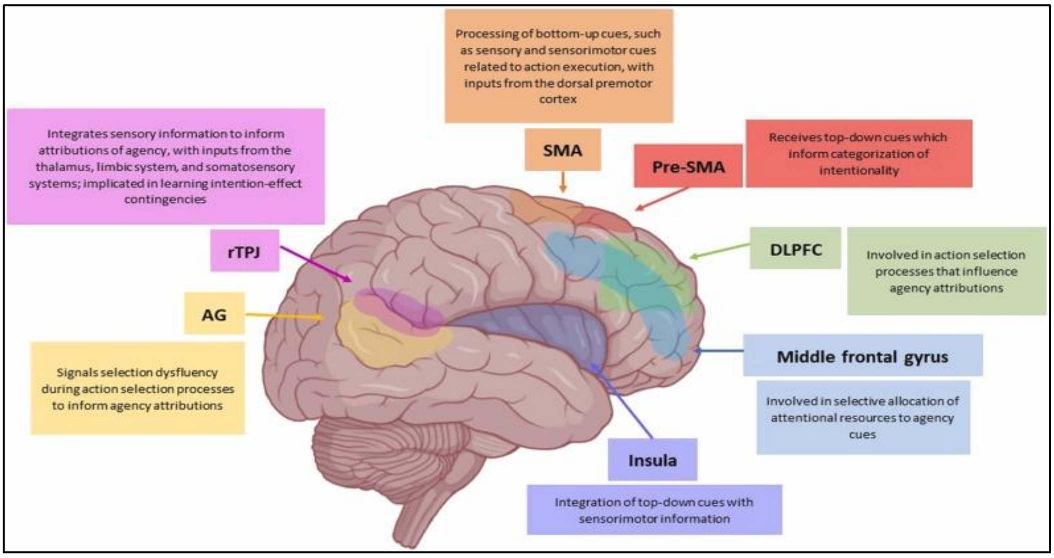

Figure 5 - Brain regions involved in the sense of agency from Malik, Galang and Finger (2022).

Furthermore, Miele et al. (2011), using fMRI, attempted to pinpoint the brain regions involved in agency judgments versus action monitoring, finding that regions like the pre-SMA, rostral cingulate zone, and dorsal striatum were more active during self-agency conditions. This contrasts with the rTPJ, which was active during unexpected outcomes. Their findings suggest that agency judgments arise from higher-level causal inferences post-event, rather than simple action monitoring. Chambon et al. (2012) further explored this by using subliminal primes to separate action-selection processes from action-outcome assessments. Their imaging studies revealed that the left angular gyrus is active in cases of discrepancies between intended actions and outcomes, correlating with a reported decrease in agency. This activity occurred prior to action execution, indicating a prospective monitoring role of the angular gyrus in the sense of agency. This aligns with views suggesting that metacognitive judgments of agency may partly rely on the computational functions associated with the TPJ (Decety and Lamm, 2007). See Figure 5 for a summary of brain regions related to the sense of agency.

Affect and the Sense of Agency

Over the past decade, numerous researchers have begun to investigate the role of emotion in the sense of agency. To our knowledge, one of the first studies that first instigated such a line of inquiry was conducted by Yoshie and Haggard (2013), who used human vocalisations as outcomes of actions. They found that temporal binding was reduced during negative outcomes (a negative emotional vocalisation), compared with neutral and positive outcomes. The results suggest that our sense of agency is sensitive to the emotional valence of the consequences of our actions. Extending these findings, Yoshie and Haggard (2017) developed a paradigm to manipulate the degree of outcome predictability. They found that temporal binding was reduced for predictable negative outcomes compared to predictable positive outcomes. However, there was no significant effect of valence when the outcome was unpredictable. The authors conclude that a specialized cognitive mechanism reduces the sense of agency when actions are known to have a negative effect. This is also supported by evidence that suggests that emotional expectation counteracts the prospective component of the sense of agency (Christensen et al. 2016). While there some conflicting evidence of the role of emotional valence on the sense of agency (Moreton & Hughes, 2017), the evidence outlined above indicates that our sense of agency is sensitive to the emotional valence of action outcomes. For social interaction, the ability to monitor how one's actions affect others is crucial. Thus, if the sense of agency evolved as a social mechanism, the fact that it is sensitive to emotional valence would not be unexpected. Researchers have sought to explain this modulation of agency through the self-serving bias (Campbell & Sedikidis, 1999). That is, the tendency for people to attribute positive outcomes to themselves and negative outcomes to others (Bandura, 1982). Such an effect has also been shown outside of emotional valence; the sense of agency has been found to be reduced when incurring monetary losses (Takahata et al. 2012). Thus, a reduction in the sense of agency for undesirable outcomes might distance oneself from one's own actions. This might serve to protect self-esteem and reduce feelings of shame or guilt.

The evidence generated so far suggests that agency and affective mechanisms can interact. There is also some research to suggest that this interaction is not limited to modulation of agency through outcome valence, but also the internal state of the actor. Christensen, Beck and Haggard (2019) found that the sense of agency was reduced when inducing angry or fearful moods into participants, compared to when their mood was neutral. Importantly, the outcome result was the same in each condition. These results suggest that the emotional state of an agent can affect the amount of control that they feel over their actions. Anger and fear are both linked with antisocial behaviour but aggression is often a necessary component of social hierarchies (Holekamp & Strauss, 2016; Stagkourakis et al., 2018). A reduction in the sense of agency during episodes of anger or fear might help to distance one's actions from undesirable, antisocial behaviour that is likely to emerge from such emotions. However, there are still gaps in the research so far. Researchers are still yet to investigate whether emotional valence modulates the sense of agency during social interaction (rather than using human vocalisations).

To help explain such data, Gentsch and Synofzik (2014) proposed an emotional dimension to the sense of agency known as affective coding. The authors argue that accounts like those outlined in this review fail to consider emotional states. They outline three primary ways that affect might affect the sense of agency. The first is prospective affective coding, where emotional and motivational states of an agent or the contextual information surrounding an action will affect agency felt before it occurs. Next, there is immediate affective coding, which are the fast and automatic emotion processes which occur during the action. Finally, retrospective affective coding suggests that the experience of agency is adapted based on the affective appraisal of the outcome. Such an account of the sense of agency is supported from research that indicates that emotional value of an action can affect sensorimotor processes relating to that action (Gentsch et al., 2015). Taking into account the research above, it has been shown that inducing different emotional states such as anger (Christensen et al., 2019) and depression (Obhi, Swiderski & Farquhar, 2013) affects temporal binding. This lends support to the idea that prospective affective coding, such as mood can affect the sense of agency. Furthermore, the fact that emotional valence of an outcome may also affect the sense of agency supports the idea of retrospective affective coding. Finally, support of the connection between action and affect could also come from clinical populations, for example, those with depression are less likely to act on their environment (Ratcliffe, 2013).

Global States of Agency

In the previous section, we touched on how emotional states might affect the sense of agency (Christensen et al. 2016; Obhi, Swiderski & Farquhar, 2013). Such research indicates that the human sense of agency is not constant across time and not simply contingent on outcomes of an action. It appears that our environment can affect the global state of an individual's sense of agency. That is, aspects outside of those specific to an action can affect the experience of volition over that action. A growing amount of evidence suggests that this is the case. Obhi and colleagues have conducted several experiments demonstrating that recalling episodes of depression, recalling memories of low power and also social exclusion reduces temporal binding (Obhi et al., 2013; Obhi et al., 2012; Malik & Obhi et al., 2012). This global effect on the sense of agency also does not seem be limited to proximal experimental manipulations. Caspar et al. (2020) demonstrated that being part of strict hierarchical organization may have an impact on how we experience agency. In their study, they looked at how temporal binding differs as a result of military experience. Among other findings, they found that privates in the military showed reduced temporal binding compared to senior cadets. Senior cadets are those who are trained to be officers in the military, who are likely to be trained to give orders and have higher accountability. This is opposed to privates, who are trained to take and follow orders without question. The implication of these results is that being on different levels of a social hierarchy has a differential impact on an individual's sense of agency. It also expands on Obhi and colleagues research by demonstrating that one's environment can affect the sense of agency even if it is not necessarily pertinent to the action.

Whether or not social hierarchies *per say* are affecting the sense of agency, or if this is related more to trait changes of an individual is up for debate. For example, it has also been shown that those who score low on the narcissistic personality inventory show reduced binding compared to those who score higher (Hascalovitz & Obhi, 2015) and those who show depressive traits produce less agency when presented with choice compared to those who do not (Scott et al., 2021). All together, these results indicate that our global sense of agency can be manipulated by proximal events unrelated to an action, social hierarchies and trait differences in narcissism and depression. This lends support to the idea that high order cognition can prospectively impact sensorimotor motor processes (Gentsch, & Synofzik, 2014).

The Sense of Agency and Human-Computer Interaction

Human-computer interaction (HCI) describes the study and design of computer technology and how we interface with it (Carroll, 1997). Since the advent of commercial and personal computers alike there has been a desire and indeed, an incentive to create machines that are intuitive for humans to use. Even in early commercial computers, the text-based interface was representative of a typewriter, where a user would type on a keyboard, and it would appear on the screen. However, users were still inputting commands to the computer through abstract commands and code. With the desire to move computers into the home, there was a need to simplify the interface further and move away from type-based commands. Graphical user interfaces (GUI) were popularised, and the adoption of the mouse and keyboard became widespread. The link between action and consequence became clearer. When you moved the mouse the cursor would move, when you held down on a window and moved the mouse the window would move as if you were holding it yourself. In more recent times, touch based devices such as smartphones take away the mouse and use your hands as the pointing device and haptic feedback provides tactile responses to our actions. Large technology companies are constantly producing new ways for us to engage with our devices. Interactions such as voice commands, augmented and virtual reality are becoming more widespread, and these interactions are becoming more frequent as computing moves into different domains like phones, automobiles, shopping and many more.

A challenge and goal of HCI is being able to translate psychological states of humans onto machines which operate functionally differently than the user and their environment (Limerick, Coyle & Moore, 2014). While our actions on computers appear to have direct consequences, this is an illusion by design. Simply typing on a keyboard produces a signal which is then encoded by the computer, interpreted, and then processed and visually rendered. Thus, there are a lot more steps that occur to create the causal link between action and effect, than say a typewriter, where the action of typing is mechanical and direct. Because of these steps, the causal link with computers can be altered or broken by aspects such as latency, reliability and congruency between the outcome and one's own action (see Limerick, Coyle & Moore, 2014 for review).

Studies typically show that the sense of agency is strongest at intervals where the time between action and effect are shortest (e.g. Farrer et al. 2003; Haggard, Clark & Kalogeras, 2002), demonstrating the impact that latency can have on the sense of agency. On the other hand, it has also been found that the sense of agency remains intact at intervals from two to four seconds (Berberian et al., 2012; Humphreys & Buehner, 2009). So, although it could be argued that latency reduces the sense of agency, even after seconds passing it does not diminish it entirely. While it is intuitive that a tight temporal link between an action and its effect would increase agency, as a larger temporal gap would increase the space for external events to be the cause, it would also be necessary for an agency system to be adaptable to longer delays as not all action effect pairings fall within millisecond time windows. Furthermore, the underlying mechanisms for how latency affects the sense of agency is currently poorly understood. In a virtual reality (VR) experiment, Waltemate et al. (2016) found that the sense of agency was reduced by latency between movement and visual feedback from their avatar. However, motor performance accuracy was a better predictor of loss of agency performance than the actual delay in feedback. In other words, when motor performance accuracy was high, the effect of latency was diminished. So, latency could reduce motor performance accuracy, leading to a reduced sense of agency and thus, delays in the same system might produce different effects on the sense of agency depending on the difficulty of the motor task.

In addition to the type of task potentially impacting the sense of agency, the type of action may also be a contributing factor in HCI. As touched on earlier, how we interact with computers is now quite uniform, with button presses (particularly a mouse and keyboard) or touch interfaces (mostly for mobile devices) dominating most of the computing market. However, other technologies such as VR, and voice and eye control are becoming adopted, which use action modalities that are distinct from button pushes or mouse tracking. It has been shown that for both voice (Pascolini et al., 2021) and eye control (Stephenson et al., 2018), we can experience agency over action-effect pairings. How these different ways of interacting with computers are cognitively or functionally different from one another is not well understood. However, it does appear that the type of action modality can indeed influence our experience of agency during HCI. Coyle et al. (2012) found that skin-based interaction, where a participant controls the computer through touching their own skin, leads to a significantly greater sense of agency (as measured by temporal binding) than when using a keyboard. Conversely, there is some evidence to also suggest that the sense of agency is weaker during voice commands compared to button presses (Winkler et al., 2020). Further understanding how these different action modalities affect the sense of agency not only could help with the development of such technologies and HCI research, but also with neurocognitive psychological research in general, where the use of mouse and keyboard is dominant. It is particularly salient for the sense of agency as an action driven phenomenon. Here, the different action modalities themselves are directly related to the sense of agency. This is opposed to other areas, where the inputs might only have the function of recording responses or responding to stimuli.

Chapter 2 – Social Agency

Introduction

Recently, researchers have begun to investigate the role that the sense of agency in social interaction (Stephenson et al., 2018; Brandi et al., 2019). In the example at the start of this review, the stone that is thrown has a predictable outcome. It will invariably land in the general direction to which I throw it. However, the outcomes in social interaction have higher variance. What I say to someone might be intended to be a compliment and to be received gratuitously, but it might also be taken the wrong way and cause irritation. This increased variability in social interactions leads to the question as to whether social agency works under the same mechanisms as when actions and their effects are inextricably linked. For example, because the outcome proceeding an action is inherently less reliable in social situations, perhaps our sense of agency relies more on retrospective mechanisms than predictive mechanisms. Here, we review literature into the social sense of agency and discuss the role that the sense of agency has within social interaction.

The sense of agency as a social mechanism

The sense of agency supposedly contributes to what we know as the ‘minimal self’, which is the experience of that ‘I am me’ as a constant through time (Gallagher, 2000). The experience of the self permeates so much of human experience that it is difficult to picture what life would be like without it. Whilst the concept of the self is beyond the scope of the review, it is worth noting that the evolution of the self might have important contributions for the development of human society. The philosopher George Mead insisted that the individual mind can only exist in relation to other minds. In other words, the self can only exist with the presence of the other. He states: “*It is only as the individual finds himself acting with reference to himself as he acts towards others, that he becomes a subject to himself rather than an object, and only as he is affected by his own social conduct in the manner in which he is affected by that of others, that he becomes an object to his own social conduct.”* (Mead, 1913). On this basis, the notion of the self is inherently social as it does not exist without the involvement of other agents. However, we must also consider the evolutionary cost of having such a social mechanism (Dunbar, 1998). In short, what is the advantage of experiencing me as me? This is a question that is expansive, so this review aims to concentrate on one aspect – the sense of agency. As a supposed contributor to the minimal self, understanding why the sense of agency formed might provide insight into the purpose and the function of it.

If the sense of agency is an aspect of the self, then we propose that it partly functions as social mechanism. In this account, the sense of agency evolved to monitor intentional action to facilitate social interaction. The advantage of having such a mechanism for social groups is the administration of appropriate rewards and punishments for actions (Frith, 2014). This in turn might encourage group cooperation and cohesion. Without a system to distinguish intentional and unintentional action, social consequence would be difficult to dispense to individuals within a group. An embodied sense of agency would allow a first-hand perspective of intentional action, which would help individuals to identify deliberate transgressions from others and penalise them accordingly. Furthermore, it could be that the sense of agency evolved to link our own actions with the response of others. Having a mechanism which monitors the effects of our intentional actions might help to adjust behaviour which is socially undesirable and reinforce behaviour which is received well by the group. The sense of agency might link with affective processes, promoting feelings of shame or guilt when transgressions towards the group occur. Finally, the sense of agency might help to afford group action through the formation of the ‘agentic we’.

Because the sense of agency is so ubiquitous and grounded within conscious experience, its social role could be overlooked. However, one approach of developing an understanding of the role of the sense of agency in social interaction is by looking at those who demonstrate abnormal agency over their actions. The most studied patients within the agency literature are those with schizophrenia. Patients with schizophrenia show severe and persistent impairments in the domains of social cognition, including emotional processing and theory of mind (Green & Horan, 2010) and often have difficulty integrating into society. Patients with schizophrenia have also been shown to have a systematic inability to attribute their own thoughts and actions to themselves, leading researchers to believe that an impaired sense of agency plays a crucial role in the aetiology of the disease (Jeannerod, 2009). Other conditions with social impairments, such as autism (Sperduti et al. 2014) and anxiety disorders (Gallagher & Trigg, 2016) also appear to have abnormal agency systems. Furthermore, evidence has suggested that higher trait narcissism results in higher temporal binding (Hascalovitz & Obhi, 2015). Narcissists are more likely to demonstrate aggressive behaviour in social situations (Twenge and Campbell, 2003) and have significant impairments in emotional empathy (Ritter et al. 2010). While there is currently no causal link between the sense of agency and social impairment, the evidence does suggest that a regulated agency system plays a role in successful and complete integration into social groups.

Agency, Social Consequence and Responsibility

One of the social functions of the sense of agency could be to enable the attribution of individual responsibility. The notion of responsibility plays a crucial role in the maintenance of society as it enables individuals to be rewarded for actions that benefit the group and punished for those which are deemed to harm it. This in turn encourages individuals to be more cooperative and altruistic by limiting the advantage of harming the group and encouraging actions which will benefit it. Indeed, in experimental settings it has been shown that participants are likely to only administer rewards or punishments if the person acting is deemed responsible for their actions (Singer et al. 2006). Taken together, this suggests the importance of responsibility in the regulation of group cooperation and cohesion. However, the result of one's actions are typically not taken in isolation. The magnitude of the social consequences for our actions are modulated through two primary determining factors – intent and freedom to act. In most legal systems across the world, intention plays a crucial role in criminal sentencing. For example, murder and manslaughter carry very different sentences because of the difference in intent, despite the outcome being the same. Likewise, a person who is forced to do something at gunpoint likely would not be sentenced in the same way if he acted on his own volition (see Frith, 2014). While linking the emergence of agency to legal systems might seem extreme, it is important to consider that having mechanisms that are able to monitor intentional action have salient implications for social animals. Without it, accidents would be punished with the severity of intentional acts and virtuousness would go unrewarded. This might result in acting selfishly being more beneficial for an individual than collaborating with the group. Thus, the role of the sense of agency could be to enable the attribution of responsibility by providing a mechanism that monitors intentional action. This system might facilitate response towards others by helping to distinguish if an action was intentional. In this position, the sense of agency evolved as a social mechanism to help group cohesion and cooperation, by providing a mechanism that enables the appropriate administration of rewards and punishments.

Modulating the sense of agency with social context

There has been accumulating evidence over the past decade that our sense of agency is sensitive to several social cues, including emotional valence (Yoshie & Haggard, 2013), monetary loss (Takahata et al., 2012), coercion (Caspar et al., 2016) and even the mere presence of others (Recht & Grynszpan, 2019). Such research indicates that agency over non-social actions can be modulated through social context.

One approach that researchers have taken is assessing how the presence of others affects the sense of agency. One such study was conducted by Beyer et al. (2017). In an adaption of a methodology developed by Kuhn, Brass & Haggard (2009), participants were given a computerised task where they were required to stop a marble rolling down a bar to stop it crashing. The task was designed so that participants lost points on every trial. The later that the ball was stopped, the less points that participants lost. However, if the ball crashed then maximum points were lost. Trials were randomly assigned, one where the participants acted alone and another where they supposedly acted with another person (the partner's behaviour was in fact controlled by a computer). After each trial, participants were asked to rate their subjective control over the ball stopping. The authors found that participants self-reports of the sense of agency were lower when the participant supposedly acted with another participant, rather than alone. They also found that FRN amplitudes were reduced when participants performed the task with a partner. This evidence links in with the well-known phenomena of 'diffusion of responsibility. That is, the tendency for people to feel less responsible for their actions when in the presence of others (Bandura, 1991; Bandura 1999; Bandura 1975). This evidence suggests that being in the presence of others actively reduces the link between our actions and their effects, possibly at the neuronal level. Extending this research, Beyer et al. (2018) used fMRI to study how social context affects the sense of agency in the brain. The results showed increased activity in the bilateral temporo-parietal junction (TPJ), precuneus and middle frontal gyrus during social interaction. Citing a study conducted by Chambon et al. (2013), who found negative relationship with the left angular gyrus and agency ratings when manipulating action fluency, the authors suggest that social context affects the sense of agency through increased load on mentalizing networks, which affect action-related processes and reduce agency. Furthermore, this seems to only occur when participants perceive themselves to be in the presence of other intentional agents. A series of experiments conducted by Ciardo et al. (2020), showed that the sense of agency is reduced when in the presence of either a robot or a human, suggesting that we ascribe agency in a similar manner to robots as to humans. Crucially, they also found that the sense of agency is not altered with a mechanical pump that could passively interact with the task in a similar manner to a human or a robot. This suggests that a core component of social context on the sense of agency is the attribution of intentionality, highlighting how the sense of agency can be uniquely modulated from other agents.

Overall, the experiments outlined above demonstrate that social context can play a major role in the formation of agency over our actions. However, there are also several considerations to take into account. The first is that the studies all used subjective self-reports of agency, rather than temporal binding. Whilst Beyer et al. (2017) and Beyer et al.

(2018) bolstered their research with brain imaging data, this is an important factor to consider as research has indicated that implicit and explicit measures of the sense of agency aren't necessarily correlated and may tap into different processes (Dewey and Knoblich,

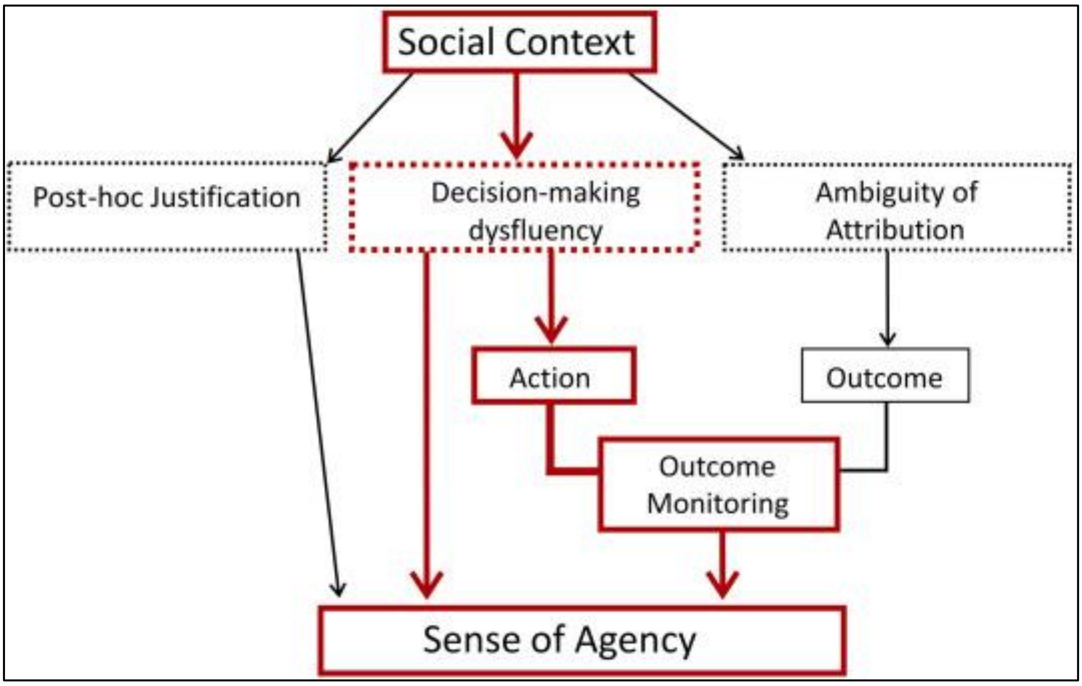

Figure 6 – Proposed model of how social context affects the sense of agency by Beyer et al. (2017). From Sidarus et al. (2020).

2014). Using explicit self-reports are also attributive in nature and thus cannot disentangle the pre-reflective and reflective sense of agency. Thus, the studies so far do not tell us how social context affects the feeling of agency at the pre-reflective level. Further, the question of whether social context affects prospective or retrospective processes is still up for debate. We also are currently unaware how different levels of intentionality play a role in this effect. In the study from Ciardo et al. (2020), the mechanical pump failing was the 'action'. This does not necessarily constitute an action but rather a non-action and also does not disentangle the varying levels of intentionality that we might ascribe to machines. While we might not ascribe intentionality to a mechanical pump, we may ascribe varying levels of intentionality to machines which appear more intelligent, such as the difference between a simple computer interface or personality-endowed avatar.

Joint Agency

Most studies looking at the sense of agency have focused on individual action. However, as hyper-social animals, we often perform actions with others to complete tasks. Pacherie (2011) notes how collective actions could be as broad as a traffic jam. Where the people causing it do not have necessarily have any goal to produce the outcome, are not cooperating and did not intend to work together to achieve it. However, it is also noted that researchers largely focus on “joint cooperative actions”, which tend to be small scale tasks which do in fact have the characteristics that a traffic jam would not. When performing an action to complete a goal with another person, you do not only have to consider your own action representations but also those who you are working with. This shared action representation has been called dyadic adjustment (Pacherie, 2014) and requires dyads or groups to predict and adjust motor actions to adjust for the people they are working with. In addition to this, it is required that dyads are required also to share the same goal to for shared outcome monitoring, known as triadic adjustments. Thus, shared action representations are not sufficient to coordinate joint action, but also that the group also have shared task representations with a shared goal (Knoblich, Butterfill & Sebanz, 2011).

Pacherie (2014) notes that joint action is a way of increasing agency scope as groups are able to perform actions that an individual cannot alone. However, she also raises the question as to whether this expansion of agency is an extension of one’s own self-agency or the merging of one’s agency with the group. One study that might help answer this was from Obhi and Hall (2011a), who tasked participants with a joint action task in which one person initiated a movement which another person joined in with. This occurred in two conditions, one where both participants intended to initiate the action and the other where only one person intended to initiate the action. In each condition, both participants experienced the effect of their actions. The results showed that in the co-intention condition, only the initiator of the action showed subjective experience of agency, but *both* the initiator and responder demonstrated temporal binding. This pattern of results was demonstrated in the assigned intention condition. The authors suggest that during joint action, an automatic, pre-reflective ‘we’ identity is formed during joint action, but this is not registered at the conscious, reflective level. Other studies use a leader-follower paradigm, where participants work together with a small-scale task to achieve a goal. The typical results from this type of research are that participants report higher levels of joint agency when on tasks together (Bolt et al., 2016; Shirashi & Shamida, 2021 & Sahai et al., 2019). Such studies suggest that working together on tasks seems to create a sense of joint agency, where participants experience share their experience of control over an outcome. Pacherie (2014) argues that the more that one’s contributions task is equal to their partner, the more joint agency we should feel over our shared actions. This arguably could be because we can extrapolate our own sensorimotor processes to the partner and use our own predictive processes to help predict the partners actions. Indeed, when performing a joint-action, one not only has to make predictions about their own actions, but also predictions about their partners actions. It has been found that predictable partners lead to a stronger sense of joint agency (Bolt, Loehr & Pacherie, 2017), suggesting that people use reliable signals from their partner to maintain joint agency and coordinate action.

One approach that researchers have taken is to use the Simon effect as means of assessing how we represent partners actions during tasks. The Simon task found the original Simon effect, which asked participants to react to targets by pressing the left and right buttons on a keyboard. The typical approach of these experiments is to use the left and right keypresses and assign them a colour or a tone pitch (Simon, 1969; Simon & Small, 1969; Simon and Craft, 1970). It was shown that if the target was not spatially mapped on the screen congruently with the mapping on the keyboard, then reaction times were slower, even though the spatial mapping of the stimuli was not relevant to the task. Interestingly, this effect remains constant even when the participants are working with a partner. Sebanz et al. (2003) asked participants to perform a similar task, in three conditions. In one, they performed the task alone and could respond to both of the coloured circles (*two-choice condition),* which acted as the control condition. In another condition, they performed the task alone again but only responded to one of the coloured circles (*individual go-nogo condition).* Finally, they performed the task with a partner, where them and their partner were assigned a colour to respond to (*joint go-nogo)*. So, the participant would respond to one colour and their partner another. They found that in the *two-choice condition*, participants were slower at reacting to targets when the stimulus were incongruent with the direction of the stimulus on the screen (the typical interference effect found in the Simon task). In the *individual- no-go* condition, they found that this interference effect disappeared, as participants only had to react to one target. This suggests that the effect arises from two action representations having to activated together. Importantly, in the *joint go-nogo* condition, this effect remerged, even though participants were still only required to respond to a single stimulus. This suggests that the action representations of the partner were incorporated into the participant's own motor system and provides the same cognitive interference as when the participant themselves are provided with two choices. This effect has since been labelled the Social Simon Effect. These results suggest that we represent other peoples' actions in our own motor system during joint action, even when the action is not task relevant. Sahai et al. (2019) extended this further, incorporating temporal binding as a measure. Along with finding the typical Social Simon Effect outlined above, they also found that temporal binding was larger during tasks where the participant performed the task with the partner (i.e. *joint no-nogo*) compared with simply observing the partner perform the action, even though the process was the same. This supports the idea that interaction bolsters agency over shared actions (Pacherie, 2014). Further, a study conducted by Sahai et al. (2023) found that while temporal binding occurred during a joint Simon task when interacting with other humans, this did not occur when interacting with a computer. Interestingly, another study from Obhi and Hall (2011b) showed that while temporal binding occurs during joint action with humans, it does not occur with human-computer partnerships. These results are consistent with the idea that social identity is crucial for the co-representation abilities during joint tasks (Stenzel et al., 2012). The 'socialness' of joint agency is also highlighted in a study by Le Bars et al. (2020), who found that the degree of joint agency is modulated by fairness. In short, the study found that participants felt stronger joint agency when monetary outcomes were equal to an individual's motor contribution during joint action.

Taken together, research into joint agency suggests that our pre-reflective agency is malleable enough to incorporate other people's actions to form an agentic 'we', but this does not necessarily form at the conscious level. This also seems to only occur with other intentional agents. The results suggest that the sense of agency seems to not only play the social role of identifying self-generated actions from those generated by others, but it also seems to have a role in facilitating group action. The indication that this only occurs for other intentional agents also points to the sense of agency being a social driven phenomenon.

Different Types of Social Interaction and Agency

In the above sections, we have established several different ways in which the social agency may form. We have noted that an individual's sense of agency can be modulated through social context, how we can feel agency over causing the actions of another person and how we can develop a joint sense of agency with groups. Here, we would like to point out that the term social agency is broad and encompasses many types of social interaction. As such, there is a need to categorise the different ways that social agency emerges. Silver et al. (2020) have attempted to do so by suggesting that social agency exists on a continuum. They have suggested that on one end of this continuum, there is the 'we' agency. This is established through cooperative, egalitarian tasks with shared goals (Obhi & Hall 2011a, Obhi and Hall 2011b; Clarke et al., 2019; Sebanz et al., 2006). When inducing 'we' agency, the boundaries between self and other become blurred and a joint agentic identity is formed. Next, we have shared agency, where there is no self-other blurring, but groups engage in shared action. This would be tasks where motor representations between the group is low, but a goal is still shared. Examples of this are joint attention, where attention is coordinated between individuals (Moore, Dunham & Dunham, 1995). When gaze leading (Stephenson et al. 2018) occurs, an individual is acting on another individual to make a change in the environment, yet there are no explicit action representations to predict or to share. As there is no evidence yet to suggest that engaging in joint attention results in shared agency, one might argue that the agency felt during gaze leading is not necessarily shared with the partner. Instead, one might conclude that it is instead simply possible to experience agency over causing another individual to perform an action. This conclusion might be seen as preferable with the current information that we have. Conversely, tasks which involve more than one agent where they are not working together have also been adopted to show effects on the sense of agency. Agents seem to be able to modulate the sense of agency of another agent without directly working with them on the task. One example is coercion, where an agent coercing the participant into performing a non-social action reduces the sense of agency of the actor (Caspar et al., 2016; Caspar et al., 2021). So, it would seem that not only are we able to feel agency over making another individual act, but also our feeling of control over our actions can be modulated by other people's actions. Silver and colleagues might suggest that this type of interaction be called "interfered agency". Although, they were referencing studies where participants are required to work with a partner with little or no cooperation (Beyer et al., 2017; Beyer et al., 2018). The authors suggest that tasks where individuals act together with little to no cooperation appear to reduce the sense of agency. Arguably, the studies into coercion also fit under this category.

Therefore, social interaction seems able to modulate our experience of agency in multiple ways and its effect will largely depend on the context. That is, social interaction seems to be able to affect our own sense of agency and this modulation seems to be dependent on the type of interaction that is occurring. Merely presenting a social context where a participant is also able to act reduces the sense of agency. However, we have also discussed how people feel joint agency in different experimental paradigms (Bolt et al., 2016; Shirashi & Shamida, 2021 & Sahai et al., 2019), how leading another to act (Stephenson et al., 2018) can affect the feeling of authorship over our actions and how the presence of social context can also increase agency (Ulloa et al., 2019; Vogel et al., 2021). Even military service can affect the how we perceive our control over our actions depending on the role that a soldier might play (Caspar et al., 2020), suggesting that positions in social hierarchies might affect the sense of agency. The results from these different studies suggest that different types of social interactions can alter the sense of agency in different ways and directions. Thus, a better understanding of the conditions which bolster or reduce agency would help clarify the complex role that social interaction has on our experience of volition.

Social interaction, Predictability and the Sense of Agency

During social interaction, we seem to be able to feel causality over how our actions affects others. Take for example, telling a joke to someone. After we tell the joke we might elicit a laugh from the listener. We are then able to infer that our joke led the listener to laugh in response our joke. We feel agency over an action whose effect was a reaction from another agent. The other person’s laugh might also lead you to laugh, and they might in turn feel agency over this. In short, we can feel agency over social interactions and the agency felt during these interactions can be reciprocal. Importantly, these interactions are not linked directly to a simple cause and effect, rather, they are dependent on two independent agents engaging with each other. As people are generally quite different to one another, this makes our experience of social interaction quite unpredictable. Consider the previous example about making a joke. You might tell the joke and the outcome might not be what you expected. If you are not as humorous as you might think, you might not receive a laugh. Worse, you might make the listener annoyed or angry. Alternatively, the person might receive the joke but not get it straight away. The person might stare blankly until it all clicks and they laugh, thus leaving a temporal delay between the joke and the response. This is not what typically happens in in a causal universe, where our actions are bound by the laws of physics and will lead to predictable results. Our actions tend to have predictable effects that are congruent with our own actions and those effects occur within an expected timeframe. Congruency and punctuality of an action’s outcome both seem to be important for our sense of agency (Moore, Wegner & Haggard, 2009; Farrer, Valentin, Hupé, 2013). Furthermore, the predictability of such cues seems to be an important component in generating the experience (Haggard, Clark & Kalogeras, 2002) and measures of agency tend to be strongest when outcomes are predictable (Sato & Yasuda, 2005; Tanaka & Kawabata, 2021). Some researchers also suggest that the contribution of external cues to the sense of agency is based on their reliability (Moore and Fletcher 2012; Synofzik et al. 2013). Indeed, Pfister et al. (2014) conducted an experiment where participants actions gave a signal for their partner to act in a “leader-follower” type paradigm. So, the outcome of the participant’s action was a reaction from their partner – a social consequence. They found that temporal binding was increased for the leader compared to the follower, suggesting temporal binding can occur for other people’s actions.

Recently, some studies have focused on the concept of gaze leading to tap into the social aspect of the sense of agency. Gaze leading is a component of joint attention, which is the phenomenon of attending to something with someone else and both being aware of this shared attention (Moore, Dunham & Dunham, 2014). Joint attention is thought to be inherently social in nature, involving the mutual reciprocity of attention of multiple agents. As such, it is a suitable mechanism for studying the sense of agency during social interaction. A pioneering study that launched this investigation was a seminal eye-tracking study conducted by Stephenson et al. (2018), which looked at the sense of agency in a well-established gaze leading paradigm. The task required participants to fixate on an on-screen face and then orient their gaze to an object when it appeared. After a varied amount of time, the face then followed their gaze. Using both temporal binding and explicit self-reports, they found that participants developed a sense of agency through leading the on-screen gaze. This paper provided evidence that the sense of agency is not limited to motor-control mechanisms and can occur within purely social outcomes. However, it is also worth noting that these are (understandably) contrived experimental settings, where outcomes are inherently more predictable than natural social interactions. To better understand this, some researchers have begun to investigate social agency using more interactive and ecologically valid paradigms. The rationale behind such methods is that research suggests that social cognition is fundamentally different when engaging with others (Schilbach et al. 2013). Indeed, what seems to separate interactions with machines and humans and avoiding the ”uncanny valley” is the attribution of intentionality or agency (Gray & Wegner, 2012), suggesting the lack of agency in that humanoid objects (robots, computer-generated characters or dolls) is what creates the eerie sensation about them.

One study that has attempted a naturalistic paradigm was conducted by Brandi et al. (2018), who used video simulation of a real face to assess agency during gaze leading. Using computer algorithms which the face responded to participants’ gaze. The study was designed in a way so that contingency and congruency of gaze response was manipulated. Explicit self-reports were used as a measure of the sense of agency. The researchers found that participants report a higher sense of agency when the gaze is highly congruent and contingent. They also tested how believable the simulation was and found that ninety-six percent of participants stated that they believed that they were engaging with a real person. The authors conclude that the results indicate that the novel experimental procedure allows studying the experience of social agency in an ecologically valid and interactive manner even without the physical presence of a second person. However, it is also worth noting that no studies so far have investigated how social agency is affected by whether participants feel that they are interacting with a real person or not. Indeed, there is evidence to suggest that ecologically valid paradigms are not strictly necessary to find effects of agency when causing another to act. For example, one study conducted by Ulloa, George and Brass (2019) used still images of faces (like that used in Stephenson et al. (2018)) to investigate the effect of eye contact on the sense of agency. They found that temporal estimates were shorter when the gaze went from averted to direct, rather than the other way around. This social effect of eye contact on agency was found despite using images that clearly lacked intentionality, suggesting that less ecologically valid paradigms can be used to modulate the sense of agency in social scenarios. Despite this, we still do not know if the experience of social agency is qualitatively different when interacting with intentional stimuli. The presence of other intentional agents has been found to modulate the sense of agency for non-social actions (Beyer et al. 2017; Beyer et al. 2018; Ciardo et al. 2020), so it is possible that this might be necessary to investigate social agency. Future studies should investigate deeper into how perceived intentionality of the other affects agency during social interaction.

Chapter 3 – A Discussion Regarding Online Experiments

COVID-19 and Online Experiments

Due to COVID-19 (Zheng, Ma, Zhang, 2020) and the resulting lockdown of many countries across the world, typical laboratory experiments largely stopped among universities. Many academics therefore sought alternative methods for studies, perhaps the most fruitful being by conducting them online. Fortunately, the capacity for online experimentation has made huge strides in the past 20 years, with developments in infrastructure and technology, making access to computers and the internet more widespread amongst the public. According to the UK government, 96% of the population have access to high-speed broadband (Ofcom, 2023), coupled with over 90% of households in the UK having access to a home computer (UK Office of National Statistics, 2022). In addition, several platforms have been developed which have enabled the building of online surveys such as Qualtrics (www.qualtrics.com) and Alchemer (www.alchemer.com) and also psychological experiments (e.g. Psychopy (www.Psychopy.org) and Gorilla (www.Gorilla.sc)). Participant panel platforms have also been developed, such as Amazon MTurk (www.mturk.com) and Prolific (www.profilic.co). Such platforms have allowed for the creation and distribution of surveys and experiments without the need for extensive knowledge of complicated programming languages or techniques. So, although COVID-19 was disruptive to laboratory experiments, the groundwork had already been laid for researchers to pivot to online experiments during the resulting lockdowns. Indeed, querying the PubMed database for the terms “Amazon Mechanical Turk” (MTurk) and “Prolific” shows a large increase from 2020 onwards (Figure 7**),** suggesting an increase in the use of online experiment distribution platforms for published research during the pandemic. As the research in this thesis has participants solely gathered from online sources, it would be beneficial to outline the implications of this. Thus, this section will first discuss implications of sampling from online panels, before discussing potential issues with online experimental protocol. Finally, we will discuss temporal binding in online contexts and the present experiment in this chapter.

Sampling in Online Experiments

One of the biggest advantages of online experiments is drawn from larger sample sizes and reach. In psychology, there has been a drive to push for larger sample sizes as it has been found that many laboratory-based experiments are underpowered (Maxwell, 2004; Vadillo, Konstantinidis, Shanks, 2014). This has led to some researchers saying that there is a “replication crisis”, particularly in psychological science as many findings do not replicate (see Shrout & Rodgers, 2018 for review). Although there are multiple facets to these arguments, including questionable research practices, publishing pressure and the vulnerability to manipulation of null hypothesis significance testing (Wiggins & Christopherson, 2019), power as a statistical construct has largely been at the core of the (and the solution to) the crisis. Increasing sample size for experiments however comes at no small cost. Moving to sufficiently powered studies increases time spent in the laboratory as the necessary sample size drastically inflates. For example, Shrout and Rogers (2018)

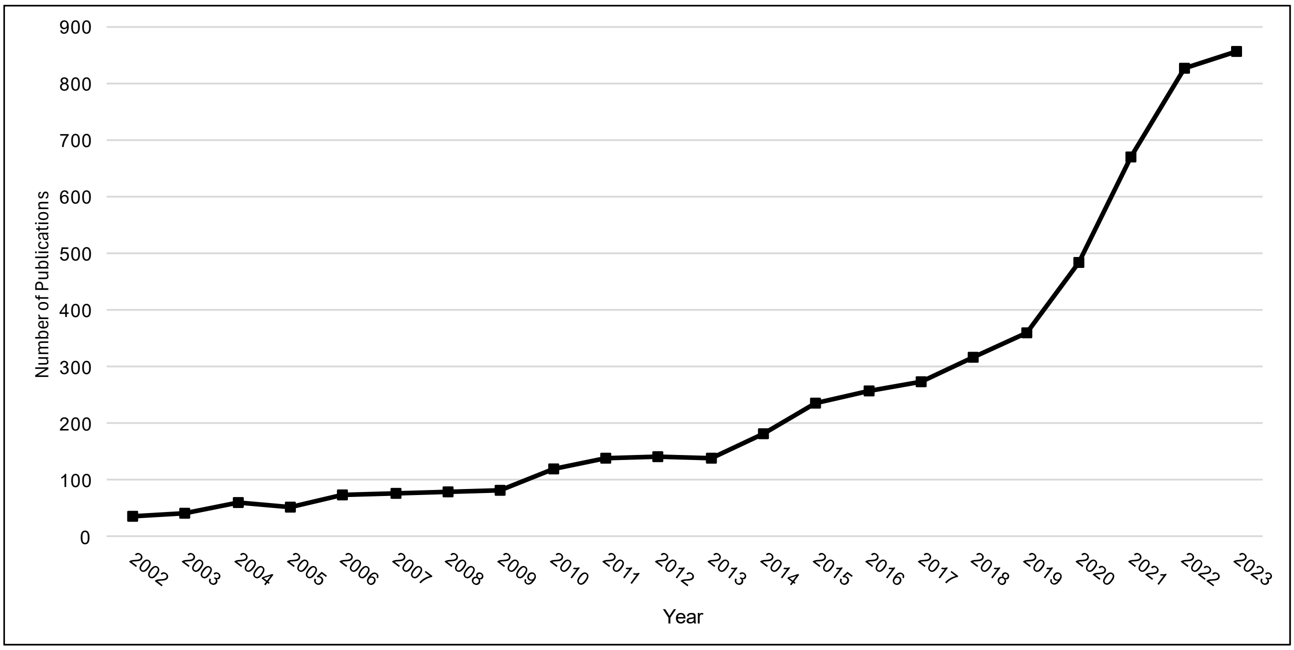

Figure 7 - Number of publications from www.PubMed.com using the terms "MTurk" or "Prolific". Data obtained from PubMeb's website. Last accessed May 4th, 2024.

discuss a study from Rossi (1990), which showed that published studies had an average of 57% of detecting a medium effect size (Cohen's d = 0.5). Shrout and Rogers argue there is no evidence to suggest that this has improved since then. To increase the power to 80% as suggested by Cohen (1988) (i.e. the study would have an 80% chance of rejecting the null hypothesis if an effect was present, or an 80% chance of avoiding a Type II error), it would require significantly more participants, this is especially true when looking at between subjects designs and those with small effect sizes. Online experiments may help increase sample size by taking away the need for the researcher to be on site in the laboratory and by allowing many subjects to be tested simultaneously, drastically cutting down time spent gathering data. Furthermore, as online studies are generally generated within browser-based software (Rodd, 2024), this allows studies to be shared more easily among researchers, potentially facilitating research and studies for replication. Running online experiments also allows for controlled and specific sampling, allowing for balanced groups and testing of hard-to-reach samples. Participant sampling platforms offer ways of obtaining specific quotas of participants and the possibility of screening by aspects such as age, occupation, education and many other demographic and socio-economic attributes which typically might be difficult to obtain (Palan & Schitter, 2018). It also creates an additional research environment for participants who might not be well equipped to attend lab experiments, such as those who have busy work lives, or those with disabilities.

On the other hand, some researchers have pointed out the problem with participant attrition in online experiments. That is, the increased likelihood that participants will drop-out in online experiments compared to lab-based studies. Zhou and Fishbach (2016) examined a number of studies for both in lab and online experiments and found that over 20% of online based studies had an attrition rate of over 30% while 96% of lab studies were free of attrition. This can be particularly problematic when assigning participants to groups, where one condition may lead to higher drop-out rates than the other, leading to attrition bias (Dumville, Torgerson & Hewitt, 2006). Furthermore, sampling platforms with minimum identity verification such as MTurk may have problems with participant identity (i.e. false demographic information) and “superworkers” - people who actively participate in a lot of studies (see Robinson et al. 2019 for review). The latter could be particularly problematic, as some researchers have suggested the participant pool follows a Pareto distribution (Arnold, 2014), with 20% of workers on MTurk accounting for 80% of the studies (Fort, Adda & Cohen, 2011 cited in Robinson et al., 2019). This may create a problem for experiment naivety, as participating in many experiments (especially those of a similar nature) may affect how the participant performs the task. While this may be the case, concerns over sample naivety have been a long withstanding issue in laboratory-based studies, which are largely sampled from undergraduate populations which repeatedly take part in studies for course credits. However, Chandler et al. (2013) noted that workers on these platforms (e.g. MTurk) are both exposed to a larger variation in experiments and for longer periods of time, as the time spent on these platforms by participants are typically longer than those of undergraduate populations. This is exacerbated by the ease with which a participant can access studies, as there is typically little to no scheduling involved nor is there any need to be on site. This could lead to them being more experienced in specific experimental paradigms and psychological research in general than typical university subject pools.

Another point of discussion between online and laboratory-based experiments is external validity. Calder, Phillips & Tybout (1982) states that “External validity examines whether or not an observed causal relationship should be generalized to and across different measures, persons, settings, and times”. Or, in other words how well can findings from an experiment be extrapolated to the general or certain populations. Laboratory based experiments have been criticized for their overreliance on convenience sampling, particularly with undergraduate populations from universities (Higbee, Millard & Folkman, 1982; Sears 1986; Nielsen et al., 2017). This creates a situation where a large amount of published research is based on a small stratum of the whole population, yet findings are discussed and interpreted as if they apply to the whole. This is a longstanding issue within psychological research that for a long time was largely unavoidable. Sampling from more varied populations is expensive and time consuming. Online research seems like a natural way to resolve this issue. As touched on previously, online research offers researchers the ability to sample specific and perhaps hard to reach populations. Furthermore, Samples through online panels are typically more heterogeneous than student populations, containing a broader range of age and socioeconomic backgrounds (Huff, & Tingley, 2015). Therefore, online sampling can provide researchers with a way to move away from strictly testing university sample pools. Landers and Behrend (2015) argue that online samples are at least as good as those gathered from typical university sampling pools and that resistance to online sampling is rooted in tradition. However, it is also worth noting that while the online experiment pool can appear more diverse than undergraduate populations, there is still a commonality between participants as they are all, in one or another, motivated to take online experiments.

Experimental Procedure in Online Experiments

One of the aspects of online research is that the task is unsupervised and, in cases of samples outside of university subject pools, anonymous. This raised concerns over the attention participants give to tasks (Palan & Schitter, 2018). The issue is that there is no experimenter to ensure that participants give the task their full attention and being outside of a laboratory setting, there are more chances for interruptions and distractions for the participant. Chandler, Mueller and Paolacci (2013) found that a proportion of participants (6%-18%) in online experiments cited that they were often engaged in other activities, such as watching TV, listening to music or instant messaging during the tasks. This will be naturally quite alarming for researchers, as inattention to the task could at best reduce the desired effect and at worst invalidate the participant’s results. Inattention can lead to spurious elevation in symptom levels when conducting psychiatric research questionnaires (Zorowitz et al., 2023) and could also lead to a reduction in treatment effects when replicating laboratory-based experiments (Peyton, Huber & Coppock, 2022). On the other hand, it has also been shown that participants gathered through MTurk pass attention checks more often than participants who were sourced from university samples (Hauser & Schwarz, 2015), suggesting they are more likely to be paying attention during the task. Or this could be a result of those gathered from paid participant sample pools being more experienced in psychological research (and thus know to look out for attention checks) and being more likely to have their compensation withdrawn because of failing them, compared to university students. Furthermore, other environmental advantages have also been outlined, such as increased external validity and reduced experimenter and demand characteristic effects (Reips, 2000). Inattention could simply be seen as an occupational hazard for online research and an inevitability of relinquishing control over the environment a subject is tested in and as such, more careful consideration should be given to screening and exclusion criteria when researching in such environments. Along with inattention, another consideration with online experiments is that the technology being used isn't controlled. Although this could be seen as a double-edged sword. On one hand, researchers may make special effort to control for the equipment that participants use to remove extraneous noise in experiments. On the other hand, finding effects without controlling for equipment may demonstrate generalisability of the effect outside of specific equipment and laboratory settings.

In the last three chapters, we have discussed the sense of agency, social agency and online experiments in detail. The next series of chapters will outline the experiments conducted for this thesis. First look at how working with partners of varying intentionality affects the explicit sense of agency in the following chapter.

Chapter 4 - Intentionality and Explicit Agency whilst Working with Others

Introduction

We might consider the authorship of our actions to be absolute. Certainly, in everyday motor events such as moving and writing we typically feel a strong sense that we are the instigators of our actions and are the result of its consequence. We have the feeling that we sit firmly in the central hub of our actions and feel very much in command. However, this feeling of agency is not necessarily an accurate reproduction of reality but rather a subjective conscious experience which is just as fallible as other aspects of perception and cognition. Consider for example superstition and games of chance. People might wear their lucky socks before a game of football or kiss the chip they place before playing roulette in a casino. Neither of these actions would lead to any logical outcome change but people might think these actions lead to a favourable outcome. Conversely, another footballer (who may or may not have also been wearing their lucky socks) may point to the sky or pray and thank God for the goal which they have just scored. Speaking empirically, the outcome (the goal scored) was from the actions of the player. However, the player might ascribe at least some of the causality to the divine. These examples are presented to illustrate how an individual's sense of agency can be separate from objective reality. In addition, they also show how we can extend the causal relationship between our actions and their consequences to incorporate our surroundings and our beliefs (Desantis, Roussel & Waszak, 2011; Feldman, 2017; Aarts, van den Bos, 2011). Therefore, the sense of agency does not seem to be a binary "all or nothing" system, but rather graded and multifaceted (Haggard, 2017). In short, we seem able to feel more agency in certain situations and less in others, irrespective of veridical control we have over the situation. Over the past decade, research has started to look at how this might occur by including social context to actions.

The Sense of Agency within Social Context

The term social context has been used to describe the general environment that can alter individual or interpersonal behaviour (American Psychological Association, 2015). Research within social psychology has long pointed towards the profound effect that social context has on behaviour (Milgrim, 1963, Asch, 1951, Sherif, 1953). It has been a cemented into the psychological literature over the past century that our behaviour is modified by our social surroundings. However, the cognitive and neural processes that underlie if, how and why social context affects voluntary action remains poorly understood. Recent research has attempted to rectify this by using social context to alter voluntary action in various ways. Joint agency tasks typically ask participants to simultaneously work together towards a goal

(Obhi & Hall, 2011), other tasks follow paradigms which include emotional valence (Yoshie & Haggard, 2013; Christensen et al., 2016; Yoshie & Haggard 2017), coercion or instruction (Caspar et al., 2016, Caspar et al., 2021; Barlas, 2019) or eye gaze (Stephenson et al., 2018; Recht & Grynznpan, 2019; Ulloa et al., 2019). These studies have led to results that aren't necessarily compatible with the idea that social context is a single entity affecting agency. Instead, its effect is context dependent. Different types of social contexts lead affect voluntary action in different ways. This can be confusing as the term has been used by some authors to describe social context in a particular way, namely the presence of another person who is salient to the task. However, other authors have used social context to describe any social effect on the sense of agency. Here, we will call social agency as the sense of agency within any social context but it's important to consider that social context is an umbrella term which can encompass many types of social interaction. Social interaction itself might enable, complement, or replace individual mechanisms of social cognition and this will likely vary by the type of interaction that is occurring (Jaegher, Paolo, Gallagher, 2010). Therefore, it is important to consider that the type of social context or interaction that is occurring might fundamentally change what happens to the sense of agency. However, semantics aside, the effect of social context on the sense of agency (or social agency) should not be seen as a uniform construct, as its effect on the sense of agency seems to be dependent on the task which is afforded.

The Different Levels of Joint Agency

Silver, Tatler, Chakravarthi and Timmermans (2020) developed a useful framework for social agency, which suggests that it exists on a continuum. They have categorised the existing data based on the amount of cooperation that occurs during the task. Arguing that social contexts where cooperation is strongest are where we should experience an enhanced sense of agency. One aspect of this is 'we' agency, where the line between the self and other is blurred. These cases would typically occur where groups share a similar goal, and their actions affect the environment in an immediate and proportionate way. An example of this would be two or more people moving a sofa. The people would work together, simultaneously to lift and reposition the sofa. They would both need to be cognisant of how high they are lifting each end of the sofa to account for the other person and coordinate movement together. Research into this type of social agency is gathered from joint-action contexts, where participants are required to coordinate action together to complete the task (Obhi & Hall, 2011; Bolt et al. 2016; Sahai et al., 2019; Shiraishi & Shamida, 2021). In these cases, who is causing the effect to occur is generally made ambiguous. In Obhi and Hall's (2011) experiment, they found that only the initiator of joint action reported explicit sense of agency over their action, but both the initiator and responder demonstrated temporal binding (an implicit measure of agency). The authors conclude that a 'we' identity is automatically formed during joint action. In contrast to joint agency (or 'we' agency) there are tasks where two or more people are performing the same task, but independently. An example of this could be a pilot and a co-pilot flying a plane. Whilst they have the same goal, they act independently from one another; their actions and assigned responsibilities for the task are distinct and are not ambiguous. Nonetheless, because they both pilots have the common goal of flying the plane, it is likely that they will experience a shared sense of agency (i.e. "*we* are flying the plane"). Indeed, it has been found that knowing what a partner is doing on their task affects an individual's planning on their own task, even if the partners planning is not relevant (Sebanz & Knoblich, 2008).

As mentioned in Chapter 2, one example of how the voluntary action might be changed in these circumstances is the bystander effect and the diffusion of responsibility (Latane & Darley, 1968). As a reminder this is the notion that an individual's sense of personal responsibility is divided by the number of bystanders present. Thus, an individual will feel less personal responsibility to intervene in a situation and less responsible for the consequences of the outcome. We discussed in Chapter 2 how this type of social context has also been described in some studies into social agency. To review, Beyer et al. (2017) found that people report less control over the outcome of their actions when working with a human partner compared to working alone. Their experiment required participants to stop a ball rolling down a slope before it crashed off the end. The key manipulation in this study was the inclusion of a partner who the participant believed could also intervene during the experiment. Their partner was in fact a confederate in the study and the partner's behaviour was controlled by the computer. The results showed that participants reported less control over the outcome when working with their partner, compared to working alone. Results from EEG also showed reduced FRN amplitude of successful actions when working with a partner compared with working alone, which Beyer and colleagues argue represents inhibition of systems which are sensitive to control when working with others. Furthering this, in a separate study Beyer et al. (2018) replicated the findings regarding explicit agency in an analogous study using an inflating balloon. Using fMRI, they also found that when participants were working with the partner, they had increased activity in areas thought to be involved in mentalising processes, notably the temporal parietal junction and the precuneus. From this data, the authors suggest that the presence of other agents lead to reduced outcome monitoring by interfering with mentalising processes. Indeed, action-selection dysfluency has been noted to affect the sense of agency (Sidarus et al., 2013; Chambon et al., 2014; Sidarus and Haggard, 2016; Sidarus et al., 2017). The theory from the authors is that social context reduces the sense of agency by inhibiting mentalising processes, decreasing action-selection fluency, which affects when and if to act, leading to a reduced sense of agency. In other words, the presence of others reduces the neural link between action and effect (Pagliari, Chambon & Berberian, 2021). It is worth noting here that social context is used to describe the presence of another agent. As mentioned previously, it is important to make a distinction between the types of social contexts that occur as they seem to have different effects on the sense of agency. Although to summarise, the presence of another agent who can act but is working independently of the participant reduces the explicit sense of agency. It has been suggested that the reduction in agency is from the presence of other agent interfering with mentalising processes, leading to action-selection dysfluency.

Social Agency and Artificial Agents

With the rapid rise in artificial intelligence and the ever-increasing capabilities of computers, robotics and virtual reality, psychological research has been looking at how social cognition interacts with artificial machines. We know that to a certain extent humans are capable of ascribing social characteristics to machines (Gray, Gray, Wegner, 2007; Wang, Quadflieg, 2015). Indeed, it's hard not to sympathise with Disney's WALL-E, as it spends its time cleaning up a desolate planet. The fact that we can sympathise with WALL-E tells us something about our ability to extend a degree of socialness to machines. Here, socialness can described as "the presence of intentional goal-directed recursive interactions with other beings" (Hortensius & Cross, 2018). The narration in the film clearly tries to attribute human-like characteristics to WALL-E and despite not being human, we see WALL-E as a timid and relatable character as if he were. We read his emotions from his expression, can tell his anxiety from his rushed movement, and can hear the admiration in the cadence of his robotic sounding voice. Indeed, humans seem to be able to perceive and respond to the emotions presented by artificial agents and can empathise and be aggressive towards them (see Hortensius, Hekele & Cross, 2018 for review). Thus, we seem to be capable of interacting and responding with machines in a way that reflects human-to-human interaction. This raises a question: at what point do we consider a machine to be an artificial agent? In other words, when do we start treating machines as intentional beings? We might agree that WALL-E is agentic. After all, he has goals and follows motivational behaviour. We therefore might ascribe mental states to him (referred to as Theory-of-Mind (Leslie, Friedman & German, 2004)). If we were to interact with WALL-E, we might treat him in a similar manner to how we would treat humans. However, we probably would not say the same about a robot arm that's only job is to packages boxes. And yet, it's difficult to not get irritated at your computer that throws errors as you try to work despite it clearly not having intentions. We might mutter disparaging comments under our breath about the computer's stupidity, yet our insults will only be replied with the quiet whirr of a computer fan or the same obstinate refusal to perform our requests. When we do this, we are to a certain extent ascribing intentional behaviour (or agency), as we are cursing at and blaming a machine as if it had the choice to do otherwise. Indeed, it has been shown that people tend to ascribe decision making capabilities and intentions to computers and apply social rules when interacting or evaluating them (Nass, Steur & Tauber, 1994). This could be from an automatic inference that when you interact with something that responds accordingly it is agentic. The "computers are social actors framework" (CASA) posits that people treat computers like real people and automatically apply social scripts and heuristics when interacting with them (Nass and Moon, 2000). This framework was built from a number of studies which looked how people interact with computers. They found that people applied gender stereotypes, reciprocal altruism, and personalities to them (Nass, Steuer & Tauber, 1994). While our relationship with machines and computers have changed over the last 20 years and CASA might be need extension to reflect this (Gambino et al., 2020), it does bring an interesting perspective to human-computer-interaction by demonstrating a willingness for people to attribute socialness to machines that might follow only a few basic aspects of social interaction.

Humanness, Intentionality and Artificial Agents

Are our interactions with computers, or artificial agents in some way analogous to interactions with humans? The research on this appears to be mixed. For example, Sahai et al. (2023) found that agency for one's own actions and for the partners actions declined when working with a servomotor partner compared to when working with a human partner in a Joint Simon task. Interestingly, this effect appeared to be attenuated when the participants worked with a humanoid robot. They found that while working with a servomotor robot sharply declined one's experience of agency, this effect was reduced when working with a humanoid robot, albeit still reduced compared to working with a human. In another series of experiments conducted by Barlas (2019), who investigated how the sense of agency was affected by instruction using human and a humanoid robot. He found that temporal binding and control ratings were greater in a condition where participants were free to choose when to act compared to when instructed. Importantly, the robot and human instructor conditions showed comparable results, suggesting that the effect on agency (namely the reduction of agency because of instruction) was the same if it was a human or a robot instructor. Control ratings were also positively correlated with how human-like participants perceived the robot to be, indicating that form of an artificial agent is important for social agency effects (Hortensius & Cross, 2018). In other series of experiments, Ciardo et al. (2018, 2020) replicated the finding from Beyer et al. (2017, 2018) using a robot. They found that when working with the robot, participants reported lower ratings of control than when working alone, like the results from the studies from Beyer and colleagues. Furthermore, Ciardo et al. (2020) demonstrated that this effect did not occur when working with a mechanical pump and that comparable results were found when participants worked with the robot and humans. These results demonstrate that perceived intentionality of a system might be crucial for the application of social agency mechanisms when working with them. Such results appear to suggest that our experience of agency when working with artificial agents is modulated differently based on both how anthropomorphic the artificial agent is and the perceived degree of intentionality to which it is ascribed. Furthermore, the extent that someone might apply socialness or intentionality to a machine might be dependent on individual traits that they might possess. This could be traits such as personality, empathy, or autism, but one that seems particularly relevant is anthropomorphism. Anthropomorphism is the tendency to apply nonhuman entities with humanlike characteristics such as motivations, intentions, or emotions (Epley, Waytz, Cacioppo, 2007). It is considered an individual trait and can predict trust in technological agents to make decisions, the amount of moral care that should be afforded to them and increase the social influence of nonhuman entities (Waytz, Epley, Cacioppo, 2010). Thus, not only should form be considered when investigating social agency with artificial agents, but also the extent to which people will attribute human-like characteristics to machines (i.e., anthropomorphism).

After presenting this research, it is probably worth presenting the question posed earlier again: at what point do we consider a machine to be an artificial agent? When do we ascribe intentionality? Ciardo and colleagues have shown that the presence of a mechanical pump is not sufficient to produce reductions in agency. However, their use of Cozmo (Anki Robotics), a robot that is capable of facial expression is also quite far on the other end of the social spectrum. It is unclear from the current research whether working with a computer, with no declared intentions or motivations can produce the effect found in the Beyer et al. studies. The research in this chapter aims to tackle this gap in the literature.

Overview of Experiments 1-3

In the three experiments in this chapter, we looked at how partners with different levels of intentionality affected the explicit sense of agency when working with others. Experiment 1 looked at replicating the Beyer et al. (2017, 2018) studies in an online environment using a similar methodology. In this experiment, participants were told that they were working with another human on the task. While we expected that the findings would be replicated, it is worth noting that this had not yet been done in an online setting. Experiment 2 followed the same methodology as Experiment 1, with the exception that participants were told that they would be working with a robotic agent named BOBBY. Finally, in Experiment 3, participants were told that they would be working with a computer, who was pre-programmed by the experimenter. Each experiment varied the levels of intentionality of the partner to see if it affected whether the sense of agency was reduced when working with them. Experiment 2 asked whether simply describing a partner as a robotic agent (Ciardo et al., 2018, 2020), rather than using an embodied social robot was sufficient to produce this effect. Experiment 3 asked whether a computer that was purposefully described as non-intentional was sufficient. We predicted that Experiment 1 would replicate the findings found in the experiments conducted by Beyer and colleagues. We also hypothesized that if descriptions of robotic avatars were sufficient for participants to treat their partner with a degree of intentionality, we would see similar results to Experiment 1. Finally, because the partner in Experiment 3 was described as pre-programmed and unintentional, we should have expected that the effect would not be present in this study. On the other hand, the fact that the partner was working on the task with the participant and was capable of acting might have been sufficient alone to produce a reduction in agency, despite how the partner was described. In this case, we would have expected that each partner would reduce the sense of agency compared to when they work alone.

Experiment 1

In Experiment 1, participants were asked to view a ball rolling down a slope, which they were required to stop through a button press. This occurred under two conditions, one where the participant worked alone and one where they were working with a partner. The two conditions were manipulated within-subjects. In this experiment, participants were told that they were working with a human partner. After the ball-stopping section, the participant was told who acted, the number of points they lost and who acted. Participants were then asked to rate how much control they had over the outcome.

Method

Participants

33 undergraduate students from the University of East Anglia took part to fulfil a course requirement. Approval was granted by the University of East Anglia ethics committee and consent was obtained from each participant before the experiment. Data was collected between 12th August 2020 and 26th August 2020. Participants were recruited through SONA (www.sona-systems.com). Sample size was based on the research from Beyer et al. (2017). Additional participants were included as we conducted the experiment online. No exclusion criteria were applied to the study and thus, all participants that completed the study were included.

Materials

Psychopy3 (Peirce et al., 2019) was used to create the experiment and the experiment was run online on Pavlovia.org. The main part of the experiment was a series of images which created an animation of a ball rolling down a slope, which was created within Blender (www.blender.org). Agency ratings were obtained using a slider created and recorded within PsychoPy3.

Procedure

The experiment followed a within-subjects design, with the independent variable of *Partner* (Alone, Together). The dependent variable to be analysed were agency ratings derived from a 1-100 slider. Trials were presented in a single randomised block. Overall, participants completed 120 trials each, 60 trials for each *Partner* condition.

The experiment began with a voice call using Skype or Microsoft Teams. The researcher explained to the experiment the general task and what they would need to do.

They were told that they would be working with another person who would be doing the experiment at the same time. The behaviour of this person was in fact controlled by a computer program. After this, the participant was instructed to click on a link to the experiment and was then presented with a screen showing “Waiting for other participant…”. This was to try and reinforce the illusion that they were working with another person. Participants were then presented with the instructions for the experiment. They were told that they were required to stop a ball from rolling down a slope before it crashes off the end. The later they stop the ball the less points they would lose, but if the ball falls off the end, then they would lose maximum points. They were also told that on each trial they would either be working alone or with their partner.

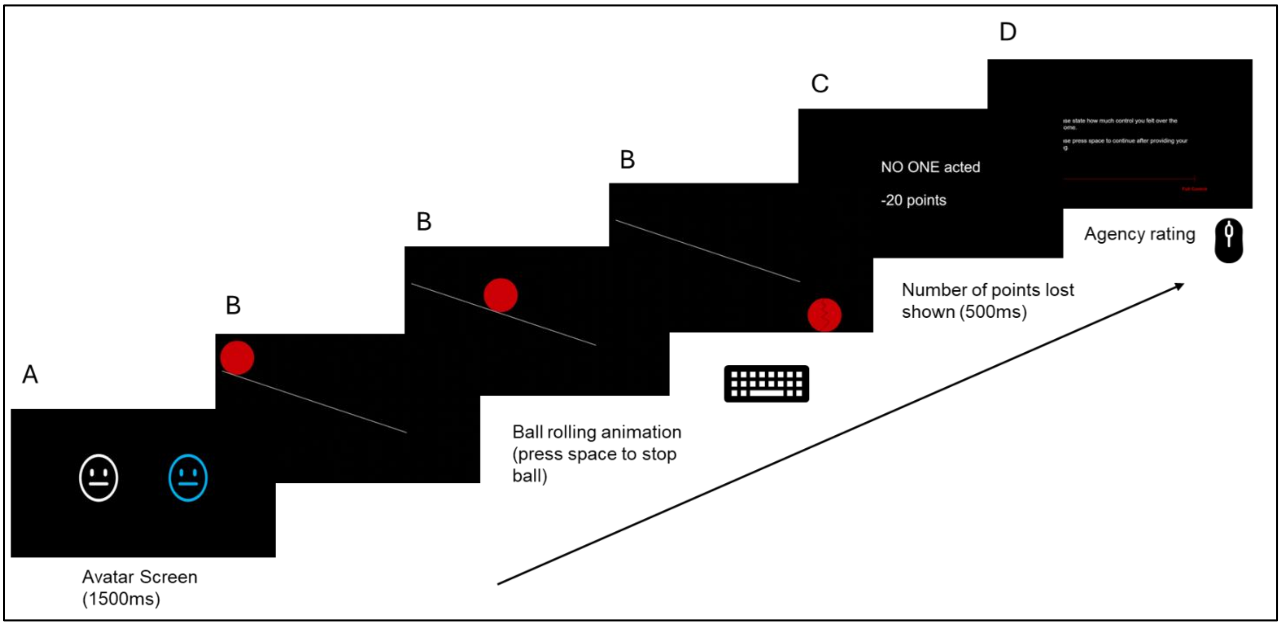

Figure 8 - Trial timeline for Experiment 2. The trial starts with the two avatars (left being the participant's and the right being their partner) for 1500ms. Only their own avatar is shown in the Alone condition. After the avatar screen is shown, the participants are presented with a ball at the top of a slope for 500ms. The participants could stop the ball by pressing space. After this, participants are shown who acted and the number of points lost. They are then asked to report how much control they felt over the outcome.

Participants were then provided with 3 practice trials for each *Partner* condition (6 in total). Each trial began with two straight-faced emoticons (created in PowerPoint) in horizontal orientation for 1500ms. Participants were told that the white emoticon on the left was their own emoticon and the blue emoticon on the right was their partners. They were instructed that if both emoticons were present then on the following trial they would be working with their partner. If only their own emoticon was present, then they would be working alone on that trial. After this screen participants saw a ball at the top of a slope for 500ms. The ball was red if the participant was working alone and blue if the participant was working with their partner. The ball then rolled down the slope and participants stopped the ball by pressing space. The speed at which the ball rolled down the slope was tied to the framerate of the computer monitor and varied on every trial. On each trial, the ball would move every 2 to 6 frames. If the participant or their partner failed to stop the ball then the ball would fall off the end of the slope and would be shown with a crack down the middle.

After the ball was successfully stopped or it crashed, the ball's final location was presented for 1000ms and then participants were presented with a screen that showed who acted. This could either be "YOU" (the participant acted), "THEY" (the partner acted) or "NO ONE" (no one acted, the ball crashed). As the purpose of this study was to understand how one's own agency is altered by working with others, only trials where the participant acted were analysed. As such, participants were explicitly told who acted to ensure there was no ambiguity around who caused the ball to stop. On the same frame as being told who acted, participants were also informed on this screen how many points they had lost on this trial. Finally, participants were shown a screen with text and a slider stating:

"Please state how much control you felt over the outcome.

Please press space to continue after providing your rating."

The slider had a label on the left side stating "No control" and a label on the right side stating "Full Control". Participants could select anywhere on the slider between 1-100, but could only see the position they were selecting, not its underlying value. The slider had no set value before the participant clicked on it. After a value was selected, participants could press space to end the trial. They were then presented with a blank screen for 1000ms before the next trial began.

Results and Discussion

Each participant returned 60 agency ratings per *Partner* condition, resulting in 120 trials across the whole experiment. Only trials where participants acted were analysed. For each participant, a median agency rating was calculated for each *Partner* condition. Results from a paired sample t-test indicate that participants rate their control over the outcome lower in the Together (M = 61.70, SD = 19.71) compared to the Alone (M = 78.70, SD = 19.31) condition, $t(32) = 4.90$, $d = .85$, $p < .001$ (Figure 10).

Consistent with results from Beyer et al. (2018), the results showed that participants report less control over the outcomes of their actions when working with a partner compared to when they work alone. This effect occurs even when online, suggesting that the social context effect on the sense of agency is not limited to situations where people are in proximity to each other. Overall, the current experiment extends research into the social context effect on the sense of agency by demonstrating that the effect is also present in online experiments.

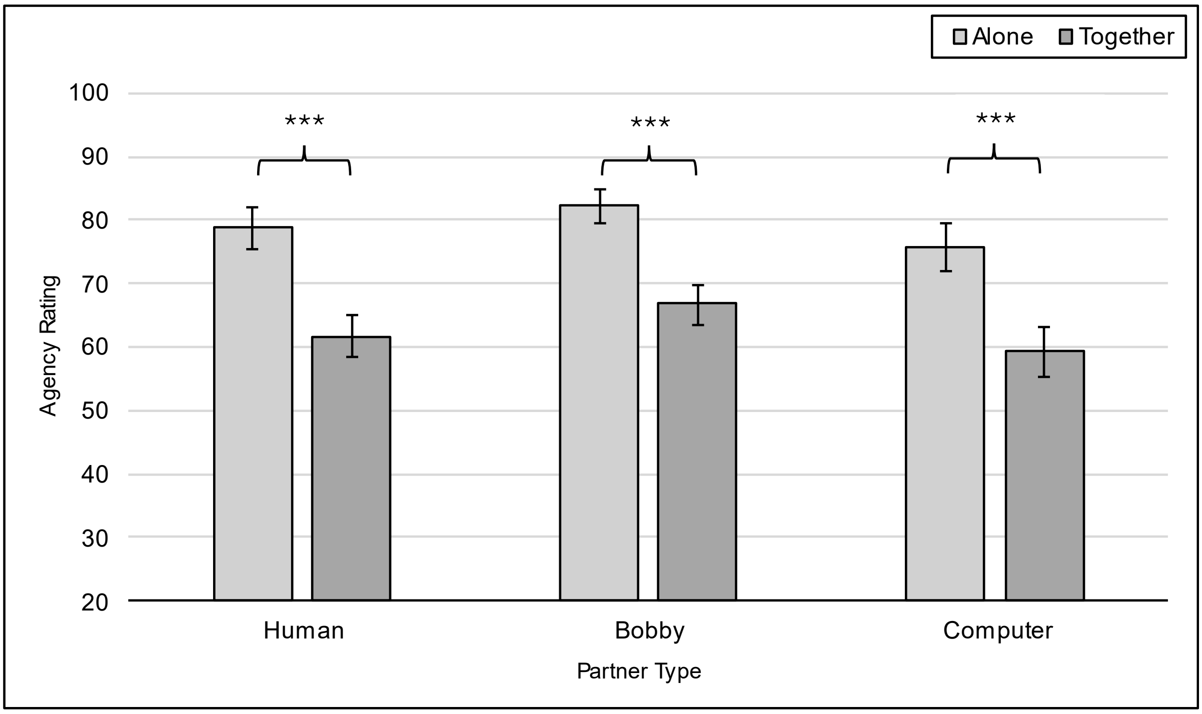

*Figure 9 - Graph showing agency ratings for Experiment 2, Experiment 3 and Experiment 4. Error bars show the standard error of the mean. Asterisks indicate significance level (*p < .05, **p < .01, ***p < .001*).

Experiment 2

As we replicated the findings from Beyer et al. (2017,2018), finding that explicit agency was reduced when working with a partner, we wanted to further examine how varying the level of intentionality of the partner affected explicit agency ratings. The purpose of this was to understand whether the reduction of agency was tied to the idea that the partner is another human. To do this, we altered the initial instructions of the experiment, telling participants that they would be working with a robotic agent named BOBBY. In the instructions, we anthropomorphised BOBBY so it had human-like characteristics, such as desire, emotion and motivation. If the effect was only due to participants working with humans specifically, we would not expect a reduction of agency when working with the partner. On the other hand, if working with an anthropomorphised robotic avatar affects agency in a similar way to working with a human partner, we would expect a reduction of agency when working with the partner, compared to working alone.

Method

Participants

31 undergraduate students from the University of East Anglia to fulfil a course requirement. Approval was granted by the University of East Anglia ethics committee and consent was obtained from each participant after the experiment. Participants were recruited through SONA (www.sona-systems.com). Sample size was based on the number of participants in Experiment 1. Like with Experiment 1, no exclusion criteria were applied and all participants who completed the study were included. However, those who took part in Experiment 1 could not take part in Experiment 2.

Materials

Like in Experiment 1, the experiment was built in Psychopy3 and data was collected between 9th September 2020 and 28th September 2020. The same materials and dependent variable were used during this experiment.

Procedure

The design and procedure in this study is largely identical to *Experiment 1*. However, in this study participants were told that they would be working with a robotic agent named "BOBBY". Participants were introduced to BOBBY and shown the text in *Figure 11*. Other than this change in the introduction, the procedure for the experiment was consistent with *Experiment 1*.

> *"This is BOBBY.*
>
> *BOBBY has been selected from the other robotic agents to work with you on this experiment. Compared with the other robotic agents in this experiment, he is unique. He has the desire to beat his own score each time he takes part in the study."*
>
> *BOBBY intends to do his best on every trial and is happy when he accomplishes this. This is what sets BOBBY apart from the other robotic agents. He is emotionally invested in the experiment and enjoys taking part. Getting a bad score in the experiment makes BOBBY disappointed, so he tries to avoid this.*
>
> *Press space to continue.*

Figure 10 - Text shown to participants introducing their BOBBY before the experiment began.

Results and Discussion

As in *Experiment 2*, each participant returned 60 agency ratings per *Partner* condition, resulting in 120 trials across the whole experiment. A median agency rating for each participant was calculated for both *Partner* conditions. Results from a paired sample t-test indicate that participants rate their control over the outcome lower in the Together (M = 66.97, SD = 20.41) compared to the Alone (M = 82.21, SD = 14.65) condition, $t(30) = 5.76$, d = 1.03, $p < .001$. Like with *Experiment 1*, participants rate their control over the outcome as lower when they are working with their partner rather than working alone.

Here, we demonstrate that the reduction in agency is present when using a partner described as a robotic avatar. Consistent with Ciardo et al. (2018, 2020), we show that non-human agents can reduce the sense of agency when working with the participant. Extending the research of Ciardo and colleagues, these results show that this effect occurs with on-screen avatars described as a robotic agent and an embodied robot is not necessary to produce this effect. This will be discussed in more detail later in this chapter.

Experiment 3

After finding that the reduction of agency persisted when working with a robotic avatar named "BOBBY", the final step was to understand whether a lack of perceived intentionality of the partner would lead to a reduction in agency when working with them. Humans often perform tasks with computers that are not anthropomorphised, like in manufacturing or with navigation systems. Understanding whether our experience of our own agency is affected by such factors would better inform the design and our own understanding of such devices and machines. Thus, this experiment introduced the partner as a pre-programmed computer. The instructions told participants that the partner had been pre-programmed and has no goals. This contrasts with BOBBY, who they were told had motivations, desires and emotion. If intentionality is indeed crucial for this effect to occur, we would expect that there would be no reduction in agency when working with the computer. However, if intentionality is not relevant, we would expect the reduction in agency to persist, even when working with a computer.

Method

Participants

30 undergraduate students from the University of East Anglia, again to fulfil a course requirement. Approval was granted by the University of East Anglia ethics committee and consent was obtained from each participant after the experiment. Participants were obtained through SONA. Sample size was based on the number of participants in Experiment 1. Like the previous two experiments, no exclusion criteria were applied. However, to ensure experiment naivety, those who took part in Experiment 1 and Experiment 2 could not take part in this experiment.

Materials

Psychopy3 was used to create the experiment it was run online using Pavlovia.org. The data was collected between 6th October and 12th October 2020. Materials and dependent variables were consistent with *Experiment 1* and *Experiment 2.*

Procedure

The design and the procedure for this study was identical to *Experiment 1* and *Experiment 2*. The exception was that like in *Experiment* 2, they were introduced to their computerised partner. Participants were introduced to their partner through the text shown in

Figure 12. However, unlike in *Experiment 2*, their partner was introduced as a non-agentic computer programme. Participants also saw a blue emoticon of a computer, rather than an emoticon of a face in this experiment.

"Your partner in this experiment is a pre-programmed computer.

The computer's actions have been pre-programmed by the experimenter. Therefore, the computer has no goals and does not record its own performance."

Figure 11 - Text shown to participants introducing their partner as a pre-programmed computer.

Results and Discussion

Like with *Experiment 1* and *Experiment 2*, each participant returned 60 agency ratings per *Partner* condition, resulting in 120 trials across the whole experiment. A median agency rating for each participant was calculated for both *Partner* conditions. Results from a paired sample t-test indicate that participants rate their control over the outcome lower in the Together (M = 59.37, SD = 21.95) compared to the Alone (M = 75.73, SD = 20.71) condition, $t(29) = 4.40$, $d = .803$, $p < .001$. As with *Experiment 1* and *Experiment 2*, participants rate their control over the outcome lower in the Together condition compared to the Alone condition.

As with the previous experiments in this chapter, we found that the sense of agency was reduced when working with a partner compared to working alone. Building on these two experiments, we demonstrate that this occurs when the partner is described as a pre-programmed computer. This could be seen as inconsistent with the results of Ciardo et al. (2018, 2020) who argue that the attribution of intentional agency is an important aspect of this effect.

Exploratory Analysis: Experiment 1 to Experiment 3

The experiments in this chapter were run consecutively, rather than as a single experiment. Therefore, so far, no direct comparisons have been made of the different partner conditions in each experiment. While the studies were not designed for between-group comparison and the sample size might be underpowered, it might be useful for analytical purposes to do so. The experiments followed an identical methodology minus how the partners were introduced and the steps taken to maintain the illusion of a real participant in Experiment 1. As such, the following section includes a short meta-analysis looking at the agency ratings of Experiment 1, Experiment 2 and Experiment 3.

Median agency ratings from *Experiment 1*, *Experiment 2* and *Experiment 3* were combined into a single dataset. This resulted in a 3x2 mixed design, with the within-subjects factor of *Partner* (Alone, Together) and the between-subjects factor of *Computer* (Human, Bobby, Computer).

A mixed ANOVA was conducted and found a main effect of *Partner* ($F(1,93) = 71.65$, $MSE = 12319.78$, $p < .001$, $\eta_p^2 = .441$), no main effect of *Computer* ($p = .27$) and no significant interaction ($p = .93$). Thus, the results indicate that participants rated their control over the outcome lower when they were in the Together condition compared to the Alone condition. The type of partner that they worked with was not shown to affect the presence or the magnitude of this effect. However, it is possible that we lack statistical power to find this effect.

Chapter Discussion

The aim of this chapter was to investigate how different levels of intentionality affect the sense of agency when working with a partner. This was achieved through manipulating how the partner was introduced at the start of each experiment. In one experiment, the partner was introduced as another real participant and measures were taken to keep up this illusion. In the next experiment, the partner was introduced as BOBBY, a robotic agent with motivations and emotion. Finally, in the last experiment the partner was introduced as a pre-programmed computer with no goals and no record of its own performance. Our hypothesis for Experiment 1 was that like in the Beyer et al. (2017, 2018) studies, we should see a reduction in agency when working with others, perhaps because of action-selection dysfluency (Sidarus et al., 2013; Chambon et al., 2014; Sidarus and Haggard, 2016; Sidarus et al., 2017; Sidarus et al. 2020). For Experiment 2, we predicted that if a robotic avatar, rather than an embodied robot (Ciardo et al. 2018, 2020) was sufficient to impress intentionality, then we should see similar results to Experiment 1. Finally, in Experiment 3 we hypothesised that we should not see a reduction in the sense of agency when working with a computer explicated stated to not have intentionality. However, if it were the case that people automatically attribute social characteristics to computers, the effect might be present. The overall outcome of these experiments showed that in every partner condition the sense of agency was reduced in conditions where participants were told that they were working with the partner compared to when they were working alone. Furthermore, as the methodology was identical between the experiments and the only change (other than efforts to keep up the human partner illusion in Experiment 1) was the introductory text and avatars, a meta-analysis was conducted comparing the results of the three experiments. This analysis showed no differences between each partner condition, suggesting that the type of partner (human, robotic agent, computer) that participants believed they were working with did not have an effect.

The results from Experiment 1 replicate the findings from Beyer et al. (2017, 2018), demonstrating that the explicit sense of agency when working with a partner is reduced compared to working alone. This finding extends such research into an online environment, suggesting that the effect is persistent outside of controlled laboratory settings. Future research may benefit from this, as online research has certain benefits compared to in-house laboratories, such as large, diverse sample sizes and cost reductions (Kraut et al. 2004). Further experiments may be inclined to capitalise on these advantages to investigate and use large scale studies to investigate whether individual differences, such as socio-economic status, job status (Caspar et al, 2020) or other demographic information might play a role in this effect on agency. Experiment 2 and Experiment 3 showed that introducing the partner as a robotic agent or simply a pre-programmed computer were both sufficient in producing the reduction in explicit agency. The results from Experiment 2 might be seen as expected. After all, Ciardo et al. (2018, 2020) found similar results used an embodied robot. The novel finding here is that this extends to partners simply described as robotic avatars in addition to physical, embodied robots. Thus, a physical robot is not necessary to reduce the sense of agency in these types of tasks. On the other hand, the results from Experiment 3 could be seen as inconsistent with the idea that intentionality is important for producing this effect. The computer was stated to be pre-programmed and thus, would not have any motivations or goals. This therefore might be seen as analogous to mechanical pump from the study from Ciardo and colleagues. However, there are a few important distinctions. In the study by Ciardo et al. (2020), the pump acting was described as a failure in the system. Therefore, it was framed in a way where when the pump stopped the trial through non-action (i.e. breaking down). As such, participants might be less likely to attribute intentional agency to it. In contrast, the computer's actions were framed in Experiment 3 in the same way as in the other experiments in this chapter. Participants were told that the computer could act to stop the ball from rolling down the slope. This could indeed affect the attribution of intentional agency. Acting and not acting are qualitatively different from one another and have different moral and social consequences (Thomson, 1984). With the mechanical pump, the action is perceived as accidental. In the same way that you would likely not apply intentionality to a boulder rolling down a cliff, you would likely not apply intentionality to a mechanical pump failing. On the other hand, a system which can act, even in a contrived experimental situation, might provide enough social information to attribute intentional agency. Simply informing the participants that the computer is pre-programmed might not be enough to counteract this effect. In other words, the behaviour of the computer acting in the experiment might be sufficient to produce the reduction in sense of agency alone.

The results here contain some important considerations for those researching and designing system with human-computer interaction. As the amount of automation in our lives increases, we begin to rely more on machines for every-day tasks. This reliance on machines will likely only increase in the future. While many will welcome the change that this brings, human activity is often impacted unintentionally by those designing such systems (see Pagliari, Chambon & Berberian, 2021 for review). A reduction in agency when working with a machine that can intervene in a task could be one of those unintended effects. One example of this is the emergence of automation systems within driving. Driving is increasingly incorporating automation to assist drivers, such as the inclusion of parking assist, adaptive cruise control and traffic sign recognition. All these systems can intervene or warn drivers to help make their experience easier. However, evidence suggests that assisted driving reduces the sense of agency (Yun et al., 2018) compared to when the drivers are fully autonomous. The results here are consistent with such results and suggest that systems that can act on their environment are sufficient to produce a reduction in agency, irrespective of whether participants are told they are intentional or not. The idea that a system that can act intervene on a task that a human is working on is an important consideration for those researching human-computer interaction.

Conclusion

The results from the studies in this chapter illustrate how the sense of agency can be reduced in conditions where people are working with a partner on a task. Despite whether the partner is described as a human, a robotic avatar or a computer, the effect is persistent. This chapter questions the level of intentionality that is required to reduce the sense of agency and suggests that systems that can act or intervene on a task are sufficient to produce this effect.

Chapter 5 - Interval Estimation Methodology in Online Settings

Introduction

How do we reflect on and quantify the amount of control that we have over our actions? When we act, it would be safe to say typically we do not think about such things. The feeling of agency is an automatic part of conscious experience that is so pervasive in our lives that we seem to take it for granted. It is of course not something a healthy person should take for granted, as disorders of the sense of agency is part of the symptomatology of those with schizophrenia (Synofzik & Voss, 2010). Nonetheless, it has been noted that the feeling of the sense of agency is phenomenally thin (Haggard, 2017). That is, it occupies a space on the edge of consciousness that isn't easily accessible. It does not have the same stability and clarity of other modalities of conscious experience such as hearing or vision. For instance, it is much easier to imagine a picture of a mountain than the feeling of climbing it. This raises a question about using explicit self-reports as a measure of the sense of agency. How do we ask people to rate or reflect on something that they themselves aren't necessarily aware of? This of course isn't just a question for experiments investigating the sense of agency, but in all domains of psychology. As Beumeister, Vohs & Funder (2007) note:

"…it is abundantly clear from their studies, other research, and everyday observation that people have not always done what they say they have done, will not always do what they say they will do, and often do not even know the real causes of the things they do. These discrepancies mean that self-reports of past behaviors, hypothetical future behaviors, or causes of behavior are not necessarily accurate".

For explicit agency judgments, there is a tendency to overestimate how much control one has over actions and misattribute causality to events which are not relevant to their own actions (Wegner & Wheatley, 1999). Nevertheless, researchers into the sense of agency use explicit self-reports as a measure in experiments. One of the reasons for this at first due to necessity – early research into the sense of agency lacked any objective behavioural measure. However, later research still does use them as they provide a retrospective account of how much control a participant felt over an action-effect pairing. There is also some evidence to suggest that explicit agency measures are sensitive to veridical feelings of control that participants experience. Metcalfe and Greene (2007) found that participants self-reports were sensitive to variables that should affect outcome-monitoring. Participants self-reports of agency reflected the amount of control that they had on each trial. These results suggest that there is a certain amount of validity in asking participants to rate the amount of control that they feel. A measure like this also provides a reflective view of the sense of agency, which could use different cognitive processes that might be lacking in objective measures. It is also used in studies where objective measures might not be convenient or possible in accordance with the experiment's design (e.g. Farrer et al., 2008). Nonetheless, self-reports are prone to criticisms such as demand characteristics and studies which use explicit agency as a measure of the sense of agency can be bolstered by objective measures. One such measure is temporal binding, which is the perceived compression of time between volitional action and its effect. The current understanding of the relationship between temporal binding and explicit agency is unclear. Whilst Dewey & Knoblich (2014) found no correlation between temporal binding and explicit agency, more recent research suggests that they do correlate on a trial-by-trial basis (Imaizumi & Tanno, 2019). Furthermore, research has shown how these two measures of agency dissociate in different paradigms (Saito et al., 2015; Dewey & Knoblich, 2014). Why these dissociations occur is currently not known. There is evidence to suggest that explicit agency relies more strongly on inferential processes (Pacherie, 2014). This is opposed to temporal binding, which appears to be modulated greater by predictive cues (Haggard & Clark, 2003). It seems that the processes between explicit and implicit agency overlap. However, the exact relationship between the two measures of agency is currently unclear.

Most research into temporal binding appears to be conducted within laboratories, rather than online. However, there is also growing evidence to suggest that online experiments can support a variety of effects found in the laboratory (Dandurand, Schultz, Onishi, 2008; Arechar, Gächter & Molleman 2017; Anwyl-Irvine et al. 2020; Peyton, Huber & Copock, 2022). For example, Crump, McDonnell and Gureckis (2013) demonstrated a wide range of paradigms within cognitive psychology (e.g. Stroop effect, Flanker task, Simon tasks) can be replicated online. Such research supports the idea that typical laboratory-based experiments can be conducted within online settings. However, to our knowledge evidence demonstrating temporal binding in online settings is limited. Some studies have demonstrated temporal binding online. Galang et al. (2021) found that typical binding effects occurred in online settings using the Libet Clock version of the temporal binding task. Other studies (Garaizar et al., 2016 & Cubillas et al., 2020) showed that action-binding effects specifically were also obtainable online. However, the Libet Clock method has some limitations. Pockett and Miller (2007) point out that experimental instructions, luminance of the clock hand and its size may affect participants' temporal estimations. Further, the presence of the clock is visually demanding and requires attention and screen space that is likely separate to the task at hand. This leads to difficulties creating tasks that involve the modulation of agency involving other visual stimuli. Further, the clock would also restrict attentional resources to other information such as the interactions of a partner (Silver, Tatler, Chakravarthi, Timmermans, 2020). This makes the Libet Clock method not suitable for social agency paradigms and thus, we look to other ways to measure temporal binding. One approach is the use of interval estimation (Humphreys & Buehner, 2009; Humphreys & Buehner, 2010), which asks participants to time the interval between their action and its effect. This has the advantage of not requiring a clock which takes up visual and attentional resources and could arguably be a more direct measure of binding as the method measures the interval between two events, rather than an event per trial in relation to baseline as with the Libet Clock method. Currently, it is currently unclear whether the conducting the interval estimation method online replicates lab-based temporal binding studies. To test this, we conducted an online experiment using a temporal binding task using the interval estimation methodology. We predict that participants will perceive the temporal gap between two events to be smaller when they are the cause of the outcome versus when the outcome is externally generated. We also expect that this effect will become larger when the temporal gap increases in accordance with previous research.

Experiment 4

In Experiment 4, we looked to understand whether temporal binding could be achieved through the interval estimation methodology online. The purpose of this study was to be a precursor to the next chapter, by first understanding whether the standard temporal binding effect could occur online, rather than within an in-person laboratory. We felt this was a necessary step, as to our knowledge, research had not yet shown temporal binding occurs online using the methodology we are using in this experiment. Further, online experiments potentially bring a number of possible extraneous variables which could lead to temporal binding not being able to be accurately recorded (see Chapter 3 for discussion around online experiments). To do this, we asked participants to estimate the time it took for a shape to change colour after an event occurred in two within-subjects conditions. This event could have either been a white circle appearing within the shape (Baseline), or the participant pressing a button (Operant). Participants would then estimate the time interval between the two events by holding down space. If temporal binding using this particular methodology can be achieved online, we would expect that participants would underestimate the time interval between the two events when they caused the shape to change colour, compared to when the shape changed colour after the circle appeared.

Method

Participants

20 undergraduate students from the University of East Anglia to fulfil a course requirement. Approval was granted by the University of East Anglia ethics committee and consent was obtained from each participant after the experiment. Participants were recruited through SONA (www.sona-systems.com). Data was collected between 28th May 2020 and 26th June 2020. Sample size was based on Buehner & Humphreys (2009) research, with additional participants to account for conducting the experiment online.

Materials

The Gorilla Experiment Builder (www.gorilla.sc) was used to create and host the experiment (Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2018) and the survey platform Alchemer (www.Alchemer.com) was used to obtain consent for data usage and the debrief. All stimuli were created in the Gorilla platform, including a green triangle, a red square and a white circle. Temporal binding was measured through the interval reproduction task (Humphreys & Buehner, 2010), where participants were required to estimate the temporal distance between an event and a shape changing colour. The timing between these two events occurring was manipulated and randomised, to avoid behavioural adaptation to a specific interval.

Procedure

The experiment followed a 2 x 4 repeated measures design, with the factors of *Agency Condition* (Operant and Observational) *and Interval Length* (500ms, 750ms, 1000ms, 1250ms). Trials were split into two blocks, one Operant and the other Observational. The order of these blocks was counterbalanced between participants. In each block, the 4 different interval lengths were presented in a randomised order. Overall, participants completed 112 trials each, 56 trials for each *Agency Condition,* and 14 trials for each *Interval Length.*

The experiment began with a voice call using Skype or Microsoft Teams. The experiment instructions were presented to the participants and they were asked to watch a pre-recorded video outlining what was needed during the study. Participants were then provided with 6 practice trials for each *Agency Condition* (12 in total). Before starting the main experiment, they were asked whether they understood and the task and were happy to proceed. Each trial began with a red square in the centre of the screen on a black background. In Operant trials this square had a white circle at its centre. In this condition, participants were asked to press the space bar at a time of their choosing, this then caused the square to turn green after either 500ms, 750ms, 1000ms or 1250ms. Participants were then presented with a screen displaying the word ‘estimate’ and were required to hold down the space bar to replicate the time between them pressing the space bar and the shape changing colour (Humphreys & Buehner, 2010). In the Observational trials, participants were first presented with a red square without a white circle in its centre and after 1000ms a white circle appeared. Then after 500ms, 750ms, 1000ms or 1250ms, the square turned green. As in the Operant condition, participants were then presented with a screen displaying the word ‘estimate’ and were asked to replicate the length of time between the circle appearing and the shape changing colour. After the end of each trial in both *Agency Conditions,* a blank screen was presented for 1000ms.

Results

Each participant returned 14 temporal estimations per *Agency Condition x Interval Length*, resulting in 112 trials across the whole experiment. Participants' temporal estimates were subtracted by the interval length to calculate the judgment error on each trial. The median judgment error is the primary unit of analysis for the results section. Median judgment errors were calculated for each Agency Condition x Interval Length. Following outlier protocol from a similar experiment conducted by Humphreys & Buehner (2009), if a participant had a median judgment error that was 2 standard deviations from the mean median judgment error on 2 or more *Agency Condition x Interval Length* combinations, they were excluded from the analysis. This resulted in one participant being excluded. A repeated measures ANOVA was conducted and found a significant main effect of *Agency Condition* ($F(1,18) = 5.09$, $MSE = .495$, $p = .037$, $\eta_p^2 = .220$). Participants overall underestimated the temporal length of the two events occurring more in the Operant compared to the Observational condition. There was also a significant effect of *Interval Length* ($F(3,18) = 49.63$, $MSE = 1.10$, $p = < .001$, $\eta_p^2 = .717$). Participants underestimated the temporal length between the two events more when the interval length was longer. Further, the *Agency Condition x Trial Type* interaction was also statistically significant, $F(3,18) = 4.56$, $MSE = .037$, $p = .006$, $\eta_p^2 = .202$). To further investigate this, paired sample t-tests for each *Agency Condition x Interval Length* combination were used to understand the interaction. The results

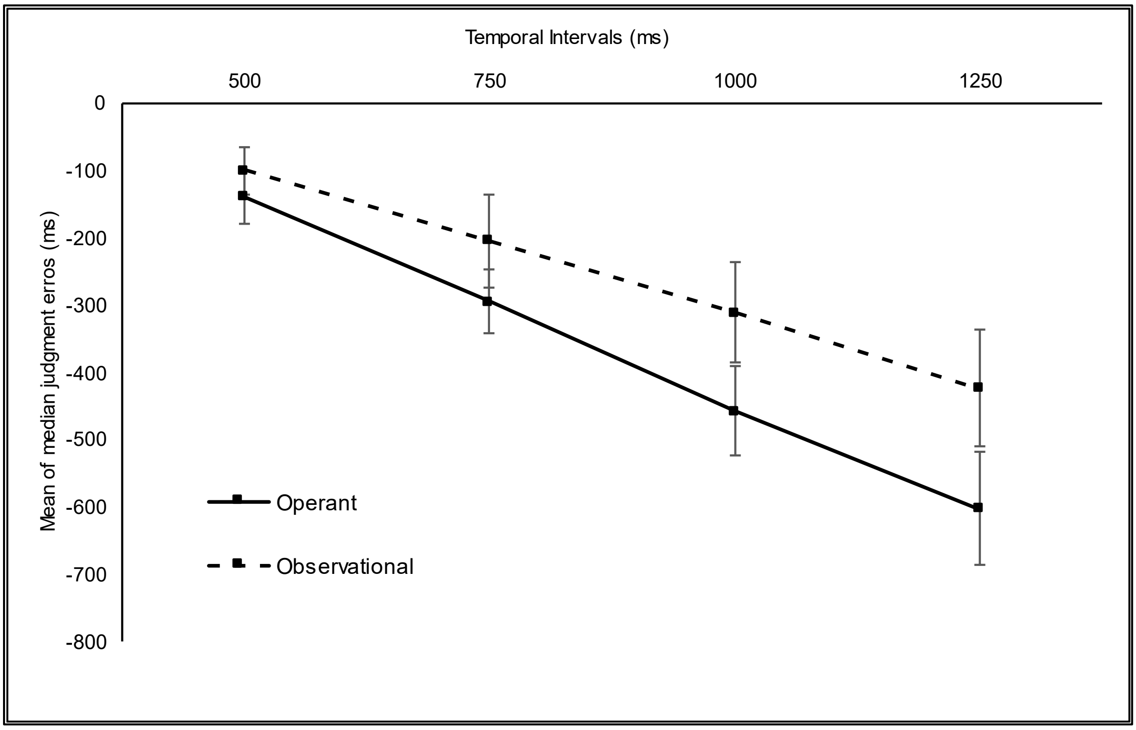

Figure 12 - Mean of median judgment errors across conditions. Error bars indicate the standard error of the mean for each Agency X Interval Length combination

showed no significant differences between the Agency (M = -.14, SD = .17) and No Agency (M = -.10, SD = .15) conditions at 500ms or Agency (M = -.29, SD = .21) and No Agency (M = -.20, SD = .30) conditions at 750ms intervals (p > .05) but a significant difference between Agency Condition at 1000ms ($t(18) = 2.78$, p = .012, d = .23) and 1250ms ($t(18) = 2.42$, p = .026, d = .32). Thus, the interaction can be explained by no significant difference in *Agency Condition* at an interval length of 500ms or 750ms, but a significant difference in *Agency Condition* at 1000ms and 1250ms.

Discussion

Consistent with previous research (Haggard, Clark & Kalogeras, 2002; Humphreys & Buehner, 2009), we found that temporal estimations were shorter when the participant caused the outcome, compared to events that were not controlled by the participant. Temporal estimations also became less accurate at longer interval lengths, with participants underestimating the interval length more at 1000ms and 1250ms compared with 500ms and 750ms. Notably, this underestimation occurred to a greater degree in the Operant condition compared with the Baseline condition.

The fact that temporal binding was only significant at longer interval lengths is interesting. Previous research has demonstrated that delaying the effect over an action reduces one’s sense of agency at a neural level (Farrer et al. 2003). Furthermore, Haggard, Clark and Kalogeras (2002) found that that the temporal binding effect decreases at longer intervals. While the current results may seem inconsistent with such research it is important to note that Haggard Clark and Kalogeras’ research used the Libet Clock method (Libet et al. 1983), whereas this experiment used interval reproduction (Humphreys & Buehner, 2010). Humphreys and Buehner (2009), suggested that the interval reproduction methodology might be more sensitive to longer interval lengths as it offers a way of directly examining the time between cause and effect. As the Libet Clock methodology does not do this directly, this could explain the longer intervals seen here. Alternatively, they also suggest that differences in pacemaker speeds in operant and Baseline conditions could cause this. Another possible reason that we found no temporal binding effect at shorter intervals is that we asked participants to estimate the interval by holding down space. At shorter intervals, participants may not have been able to accurately estimate the interval at a short time frame, as there is a moment where the participant is pressing down the button, along with releasing it which could take up the total time that the keypress is being used to estimate. In other words, the participants could have included the time of the keypress occurring and releasing into their estimation, leaving only a short time window where they are mentalising their estimation. At shorter intervals, this may have affected how they perceived the event, reducing the likelihood to find the typical temporal binding effect.

Whether or not temporal binding is a measure of the sense of agency is up for debate. Some researchers argue that it is an implicit measure of the sense of agency (Haggard, Clark & Kalogeras, 2002; Tsakiris & Haggard, 2003). Specifically, it is thought of as being a measure of, or a component of the pre-reflective agency. That is, the “background buzz” of feeling that you are an author of your actions, which is not directly accessible to conscious thought. However, research also suggests the temporal binding can occur outside of intentional action. For example, Buehner (2012) demonstrated that a machine depressing a participant’s finger onto a button still resulted in the temporal binding effect, even when it was clearly the machine that was causing the event to occur, not the participant. This lead Buehner (2012) to suggest that temporal binding is not specific to intentional action and rather a mechanism related to causality (“*causal binding*”). As our results are exclusively related to intentional action, and thus do not separate out these factors. However, as the sense of agency is seemingly affected by temporal contiguity (Farrer et al., 2003), we might wonder if the results we found here are supportive of the claim that temporal binding is not directly related to the sense of agency. Nonetheless, overall, the results indicate for the first time that temporal binding, obtained through interval estimation, can be obtained through interval reproduction online. This research may pave the way for future research into temporal binding online and may aid in the development of more complex paradigms in the future.

Chapter 6 -Temporal Binding when working with others

Introduction

An important discovery from studies investigating temporal binding is its apparent sensitivity to factors beyond self-action attribution. Temporal binding has been shown to be sensitive to a whole host of different factors that might affect how we experience action (see Moore & Obhi, 2012 for review). As a supposed marker of the sense of agency, this is important as it suggests through proxy that our feeling of being the host of our own volitional actions changes as a function of circumstance. Such factors might be external cues, like action-outcome contingencies such as congruency (Ebert & Wegner, 2010), interval length (Moore, Wegner & Haggard, 2009) or longstanding global effects like mood induction or social hierarchies (Christensen et al., 2019; Caspar et al., 2020). An extension of such research is the investigation of how temporal binding is affected by social cues. The research into this aspect of the sense of agency is varied. There are experiments where participants are asked to perform tasks where they receive social information before they act or receive it as a result of their action. Such paradigms have used items such as emotional valence, eye contact, or coercion (Yoshie & Haggard, 2012, Ulloa et al., 2019, Caspar et al., 2018). In these experiments, the actual outcome of the action can also have social or non-social consequence. For example, eye contact was used and showed that this led to an increase in temporal binding to a non-social outcome (Ulloa et al., 2019). While in the studies investigating the effect of coercion, the participant also provided a shock to their partner. In this case, the participants' actions had a social consequence where their action caused pain to another individual. Therefore, social cues can modulate temporal binding irrespective of whether the action has social or non-social consequences. Other studies have focused on how social interaction *per se* affects temporal binding. Such studies have shown that gaze leading induces temporal binding, demonstrating that we feel a sense of agency over the experience of directing the gaze of another (Stephenson et al., 2018). This study demonstrates that an event with social consequence (i.e. leading another agent to act) in isolation is sufficient to produce temporal binding. In addition, joint-agency paradigms ask participants to perform actions together, where the participants coordinate actions to complete a task (Obhi and Hall, 2012a). These experiments typically include a level of ambiguity, where participants are not sure whether it is them or the other participant causing the outcome. The results from such studies show that temporal binding occurs both when the participant acts and when their partner acts, suggesting that participants feel a sense of 'we' agency (Obhi and Hall, 2012a; Obhi and Hall, 2012b). Or, a joint sense of agency that occurs automatically during joint action (Pacherie, 2014). In total, the emerging research is suggesting that temporal binding can be modulated via social cues. These social cues can vary and do not necessarily have to be task relevant.

Whether or not these social cues need to come from human agents to produce effects on temporal binding is currently unclear. Some studies have shown that the agentic ‘we’ in joint-agency paradigms can only be produced by human co-actors, not machines (Obhi & Hall, 2011b; Sahaï et al., 2019). Furthermore, Grynszpan et al. (2019) found that even when participants were not aware of when they were working with a human or a robotic partner in a joint-action task, the haptic information alone was sufficient to produce temporal binding only when working with another human agent. The results from these findings suggest that just the haptic feedback between human-human interaction is enough to produce temporal binding in joint action, but this does not occur with machines. On the other hand, these studies are specifically related to joint action, which is only one way that social cues can affect temporal binding. Furthermore, we may also ask how anthropomorphic form might affect such measures. On in other words, whether more human-like representations of machines would affect the sense of agency differently to less human-like representations. Little research has been done to investigate this and there is a need to understand when and how machines can affect temporal binding and indeed the sense of agency in general.

A natural extension of such research is understanding whether the presence of another agent can affect temporal binding in the same way that it affects explicit agency (Beyer et al., 2017; Beyer et al., 2018). Such research (including the results from the previous chapter) found that participants rate their feeling of control over the outcome as lower when working with a partner compared to working alone. The action and outcome in this case was stopping a ball from rolling down a slope. Thus, the explicit sense of agency is lowered on non-social outcomes as a result of a social cue. What has yet to be explored in this research is whether this occurs for temporal binding. The current suggestion for why the sense of agency is reduced in this manner is that the partner interferes with action-selection fluency, leading to the sense of agency being reduced (Beyer et al., 2018). However, we currently do not know whether working with a partner has on temporal binding specifically. While we might expect temporal binding to decrease in accordance with explicit agency ratings as they both supposedly reflect the sense of agency, this might not be the case. Aside from the research indicating that temporal binding and explicit agency often dissociate (Dewey & Knoblich; Pfister et al., 2021), temporal binding has been shown to increase when presented with social context (Vogel et al., 2021).

Overview of Experiments 5-8

In the experiments in this chapter, we looked to expand upon our results from Chapter 3 by including temporal binding as a measure along with agency ratings. In the first experiment, we investigated a few lines of inquiry. First, we wanted to understand how temporal binding was modulated because of working with a partner. In addition to this, we explored whether the intentionality of the partner affected the sense of agency. The other experiments made improvements to the design, along with assessing whether the modulation of the sense of agency was related to the partner being able to act, or if it was a mere presence effect.

Experiment 5

In Experiment 5, we adapted the ball-rolling task from the previous chapter to further investigate how agency is manipulated when working with others by measuring temporal binding. In this experiment, instead of a ball rolling down a slope, participants were asked to observe a circle increasing in size. They were asked to stop the circle increasing in size before the circle turned red but were given more points the later they stopped it. Participants could stop the circle in two conditions, one where they were working alone (Alone condition) and another where they were working with a partner (Together condition). An additional condition was included where the participant would observe the circle stopping on its own (Baseline condition). The Baseline condition was included so we could ensure the typical temporal binding effect was indeed occurring. This experiment is analogous to the ball-rolling experiments in Chapter 3, where the participant was asked to stop the ball from falling off the slope, but were given more points the further to the end they stopped it. We adapted the balling-rolling experiment as the paradigm in Chapter 3 was not suitable for temporal binding, as it did not involve sufficient cues for the interval estimation methodology. Namely, an event which could either be caused externally or by the participant (the point at which the participant starts timing their estimation), and another which is produced after a short duration (the point at which the participant stops timing their estimation). Here, we tied the first event to a tone which was either produced when the participant stopped the circle from increasing in size (in the Alone and Together conditions), when the partner stopped the ball (in the Together condition) or was produced when the circle stopped on its own (in the Baseline condition). The second event, which was when the participant would stop their time estimation, was produced through the circle changing colour. Participants would then be asked to estimate the time interval between the tone being produced and the circle changing colour, to generate our dependent variable of temporal binding. After this, the participant was shown the same scale as in Chapter 4, which asked them to rate how much control they had over the outcome, which produced the dependent variable of explicit agency. In line with the results of the previous chapter, we expect that participants would report their feeling of control over the outcome to be lower when working with a partner. As participants do not control the outcome in the Baseline condition, we would expect agency ratings to be lowest here. If implicit agency was affected in a similar way to explicit agency, we would expect temporal binding to decrease in the Together condition compared to the Alone condition. Following the typical temporal binding effect (Haggard, Clark & Kalogeras, 2002), we would expect temporal binding to be lowest in the Baseline condition.

Method

Participants

A power analysis (conducted via G*Power; Faul et al., 2007) suggested that to find a medium to large effect size with a between-subjects factor, we would require a minimum sample size of 100 participants ($f = 0.3$, $1 - \beta = 0.95$, $\alpha = 0.05$, number of groups = 2, number of measurements = 3, correlation among rep measures = 0.5).

34 undergraduate students from the University of East Anglia were recruited via SONA in this study to fulfil a course requirement. 101 participants were also recruited from Prolific (www.prolific.co) and paid £2.70 for their participation (at a rate of £8 an hour). In total, 135 people were recruited to take part in this study. This sample can detect effect sizes of $f = 0.26$, $1 - \beta = 0.95$, $\alpha = 0.05$. Data was collected between 14th May 2021 and 11th June 2021. Participants were recruited through SONA (www.sona-systems.com). Ethical approval was granted by the ethics committee at the University of East Anglia.

Materials

Psychopy3 (Peirce, Gray, Simpson et al. 2019) was used to create the experiment and the experiment was run online on Pavlovia.org. A 50hz tone was created in Psychopy3 and used when the circle stopped, irrespective of whether the participant caused the circle to stop or this was caused by the computer. Temporal binding was measured through the interval methodology (Buehner & Humphreys, 2010), where participants replicated the temporal interval between two events (a tone and a shape changing colour). Explicit agency ratings were measured through a slider, created in Psychopy3.

Procedure

The experiment followed a 2 x 3 mixed factorial design, with the between-subjects factor of *Computer* (BOBBY, The Computer) and the within-subjects factor of *Partner (*Baseline, Alone, Together). Participants were randomly assigned to each *Computer* condition before the experiment began and the between-subjects factor was presented in three separate blocks which were fully counterbalanced. Participants completed 81 trials in total, 3 practice trials in each block and 24 trials for each block in the main experiment. Randomised block design was used, rather than the presenting the conditions randomly in a single block like that used in Chapter 3 as we felt it would be easier for the participant to manage and understand the differences between each condition, especially with the inclusion of the Baseline condition, where participants would be required to not act at all. It also allows participants multiple breaks at logical points during the experiment, providing natural rest times for them. As the blocks themselves are randomised and fully counterbalanced, we would not expect any order effects to be present in the data. Likewise, any learning effects within a block would be systemic within the experiment.

We measured two dependent variables in the experiment. One was the subjective temporal compression between two events (so called temporal/intentional binding). Participants were asked to estimate the time interval between two events occurring by holding down space. These two events were a tone occurring and a circle changing colour. We then calculated their proportional reproduction error by dividing their time estimation by the veridical time interval on each trial. As a result, cases where the time interval is 100% indicates that the participant perfectly reproduced the actual time interval. Scores of less than 100% indicate the participant underestimated the actual time interval and scores over 100% show the participant overestimated the interval length. We're using proportional reproduction errors as our dependent variable, rather than the actual intervals produced by the participant (like in Experiment 4), as there are multiple stimulus onset asynchronies which participants would be required to time and it was determined that we do not have a sufficient number of trials to analyse these individually. Therefore, normalising participants' estimations to a proportion of the actual time between the two events allows us to account for differences between the different stimulus onset asynchronies into a single measure and also allow for easier interpretation of results.

As with the previous experiments, participants were also asked to rate how much control that they felt over the outcome on a 1-100 slider. Where the far left of the slider had the label "No Control" and the far-right label stated "Full Control". The order that the two dependent variables were presented in was not counterbalanced as if participants were required to perform the interval estimation task after reporting how much they control they felt on the slider, they would be required to maintain the estimation in memory for a longer period, along with likely experiencing increased cognitive load of reporting explicit agency first. This would likely decrease the accuracy of their estimation. Therefore, participants always performed the interval estimation before reporting how much control they felt on the slider.

Before the experiment began, participants were brought to a survey presenting the information sheet and the consent form. Once the consent form was signed the participants were then brought to a page to test for sound on their computer. Participants were presented with a 50hz tone with a duration of 500ms, repeating every 1000ms. They were asked to adjust the volume on their computer to a comfortable listening volume and then press space when they had done so. This was to ensure that participants did not need to adjust their volume during the experiment. Participants were then brought to the instructions page which explained the task. This was where they were introduced to the avatars of their partner. If participants were in the BOBBY condition, they saw a blue avatar of a smiling emoji (created in PowerPoint). If participants were in The Computer condition, they saw a blue icon of a computer (created in PowerPoint). Their own avatar was a white, straight-faced emoji. They were also given introductory texts for each condition. If they were working with BOBBY participants were shown the text in *Figure 11.* If participants were working with The Computer, they were shown the text in *Figure 12*.

In each condition, participants were first presented with the avatars for 1500ms, which told participants who they would be working with on each trial. In the Alone condition, participants would only see their own avatar on the left of the screen. In the Together condition, participants would see their own avatar on the left and their partner's avatar on the right. In the Baseline condition, participants would only see their partner's avatar on the right side of the screen. After this, a white circle appeared in the centre of the screen for 1000ms and after this period began increasing in size. The participants were instructed to stop this circle enlarging by pressing space. They were told that the later they stopped the circle the less points they will lose. However, if they stopped the circle too late and the circle turned red, they will lose maximum points. To make the task more difficult, the point at which the circle turned red varied randomly within the last 14% of the circle enlarging. In the Alone condition, participants work alone without the possibility of their partner stopping the circle. If participants stop the circle successfully a 440hz tone will occur at the time of their keypress. The circle would then turn green after a stimulus onset asynchrony of 750ms, 1000ms or 1250ms. These interval lengths occur at random and are evenly distributed among blocks so that each interval length is presented to participants 8 times per condition. We chose these interval lengths, rather than the interval lengths in Chapter 5, because we did not find temporal binding to be significant at shorter intervals. If participants did not stop the circle a 131hz tone would occur and the circle would turn red after the same interval periods. In the Together condition, the participants partner (BOBBY or The Computer) could also act during the trial. In both *Partner* conditions, the programme was set to respond on 30% of trials and only if the circle had increased to above 80% of its size. If the participant's partner acted, a 220hz tone would be produced and the circle would turn blue after the set interval lengths. In the Baseline condition, participants observe their partner perform the task. In these trials, the partner successfully stopped the circle on 100% of trials and when it stopped the circle, a 440hz tone occurred and the circle's colour turned green. This was to ensure that the Baseline condition matched the trials where the participant successfully acted in the Alone and the Together conditions.

Irrespective of whether the participant acted, the partner acted or neither acted, participants were presented with a screen that stated who acted and were asked to estimate the time between the tone occurring and the shape changing colour by holding down space. While holding down space, participants were presented with a screen stating “estimate…”. If the participant’s estimation was lower than 100ms, then they would be presented with a screen stating that their estimation indicated that they did not hold down space and explained further that they need to estimate the interval between the tone occurring and the shape changing colour. After the temporal estimation, participants were asked to rate how much control they felt they had over the outcome. Participants were then told how many points they had lost for 1000ms at the end of each trial. After the main experiment finished, participants were brought to a brief survey which asked demographic questions, the Individual Differences in Anthropomorphism Questionnaire (IDAQ) and the debriefing sheet.

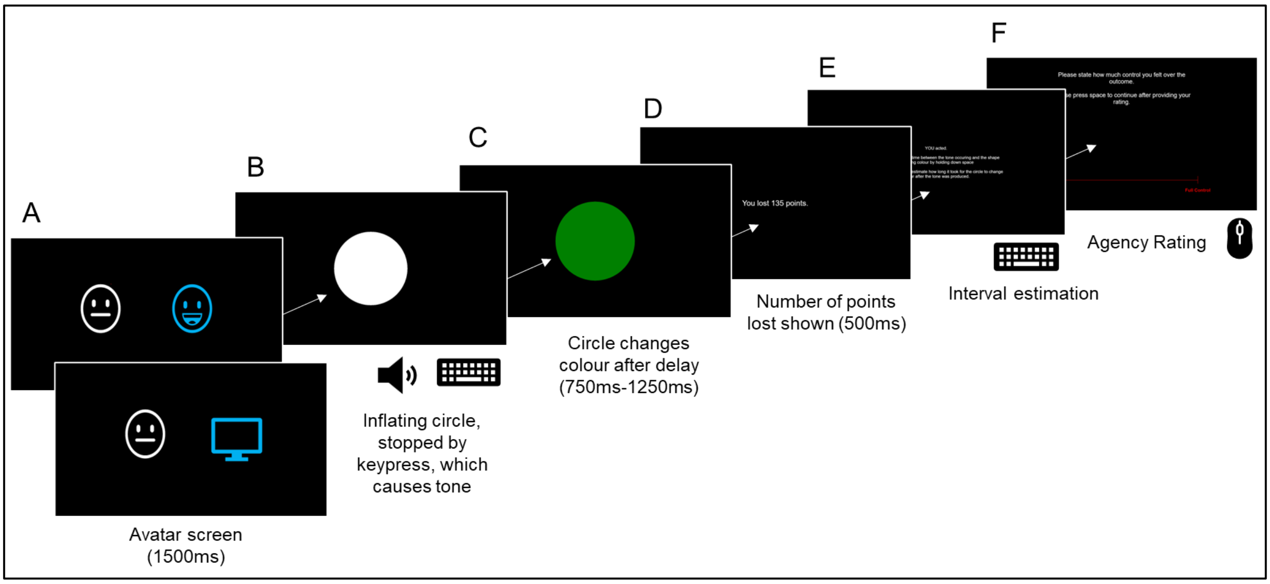

Figure 13 - Trial timeline. Trial starts with the two avatars (left being the participant's and the right being their partner) for 1500ms. In the alone condition only their avatar is shown and in the baseline condition only their partner’s avatar is shown. The circle is then shown, which begins increasing in size after 100ms. Once the circle stops, a tone will occur and it will change colour after a fixed delay of either 750ms, 1000ms or 1250ms. If the participant successfully stops the circle before it turns red, it will turn green. If the partner stops the circle it will turn blue. Participants are then shown how many points they have lost for 500ms before being asked to estimate the time between the tone occurring and the circle changing colour. Once completed, participants were asked to rate how much control they felt they had over the outcome on a slider.

Results

Exclusion Criteria

As the speed of the circle increasing size was tied to the framerate of the monitor and the experiment was conducted online, participants were excluded if they were not using a 60hz monitor as this was the framerate that the experiment was designed in. This led to 18 participants being excluded from the analysis. Next, the following exclusion criteria was applied to the data: Trials where temporal estimations varied by 3 SDs from an individual subject's mean across all experimental conditions or were less than 100ms were removed from analysis (202 trials, 1% of total trials). Trials where the participants did not act in the Alone and Together condition were then removed (2022 trials, 11% of total trials). Finally, 2 participants were removed from the analysis because they contributed less than 8 trials in one of the *Partner* conditions after the above exclusion criteria was applied. Remaining participants took part in at least 8 valid trials in each condition. This left 115 participants in total, 60 in the BOBBY group and 55 in The Computer group.

Temporal Reproductions and Stimulus-Onset-Asynchrony

A repeated measures ANOVA was used to assess whether participants could detect changes in differing lengths between the effect of their action. The ANOVA used participants' interval reproductions (i.e. the time that they held the space bar) as the dependent variable and the three different stimulus-onset-asynchronies within the factor of *Time* as the independent variable. These were 750ms (M = .73, SD = .31), 1000ms (M = .78, SD = .32) and 1250ms (M = .86, SD = .35). Results from the ANOVA showed a significant main effect of *Time*, demonstrating that participants' interval estimations increased as the length between their action and its effect increased ($F(2,228) = 57.90$, MSE = .49, $p < .001$, $\eta_p^2 = .397$).

Proportional Temporal Reproductions

For each of the remaining participants, mean proportional reproduction errors were calculated by dividing the veridical stimulus-onset-asynchrony between action and effect by the interval estimation from the participant. These were then inserted into a mixed ANOVA with the between-subjects factor of *Computer* and the within-subjects factor of *Partner*. Results showed a main effect of *Partner* (F(2,226) = 26.23, MSE = .76, $p < .001$, $\eta_p^2 = .187$), no main effect of *Computer* ($p = .21$) and no interaction effect ($p > .45$). Therefore, the mean differences between the *Partner* conditions were statistically significantly different from one another, but there were no differences between those who worked with BOBBY and The Computer at any level of the *Partner* condition.

Planned comparisons T-tests were conducted to investigate the main effect of *Partner* further. As these were determined before conducting analysis and was part of the experimental design, no multiple comparison corrections were made. Results showed that participants mean proportional reproduction errors were significantly more accurate in the Baseline (M=88.88%, SD = 38.67%) compared with the Alone (M=78.16%, SD = 34.57%, $t(114)$, = 4.25, d = .40, p < .001) and Together (M=73.00%, SD = 28.42%, $t(114)$ = 6.94, d = .66, p < .001) conditions. Furthermore, mean proportional reproduction errors were significantly more accurate in the Alone condition compared with Together condition ($t(114)$ = 2.71, d = .25, p = 008). Thus, the main effect of partner can be explained by participants underestimating temporal intervals more in the Alone and Together conditions when compared with the Baseline condition, but also underestimating more in the Together condition compared with the Alone condition.

Explicit Agency Ratings

Median agency ratings were computed, and a mixed ANOVA was conducted. The results showed a significant main effect of *Partner (F*(2,226) = 197.89, MSE = 102863.04, p < .001, η_p^2 = .637) and neither a main effect of *Computer (p* > .33) and no interaction (p > .73). The ANOVA therefore indicated that there is a significant difference between the group means of the *Partner* condition. To explore this further, planned comparisons t-tests were administered with no multiple comparison corrections and found that participants rated their control significantly lower in the Together condition (M = 70.00, SD = 19.31) than in the Alone condition (M = 76.41, SD = 20.36), t(114) = 3.39, d = .32, *p < 001.* Further, participants rated their control over the outcome in the Baseline condition (M = 21.85, SD = 28.44), lower than in both the Alone and Together conditions: t(114) = 14.86, d = 1.39, p < .001, t(128) = 15.07, d = 1.40, p < 001. Participants rated their feeling of control over the outcome lower in the Baseline compared to the Alone and Together conditions and rated their control as lower when they are in the Together condition compared to the Alone condition.

Individual Differences in Anthropomorphism

For each participant, an individual IDAQ score was computed, following the instructions from (Waytz, Cacioppo & Epley, 2010). As a precautionary check, an independent samples T-Test was administered to assess whether there were any differences between IDAQ scores between the two *Computer* groups. Results showed no differences between the two groups (p = .59).

To test the relationship between anthropomorphism and temporal binding, we subtracted mean proportional reproduction errors from the *Together* conditions by the *Baseline* condition for each participant. This was to control for individual differences in baseline temporal judgments. Pearson's correlations were then administered for the Together-Baseline subtraction for each *Partner* group against IDAQ computations. One participant was removed from analysis as it was a clear outlier. Results showed no correlation between Baseline-Together subtraction and IDAQ scores (p = .075).

Other Behavioural Analysis

To understand participants behaviour in different conditions, we computed the proportion of times that they acted in the Alone and Together. Proportions were not computed for the Baseline condition as participants were unable to act in it to stop the circle. Means for both the Alone (M = 84.19%, SD = 9.10%) and Together (M = 67.76%, SD = 15.92%) were then calculated and a paired sample t-test was conducted to assess whether participants acted less in the Together condition than the Alone condition. As expected, results from this showed that participants acted less in the Together condition compared to the Alone condition (t(114) = 13.19, d = 1.23, p = <.001). To check whether there was a correlation between the number of times a participant acted and temporal binding, we ran a Pearson's correlation for the number of times acted in the Alone condition by interval estimations and the same for the Together condition. Results showed that there was no significant relationship between the number of times acted and interval estimations in the Alone condition (p = .43) or the number of times acted and interval estimations in the Together condition (p > .11). Thus, it is unlikely that differences in temporal binding between the Alone and Together condition are due to differences in the number of times acted *per se*.

To understand whether the number of times that the partner acted affected the sense of agency, we ran a Pearson's correlation between the number of times that the partner acted in the Together condition and the mean proportional reproduction error subtractions between the Baseline and the Together conditions outlined in *Individual Differences in Anthropomorphism*. The results showed no correlation (r(114) = .068, p > .40). Thus, there is no relationship between the number of times the partner acted in the Together condition and interval estimations. To check whether this relationship also existed for explicit agency, a Pearson's correlation was conducted between the number of times the partner acted and agency ratings in the Together condition. No relationship was found (p = .919). Thus, the number of times the partner acted had no small association with temporal binding, nor explicit agency scores.

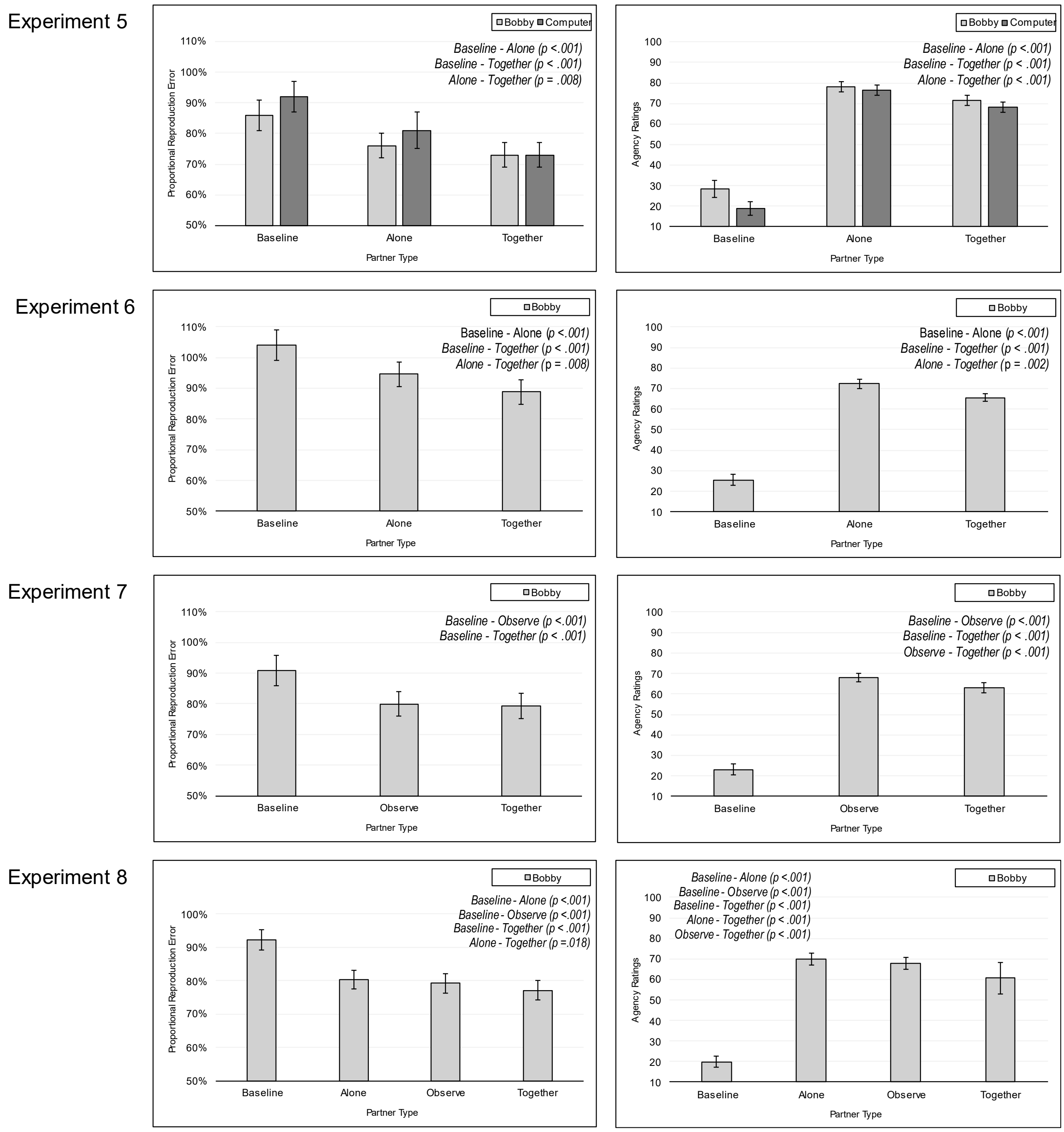

Figure 14 - Results for Experiments in Chapter 5. Graphs on the left illustrate proportional reproduction errors (temporal binding) and graphs on the right show agency ratings. Statistically different differences within an experiment and measure has been denoted on the graph.

Interim Discussion

In this study, we asked participants to stop a circle from increasing in size to gain points. Participants completed the study in three conditions, one where they were working alone, one where they were working with a partner and one where they were observing their partner complete the task. Participants were split into two groups, one where they were told that their partner was a robotic agent named BOBBY and the other where they were told that they were working with a pre-programmed computer. Like the experiments in *Chapter 2* and other studies (Beyer et al., 2017,2018; Ciardo et al., 2020), we found that participants agency ratings were lower when working with their partner compared to working alone. However, this was irrespective of whether they were told that they were working with BOBBY or The Computer. As expected, agency ratings were lowest in the Baseline condition, where they were told to observe their partner performing the task and could not stop the circle themselves. These results suggest that explicit agency is reduced when working with a partner and is lowest when participants cannot control the outcome of an event. In contrast to this, we also found that participants underestimated their temporal reproductions in the Together condition compared to the Alone condition, suggesting that implicit agency was in fact higher when working with a partner. People most accurately reproduced temporal intervals when observing their partner perform the task, suggesting agency is again lowest when the participant cannot act. We did not find any correlation between IDAQ scores and either explicit agency ratings or temporal binding in the Together condition, suggesting that individual differences on this measure do not play a role[1]. Thus, the main findings in this experiment are lowered explicit agency ratings when working with a partner compared to working alone but increased temporal binding.

The inversion of explicit agency and temporal binding could be seen as unexpected. After all, the former reflects a decrease in agency, whilst the latter would indicate an increase. However, implicit and explicit measures of the sense of agency do not always correlate and are often seen to dissociate (Dewey & Knoblich; Pfister et al., 2021). This has led researchers to suggest that implicit and explicit agency might be tapping into different cognitive and neural processes (Moore et al., 2012). Whilst this may be the case, the current explanation for why the presence of another actor reduces the sense of agency is because it affects action-selection dysfluency (Beyer et al., 2017; Beyer et al., 2018). Whilst the explicit agency results from this study are consistent with this idea, the temporal binding results are

[1] In our initial analysis we found the correlation was significant, which is why the measure was used for the rest of experiments. However, after quality control checks we realised that the correlation was a result of an incorrect outlier exclusion.

inconsistent. If the presence of another actor affects action-effect dysfluency, we should expect lower temporal binding in the Together condition.

It is worth noting that an initial explanation of the data could be that because participants had less of a chance to act in the Together condition (as the partner would sometimes act). This might in some way make their estimations less accurate, which in turn might explain why they are more likely to underestimate their temporal reproductions. However, the results showed that while people did indeed act less in the Together condition compared to the Alone condition, there was no relationship between the number of times that they acted in either the Together condition or the Alone condition. Therefore, it is unlikely that the difference between the Alone and Together conditions can be explained by participants acting less when working with the partner.

Finally, another consideration from the data is that temporal estimates in the Baseline condition were lower than we expected. Temporal estimations are typically accurate when a participant is not acting (Haggard, Clark & Kalogeras, 2002). Thus, it is surprising that the average estimation is 10% from the veridical stimulus-onset-asynchrony. One explanation for this is that we told the participant that they were observing their partner performing the task. This might induce effects of vicarious agency and therefore reduce temporal estimations (Sahai et al., 2019). Alternatively, the intervals between action and effect might have been too long, leading to underestimations. As there is a growing need for replication of results in Psychology (Ioannidis, 2005), we will look to amend these considerations whilst replicating the current results in the next experiment in this chapter.

Experiment 6

In Experiment 6, we aimed to replicate the results from Experiment 5 and make some adjustments to the procedure. Typically, temporal estimations are relatively accurate when estimating the interval of two externally caused events (e.g. Stephenson et al., 2018). Because the temporal estimations in the Baseline were 10% lower than the veridical interval time, we looked for potential reasons why this might be the case. One potential reason is that participants were told that they would be observing the partner stopping the ball in the Baseline condition. Previous research has shown that it is possible for humans to experience agency over the actions of another (Sahai et al., 2019). Therefore, it is possible that the way that we described the condition led participants to feel vicarious agency over the partner’s actions, thus leading to a decrease in temporal binding. To account for this, in Experiment 6 participants were simply told that the circle would stop on its own, rather than stop through the actions of their partner. Another reason why temporal estimations were lower than expected in Experiment 5 could be that the stimulus onset asynchronies were too long for accurate estimation. We therefore reduced the time between the tone occurring and the circle changing colour in each condition from 750ms to 1250ms to 600ms to 1000ms. Adjusting these intervals also allows us to investigate whether the effect of working with a partner is still present at shorter intervals. If the results from Experiment 5 are reliable and do indeed persist at shorter intervals, we would expect temporal binding to increase when working with a partner, compared to working alone and expect higher temporal binding in the Alone and Together conditions compared to Baseline. We would also expect agency ratings to be lower in the Together condition compared to the Alone condition and lowest in the Baseline condition. Finally, as we found no difference between the BOBBY and The Computer conditions, we proceeded with only including the BOBBY condition for the rest of the experiments within this chapter.

Method

Participants

Before sampling, an additional power analysis was conducted via G*Power (Faul et al., 2007) to assess necessary sample size for a difference between two dependent means. We used the Cohen's *d* calculation from Experiment 5 for the effect size between the Alone and Together condition (d = .29). The results of this analysis suggested a sample size of 96 participants would be required to detect effects of two-tailed t-test with the current power (t = 1.99, 1 – β = 0.8, α = 0.05). 142 participants were recruited from Prolific (www.prolific.co) and paid £2.70 for their participation (at a rate of £8 an hour). Ethical approval was granted by the ethics committee at the University of East Anglia. Participants were recruited through SONA (www.sona-systems.com).

Materials

Psychopy3 (Peirce, Gray, Simpson et al. 2019) was used to create the experiment and the experiment was run online on Pavlovia.org. Data was collected between 14th July 2021 and 21st July 2021.

In this experiment, we removed the between-subjects factor of *Computer.* Each of the previous conditions remained, except that the *Baseline* condition was altered. This is to participants were told that they are simply observing the circle stopping on its own, rather than their partner stopping it. Therefore, we have a single factor, repeated measures design, with the independent variable being *Partner* (Baseline, Alone, Together). Participants completed 81 trials in total, 3 practice trials in each block and 24 trials for each block in the main experiment. Like in Experiment 5, we measured two dependent variables in this experiment: temporal binding and explicit agency ratings. The dependent variables were measured and calculated identically to Experiment 5.

Procedure

The procedure was identical to Experiment 5, with a few exceptions. *The Computer* condition was removed. Participants were only told that they were working with BOBBY. The baseline condition was also altered. Whilst in actuality the experience of this condition was the same, the way that we described the condition to the participant was different. Participants were now told that they were observing the circle from stopping, rather than watching their partner stopping the circle. Finally, the stimulus-onset-asynchrony after the participant acts to stop the circle was changed from 750ms, 1000ms, 1250ms to 600ms, 800ms, 1000ms. Like before, when the main experiment finished, participants were brought to a survey and were asked to provide demographic information and answers to the IDAQ, along with the debriefing sheet.

Results

Exclusion Criteria

The same exclusion criteria that were used in Experiment 5 was applied here. Trials where temporal estimations were less than 100ms or 3 SDs above an individual subject's mean were excluded from the analysis (4% of total trials). Trials where the participant did not act were also removed (18% of total trials). Finally, 6 participants were excluded from the analysis as they contributed less than 8 trials in any condition after the above trial exclusion occurred. 13 participants were also removed as they did not use a 60*hz* monitor. Therefore, 123 participants remained for analysis.

Temporal Reproductions and Stimulus-Onset-Asynchrony

To see if participants' interval estimations increased as the stimulus-onset-asynchrony after their action increased, a one-way repeated measures ANOVA was conducted, using their temporal interval estimation as the dependent variable and the factor of *Time* as the independent variable. Results showed that interval estimations increased as the temporal gap between their action and its effect increased ($F(2,244) = 39.21$, MSE = .19, $p < .001$, $\eta_p^2 = .243$). Thus, participants interval estimations increased as the stimulus-onset-asynchrony increased from 600ms (M = .74, SD = .29), 800ms (M = .77, SD = .28) and 1000ms (M = .82, SD = .30).

Proportional Temporal Reproductions

As in Experiment 5, each of the mean proportional reproduction errors were calculated by dividing the veridical stimulus-onset-asynchrony between action and effect by the interval estimation from the participant. A one-way repeated measures ANOVA was then conducted to assess differences between the means and showed a significant main effect of *Partner* (F(2,244) = 25.51, MSE = 1.10, $p < .001$, $\eta_p^2 = .173$).

Planned comparisons T-tests were then used, with no multiple comparison corrections to assess the individual components of the *Partner* condition. Results indicate that participants mean proportional reproductions were significantly higher in the Baseline (M=109.91%, SD = 44.48%) condition compared to the Alone (M=97.68%, SD = 39.18%, $t(122) = 4.52$, $d = .41$, $p < .001$) and Together (M=91.34%, SD = 35.71%), $t(122) = 2.69$, $d = .59$, $p < .001$ conditions. Furthermore, the mean proportional reproductions were also higher in the Alone condition compared to the Together condition, $t(122)$, = 2.69, $d = .24$, $p = .008$. The main effect of *Partner* can therefore be explained by lower temporal estimations in the

Alone and Together conditions compared to the Baseline conditions and also lower temporal estimations in the Together condition compared to the Alone condition.

Explicit Agency Ratings

Median agency ratings were computed, and a one-way repeated measures ANOVA was conducted, which showed a significant main effect of *Partner (F*(2,244) = 91.37, MSE = 48185.80, *p* < 001, η_p^2 = .428) To explore this further, planned comparisons T-tests with no multiple comparison adjustments were administered and found that participants rated their control in the Baseline (M = 32.07, SD = 32.23) condition lower than both the Alone (M = 63.66, SD = 26.35, t(122) = 10.23, *d* = .92, *p* < .001) and Together (M = 63.45, SD = 22.26, t(122) = 9.85, *d* = .88, *p* .< 001) conditions. Participants' agency ratings were also lower in the Together condition compared to the Alone condition (t(122) = 3.12, *d* = .28, *p* = .002). Thus, participants rated their control in the Baseline condition lower than in the Alone and Together conditions. Participants also rated their control as lower in the Together condition compared to the Alone condition.

Individual Differences in Anthropomorphism

Like with Experiment 5, an IDAQ scores were computed for each experiment following guidance from Waytz, Cacioppo & Epley (2010). To test whether there was a relationship between proportional reproduction errors and IDAQ scores, temporal estimation errors in the Together condition were subtracted from estimation errors in the Baseline condition. A Pearson's correlation was then used to assess the relationship between this computation and IDAQ scores. No relationship was found (*p* = .47). Agency ratings in the Together condition were also inserted into a Pearson's correlation with the IDAQ scores and no significant relationship was found (*p* = .68).

Other Behavioural Analysis

To test the number of times that participants acted between conditions, a paired sample t-test was conducted looking at the proportion of times that participants acted in the Alone (M = 82.47%, SD = 15.49%) and Together (M = 68.35%, SD = 13.04%) conditions. Results showed that participants acted more in the Alone condition compared to the Together condition (*t*(122) = 10.60, *d* = .95, *p* < .001). Next, we checked whether there was a correlation between the number of times acted and temporal reproduction errors in the Alone condition and the same for the Together condition and found no significant relationship in either *(p* = .75, *p* = .97). Next, looked at whether the number of times that the partner acted in the Together condition related to temporal reproduction error and found no relationship (p = .52). Thus, neither the number of times that the participant acted or the number of times that the partner acted affected temporal binding in this experiment.

Interim Discussion

The aim of this study was to replicate the findings from Experiment 5 and make some adjustments to some potential small issues with the experiment. We changed the description of the Baseline condition so that rather than informing the participants that they would be observing their partner, they would be watching the circle stop on its own. We also adjusted the time that the circle changed colour after the participants action to 600ms, 800ms and 1000ms. The results from this experiment replicate those from Experiment 5, showing that temporal binding is increased when working with their partner, but explicit agency ratings are reduced. We also found that temporal binding and agency ratings were lowest in the Baseline condition as expected. We found no relationship between either temporal binding or explicit agency ratings with IDAQ scores. In addition, we found that whilst participants acted more in the Alone condition compared to the Together condition, this did not correlate with interval estimation reproductions.

Experiment 7

As Experiment 6 replicated the results from Experiment 5, we looked to further investigate the driving factors behind the results. One aspect that we were interested in further understanding was whether an individual's agency being altered by working with a partner was contingent on the partner being able to act or could be achieved simply through the fact that the partner is present during the task. The mere presence effect describes the effect of social facilitation when the influences of other persons are removed, minus the fact that they are present in the situation (Guerin, 1985). In this chapter, the question remains as to whether the increase in temporal binding and reduction in explicit agency is specifically tied to the partner being able to act on the task, or is simply a result of their presence during the trial and therefore due to a mere presence effect. Furthermore, other studies have found that socially driven stimuli increase binding (Vogel et al., 2021; Silver et al., 2024). Thus, it is possible that the modulation of the sense of agency when working with a partner is driven by the fact that it is more social of a condition than when working alone. To further understand this, we replaced the Alone condition with a condition with called "Observe". Here, participants were told that they would perform the condition alone, but their partner would be observing them. The goal here was to create the idea that the partner was present, but they would not be able to act on the task. To reinforce this idea, their avatar was presented in the top right-hand corner of the screen during the trial. The Together and Baseline conditions remained the same. If the effect found in the previous experiments (i.e. an increase in temporal binding and a reduction of explicit agency) was contingent on the partner being able to act, then we would expect no differences between the Observe and the Together condition. Alternatively, if the effect is from the partner simply being present during the task, we should expect the effect to be present in this experiment. In addition to this change, we reverted the stimulus onset asynchronies back to the intervals from Experiment 5 (750ms, 1000ms and 1250ms), rather than those provided in Experiment 6 (600ms, 800ms, 1000ms). We did this as we found that in Experiment 6, participants were vastly over-estimating the time interval at 600ms.

Method

Participants

147 participants were recruited from Prolific (www.prolific.co) and paid £2.70 for their participation (at a rate of £8 an hour). The sample size of this study was based on the

samples from Experiment 5 and Experiment 6. Ethical approval was granted by the ethics committee at the University of East Anglia.

Materials

Psychopy3 (Peirce, Gray, Simpson et al. 2019) was used to create the experiment and the experiment was run online on Pavlovia.org. This study followed a similar design to *Experiment 6*. However, the Alone condition is now replaced with an Observe condition. Thus, the factor of *Partner* is now made up of three conditions: Baseline, Observe and Together. We therefore have a single factor, repeated measures design with the independent variable of *Partner* (Baseline, Observe, Together). As before, participants completed 81 trials, 3 practice trials for each condition and 24 trials in each condition of the main experiment. As with the other experiments, the dependent variable was temporal binding and explicit agency ratings.

Procedure

The procedure for this experiment was largely the same as Experiment 6. The main change in this experiment to Experiment 6 was that the Alone condition was replaced with the Observe condition. In this experiment, participants acted alone but they were told that BOBBY would be observing them as they performed the task. To reinforce this idea, BOBBY's avatar was shown in the top right of the screen while the circle was increasing in size. This disappeared after the participant stopped the circle. Notably, BOBBY's avatar was not shown in the Together condition while the circle was increasing in size. Another change was the reversion of the stimulus-onset-asynchrony between the participant's action and its effect to the delays in Experiment 5 (750ms, 1000ms, 1250ms). We reverted this change as it seemed that participants were vastly overestimating their temporal estimations when the delay was at 600ms.

Results

Exclusion Criteria

The exclusion criteria followed those from the previous experiments in this chapter. Trials where temporal estimations were less than 100ms or 3 SDs above an individual subject's mean were excluded from the analysis (5% of total trials). Trials where the participant did not act were also removed (15% of total trials). 3 participants were removed from the analysis as they contributed less than 8 trials in any condition after trial exclusion and 12 participants were removed for not using a 60*hz* monitor. Therefore, 132 participants remained for analysis.

Temporal Reproductions and Stimulus-Onset-Asynchrony

As before a one-way repeated measures ANOVA was conducted to see if temporal estimations increased as interval delay increased. Results showed a main effect of *Time*; interval estimations increased as the delay after the participants action increased ($F(2,262) = 64.00$, MSE = .61, $p < .001$, $\eta_p^2 = .33$). The participants' temporal estimations increased as the interval delay increased from 750ms (M = .76, SD = .33), to 1000ms (M = .82, SD = .35) to 1250ms (M = .90., SD = .39).

Proportional Temporal Reproductions

Proportional temporal reproductions were calculated in the same way as the previous experiments. Proportional temporal reproductions were then inserted into a one-way repeated measures ANOVA with the factor of *Partner* (Baseline, Observe, Together). Results showed a main effect of *Partner* (F(2,262) = 17.77, MSE = .46, $p < .001$, $\eta_p^2 = .119$). Planned comparison T-tests were then administered with no multiple comparison adjustments and showed a significant difference between the Baseline (M = 91.00%, SD = 38.06%) and Observe (M = 81.28%, SD = 39.18%, $t(131) = 4.57$, $d = .40$, $p < .001$) conditions. Also, a significant difference between Baseline and Together (M = 80.29%, SD = 37.73%, t(132)=4.96, $d = .43$, $p < .001$) conditions. Finally, there was no significant difference between Observe and Together conditions ($p = .54$).

Explicit Agency Ratings

A one-way repeated measures ANOVA was conducted to assess differences in agency ratings in the *Partner* condition. Results showed a significant main effect of *Partner* (F = 144.55, MSE = 84570.55, $p < .001$, $\eta_p^2 = .525$). Planned comparisons t-tests were also conducted with no multiple comparison adjustments and showed a significant difference between the Baseline (M = 22.19, SD = 28.77) and Observe conditions (M = 68.41, SD = 27.45, t(131) = 12.50, *d* = 1.09, *p* < .001), a significant difference between Baseline and Together (M =63.19, SD = 25.78, t(131) = 12.35, *d* = 1.08, *p* < .001) and Observe and Together (t(131) = 3.77, *d* = .33, *p* < .001) conditions.

Individual Differences in Anthropomorphism

Like in the previous experiments IDAQ scores were computed and a Pearson's correlations was used to assess whether a relationship existed between proportional reproduction errors and IDAQ scores. Temporal estimation errors in the Together condition were subtracted from estimation errors in the Baseline condition. A Pearson's correlation was then used to assess the relationship between this computation and IDAQ scores. No relationship was found (*p* = .56). Agency ratings in the Together condition were also inserted into a Pearson's correlation with the IDAQ scores and no significant relationship was found (*p* = .50).

Other Behavioural Analysis

A paired sample t-test was used to see whether there was a significant difference between the number of times acted in the Observe (M = 87.51%, SD = 13.22%) and Together (M = 73.93%, SD = 15.95%) conditions. Results showed a significant difference between Observe and Together conditions (t(131) = 12.40, *d* = .86, *p* < .001). We then used a Pearson's correlation to see whether there was a relationship between the number of times acted in the Observe and Together conditions and their respective proportional reproduction errors and found no significant correlations (*p* = .71, *p* = .45).

Interim Discussion

In this experiment, the Alone condition was changed for the Observe condition. The difference between the Observe and Alone condition is that in the Observe condition, participants were informed that their partner would be observing them during the trial. Their avatar was presented in the top-right of the screen which disappeared after they stopped the circle. The main findings from this experiment showed that while participants rated their explicit agency lower in the Together condition compared to the Observe condition, there was no difference in temporal binding between the two conditions. This is surprising considering the findings from the previous two experiments, as the participants still acted alone during the task. However, there are a few explanations for why there was no difference here.

The first is that the presence of another agent alone might be sufficient to increase temporal binding. This would be akin to the mere presence effect, whereby the presence of other agents alone produces social facilitation (Zajonc, 1965; Schmitt et al., 1986). On this basis, the presence of another person might be sufficient to increase temporal binding in the Observe condition and thus produce similar levels to that found in the Together condition. On the other hand, the partner's face was shown while the circle was increasing in size only in the Observe condition and not the Together condition. Eye gaze carries important social information that can affect social cognition (see Itier & Batty, 2009 for review) and has been shown to increase temporal binding (Ulloa et al., 2019). Therefore, the avatar presented during the Observe trials might increase temporal binding because of eye gaze, rather than the intended observational effect. Because the avatar was not shown during the trial in the Together condition, it is not easy to untangle these two potential effects. The difficulty of separating these two variables is further compounded by there not being an Alone condition, as we do not know whether observation effects the sense of agency compared to working alone. The next experiment in this chapter will look to reconcile these issues by including the avatar during the trial in the Together condition and including the Alone condition.

Experiment 8

In Experiment 7, we found that during that while explicit agency ratings were reduced in the Alone condition compared to the Observe condition, there was no difference between the Alone and Observe conditions for temporal binding. As we just discussed, as the avatar was presented in the top right-hand corner during the trial in the Observe condition but not the Together condition, we cannot account for the effects of this in the experiment. As eye gaze has been shown to affect temporal binding (Ulloa et al., 2019) and having the avatar on screen could potentially present additional cognitive and attentional loads during the task, it is necessary to isolate this effect from the main experimental findings. Therefore, in Experiment 8 a couple of changes were made. The first is that in both the Observe and Together conditions, the partner's avatar was presented in the top right-hand corner of the screen. In addition to this, we reincluded the Alone condition. In the Alone condition, the avatar was not present. Therefore, we can now individually isolate the effects of the presence of the partner (through the Observe condition) and the partner being present and being to act (in Together condition) against working alone (in the Alone condition). If the effect is due to the partner being able to act, we should expect agency ratings to be higher in the Alone condition compared to the Together and Observe conditions. Likewise, we should expect temporal binding to increase in the Together and Observe conditions compared with the Alone condition. Alternatively, if the effect is driven by the presence of the partner, rather than the partner being able to act, the difference should be present between the Together condition against the Alone and Observe conditions, but there should be no difference between the Alone and Observe conditions.

Method

Participants

Before sampling, an additional power analysis was conducted via G*Power (Faul et al., 2007) to assess necessary sample size for a difference between two dependent means. We used the Cohen's *d* calculation from Experiment 7 for the effect size between the Alone and Together condition (d = .29). The results of this analysis suggested a sample size of 96 participants would be required to detect effects of two-tailed t-test with the current power (t = 1.99, $1 - \beta = 0.8$, $\alpha = 0.05$). 110 participants were recruited from Prolific (www.prolific.co) and paid £2.70 for their participation (at a rate of £8 an hour).. Ethical approval was granted by the ethics committee at the University of East Anglia.

Materials

Psychopy3 (Peirce, Gray, Simpson et al. 2019) was used to create the experiment and the experiment was run online on Pavlovia.org.

The design of this study was based on Experiment 5*.* The difference in this study is that the Alone condition was reincluded and thus there are four conditions in this study. Therefore, we have a single factor repeated measures design with the independent variable of *Partner* (Baseline, Alone, Observe, Together) Participants completed 94 trials in total, 3 practice trials for each condition and 18 trials for each condition in the main experiment. The dependent variable for each condition was proportional reproduction errors and explicit agency ratings.

Procedure

The procedure for this experiment was the same as Experiment 7, with the addition of the Alone condition that was present in Experiment 5 and Experiment 6.. When the experiment was complete, participants were then brought to a survey which contained the IDAQ questionnaire and the debriefing sheet.

Results

Exclusion Criteria

The same exclusion criteria that were used in Experiment 5 was applied here. Trials where temporal estimations were less than 100ms or 3 SDs above an individual subject's mean were excluded from the analysis (3% of total trials). Trials where the participant did not act were also removed (18% of total trials). Finally, 1 participant was excluded from the analysis as they contributed less than 8 trials in any condition after this trial exclusion occurred and 7 participants were removed as they did not use a 60*hz* monitor. Therefore, 102 participants remained for analysis.

Temporal Reproductions and Stimulus-Onset-Asynchrony

To see if participants interval estimations increased as the stimulus-onset-asynchrony after their action increased, a one-way repeated measures ANOVA was conducted, using their temporal interval estimation as the dependent variable and the factor of *Time* as the independent variable. Results showed that interval estimations increased as the temporal gap between their action and its effect increased ($F(2,202) = 52.02$, MSE = 51, $p < .001$, $\eta_p^2 = .40$). Thus, participants interval estimations increased as the stimulus-onset-asynchrony increased from 750ms (M = .81, SD = .35), 1000ms (M = .88, SD = .36) and 1250ms (M = .95, SD = .37).

Proportional Temporal Reproductions

Proportional reproduction errors were calculated and were inserted into a one-way repeated measures ANOVA with the factor of *Partner* (Baseline, Alone, Observe, Together). The results from this test showed a significant main effect of *Partner* ($F(3,303) = 15.66$, MSE = .36, $p < .001$, $\eta_p^2 = .134$). Planned follow up t-tests were conducted with no multiple comparison adjustments and showed a significant difference between the Baseline (M = 97.57%, SD = 40.05%) condition and the Alone (M = 89.28%, SD = 39.84%), Observe (M = 87.25%, SD = 38.18%) and Together (M = 83.66%, SD = 36.47%) conditions: $t(101) = 3.75$, $d = .37$, $p = <.001$, $t(101) = 5.17$, $d = .51$, $p = < .001$, $t(101) = 6.11$, $d = .61$, $p < .001$. Results also showed no significant difference between the Alone and Observe condition ($p = .078$) and a significant difference between the Alone and the Together conditions ($t(101) = 2.39$, $d = .24$, $p = .018$). Finally, no significant difference was found between the Observe and Together conditions ($p = .245$).

Explicit Agency Ratings

Median agency ratings were computed, and a one-way repeated measures ANOVA was conducted, showing a significant main effect of *Partner* (F(3,101) = 144.89, MSE = 56697.49, p < .001, η_p^2 = .589) To explore this further, we conducted planned comparisons T-tests with no multiple comparison adjustments and found that participants rated their control less in the Baseline (M = 21.35, SD = 28.13) and the Alone (M = 71.56, SD = 26.12), Observe (M = 69.74, SD = 25.68) and Together (M = 61.98, SD = 23.61) conditions: t(101) = 13.97, d = 1.38, p < .001, t(101) = 13.28, d = 1.37, p < .001, t(101) = 11.88, d = .1.18, p < .001. Next, there was no significant difference in agency ratings between the Alone and Observe conditions (p = .19) but a significant difference between Alone and Together conditions (t(101) = 5.49, d = .54, p < .001). Finally, participants rated their agency ratings lower in the Together condition compared to the Observe condition (t(101) = 4.34, d = .43, p < .001).

Individual Differences in Anthropomorphism

Like with Experiment 5, an IDAQ scores were computed for each experiment following guidance from Waytz, Cacioppo & Epley (2010). To test whether there was a relationship between proportional reproduction errors and IDAQ scores, temporal estimation errors in the Together condition were subtracted from estimation errors in the Baseline condition. A Pearson's correlation was then used to assess the relationship between this computation and IDAQ scores. No relationship was found (p = .76). Agency ratings in the Together condition were also inserted into a Pearson's correlation with the IDAQ scores and no significant relationship was found (p = .57).

Other Behavioural Analysis

To test the number of times that participants acted between conditions, we first compared whether there was a significant difference between the Observe (M = 83.15%, SD = 11.17%) and Alone (M = 82.52%, SD = 12.49%) conditions using a paired sample t-test. Results showed no significant difference (p = .67). Next, we checked whether the typical difference between the Alone and Together (M = 72.15%, SD = 15.32%) condition was found. Results showed that participants acted more in the Alone condition compared to the Together condition (t(101) = 11.21, d = 1.11, p < .001). Next, we checked whether there was a correlation between the number of times acted and temporal reproduction error subtractions in the Alone condition and the same for the Together condition and found a small correlation between percent of time acted in the Alone condition and the judgment error for their temporal reproduction (r = .21, p = .035) and no significant relationship for the Together condition (p = .14). Next, looked at whether the number of times that the partner acted in the Together condition related to temporal reproduction error and found no relationship (p = .52).

Interim Discussion

The primary goal of this experiment was to understand whether the effect of a partner on the sense of agency is contingent on them being able to act. To do this, we followed a similar procedure from the previous experiments in four conditions. In the Baseline condition, participants were asked to observe the circle stopping on its own. In the Alone condition, participants stopped the circle on their own. In the Observe condition, participants were told that they would stop the circle on its own but with their partner observing them. This idea was reinforced by having their avatar in the top right during the trials. Finally, in the Together condition the participant and their partner could both stop the circle and the partner's avatar was shown in the top right on each trial. The main results show that temporal reproductions in the Baseline condition were longer than each other condition, replicating the typical temporal binding effect. Furthermore, whilst the Alone and Observe conditions did not show any difference in temporal binding or explicit agency ratings, temporal binding was greater in the Alone condition but not the Observe condition. Explicit agency ratings for both the Observe and the Alone conditions were greater than in the Together condition. The results here suggest that the effect of working with a partner is not driven solely by the mere presence effect (Zajonc, 1965) but is instead contingent on the partner being able to act during the task. We will discuss this further in the next section of this chapter.

Chapter Discussion

This chapter investigated how the sense of agency can be modulated by working with a partner on a task. Specifically, we looked at how working with a partner can modulate explicit ratings of control and the subjective temporal compression of actions and their resulting consequence (known as temporal binding). Across four experiments, we found that working with a partner reduced explicit ratings of control, consistent with previous findings (Beyer et al., 2017; Beyer et al., 2018; Ciardo et al., 2020). In contrast, we also found in three of the experiments temporal binding increased when working with a partner. Experiment 8 suggests that the mere presence of the partner alone is not sufficient to cause either the reduction in explicit agency or the increase in temporal binding. This effect therefore seems contingent on the partner being able to act.

The reduction of explicit agency when working with a human or robotic partners has been found multiple times in other experiments, including those in the previous chapter. Although little is known regarding the cause of this, Beyer et al. (Beyer et al., 2017; Beyer et al., 2018) suggest that it is from the partner interfering with mentalizing processes that effect action selection fluency. They point to neural evidence in their study that show precuneus activity was both increased when working with a partner and negatively correlated with explicit agency ratings. Furthermore, they found a negative relationship between the left angular gyrus and agency ratings, consistent with research from Chambon et al. (2013) and increased activation in the temporal parietal junction (TPJ) and right middle frontal gyrus (MFG) when working with a partner compared to working alone. Taking this all together, the authors suggest that the reduction in explicit agency ratings are from context related processing in the TPJ, MFG and precuneus interfering with action-related processing, which they argue reduces the sense of agency. Whilst the results in this chapter do not say anything specific about neural processes, we replicate the behavioural findings and show that this effect on explicit agency is contingent on the partner being able to act. Experiment 8 demonstrated that while agency ratings were lower in the Together condition compared to the Alone condition, there was no difference between the Alone condition and the Observe condition. Furthermore, agency ratings were lower in the Together condition compared to the Observe condition. As the avatar was present during the action stage in both the Together and the Observe conditions, this would suggest that the presence of a partner is not sufficient to produce this effect. Considering the model from Beyer and colleagues, we would suggest that context related processes are only salient to the sense of agency if the other agent is able and willing to act on the environment. That is, the partner needs to be task-relevant to affect and individual’s sense of agency in this manner. We therefore postulate that merely being in the presence of someone while performing an action is not sufficient to affect the action-related processes which in turn modulate the sense of agency.

In contrast, temporal binding increased when working with the partner. Temporal binding is meant to be a proxy for the sense of agency and increases in it are also meant to reflect an increased feeling of authorship over actions (Haggard, 2017). Indeed, there does seem to be emerging evidence that temporal binding increases when presented with social stimuli. Vogel et al. (2021) found that temporal binding is greater when an action has social consequence, rather than merely a physical consequence. It has also been shown to increase with eye contact (Ulloa et al., 2019) and working with a human partner on a joint-simon task (Sahai et al., 2019) and decrease when recalling episodes of social exclusion (Malik & Obhi, 2019). These results and the current findings suggest that temporal binding is sensitive to social context. More specifically, its magnitude increases when presented with social stimuli. Analogous to such findings, the results here demonstrate that temporal binding increases when working with a partner, suggesting that joint efforts on a task induce temporal hyper-binding. One explanation for why temporal binding might increase is to regulate and maintain dissociations between action representations from the self and the partner. Or in other words, the sense of agency increases when working with a partner to maintain the distinction between the self and the other. The idea that the sense of agency has a purpose of identifying self-generated action representations from externally generated action is not a novel idea. The notion of the “Who” system is one of the foundational elements of our understanding of the sense of agency (Georgieff & Jeannerod, 1998; Vignemont & Fourneret, 2004). As social animals, it would be beneficial to be able to distinguish self-generated action from others in the group to bolster social facilitation and responsibility (Frith, 2014). Therefore, the hyper-binding in this chapter could reflect an increased sense of agency to account for potential action representations of the partner during the task. This is in line with the findings from Experiment 8, which show that temporal binding is increased when the partner can act, compared to when it is simply observing. The participant would not need to account for the actions of the partner if the actions were not present, thus temporal hyper-binding does not occur. Another possible reason for temporal binding increasing was that participants took the lead in performing the actions and stopping the circle. Participants acted to stop the circle themselves in most trials and the partner was programmed to only act in a third of total trials in the Together condition. It is possible that taking the role of the leader and taking control of events affects the sense of agency. There has been emerging research showing that temporal hyper-binding because of leadership (Pfister et al., 2014). Whilst the number of times that the participant or the partner acted did not having any bearing on temporal binding, participants who contributed less than eight trials in any condition were removed for inter-trial reliability purposes, this might have left only participants who were willing to take control and ownership of the trials themselves. Indeed, participants typically did in fact act in the majority of trials in the Together condition in the experiments presented here. Hyper-binding in such cases could be from participants taking the lead role as the actor, so the participants saw that the onus was on them to act could be a reason for the hyper-binding we see here.

The experiments presented here demonstrate a peculiar juxtaposition of implicit and explicit agency. Whilst explicit agency decreased, temporal binding (or implicit agency) increased. The former reflects a decrease in the sense of agency while the latter suggests an increase. There is conflicting evidence on whether temporal binding and explicit agency correlate (Dewey & Knoblich, 2014; Imaizumi et al., 2019). However, the two measures of agency do dissociate. leading researchers to suggest that implicit and explicit measures of the sense of agency measure different cognitive processes (Moore et al., 2012). Therefore, it is not necessarily unusual that we see a dissociation. Here, we will attempt to offer an account for why such results occurred in this instance. One explanation is that the two the different metrics that we used to measure the sense of agency are measuring two different aspects of action-outcome monitoring. In this experiment and in similar studies (Beyer et al., 2017; Beyer et al., 2018; Ciardo et al., 2020), participants have been asked, “How much control did you feel over the outcome”. This is qualitatively different from asking participants, “How much control did you feel over your action”. While the former is relating specifically to the participants action, the latter is asking about the outcome of their action. As judgments of agency are open to rationalization and contextual information (Synofzik et al., 2014), one potential effect from this is that participants ratings are based the shared responsibility to act to stop the circle in the trial. In other words, participants rate their agency as lower because the question being asked is relating to the outcome, rather than the action. Because the partner can act on each trial, the participant has to account for their partner acting, which in turn will take away their ability for themselves to intervene in the trial and be causal in the outcome of the event. On the other hand, temporal binding might not be sensitive to the contextual information of how the question is asked for explicit agency. Instead, it might simply reflect the need to account for the partner as mentioned previously. Future research could look at adapting the explicit agency question to relate to the action rather than the outcome. Alternatively, both the reduction in explicit agency and increase in temporal binding could be indicative of a shared sense of agency. Tasks relating to joint agency typically find that participants report less control over their actions and more towards a shared sense of control. The findings here could be analogous to these findings, whereby a reduction in explicit agency is illustrating a shared sense of control over the outcome. Finally, in the

Alone condition, as participants are told that they are the only one’s who are acting, one might suggest that this artificially raises their expectation of their own agency in comparison to other conditions. While this may not impact temporal binding specifically, this may raise agency ratings in the Alone condition and thus, we see increased agency ratings in the Alone condition, rather than a reduction of agency in the Together condition.

When discussing the results of this chapter, we have used the term partner to describe the behaviour of the preprogrammed behaviour of the software. Participants in this series of experiments were aware that their partner was not another person. The first experiment assessed whether temporal binding and explicit agency were affected by the level of intentionality that the partner had attributed to them in the experiment. The results showed that there was no main effect of intentionality, the decrease in explicit agency and increase in temporal binding were independent of whether the participant was told they were working with BOBBY or a preprogrammed computer. The results from this chapter suggest that working with a computerised partner affects the sense of agency in ways that we might expect from working with other human agents. Whether we treat computers as social agents is up for debate and research into the social effects of computers on the sense of agency is equally as conflicted. While some research suggests that some social effects on the sense of agency are exclusive to human interactions, there is also a wealth of research that shows that the sense of agency can be modulated with social stimuli using no explicit claims of working with another human agent (Barlas, 2019; Roselli et al., 2022). Indeed, the reduction of explicit agency with robotic partners have been demonstrated before in similar tasks (Ciardo et al., 2020). We might then take that as evidence that non-human partners can produce social effects in a similar way to their human counterparts. Perhaps the behaviour of the partner is more important when manipulating the sense of agency rather than the belief that the partner is a real person. The behaviour of the partner being able to stop the circle from increasing in size alone might produce the effects on the sense of agency. However, with the aforementioned in mind, it would still be beneficial to understand whether the temporal hyper-binding shown here is also present if the partner believed that they were working with a human partner as this study lacks that data to produce that conclusion.

Conclusion

The goal of this chapter was to understand whether temporal binding is modulated when working with a partner. Specifically, when working with an artificial agent where cooperation was minimal. The results showed that temporal binding increased when working with a partner compared to working alone, irrespective of whether the partner was described as a computer or a robotic agent. Furthermore, we found that this increase in agency was contingent on the partner being able to act rather than just mere presence, suggesting that the increase in temporal binding is a result of the participant managing action representations of the partner. The results here have implications within social agency and human-machine interaction.

Chapter 7 – Outcome Relevance, Moral Contexts and Explicit Agency

Experiment 9 – Outcome Relevance and Explicit Agency

Introduction

When working in cooperative settings, the outcome of the joint effort towards a task can change in who it benefits. Sometimes, a group effort towards a task will only benefit an individual within the group. Imagine building a piece of furniture with a partner and in one scenario it is your own furniture, compared to building it for someone else. In the first circumstance, building the furniture directly benefits you, as you will have a new piece of furniture. In the second scenario, the outcome only benefits your partner. How might our experience of agency change over actions which we are not the benefactor? Little is known on this, however research does suggest the sense of agency is sensitive to a number of external cues, whether that be before the action occurs or post-action. Furthermore, evidence has shown that prosocial effort is linked to activation of neural networks in the brain that are absent for self-benefitting behaviour (Lockwood et al., 2022), suggesting distinct neural processing between actions which benefit oneself versus those which benefit others. We as humans do engage in activities that do not directly benefit us as part of society. We are unique among mammals in that we tend to engage in altruism on a broad scale, not just isolated within small or kin-based groups (Fehr & Fischbacher, 2003). Tankersley, Stowe and Huettel (2007) conducted an experiment and showed that regions in the brain that are related to agency recognition are also related to altruism. As noted by Montague and Chiu (2007), altruism and ideas of fairness are something shared between agents, so it is logical that detection of agency is necessary for altruistic behaviour. However, we do not know whether agency over altruistic actions are experienced differently to those which are self-benefitting.

One reason why our experience of agency might differ between actions with egoistic actions compared with non-egoistic outcomes is because of self-referential processing. Self-referential processing (or self-related processing) refers to the differential and putatively more efficient processing of stimuli which we would regard as relating to oneself (Craik, Moroz & Moscovitch, 1999; Rameson, Satpute & Liberman, 2010; Northoff et al., 2006). For example, Roger, Kuiper and Kirker (1977) found that recall of adjectives is better when the participant is thinking whether they describe themselves, rather than whether they consider the meaning or their appearance. The theory of self-referential processing posits that there are distinct cognitive and neural activities which are engaged with stimuli regarding the self. This spans across many domains of cognition, including memory, attention, agency, and ownership, typically engaging cortical midline structures in the brain (Northoff et al., 2006).

Emerging evidence has shown that many aspects of cognition seem to prioritise self-referential information (Sui & Humphreys, 2012). Sui, He & Humphreys (2012) found that matching geometrical shapes with oneself, compared to others improved performance in a shape matching task. In their task, they asked participants to code geometric shapes to concepts such as “self”, “friend” or “stranger. Participants were then asked to confirm whether pairings of the shapes to the concepts were correct as quickly as possible. They found that shapes relating to oneself were processed faster than those not relating to the self. This has also been extended to the sense of agency. In their study, Makwana and Srinivasan (2019) first employed a shape-label matching task, where participants were asked to match shapes to concepts such as “self”, “friend” or “stranger”. Participants then performed a temporal binding task which showed that interval estimates were shortest for the shape which had been associated with oneself. Meaning, people felt a stronger sense of agency for actions which produced objects relating to the self, rather than those relating to the concept of a friend or a stranger. Thus, during joint action, we might experience a stronger sense of agency for outcomes that benefit us because they are self-referential to us or alternatively feel less agency over actions unrelated to us.

To test this, we asked participants to take part in an experiment where they obtained points while performing a task. The study adopted a similar paradigm to Experiment 2, which asked participants to perform a task where a ball rolled down a slope and they were required to stop it before it fell off the edge, which they did alone in one condition and worked with a partner in another. Here though, participants were told that through the experiment, they were working towards their partners' points. The goal here was to understand whether people felt as much agency towards points which were not their own. If agency was indeed affected by responsibility, we should have expected explicit reports to be reduced when working for their partners' points compared to working alone. As we had found that explicit agency was reduced when working with a partner, we would have expected agency to be reported lower when working together compared with working alone.

Method

Participants

35 participants (28 female, 28 right-handed) from the University of East Anglia took part in this study to fulfil a course requirement. Consent was obtained prior to the participant taking part in the study and ethical approval was granted by the University of East Anglia's ethics committee. Participants were recruited through SONA. Data was collected between 4th January 2021 and April 13th 2021. Sample size was based on the number of participants in Chapter 4. No exclusion criteria were applied, therefore all participants who completed the experiment were included in analysis.

Materials

The experiment was created on Psychopy3 and was administered online. As with *Experiment 2*, for the core part of the experiment, an animation was created from a series of images, which form a ball rolling down a slope, created in created within Blender (www.blender.org). Agency ratings were obtained using a slider created and recorded within PsychoPy3.

Procedure

The experiment followed a 2x2 repeated measures design, with the within-subjects factor of *Partner* (Alone, Together) and the within-subjects factor of *Pot* (Your Pot, Partner Pot). Trials were presented in a single randomised block, with 120 trials in total. This resulted in 40 trials for each *Partner*, *Pot* pairing.

At the start of the experiment, participants were presented with an instructional video, outlining what would be required during the task. As in *Experiment 2*, participants were told that they would be working with another participant. This partner was in fact pre-programmed behaviour coded into the experiment. To bolster this illusion, participants were presented with a screen "Waiting for participant…" at the beginning of the experiment and randomly throughout the trials. Like in *Experiment 2*, participants were instructed that they were to stop a ball from rolling off the end of a slope. The later they stop the ball, the less points they will lose. However, if the ball falls off the slope they will lose maximum points. The procedure for the experiment was the same as in *Experiment 2* with the exception that in addition to being shown whether they will be working alone or their partner at the start of each trial, participants were also told whose points will be lost. This could either be their own points, or their partners points. If it was their points they would be working from, the screen which shows the avatars would also have the text "Your Pot". If it were the partner's points, the same text would state “Partner Pot”. The rest of the experiment was consistent with *Experiment* 1, with the addition of being asked the question “Did you believe you were working with a real participant*?”* after the main experimental trials had finished. Participants could answer yes or no to this question. We asked this as participants were not asked this in Experiment 1 and this gave us an opportunity to assess whether participants believed they were working with another human participant and whether this would affect results.

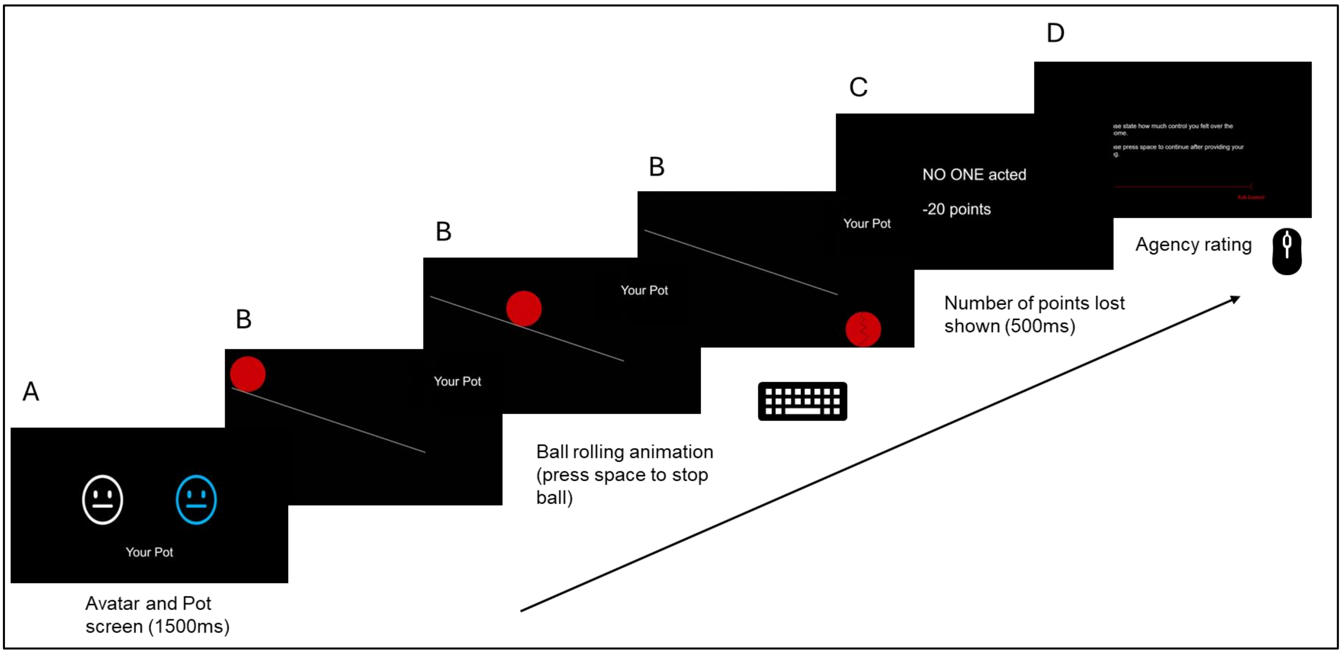

Figure 15 – Trial timeline for Experiment 9. Participants are first shown the avatar screen, which also shows whether they will be working for their own or their partner’s points. Participants are then required to stop the ball from rolling down the slope. After this, they will be told who acted and how many points was lost before being asked to rate how much control they felt over the outcome.

Results

Agency Ratings

Each participant returned 120 trials across the whole experiment. Trials where the participant did not act, or their partner acted were treated as invalid trials and were removed from the experiment. One participant was removed for not providing any valid trials in one of the conditions of the experiment. For the remaining 34 participants, a median agency rating was calculated for each *Partner x Pot* condition. The medians were then used to conduct a 2x2 repeated measures ANOVA with the factor of *Partner* (Alone, Together) and *Pot* (Your Pot, Partner Pot). Results showed a main effect of *Partner* ($F(1,33)=15.82$, $MSE=2507.77$, $p < .001$, $\eta_p^2 = .324$) and a main effect of *Pot* ($F(1,33)=17.11$, $MSE=349.44$, $p <.001$, $\eta_p^2 = .341$). There was no significant interaction between *Partner* and *Pot* ($p > .43$). Participants’ agency ratings were significantly lower in the Together condition compared to the Alone condition and were significantly lower when working with their partners points compared to when working with their own points.

Other Behavioural Analysis

To assess whether the participants actively tried to gain points, we looked at which frame on average they typically stopped the ball. The ball rolling animation consisted of 10 frames, including when the ball falls off the edge of the slope. Therefore, the participant had the opportunity to stop the ball within 9 frames. We took the median of all trials for each participant and took the mean result, finding that people stopped the ball on average at the 7^{th} frame (mean = 7.32, SD = .53). The results show that participants stopped the ball towards the end of the slope, suggesting that they were trying to reduce the number of points they lost on each trial. The participants therefore tried to perform the task correctly and achieve as many points as they could. To assess whether this differed between *Pot* conditions, we ran two paired sample T-tests, one for the *Alone* condition and one for the *Together* condition. Results showed the time that participants acted for their own points compared to their partner's points did not differ in either the *Alone* ($p > .75$) or the *Together* ($p > .58$) conditions. Thus, we found no difference in the point when participants stopped the ball on the slope between *Pot* conditions. To test whether participants were more likely to act when playing for their own points compared to their partners points, we conducted two paired sample T-tests to test the number of times participants acted when playing for their own points compared their partners, in both the *Alone and Together* conditions. We found that the number of times participants acted did not differ between *Pot* conditions in either the *Alone* condition ($p > .05$) or the *Together* condition ($p > .10$). In summary, these results show that participants behaviour when stopping the ball did not differ as a function of whether they were working for their own points or not.

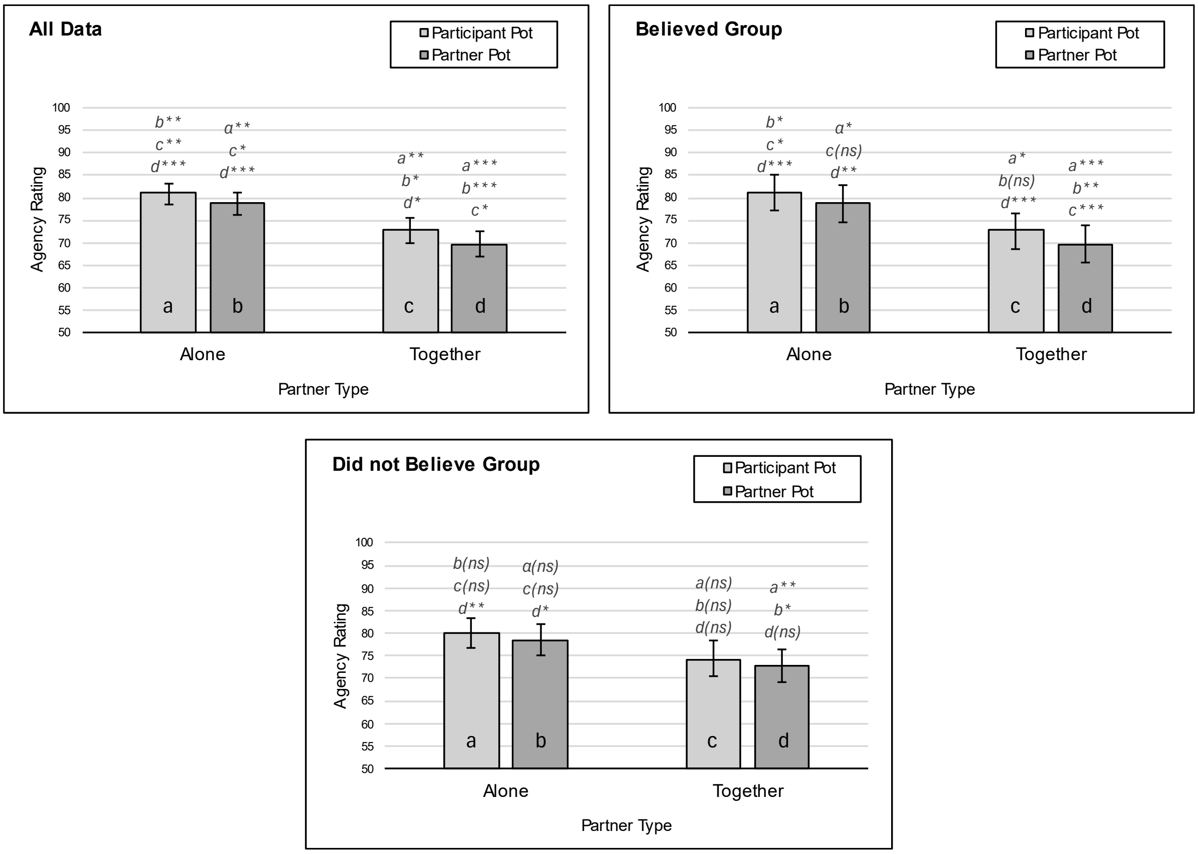

*Figure 16 - Mean agency ratings for Experiment 9. Error bars show standard error of the mean. Each bar is labelled a,b,c,d. Statistical significance between conditions is denoted by the corresponding letter for the condition. Asterisks indicate significance level (*p < .05, **p < .01, ***p < .001, ns = p >* .05).

To test whether participants' belief that they were working with a real participant affected agency ratings, participants were split into groups based on whether they answered, 'Yes' (N = 18) or 'No' (N= 16) to the question "Did you believe you were working with a real participant". A 2x2x2 repeated measures ANOVA was then conducted with the within-subjects factors of *Partner* (Alone, Together) and *Pot* (Your Pot, Partner Pot) and the between-subjects factor of *Belief* (Believed, Did Not Believe). Results showed a significant main effect of *Partner* (F(1,32) = 15.43, MSE = 2411.14, p < .001, η_p^2 = .325), a main effect of *Pot* (F(1,32) = 17.83, MSE = 328.31, p < .001, η_p^2 = .358) and a significant interaction between *Pot* and *Belief* (F(1,32) = 4.60, MSE = 84.76, p = .040, η_p^2 = .126). No other interactions were significant. To explore the significant interaction, separate repeated measures ANOVAs were administered for each belief group using the same within-subjects factors outlined above. For the *Believed* group, the results showed a main effect of *Partner* (F(1,16) = 9.56, MSE = 2202.49, p = .007, η_p^2 = .374) and a main effect of *Pot* (F(1,16) = 21.71, MSE = 371.78, p < .007, η_p^2 = .576) and no significant interaction. In the *Did Not Believe* group, results showed that there was a significant main effect of *Partner* (F(1,32) = 6.42, MSE = 546.39, p = .023, η_p^2 = .300) but no main effect of *Pot* (p > .20) nor a significant interaction ($p > .90$). Thus, the interaction of the three-way ANOVA can be explained by a significant difference in agency ratings when people were working with their own points compared to when working with their partners, but only when participants believed that they were working with a real participant. T-test results are shown in Figure 17.

		M	*SD*	*SEM*	*t*	*df*	*p*	*d*
Did not Believe group	Alone Participant Pot - Alone Partner Pot	1.69	5.34	1.33	1.26	15.00	.225	.32
	Alone Participant Pot - Together Participant Pot	6.00	13.60	3.40	1.76	15.00	.098	.44
	Alone Participant Pot - Together Partner Pot	7.38	9.83	2.46	3.00	15.00	.009	.75
	Alone Partner Pot - Together Participant Pot	4.31	10.76	2.69	1.60	15.00	.130	.40
	Alone Partner Pot - Together Partner Pot	5.69	8.65	2.16	2.63	15.00	.019	.66
	Together Participant Pot - Together Partner Pot	1.38	10.14	2.54	0.54	15.00	.596	.13
Believed Group	Alone Participant Pot - Alone Partner Pot	2.94	4.68	1.14	2.59	16.00	.020	.63
	Alone Participant Pot - Together Participant Pot	9.65	15.33	3.72	2.60	16.00	.020	.63
	Alone Participant Pot - Together Partner Pot	16.06	15.90	3.86	4.16	16.00	<.001	1.01
	Alone Partner Pot - Together Participant Pot	6.71	15.56	3.77	1.78	16.00	.095	.43
	Alone Partner Pot - Together Partner Pot	13.12	16.71	4.05	3.24	16.00	.005	.79
	Together Participant Pot - Together Partner Pot	6.41	8.10	1.96	3.26	16.00	<.001	.79

Figure 17- t-test table showing results from Experiment 9. Results are split between the group of participants that believed they were working with a human partner and those who didn't.

Discussion

Like in Chapter 2, this experiment asked participants to engage in a task where they earn points by stopping a ball rolling down a slope. Participants earned points by stopping the ball as late as they could. The key manipulation here was that on some trials, participants earn points for themselves, while in others, they earn points for their partner. Participants were then asked to rate how much control they felt they had over the outcome. The results of this experiment showed the typical effect shown throughout this thesis that participants report a higher sense of agency when working with alone compared with a partner. Furthermore, participants also reported higher agency when working with their own points compared with their partner's points. Here, we extend the results of the previous chapter to show that explicit self-reports of agency are affected by whether the participant is working towards their own points of their partner's. We will discuss the results in the context of responsibility before addressing other possible causes in addition to limitations of the study.

As we have already discussed at length the effect of working with a partner compared to working alone, for brevity this discussion will focus on the difference in agency ratings when the participant was working for their own points compared to the partners. The reduction in agency seems to suggest that outcome relevance plays a role in how much control the participant felt over the action. That is, participants report feeling less agency when they are working for their partner's points compared to their own. One way to explain this data is that when participants are playing for their own points, they see the task as more relevant to themselves. This would be in line with research that suggests we prioritise self-referential information (Sui & Rotshstein, 2019; Keil et al., 2023). This effect has been shown across various cognitive domains such as memory (Symons & Johnson, 1997) and motor responses (see Northoff et al., 2007 for review). Evidence has also shown that actions which produce outcomes related to oneself does increase agency compared with actions that produce outcomes that relate to others (Makwana and Srinivasan, 2019). When the participants are playing for their own points, they might adopt the view that it is more relevant to them, which in turn bolsters agency. Or, on the other hand, when it is their partners points, they see it as less relevant to themselves. Thus, we suggest that a similar mechanism is working here. The assigned points may reinforce the feeling of personal relevance and are thus given higher cognitive priority which in turn provides a greater sense of agency. We found no evidence to suggest that the participants performed the task any differently between when the points were their own or were their partners. Participants stopped the ball around the same point for their own points and their partners points and the number of times they acted did not differ either. Thus, we suggest that the reduction in agency isn't related to task performance, or a lack of motivation or willingness to work for their partners points, but perhaps the specific effect of self-relevance on agency itself.

As participants were told they would be working with a human partner, we also evaluated whether participants' belief of this affected results. Consistent with the other research in this thesis, the effect of working alone and with a partner was stable across whether the participant believed they were working with a human partner or it was preprogrammed. As participants who did not believe they were working with a human must have therefore believed they are working with a programme, this is expected. We have repeatedly seen that participants report lower agency whether working with a human partner or a computer throughout this thesis and in other research (e.g. Beyer et al., 2018; Ciardo et al. 2020). However, we did find that the effect of the points assignment was only present in those who believed that they were working with a human partner. Participants only rated their agency as lower when working for their partners points when they believed their partner was human. As the effect on agency of the partner is still present in both groups, we believe it is unlikely that participants are simply disengaging with the task. Believing they are working with a real participant might enable cognitive processes or postdictive rationalisation relating to social contexts unique to human-to-human interaction. For example, participants who believed that they were working with a human partner would also believe that it is likely that the participant would be performing the task with similar goals and intentions that they do (Leslie, Friedman & German, 2004). This could increase cognitive load through mentalising, which has been suggested to reduce the sense of agency (Sidarus, Travers, Haggard & Beyer, 2020). Those who believed they were working with a computer might not experience a deleterious effect on their sense of agency as there might not be any attribution of mental state. As participants were only asked if they believed they were working with a human at the end of the experiment with a simple yes/no answer option, we do not know when participants stopped believing they were working with a human (if they did so at all). We believe that more work is needed to understand how mental states would affect how outcome relevance affects the sense of agency as the experiment was not inherently designed to test this.

The current experiment only investigates explicit self-reports of agency and as such, we do not know how outcome relevance affects implicit agency. As we were asking participants to rate their control, they were performing judgments of agency. As mentioned previously, judgments of agency are open to rationalisation and demand characteristics. However, also appear to be sensitive to variables which would affect agency monitoring. Nonetheless, as temporal binding is supposed to reflect the feeling of agency, an adaptation of this task using temporal binding might prove insightful. As some evidence suggests that temporal binding is sensitive to outcomes relating to oneself, rather than others (Makwana and Srinivasan, 2019), we might expect results to follow those of explicit agency.

Conclusion

The results of this experiment show the effect of working with a partner and outcome relevance on the sense of agency. Here, we show the typical effect of working with a partner on explicit agency, demonstrating that working with a partner reduces the sense of agency. We also show that people report their agency as lower when working for their own points compared to their partners points. This suggests that outcome relevance has an impact on the explicit sense of agency. We discuss this research in the context of self-referential processing, arguing that outcomes which do not relate to the participant reduce the sense of agency.

Experiment 10 – Moral Contexts and Explicit Agency

Introduction

Moral responsibility is an aspect of human behaviour that seems to modulate how we act on the world. For example, someone who feels responsible for the environment typically engages in ecological behaviour more often than those who do not (Kaiser & Shimoda, 1999). Furthermore, reducing beliefs in free will appear to encourage vindicative and antisocial behaviour (Vohs & Schooler, 2008; Caspar & Vuillaume, 2017) and narcissistic behaviour relates to lower social responsibility (Watson & Morris, 1991). We as humans typically behave in the knowledge that our actions have consequences and within the boundaries of what society is deemed acceptable. As humans, being able to recognise one's own and other's responsibility would be necessary to keep structured societies.

In psychology, morality is often assessed through moral reasoning tasks. These tasks employ hypothetical scenarios that lead to moral judgments from participants (see Ellemers, Van Der Toorn, Paunov & Van Leeuwen, 2019 for review). One example of this is the trolley problem, which is a thought experiment regarding utilitarian decision making suggested by Foot (1967) and later adapted by Thomson (1985). You are asked to imagine a runaway train on a track running towards a fork on the rails. The train is currently heading towards five people, but you can pull a lever to change the direction to the other rail. However, on this rail there is also a single person, which will ultimately die if you pull the lever. The thought experiment asks whether acting to pull the lever is ever morally permissible. Or in other words, is it okay to kill one person to save five?

Psychological experiments investigating decision making in the trolley problem typically show that people make utilitarian decisions in the experiment, choosing to switch the train to the side of the track to save lives (Navarette et al., 2012). However, little is known how our experience of agency is affected by making such decisions. Our agency system seems to be sensitive to responsibility. Studies suggest that priming beliefs of a deterministic world (Lynn, Muhle-Karbe, Aarts & Brass, 2014) induces lower implicit agency. A small study conducted by Moretto, Walsh and Haggard (2011) suggest people experience stronger implicit agency for moral compared to non-moral outcomes. In their experiment, they asked participants to complete an experiment which varied on whether it contained economic or moral outcomes. The moral outcomes included variations of the trolley problem, including adaptions relating to poisoning and war. However, the utilitarian judgment is the same; people decide whether to kill five people to save the life of one. In the economic condition, participants were asked to make judgments regarding outcomes which hold no moral significance. For example, whether to destroy a bushel of turnips over five. They found temporal binding was increased in moral contexts compared to economic contexts and was further increased when outcomes were more severe. Although a small study, this may illustrate that our sense of agency can be modulated through moral context.

Another factor that has not been explored in these experiments is whether we feel agency over outcomes where we didn't perform an action. If we take the trolley problem and reverse the number of people on each side of the track so that turning the switch leads to the train heading towards five people and doing nothing leads to the utilitarian outcome of one person dying, most people likely will choose to do nothing. However, that decision *to do nothing* still leads to an outcome which was from a conscious choice. The question is do we still experience agency over that outcome? The idea of non-action in agency is relatively unexplored but is at the crux of the trolley problem and responsibility, as there is clearly a distinction in both law and in the trolley problem itself between the act of killing and letting one die. Research into the sense of agency typically involves some form of motor action which then produces an outcome. However, we seem to feel agency over outcomes which extend beyond the link between motor control and outcome monitoring. An experiment looking into whether we experience agency over non-action was conducted by Weller et al. (2020). In their study, participants could control a virtual pinball machine which launched the ball either left or right. The pinball machine was set in a particular direction by default and participants could choose to change the direction of the ball's launch trajectory. They found that participants felt agency over outcomes when both deciding to act and not to act. The results suggest that deliberate intention alone is sufficient to produce agency, even without action itself. However, as the authors point out, the outcomes in the experiment were morally and motivationally neutral. As such, it is separated from accounts of moral reasoning. As there is some evidence to suggest agency differs in situations relating to moral and non-moral outcomes (Moretto, Haggard & Walsh, 2011), this experiment looks to address this.

Finally, one question that is yet to be answered regarding responsibility is how working with another in a cooperative environment might affect our sense of agency. Studies investigating moral responsibility and agency has tested individuals (Moretto, Walsh & Haggard, 2011) so little is known how this interaction works in cooperative group situations. A series of studies from Caspar et al. (2016, 2018) demonstrated that our experience of agency over outcomes which would be seen as immoral can be modulated through coercion, demonstrating the effect that social interaction can have on our experience of agency in morally relevant situations. We have also shown through this thesis how the sense of agency can be altered through working on a task which has no moral implications. What is not clear is how one’s own agency might be affected when making decisions in moral outcomes.

In this experiment, we asked participants to be tasked with guiding trains from a control room. Participants viewed a train moving along a track which could lead to a fork in the rails. Each side of the rail had people working on the track and the number of people on each side varied by trial. There were six people working in total on both sides of the rails and these were distributed differently on each trial (for example, one worker on one track, five on the other). In one set of trials, three workers were on both sides of the tracks, so participants' action or non-action was inconsequential. Participants had the option to switch the direction of the train on the track before being asked how much control they felt over the outcome. Following other experiments, we expected participants to make utilitarian decisions regarding the outcomes in the experiments, so that they would choose to act and not act to save the most lives. If people did feel agency over non-action, we would expect agency ratings to be similar in trials where they chose to act and not act to save lives. In other words, we expected agency ratings to not differ when they pressed the button to switch the train to save lives and not pressed the button to save lives. Crucially, we would also expect both action and non-action to be greater than when there were three workers on each side of the track and their action or non-action had no effect on the outcome. Finally, as participants were told they were working with another participant, we looked to investigate whether the reduction in agency when working with a partner was also present in morally salient situations. If this was the case, we would expect agency ratings to be lower when working with the partner compared to working alone. If this effect was related to the action itself, we would expect agency ratings to be lower only in conditions where participants chose to act to save lives.

Method

Participants

30 participants (29 female, 27 right-handed, mean age of 20) took part in this study from the University of East Anglia as part of a course requirement. Participants were recruited through SONA. Data was collected between the 16th March 2021 to the 24th March 2021. Sample size was based on the experiments from Chapter 6. No exclusion criteria were applied. All participants who completed the experiment were included in analysis.

Materials

The experiment was conducted online via Pavlovia and was programmed using Psychopy3. The main aspect of the experiment consisted of a series of images of a train moving along the tracks. These images were created in Blender and then sequenced to create an animation of the train moving. Agency ratings were obtained via a slider created in Psychopy3.

Procedure

The experiment used a 2x5 repeated measures design, with the within-subjects factor of *Partner* (Alone, Together) and *Workers* (1 worker, 2 workers, 3 workers, 4 workers, 5 workers). All experimental conditions were presented in a single, randomised block and contributed 10 trials per condition, 100 trials in total.

Before starting the experiment, participants were presented with an information sheet outlining what they would be doing during the task. Consent was obtained prior to the participant beginning the experiment. Once the experiment began, participants were presented with a screen stating “Waiting for other participant” for one minute to reinforce the illusion that they would be working with another participant. After this period, they were presented with a series of screens instructing what they would be doing during the task. Participants were told that they would be play the role of a train operator, whose job is to set the direction of the train along the tracks. They were told that they will work with a partner who is in another station, both the participant and their partner are able to monitor workers along the tracks and are able to change the direction of the train. The partner was in fact behaviour pre-programmed into the experiment. After the instructions were presented, participants had the opportunity to get used to the experiment in four practice trials.

As with the previous experiments, participants were presented with avatars, the left avatar representing their own avatar and the right avatar representing their partner’s avatar.

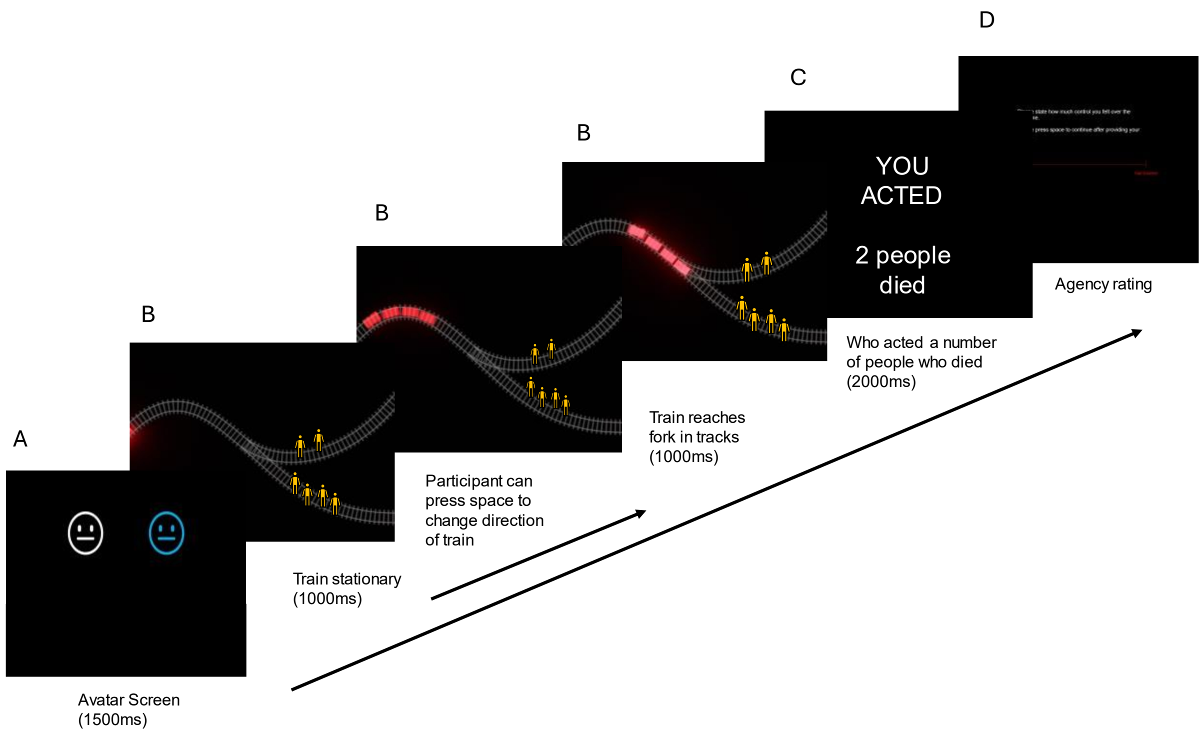

Figure 18 – Trial timeline for Experiment 10. Participants were first presented with the avatar screen. They then viewed a train move along a track, which they were told would hit workers. Participants could press space to change the direction of the train before it hit the fork on the track. Participants were then told who acted and how many people died before being asked to report how much control they felt they had over the outcome.

If they are working alone, they will only be presented with their own avatar. If they are working with their partner, they will be presented with their partner's avatar in addition to their own. These avatars were presented to the participant for 1500ms. After this, participants were presented with a screen which shows a train that forks towards the end of the track. The tracks were coloured white and the train was coloured pink. One side of the track branched off towards the upper part of the screen and one part branched off towards the bottom of the screen. On each trial, participants will see yellow stick-figures (created in PowerPoint) on each side of the track, which participants are told represent the number of workers on each side of the track. On each trial, the number of figures on each side of the track changed. In total, there were six figures that could be distributed on each side of the track. If there was one worker on the bottom track, there would be five workers on the top track. If there were two workers on the bottom track there would be four workers on the top track. Three workers on the bottom track would lead to three workers on the top track. If the participant did not act, the train would always go to the bottom fork of the track. The participant could change the direction of the train by pressing the space button. At the start of the trial, participants were presented with the train at its starting location on the track. After 1000ms, the train would begin moving. Participants had 2000ms before the train reached the fork in the track and could change the direction of it before then. After the train had passed the fork in the track, but before it was shown to hit the figures, participants were presented with a blank screen for 1000ms, after which they were presented with a screen stating who acted (“YOU”, “THEY”, “NO ONE”) and how many people died for 2000ms. Participants were then asked to rate how much control they felt they had over the outcome on a 1-100 slider as in the previous experiments. After this, they were presented with a blank screen for 1000ms before the next trial began. In the condition where participants were working with their partner, it was programmed so that the partner would only acted on 30% of trials.

Results

2 participants were excluded as they never acted on any of the trials. As participants could make a choice between acting and not acting on each trial and these both constituted to a meaningful choice, we analysed the number of times participants acted in each condition. The total number times a participant acted was calculated for each *Worker* condition and this was divided by the total number of trials in each condition to provide the proportion of times acted. A repeated measures ANOVA was then conducted with the within-subjects factor of *Partner* (Alone, Together) and *Worker* (1 worker, 2 workers, 3 workers, 4 workers, 5 workers). The *Worker* levels refer to the number of people on the bottom track. For example, if the participant did not act in the 1 worker condition, the train would hit 1 person and 5 people on the other track would survive. Results showed there was a significant main effect of *Partner* ($F(1,27) = 48.34$, MSE = .592, $p < .001$, $\eta_p^2 = .642$) and *Worker* ($F(4,27) = 342.38$, MSE = 10.73, $p < .001$, $\eta_p^2 = .927$) and a significant interaction ($F(1,27) = 10.36$, MSE = .116, $p < .001$, $\eta_p^2 = .277$).

To explore this interaction observed in the ANOVA, we employed a series of paired sample t-tests, looking at each *Partner* level at within *Worker* level. Results indicated that at *1 Worker*, there was a significant mean increase from the *Alone* condition (M = 93.14%, SD = 17.28%) to the *Together* condition (M = 75.11%, SD = 22.21%), $t(27) = 4.91$, $d = .93$, $p < .001$. Similarly, at the *2 Worker condition*, participants acted significantly more in the *Alone* condition (M = 94.43%, SD = 15.25%) compared to the *Together* condition (M = 74.79%, SD = 24.07%), $t(27) = 5.73$, $d = 1.08$, $p < .001$ .In the *3 Worker* condition, participants acted significantly more in the *Alone* condition (M = 10.75%, SD = 18.13%) compared to the *Together* condition (M = 4.61%, SD = 7.92%), $t(27) = 2.18$, $d = .41$, $p = .038$. Finally, there were no significant differences between the proportion of times participants acted between the two *Partner* conditions in either the *4 Worker* or the *5 Worker* conditions. Therefore, the interaction can be explained by participants acting more when working alone compared to working with their partner, but only in *1 Worker, 2 Worker* and *3 Worker* conditions. This interaction effect is likely by design of the experiment, as in the *Together* condition, the participants partner could also act. Because *4 Worker* and *5 Worker* led to utilitarian outcomes by not acting, participants non-action would not be affected by the response of the partner. We will discuss this further in the discussion section.

The data indicates that on average participants made utilitarian choices when deciding whether to act or not. On average, participants chose to change the direction of the train when it would save more lives and did not change the direction of the train when it would lead to more deaths. When the outcome was the same irrespective of their action, on average participants chose not to act.

Explicit Agency Ratings

Before analysing agency ratings, trials where the partner acted were removed (103 trials, 4% of total trials). As the vast majority of participants' decisions regarding the outcome in the experiment were to save the lives of the workers, we only took trials where utilitarian choices were made when there was an option to do so. That is, we only took trials where participants' action would lead to saving lives and the same for non-action. We did this to

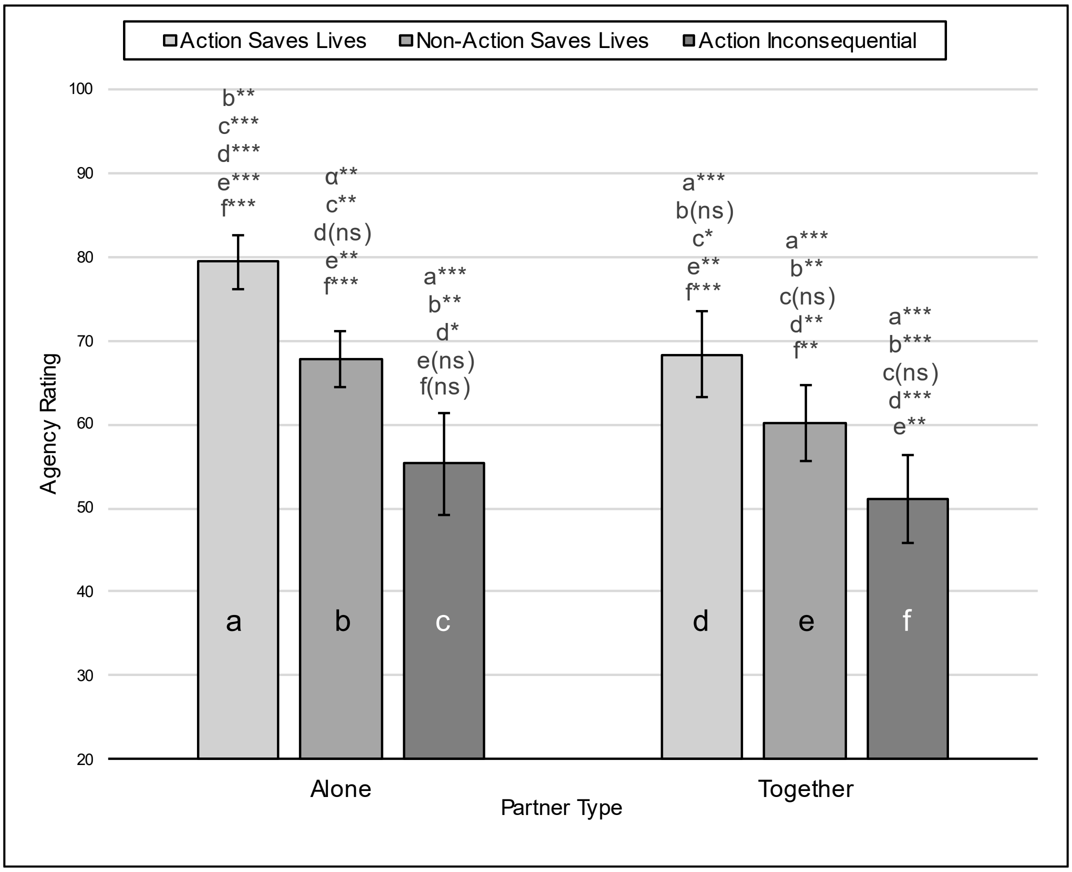

*Figure 19 - Mean agency ratings for Experiment 10. Error bars show standard error of the mean. Each bar is labelled a,b,c,d,e,f. Statistical significance between conditions is denoted by the corresponding letter for the condition. Asterisks indicate significance level (*p < .05, **p < .01, ***p < .001, ns = p > .05).*

avoid confounding the effect of non-action versus action and the number of workers on each side of the track, That is, on some trials participants might have acted in a non-utilitarian manner, inconsistent with their general performance in the experiment. Where there were three workers on each side of the track, we kept all trials as participants choice had no consequence and will serve as a baseline. As a result, 124 trials were removed before analysis (5% of total trials). Before joining trials where the number of lives saved differed, we first checked that there were no differences between such trials. We found that agency ratings did not differ between saving 1 or 2 lives both during action and non-action (*t*s < 1.3, *p*s > .23). Thus, worker conditions were collapsed together which leaves the factors of *Action* (Action saved lives, Non-action saved lives, Action inconsequential) *Partner* (Alone, Together). Median agency ratings were computed for each participant and we then conducted a 2x3 repeated measures ANOVA to assess whether agency ratings differed between conditions using the factors of *Action* and *Partner*. Results showed a significant main effect of *Partner* $F(1, 27) = 14.21$, $p < .001$, $\eta_p^{2} =$ .345 and a significant main effect of *Action* $F(2, 27) = 18.33$, $p < .001$, $\eta_p^{2} =$ .404 and a significant interaction $F(2, 27) = 6.31$, $p =$ .003, $\eta_p^{2} =$ .189. To assess this interaction, follow up tests using paired sample t-tests showed a significant difference between *Alone* (M = 79.39, SD = 16.61) and *Together* (M = 68.39, SD = 17.93) when participants acted, $t(27) = 5.08$, $d = .96$, $p < .001$, a significant difference between *Alone* (M = 67.86, SD = 26.96) and *Together* (M = 60.17, SD = 23.75) when participants did not act, $t(27) = 3.07$, $d = .58$, $p = .005$ but no significant difference between *Alone* (M = 55.32, SD = 32.08) and *Together* (M = 51.07, SD = 28.01) when participants actions were inconsequential ($p = .068$). Thus, the interaction can be explained by a significant difference of *Partner* when participants acted to save lives and did not act to save lives, but not at baseline. Finally, to test whether agency ratings differed based on action specifically, we conducted a series of t-tests at each level of the *Action* factor at each level of the *Partner* factor. Results showed that when people acted when their actions saved lives, they reported higher agency ratings compared with when their non-action saved lives in both the *Alone* ($t(27) = 3.48$, $d = .66$, $p = .002$) and *Together* conditions ($t(27) = 3.32$, $d =$ .63, $p = .003$). Furthermore, agency ratings when acting to save lives was greater than when actions were inconsequential in both the *Alone* ($t(27) = 4.90$, $d = .90$, $p < .001$) and *Together* ($t(27) = 4.62$, $d = .87$, $p < .001$). Likewise, agency ratings when non-action saved lives was also greater than when actions were inconsequential in both the *Alone* ($t(27) = 3.17$, $d = .70$, $p < .001$) and *Together* ($t(27) = 3.06$, $d = .58$, $p = .005$) conditions. Overall, these results show that working with a partner reduces agency ratings and that agency ratings are highest when the participant performed an action, however agency ratings for non-action was still higher than action outcomes were inconsequential, suggesting agency is still felt over non-action.

Discussion

The current experiment investigated how explicit agency reports were affected by moral outcomes when working with a partner. Throughout the experiment, participants were told that they could either be working with a partner or working alone. We asked participants to view a train moving along a track which would eventually reach a fork in the rails. By default, the train would move across the bottom fork in the track. Participants were able to choose whether the train could change its default direction by pressing a button. There were six workers in total distributed across two tracks which varied by trial and the train will ultimately hit one set of them. So, in some trials, participants changing the direction of the train would lead to less people getting hit by the train, in others their non-action would save lives and in others both would lead to three people being hit. First, we established that participants did indeed make utilitarian decisions regarding which direction the train took at the fork on the rails. As predicted, we found that participants on average acted when their action would save more lives and did not act when acting would lead to more lives lost, consistent with previous research on moral reasoning (Navarette et al., 2012). This meant that we had an equivalent of action and non-action where the outcome was the same. As such, we were able to isolate the effect of action specifically in a free-choice, moral scenarios. As there was no difference between one or two workers being on the bottom track and four or five workers being on the bottom track, to make the results more interpretable, we collapsed these categories, so that we have three worker conditions: "*Action saves lives*", "*Action inconsequential*" and "*Non-action saves lives*". This was then split by each Partner condition (*Alone* and *Together*).

In this experiment, participants rated their experience of agency over the outcome lower when working with a partner compared to when working alone. This extends the finding shown throughout this thesis and within other research (e.g. Beyer et al., 2017), by showing this effect persists within moral contexts. While this is not unexpected, it has not yet been explored. Experiments which have found a reduction in agency ratings have been conducted in paradigms in which morality is not a relevant factor (Weller et al., 2020). Here, for the first time, we show that this effect is still present in moral scenarios. There are a few implications of this. The first is that these results suggest that the effect of working with a partner reduces one's feeling of agency even when the outcome has higher stakes. Many teams and groups make important decisions in groups, and these decisions can at times have large societal or environmental impact. An individual of that group feeling less agency over their actions may impact the overall decisions made in that group. If indeed this reduction of agency results from action-selection dysfluency (Chambon et al., 2014), then the provision of moral contexts are not sufficient to overcome this cognitive effect. This supports the idea that this reduction of agency could be a mechanism underpinning the bystander effect (Latane & Darley, 1968), where people often do not act even in emergency situations. One limitation of this experiment however is a lack of a non-moral or economic context. This raises the question of whether participants truly experience a sense of moral decision making during the experiment. Future research could include a further condition which requires participants to collect coins with the train to provide a point of comparison for the moral context.

The next key finding of this experiment relates to whether the participant acted or not to save lives, and how this interacts with working with a partner. Participants rated their control higher when they acted to save lives, rather than when they did not act to save lives. However crucially, participants not acting to save lives was higher than when their action or non-action resulted in no change in consequence. This is consistent with research from Weller et al. (2020), who found that participants experience agency even when no action is performed and with the idea that agency independent of motor contributions (Wegner & Wheatley, 1999; Wegner, 2003). The absence of motor cues highlights how the comparator model (Blakemore, Wolpert & Frith, 1998; Frith et al., 2000; Blakemore et al., 2002) alone is likely not sufficient in explaining agency. Instead, this research supports the idea that the sense of agency arises from both internal and external cues and the use of both are dependent on their availability and reliability (Moore et al., 2009). However, interestingly agency ratings were lower when non-action saved lives compared with when action saved lives. This could suggest that performing an action, rather than making the decision to not act carries with it a stronger sense of agency. We would suggest that the contribution of motor regions enhanced the sense of agency during the task. This finding is also relevant to discussions relating to ethical dilemmas relating to moral permissibility regarding letting someone die and active euthanasia. Whether or not there is a real ethical distinction between the two (see Tooley, 1980), legal systems seem to be more stringent towards active euthanasia, compared to allowing people to die from natural causes, such as turning off life support machines. If we subtract the idea of killing someone from letting someone die, we are only left with action. So, the action itself seems to be the heart of the dilemma. While this is an ethical issue beyond the scope of this experiment, because responsibility and agency are thought to be intertwined (Frith, 2014; Haggard & Tsakiris, 2009) and also are sensitive to moral outcomes (Moretto, Walsh & Haggard, 2011), we could perceive non-action that causes an undesirable outcome to be less blameworthy compared to an outcome resulting from an action because of the weakened agentic link between action and effect. In other words, we apply our own experience of agency to our judgments of others. Indeed, this has been discussed in terms of intentional versus unintentional action (Frith, 2014) and is supported by research showing that we attribute more blame to intentional action compared with unintentional action (Lagnado & Channon, 2008). In a similar vein, non-action might lead to less blame as we experience less agency during it compared to when acting. This is speculative and the link between blame, ethical dilemmas and agency are poorly understood. Future research could investigate how our own experience of agency affects the judgments of others and under what conditions we feel agency over non-action.

Conclusion

The results of this experiment show how explicit agency is affected during a moral dilemma task. We found that agency ratings were lower when working with a partner compared with working alone. Furthermore, people report feeling agency over both action and non-action when both have consequence. The results expand upon research that suggests that agency can be felt even without any motor contributions, highlighting the role of external cues in the sense of agency.

Chapter 8 - General Discussion

Introduction

Questions regarding the sense of agency range broadly within history. These questions have covered both philosophical and psychological concepts and sits as one of the core aspects of conscious experience. These broad questions regarding the qualia of action were further drilled into by neural imaging and novel experimental paradigms. Once researchers believed they had a way of measuring agency coherently, the amount of literature into the topic exploded, which in some ways illustrates the interest into the topic but also its application into society. Unsurprisingly, the notion that the conscious experience of our own actions is malleable and context dependent has broad and important applications into many fields of research, therapeutics, technology development and social policy. Only in the past decade though have researchers began to look at how the sense of agency is affected through social interaction and social context. The term social agency has begun to be used specifically in the terms of sense of agency research, which highlights the differential effect that socially charged stimuli have on the sense of agency. The idea of social agency is somewhat baked into our experience of being social animals. Our actions are often related to social interaction or are within some form of social context if we are aware of it or not. This is what earlier research tended to miss out and what newer research, including that within this thesis often covers.

When starting this thesis, the broad goal was to better understand how our experience of agency is altered within social context. Specifically, we wanted to investigate how the sense of agency is affected by working with a partner (although this admittedly was a result of a pivot resulting from COVID-19, lab closure and the need to develop a working online paradigm). We did this by first expanding on the research Beyer et al. (Beyer et al., 2017, Beyer et al., 2018), by first showing that the effect is present in online situations and through descriptions of robotic avatars, and computers. We then looked at temporal binding as a measure for both robotic avatars and computer descriptions. Finally, we produced research showing how these effects might occur when manipulating the outcome of the action and within moral contexts. We will describe provide a summary of the results of each chapter, before discussing their implications, possible limitations and future directions.

Overview of Results

Intentionality and Explicit Agency Whilst Working With Others

How do we experience our agency over our own actions when working with others? Is this different if we were to work with a human, a computer, or a robot? Chapter 4 looked to investigate this by conducting 3 experiments (*Experiment 1, Experiment 2*, *Experiment 3*) where we asked participants to work with a partner to stop a ball from rolling down a slope. The later they stopped the ball, the less points they would lose but would lose the greatest number of points if they stopped the ball too late. This was a design used by Beyer et al. (Beyer et al., 2017, Beyer et al., 2018) which this research was built from. The experiment was split into two conditions, one where the participant was working alone and one where the participant was working with a partner. In *Experiment 1,* participants were told that they were working with a human partner. in *Experiment 2*, they were told that they were working with a robotic avatar named “BOBBY”. Finally, In *Experiment 3*, participants were told that they were working with a computer. Finally, The goal of *Experiment 1* was to understand if the effect found in the Beyer et al. (Beyer et al., 2017, Beyer et al., 2018) studies would be repeated online. We found the typical effect, where participants report less agency when working with a partner compared to when working alone. As no person is physically present during online tasks, this would suggest that the effect of the partner is related to the contextual information regarding the task itself, rather than the physical presence of the partner. Thus, we then looked to manipulate the type of partner through varying the level of intentionality that the partner has. Therefore, we then conducted two further experiments, one where the participant was working with a robotic avatar named “BOBBY”, who was described as having emotions, goals and desires and another where the partner was described as a computer programme. We found that in both cases, participants reported less agency when working with their partner, compared to working alone.

Interval Estimation Methodology in Online Settings

The primary goal of *Experiment 4* was to understand if the measure of temporal binding using interval estimation can be obtained online, using the methodology set out by Buehner & Humphreys (2010). This methodology used the temporal estimation methodology by depressing the spacebar to replicate the time interval between two events. As online experiments are inherently noisier than lab-based experiments, we thought it prudent to test this before conducting larger scale studies involving this methodology. The results showed that this methodology can detect temporal binding online, providing a green light for using this to explore further research questions, and demonstrating its robustness within online modalities.

Temporal Binding When Working With Others

Following on from the previous experiments, the goals of the next chapter was to develop our understanding of this further, using temporal binding as an additional measure. As an overview, the results showed a peculiar juxtaposition between explicit agency ratings and temporal binding. Whilst agency ratings decreased when working with a partner compared with working alone, temporal binding increased. This was somewhat unexpected, as both are supposedly markers of agency. However, they are thought to represent different processes (Moore et al., 2012) and often do dissociate (Dewey & Knoblich, 2014). The current rationale for reduced agency when working within a partner is this type of paradigm is that the partner affects mentalising processes which affect action-fluency (Beyer et al., 2017). However, this does not explain why temporal binding is increased. Indeed, other experiments have shown that temporal binding is increased when working within social contexts (Ulloa et al., 2019; Vogel et al., 2021). One way to explain these results is that the increase in temporal binding reflects processes relating to the bolstering of self-agency when working with another. That is, agency is increased to distinguish one's actions from the partner's. On the other hand, the reduction in explicit agency ratings reflects the participants feeling of shared responsibility on the task. We will go through the particulars of each experiment here.

Experiment 5 looked to answer two key questions. First, was how temporal binding in addition to explicit agency is affected by working with a partner. The other question was whether the intentionality of the partner plays a role in reducing explicit and implicit agency. We therefore used BOBBY and "The Computer" as partners for two independent groups. As mentioned above, we found that temporal binding increased but explicit agency decreased, but this did not differ between the different intentionality groups. As this finding was unexpected, we then looked to replicate the results in *Experiment 6* and make some adjustments that we felt could be affecting results. This experiment only used BOBBY as the partner and *Experiment 6* replicated results of *Experiment 7.* To assess whether this effect is due to a mere presence effect (Zajonc, 1968), we then created another manipulation of the experiment where the Alone condition (where participants performed the task without the partner) was replaced with one where they were told the partner was observing them. Here, we found no difference between when the partner was working with their partner and when the participant was told that the partner was observing them. However, in this experiment, in the condition where the partner was observing them, the partner's avatar was present but was not present in the Alone condition. We felt that this was a possible confound, so a final experiment was created, which included the Alone, Observe, Together and Baseline conditions. Here, the partner's avatar is present in both the Observe and Together conditions. The results from this study showed the effect we saw in previous experiments, where temporal estimations were shorter in the Alone condition compared to the Together condition. We also found no significant difference between the Alone and the Observe conditions, but a significant difference between the Observe and Together conditions. These results suggest that the findings in this chapter are not related specifically to the mere presence effect, but rather, related to the ability for the partner to act during the task.

As mentioned previously, while temporal binding increased when working with a partner (indicating increased agency), explicit agency ratings decreased, suggesting that judgments of agency were lower when working with a partner when working alone. Throughout the experiments, this effect was only present when the partner could act during the task. That is, we found no differences when the participant was told they were being observed by the partner and when the participant was working alone. This suggests that this effect is related specifically to the partner being able to act, rather than the effect of the partner being present.

Outcome Relevance and Explicit Agency

One aspect that has not been explored in these types of social context paradigms is investigating how agency is affected by outcomes that benefit oneself, compared to benefitting another. To investigate this, we adopted a task from *Experiment 1, Experiment 2* and *Experiment 3*, where participants were required to stop a ball from rolling down a slope. Here, we asked participants on each trial how much control they felt over the outcome. The key manipulation here was that on some trials, participants were told they were working towards their partner’s points and some they were told they were working for their own. The results showed that participants rate their experience of agency lower when they were working with their partner’s points compared to their own, irrespective of whether they were working with their partner or not. The results here seem to suggest that our judgments of agency when working with others are dependent on the relevance of the outcome to ourselves. We speculate that perhaps this is an extension of self-referential processing, where information regarding the self is prioritised and has been shown to be present in various domains (Northoff et al., 2006). As there is some evidence to suggest self-referential processing extends to the sense of agency (Makwana and Srinivasan, 2019), this could be what is underpinning the results here.

Moral Responsibility and Explicit Agency

The link between responsibility and agency has been discussed numerous times in both philosophy and psychology. However, the link between the sense of agency and moral judgment is poorly understood. To assess this, we created a task in the style of the trolley problem (Thomson, 1984), where participants were required to view a train moving along a track which reaches a fork while working with a partner. On either side of the fork participants were told there were workers. Participants could press a button to move change the direction of the train. On each trial, the number of participants on each side of the track changed so that sometimes an action to change the direction of the train would result in more lives saved and in others their non-action would save lives. We then recorded agency ratings on each trial. The results showed that working with a partner reduced the sense of agency, consistent with the rest of the research in this thesis. We also found that while agency ratings were higher when the participant acted, people still reported agency over non-action. The results from this experiment demonstrate that the effect of working with a partner reducing explicit agency is persistent even in moral scenarios and demonstrates that people still report agency over outcomes that they did not act in. This supports the notion that the sense of agency is not solely driven by internal motor signals and that we can still feel agency over outcomes that are produced without an action.

The Results of this Thesis and Existing Models of Agency

What do these results in this thesis tell us about existing models of agency? Broadly speaking, the main accounts of the sense of agency are that it arrives either through forward predictive processes (i.e. the comparator model, Blakemore, Wolpert & Frith, 2002), because of postdictive rationalisation (i.e. the theory of mental causation, Wegner & Wheatley, 1999), or from a combination of internal sensorimotor cues and external, contextual cues (Synofzik, Vosgerau & Newen, 2008; Moore & Fletcher, 2012).

In terms of the comparator model, our results in this thesis continue to expand on research which suggest that the matching of predicted sensory feedback to actual sensory feedback alone is likely not the whole story (see Chapter 2). Instead, the results have shown that the sense of agency can be modulated by external cues that do not relate specifically to the amount of veridical control the participants experienced. In Chapter 4, Chapter 6 and Chapter 7, we repeatedly show that explicit self-reports of agency are reduced when working with a partner, along with an increase in temporal binding in Chapter 5. When working alone or together with the partner, participants control over the ball stopping was the same. As we only analysed trials where the participant themselves stopped the ball, the match between predictive sensory feedback and actual sensory feedback would likely have been the same between conditions. The comparator model would therefore predict that the sense of agency would remain constant. However, in our studies we have found that working with a partner modulated the sense of agency, which the comparator model fails to account for. Accounts of action-selection dysfluency (Chambon, Sidarus, Haggard, 2014) might suggest that working with a partner makes it more difficult to select when to act on the trial, disrupting the prospective processes involved in the comparator model which generate the sense of agency. Indeed, this is the explanation provided for the results of (Beyer et al., 2017; Beyer et al., 2018), who also found a reduction in agency when working with a human partner. However, the results from Chapter 5 found that temporal binding increased when working with a partner, which would contradict this notion. If working with a partner modulates agency through affecting prospective action-effect dysfluency, we should expect agency to decrease in both situations. The juxtaposition of an increase in implicit agency along with a reduction in agency cannot be accounted for by these accounts alone. Furthermore, in Experiment 10, we found that people report agency over actions even when they did not perform an action when that non-action had a consequence (when it led to saving lives). The idea that we can experience agency over non-action is also supported by other research that investigated this through temporal binding (Weller et al., 2020). This again cannot be accounted for by the comparator model alone, as their non-action involved no motor contributions to derive agency from. On the other hand, the results from explicit agency ratings could all be accounted for as top-down, retrospective inferences of causality after the event occurred (Wegner & Wheatley, 1999; Wegner, 2003). Regarding the reduction of agency, participants may infer after the event has occurred that because their partner is present, they are less in control of the outcome. In this case, the contextual information is being fed into their experience of agency postdictively and in turn, is a reflective judgment of their own agency after the event has finished, rather than their experience of agency during the trial. In addition, the theory of mental causation is also sufficient to explain the results from Experiment 10, showing people feel more agency over non-action, when the action has consequence than non-action that has no consequence. Here, the intention itself would be sufficient to produce causal attribution. It is consistent with experiments showing that agency can be derived outside of motor control outcome monitoring providing that the intention (without motor region contribution) is matched by the outcome (Wegner, 2004, Pronin, Wegner & Rodriguez, 2006). Therefore, the mental decision participants make to not act could be seen as sufficient to retrospectively produce agency, as it satisfies the needs of priority, consistency and exclusivity that the model suggests is necessary to produce agency. For example, the participant makes the decision to not act, shortly after the train continues along the track (exclusivity). As expected, the number of workers on the track are therefore hit by the train (consistency), and there are no other explanations for the outcome on the trial (exclusivity). Therefore, the participant deduces that their decision to not act is the cause of the event and as such attribute agency to themselves, because their intention to save as many workers on the track as possible is consistent with the outcome.

While the theory of mental causation is seemingly sufficient in explaining the explicit agency results, it alone fails to explain the results of Chapter 5, where temporal binding increased while explicit agency decreased. This could be in part due to the fact that the theory does not provide a distinction between the feeling of agency and judgments of agency. In their model, Synofzik, Vosgerau and Newen (2008) propose that the sense of agency arises from both the feeling of agency, which is the automatic agentic experience based on low-level sensorimotor processes and judgments of agency, which are reflective post-hoc attributions, based on higher level cognitive processes such as beliefs, social cues and contextual information. These two constructs of agency are distinct and therefore could arise from different processes, be affected by different factors and therefore be dissociable. It is possible therefore that temporal binding and explicit agency are sensitive to different cues, or that they are affected in different ways. For example, temporal binding may increase when working with a partner as a way of distinguishing self-generated action from the partner's during the trial. Emerging evidence is suggesting that temporal binding increases during social context (Ulloa et al., 2019; Vogel et al., 2021; Chen et al., 2024; Silver et al., 2024). Thus, it is possible that the increase in temporal binding reflects a low-level mechanism to detach one's own actions from other agents. We suggest that this occurs in a low-level, automatic manner and would be a function of the feeling of agency that Synofzik and colleague suggests. A mechanism like this may be particularly important in situations like those in this thesis, where participants are tasked with working together but their actions are independent from one another. Alternatively, an increase in temporal binding could also reflect a summation of agency of the dyad, like that posited through research into joint agency (Pacherie, 2012), whereby participants are simultaneously experiencing self-agency and vicarious agency of the partner. On the other hand, explicit judgments of agency might be more prone to modulation through higher order aspects of social cognition, such as the diffusion of responsibility (Bandura, 1991) or other rationalisations, perhaps related to the task itself, which led participants to judge that their agency is reduced. For example, objectively, the amount of control that the participants had over the outcome is reduced when working with the partner as they could stop the ball. Thus, the total potential for the participant to stop the ball is reduced in the Together condition compared to the Alone. This could be consciously detected and rationalised by the participant, leading them to deduce that they did in fact have less control over the outcome.

Limitations and Future Directions

As a result of COVID-19 and the lockdowns that occurred due to it, the studies in this thesis were all conducted online. While research has suggested that effects found in laboratory-based paradigms can be replicated online (Dandurand, Schultz, Onishi, 2008; Arechar, Gächter & Molleman 2017; Anwyl-Irvine et al. 2020; Peyton, Huber & Copock, 2022), it is particularly worth noting as the experiments here are focussing on the social component of the sense of agency. Evidence seems to show that how we interact with each other online compared with in person are distinct (Lieberman & Schroeder, 2020). Furthermore, as the participant cannot see the partner, action representations or other social cues are not available from them. This creates an experience of working together that is not typically seen in many in-person cooperative tasks. On the other hand, the online domain is increasingly becoming a host to various cooperative and competitive tasks, where social interaction takes place. This has been present in video games for many years and is becoming increasingly common in the workplace. Nonetheless, future research could expand upon the current findings by conducting them in the laboratory using confederates to perform the task with the participant. Another point potentially relating to the experiments conducted online is the relatively small effect sizes found from temporal binding when comparing conditions where the participants were working alone, compared to with their partner. These effect sizes ranged between $d = .2$ and $d = .27$, which would be considered a small effect size (Fritz, Morris & Richler, 2012). Online research might lead to smaller effect sizes compared with laboratory-based experiments, due to an increased noise in the data. However, without laboratory-based experiments as a comparison we do not know if the effect is a result of the diluted social interaction that takes place online, due to a lack of attention or other variables that are not controlled for in online experiments. Future research would benefit by assessing these findings further in the laboratory, to better determine the true population effect size.

The studies in this thesis were based on the paradigm from Beyer et al. (2017,2018), where participants were required to stop a ball from rolling down a slope to earn point. As all the studies followed this approach, or an approach analogous to it, it is worth discussing whether this type of paradigm is fit for purpose. As the participants are required to stop the ball within a timeframe (ideally as late as possible), their actions are naturally performance orientated. This means that if they perform the task as it is intended, their choice on the trials is limited to stopping the trial as late as possible. This creates a contrived situation where participants are incorporating strategies in a gamified approach, rather than making free choices regarding their actions. Perception of choice seems to be a factor that modulates the sense of agency (Barlas & Obhi, 2013). Therefore, it could be argued that the results here are specific to outcome-monitoring limited to performance-based action. We therefore do not know whether the effects found in this thesis are limited to such action, or whether they would persist where participants have more freedom to choose if, and when to act. On the other hand, it could be argued that Experiment 10 did in fact offer participants a choice in their actions and participants still reported less agency over their actions (even if participants opted for the utilitarian choice the vast majority of the time), suggesting that the effect still occurs when choice is given. Future research could investigate this further by assessing how working with a partner affects agency where participants have more freedom to choose where and when to act.

As a subjective measure, agency ratings are prone to rationalisation or demand characteristics. While evidence has shown that people are better at judging their sense of agency better than we might expect (Metcalfe and Greene, 2007), one might wonder the extent to which the findings of explicit agency can be attributed to the methodology that we used. We asked participants “How much control did you feel over the outcome”; a qualitatively different question from asking “Did you cause this outcome”. Typically, discussions of outcome monitoring and the sense of agency concerns whether people feel that they are the cause of their actions and the resulting consequence (Gallagher, 2000; Haggard, 2017). Control as a variable certainly could be an aspect of the sense of agency, but perhaps does not capture the definition in full. For example, someone with obsessive-compulsive disorder may not feel that they have control over their compulsions but would experience agency over them as there is no causal ambiguity. On the other hand, without control over outcomes entirely our sense of agency would likely not exist. Thus, it is likely that our sense of control contributes to the sense of agency but does not explain its existence entirely. This is a natural limitation of these sorts of paradigms as we told participants who acted on the trial and therefore, they are aware of the binary outcome of who caused it. By design, we ensured that the cause was unambiguous as we wanted to assess the effect of working with a partner on self-generated actions specifically. Thus, asking the participant “Did you cause this outcome” would likely lead to participants to report that they were in fact the cause, as the cause and effect intentionally lacked ambiguity. However, this naturally means that the paradigm this study used and those used by Beyer et al. (2017,2018) are focussing on the sense of control, which is perhaps only part of agentic experience. This could be a reason for why temporal binding and explicit measures dissociate, as they are capturing different elements that contribute to the sense of agency. As the sense of control could be seen as a dissociable element from general agentic experience (Pacherie, 2007), perhaps this is why we see the results in Chapter 5. Furthermore, as action was not necessary to produce agency in Experiment 10, we may question whether these explicit agency ratings are relating to the experience of action at all, or are more related to how much control participants believed that had over the outcome specifically. Future studies investigating this further could attempt to capture other aspects of agentic experience, like intention or cause (rather than control). For example, an experiment could ask participants to perform the same task but also ask “Did you intend to stop the ball?”, “Did you cause the ball to stop?” and “How much control did you feel over the outcome?". Assessing these factors individually, along with their interactions may help untangle whether the sense of agency is reduced as a whole, or whether it is specifically related to feelings of control. Finally, as we asked specifically about control regarding the outcome (rather than the action), the scale we used could be seen as missing the experience of action itself. Thus, it could be argued that the rating scale does not fully capture the full experience of agency. An alternative approach for future research could use a scale that focusses on the action, rather than the outcome in similar style experiments.

Only in Chapter 5 did we use temporal binding. Explicit agency ratings are supposedly reflective of judgments of agency, while temporal binding is a putative marker for the feeling of agency. Evidence has suggested that these are two distinct processes, which may involve different cognitive and neural mechanisms (Moore et al., 2012). Therefore, the studies that do not include temporal binding as a measure do not necessarily tell us about the feeling of agency that the participants experience. Therefore, the findings from Chapter 7 would benefit from the inclusion of temporal binding, especially considering the contrasting results of temporal binding and explicit agency ratings from Chapter 6. Future experiments could explore this line of research.

Our means of measuring temporal binding was through the interval estimation methodology outlined by Humphreys & Buehner (2010). This method holds some advantages over the Libet clock methodology (Haggard, Clark and Kalogeras, 2002), particularly in social agency paradigms, where often attention is directed towards other aspects of the experiment rather than the clock (Silver et al., 2021). However, as the methodology relies on the timing between two events. In the baseline condition, a tone is presented which leads to a stimulus change (in this case the shape changing colour). In the operant conditions (where the participant causes the stimulus change), the participant performs an action, which causes a tone to occur and then stimulus change. As the tone doesn't occur until after the button press, the latency of the keyboard could offset the perception of the action, leading to shorter perceived estimations when acting. However, keyboard latencies are often very short (around 4ms, to 20ms) and this alone would not be able to account for the differences between baseline conditions and operant conditions. Furthermore, our particular methodology asked participants to estimate the time interval by holding down a button, which was the approach taken by Stephenson et al. (2018). This is however not the only approach that can be taken as other research has asked participants to report their estimations verbally (Humphreys & Buehner, 2009) or through a time estimation scale (Moreton, Callan & Hughes, 2017). As we only took one approach to measuring temporal binding, one might wonder whether it incorporates specific biases to the task. We have outlined one potential issue above and in Chapter 5, we discuss how the timing could be affected by the onset of the key press and its release (i.e. the key press and release are not accounted for in the timing estimation). We would argue that this is unlikely, as the way we measured time estimates was consistent across the conditions and is therefore any bias that may occur would be systemic in the experiment. Furthermore, one could also argue that other methodologies have their own biases. If participants are asked to report on a scale or verbally estimate, this may lead to rationalisations or demand characteristics that self-reports are typically prone to. The methodology used in this thesis is arguably less prone to such biases. Nonetheless, as we only used the button press methodology, it might be useful for researchers in the future to investigate whether we come to the same conclusions as those found in this thesis with other interval estimation methodologies or with the Libet clock method.

Summary of Thesis

The focus of this thesis was to understand how the sense of agency is affected when working with others in an online environment. First, in Chapter 4, we looked to expand on research that suggests that the explicit sense of agency is reduced when working with others. We followed a paradigm from Beyer et al. (2017,2018), replicating their results when participants were told they were working with a human participant. These results demonstrate that a physical representation of the participant is not necessary for the effect to occur and the effect is specifically related to the task. We then wanted to understand how the perceived intentionality of the partner might affect the explicit agency ratings and found that the effect is present for both a robotic avatar and a computer, suggesting that intentionality of the partner plays no role in the reduction of explicit agency when working with a partner. Next, in Chapter 5, we investigated whether the interval estimation methodology (Humphreys & Buehner, 2010) was obtainable online. To our knowledge, this is the first time that this methodology was used and an effect was found in an online environment. Chapter 6 adapted the methodology from Chapter 4 to make the task suitable for temporal binding and looked at whether intentionality could be a factor in either temporal binding or explicit agency. We found no effect of intentionality but found an interesting juxtaposition between temporal binding and explicit agency ratings, where temporal binding increased but explicit agency decreased when working with a partner. This finding expands on work suggesting that social contexts increase our implicit sense of agency. In addition to this, we found that these results were not simply due to a mere presence effect, but rather a function of the partner being able to act on the task. Finally, we found that explicit agency ratings can be modulated by whether the outcome is related to the self, compared to the partner. This study potentially highlights the role of self-referential processing in the sense of agency.

The results here show the complex relationship of social cues on the sense of agency. The thesis highlights how working with a partner seemingly reduces judgments of agency (or at least, our feeling of control) over the outcomes of our actions. However, it also furthers research that suggests that how we experience agency is perhaps not derived from a single mechanism. The research outlined here seems to suggest that implicit and explicit agency do not arise from the same mechanism and are sensitive to different social cues. Furthermore, this thesis questions the degree to which social cues necessarily need to come from human actors, or whether they can be derived from machines or robotic avatars. Indeed, the results from Chapter 4, Chapter 6 and Chapter 7 all seem to indicate that agency is reduced in the same way as when they believe they are working with a machine or with a partner. Finally, this thesis questions the extent to which agency is driven through motor control regions, as it appears that agency can be felt irrespective of whether a participant acted or not. We hope the results here can contribute to research across various domains, including psychology and human-computer interaction. A better understanding of how working together with both humans and machine partners could help us better navigate this technologically driven world that we have created.

References

Aarts, H., & van den Bos, K. (2011). On the foundations of beliefs in free will: Intentional binding and unconscious priming in self-agency. *Psychological science*, *22*(4), 532-537.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388-407.

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental economics*, *21*, 99-131.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men; research in human relations* (pp. 177–190). Carnegie Press.

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

Bandura, A. (1991). Social cognitive theory of self-regulation. Organizational Behavior and Human Decision Processes, 50(2), 248–87

Bandura, A., Underwood, B., & Fromson, M. E. (1975). Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of research in personality*, *9*(4), 253-269.

Barlas, Z. (2019). When robots tell you what to do: Sense of agency in human-and robot-guided actions. *Consciousness and cognition*, *75*, 102819.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?. *Perspectives on psychological science*, *2*(4), 396-403.

Berberian, B., Sarrazin, J. C., Le Blaye, P., & Haggard, P. (2012). Automation technology and sense of control: A window on human agency. *PLOS ONE*, 7(3), e34075.

Beyer, F., Münte, T. F., Krämer, U. M., & Hinrichs, H. (2018). Emotional reactivity to threat modulates activity in mentalizing network during aggression. *Social Cognitive and Affective Neuroscience*, 13(5), 507-516.

References

Beyer, F., Sidarus, N., Bonicalzi, S., & Haggard, P. (2017). Beyond self-serving bias: Diffusion of responsibility reduces sense of agency and outcome monitoring. *Social Cognitive and Affective Neuroscience*, 12(1), 138-145.

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

Bolt, N. K., Loehr, J. D., & Pacherie, E. (2016). The agentic we: Collective intentionality and the sense of agency. *Philosophical Explorations*, 19(3), 217-234.

Brandi, M. L., Kaifel, D., Bolis, D., & Schilbach, L. (2019). The interactive self–a review on simulating social interactions to understand the mechanisms of social agency. *i-com*, *18*(1), 17-31.Buehner, M. J., & Humphreys, G. R. (2009). Causal binding of actions to their effects. *Psychological Science*, 20(10), 1221-1228.

Brandi, M. L., Kaifel, D., Lahnakoski, J. M., & Schilbach, L. (2020). A naturalistic paradigm simulating gaze-based social interactions for the investigation of social agency. *Behavior Research Methods*, *52*, 1044-1055.

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.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1982). The concept of external validity. *Journal of consumer research*, *9*(3), 240-244.

Campbell, W. K., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of general Psychology*, *3*(1), 23-43.

Carroll, J. M. (1997). Human-computer interaction: psychology as a science of design. *Annual review of psychology*, *48*(1), 61-83.

Carruthers, G. (2012). The case for the comparator model as an explanation of the sense of agency and its breakdowns. *Consciousness and Cognition*, 21(1), 30-45.

Carruthers, P. (2012). The architecture of the mind: Massive modularity and the flexibility of thought. Oxford University Press.

Caspar, E. A., Beyer, F., Cleeremans, A., & Haggard, P. (2021). The obedient mind and the volitional brain: A neural basis for preserved sense of agency and sense of responsibility under coercion. *PLoS One*, *16*(10), e0258884.

References

Caspar, E. A., Lo Bue, S., Magalhães De Saldanha da Gama, P. A., Haggard, P., & Cleeremans, A. (2020). The effect of military training on the sense of agency and outcome processing. *Nature communications*, *11*(1), 4366.

Caspar, E. A., Vuillaume, L., Magalhães De Saldanha da Gama, P. A., & Cleeremans, A. (2017). The influence of (dis) belief in free will on immoral behavior. *Frontiers in psychology*, *8*, 226717.

Chambon, V., Haggard, P., & Moore, J. W. (2012). Agency, choice and the role of the angular gyrus in the sense of agency. *Frontiers in Human Neuroscience*, 6, 204.

Chambon, V., Sidarus, N., & Haggard, P. (2014). From action intentions to action effects: how does the sense of agency come about? *Frontiers in Human Neuroscience*, 8, 320.

Chambon, V., Wenke, D., Fleming, S. M., Prinz, W., & Haggard, P. (2013). An online neural substrate for sense of agency. *Cerebral Cortex*, *23*(5), 1031-1037.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*(1), 112-130.

Chen, Y., Zou, X., Wang, Y., He, H., & Zhang, X. (2024). The enhancement of temporal binding effect after negative social feedback. *Cognition and Emotion*, 1-18.

Christensen, J. F., Di Costa, S., Beck, B., & Haggard, P. (2019). I just lost it! Fear and anger reduce the sense of agency: a study using intentional binding. *Experimental brain research*, *237*, 1205-1212.

Christensen, J. F., Yoshie, M., & Haggard, P. (2016). Emotional valence, sense of agency and responsibility: A study using intentional binding. *Consciousness and Cognition*, 70, 1-10.

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). Reduced sense of agency in human-robot interaction. In *Social Robotics: 10th International Conference, ICSR 2018, Qingdao, China, November 28-30, 2018, Proceedings 10* (pp. 441-450). Springer International Publishing.

Clarke, S., McEllin, L., Francová, A., Székely, M., Butterfill, S. A., & Michael, J. (2019). Joint action goals reduce visuomotor interference effects from a partner’s incongruent actions. *Scientific Reports*, *9*(1), 15414.

Cleeremans, A. (2021). The impact of agency on trust in human-robot interaction. *Experimental Brain Research*, 239(1), 289-303.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Coyle, D., Moore, J., Kristensson, P. O., Fletcher, P., & Blackwell, A. (2012, May). I did that! Measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2025-2034).

Craik, F. I., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., & Kapur, S. (1999). In search of the self: A positron emission tomography study. *Psychological science*, *10*(1), 26-34.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE, 8*(3), e57410.

Cubillas, C. P., Landáburu, Í., & Matute, H. (2020). Methodological factors involved in the study of temporal binding using the Open Source Software Labclock Web. *Frontiers in Psychology*, *11*, 521657.

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. 40(2), 428–434. http://dx.doi.org/10.3758/brm.40.2.428.

Daprati, E., Sirigu, A., Pradat-Diehl, P., Franck, N., & Jeannerod, M. (2000). Recognition of self-produced movement in a case of severe neglect. *Neurocase*, *6*(6), 477-486.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of personality and social psychology*, *8*(4p1), 377.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition?. *Trends in cognitive sciences*, *14*(10), 441-447.

De Vignemont, F., & Fourneret, P. (2004). The sense of agency: A philosophical and empirical review of the “Who” system. *Consciousness and cognition*, *13*(1), 1-19.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *Neuroscientist*, 13(6), 580-593.

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.

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science*, 324(5928), 811-813.

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.

Dumville, J. C., Torgerson, D. J., & Hewitt, C. E. (2006). Reporting attrition in randomised controlled trials. *BMJ, 332*(7547), 969-971.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, 6(5), 178-190.

Ebert, J. P., & Wegner, D. M. (2010). Time warp: Authorship shapes the perceived timing of actions and events. *Consciousness and cognition*, *19*(1), 481-489.

Eddy, C. M. (2016). The junction between self and other? Temporoparietal junction dysfunction in neuropsychiatry. *Neuropsychologia*, 89, 465-477.

Ellemers, N., Van Der Toorn, J., Paunov, Y., & Van Leeuwen, T. (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review*, *23*(4), 332-366.

Engbert, K., & Wohlschläger, A. (2007). Intentions and expectations in temporal binding. *Consciousness and Cognition*, 16(2), 255-264.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, 5(1), 16-25.

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

Farrer, C., Bouchereau, M., Jeannerod, M., & Franck, N. (2008). Effect of distorted visual feedback on the sense of agency. *Behavioural neurology*, *19*(1-2), 53-57.

Farrer, C., Franck, N., Georgieff, N., Frith, C.D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: A positron emission tomography study. *NeuroImage*, 18(2), 324-333.

Farrer, C., Franck, N., Paillard, J., & Jeannerod, M. (2003). The role of proprioception in action recognition. *Consciousness and Cognition*, 12(4), 609-619.

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.

Farrer, C., Valentin, G., & Hupé, J. M. (2013). The time windows of the sense of agency. *Consciousness and Cognition*, 22(4), 1431-1441.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Fehr, E., Fischbacher, U., & Kosfeld, M. (2005). Neuroeconomic foundations of trust and social preferences: initial evidence. *American Economic Review*, *95*(2), 346-351.

Feldman, G. (2017). Making sense of agency: Belief in free will as a unique and important construct. *Social and Personality Psychology Compass*, *11*(1), e12293.

Foot, P. (1967). *The problem of abortion and the doctrine of double effect* (Vol. 5, pp. 5-15). Oxford.

Fort, K., Adda, G., & Cohen, K. B. (2011). Amazon Mechanical Turk: Gold mine or coal mine?. *Computational Linguistics*, *37*(2), 413-420.

Frith, C. (2014). Action, agency and responsibility. *Neuropsychologia*, 55, 137-142.

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*(1404), 1771-1788.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General*, *141*(1), 2.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General*, *141*(1), 2.

Galang, C. M., Malik, R., Kinley, I., & Obhi, S. S. (2021). Studying sense of agency online: Can intentional binding be observed in uncontrolled online settings?. *Consciousness and Cognition*, *95*, 103217.

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

Gallagher, S., & Trigg, D. (2016). Agency and anxiety: Delusions of control and loss of control in schizophrenia and agoraphobia. *Frontiers in Human Neuroscience*, 10, 459.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 71-85.

Garaizar, P., Cubillas, C. P., & Matute, H. (2016). A HTML5 open source tool to conduct studies based on Libet's clock paradigm. *Scientific reports*, *6*(1), 1-9.

Gentsch, A., & Synofzik, M. (2014). Affective coding: The emotional dimension of agency. *Frontiers in Human Neuroscience*, 8, 608.

Gentsch, A., Weiss, C., Spengler, S., Synofzik, M., & Schütz-Bosbach, S. (2015). Doing good or bad: How interactions between action and emotion expectations shape the sense of agency. *Social neuroscience*, *10*(4), 418-430.

Georgieff, N., & Jeannerod, M. (1998). Beyond consciousness of external reality: A "who" system for consciousness of action and self-consciousness. *Consciousness and Cognition*, 7(3), 465-477.

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

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125-130.

Green, M. F., & Horan, W. P. (2010). Social cognition in schizophrenia. *Current Directions in Psychological Science*, 19(4), 243-248.

Grynszpan, O., Sahaï, A., Hamidi, N., Pacherie, E., Berberian, B., Roche, L., & Saint-Bauzel, L. (2019). The sense of agency in human-human vs human-robot joint action. *Consciousness and cognition*, *75*, 102820.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 196-207.

Haggard, P., & Clark, S. (2003). Intentional action: Conscious experience and neural prediction. *Consciousness and cognition*, *12*(4), 695-707.

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

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

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

Hascalovitz, A., & Obhi, S. S. (2015). Personality and intentional binding: an exploratory study using the narcissistic personality inventory. *Frontiers in human neuroscience*, *9*, 13.

Hascalovitz, S. F., & Obhi, S. S. (2015). The impact of narcissism on agency detection and belief in free will. *Personality and Individual Differences*, 79, 172-177.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*(1), 400-407.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior research methods*, *48*, 400-407.

Higbee, K. L., Millard, R. J., & Folkman, J. R. (1982). Social Psychology Research during the 1970s: Predominance of Experimentation and College Students. *Personality and Social Psychology Bulletin*, *8*(1), 180-183.

Holekamp, K. E., & Strauss, E. D. (2016). Aggression and dominance: an interdisciplinary overview. *Current Opinion in Behavioral Sciences*, *12*, 44-51.

Hortensius, R., & Cross, E. S. (2018). From automata to animate beings: the scope and limits of attributing socialness to artificial agents. *Annals of the new York Academy of Sciences*, *1426*(1), 93-110.

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.

Huff, C., & Tingley, D. (2015). “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, *2*(3).

Hughes, G. (2018). The role of the temporoparietal junction in implicit and explicit sense of agency. *Neuropsychologia*, *113*, 1-5.

Humphreys, G. R., & Buehner, M. J. (2009). Magnitude estimation reveals temporal binding at super-second intervals. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1542.

Humphreys, G. R., & Buehner, M. J. (2010). Temporal binding of action and effect in interval reproduction. *Experimental Brain Research*, 203(2), 359-369.

Imaizumi, S., & Tanno, Y. (2019). Intentional binding coincides with explicit sense of agency. *Consciousness and cognition*, *67*, 1-15.

Ioannidis, J. P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Jama*, *294*(2), 218-228.

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neuroscience & Biobehavioral Reviews*, *33*(6), 843-863.

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

Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: A reappraisal. *Experimental Brain Research*, 192(3), 527-532.

Keil, J., Barutchu, A., Desebrock, C., & Spence, C. (2023). More of me: Self-prioritization of numeric stimuli. *Journal of Experimental Psychology: Human Perception and Performance*.

Khalighinejad, N., & Haggard, P. (2015). Modulating human sense of agency with non-invasive brain stimulation. *Cortex*, *69*, 93-103.

Kirsch, W., Kunde, W., & Herbort, O. (2019). Intentional binding is unrelated to action intention. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3), 378-385.

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

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

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754-770.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754-770.

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology*, *8*(2), 142-164.

Langer, E. J. (1975). The illusion of control. *Journal of personality and social psychology*, *32*(2), 311.

Le Bars, S., Devaux, A., Nevidal, T., Chambon, V., & Pacherie, E. (2020). Agents' pivotality and reward fairness modulate sense of agency in cooperative joint action. *Cognition*, *195*, 104117.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in cognitive sciences*, *8*(12), 528-533.

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): The unconscious initiation of a freely voluntary act. *Brain*, 106(3), 623-642.

Lieberman, A., & Schroeder, J. (2020). Two social lives: How differences between online and offline interaction influence social outcomes. *Current opinion in psychology*, *31*, 16-21.

Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions: a review. *Frontiers in Human Neuroscience*, 8, 643.

Lockwood, P. L., Wittmann, M. K., Nili, H., Matsumoto-Ryan, M., Abdurahman, A., Cutler, J., ... & Apps, M. A. (2022). Distinct neural representations for prosocial and self-benefiting effort. *Current Biology*, *32*(19), 4172-4185.

Lynn, M. T., Muhle-Karbe, P. S., Aarts, H., & Brass, M. (2014). Priming determinist beliefs diminishes implicit (but not explicit) components of self-agency. *Frontiers in psychology*, *5*, 123385.

Makwana, M., & Srinivasan, N. (2019). Self-associated stimuli produce stronger intentional binding. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(11), 1436.

Malik, A., & Obhi, S. S. (2012). Recalling episodes of social exclusion affects perceived levels of social pain and subsequent social choices. *Journal of Social and Personal Relationships*, 29(7), 853-865.

Malik, R. A., Galang, C. M., & Finger, E. (2022). The sense of agency for brain disorders: A comprehensive review and proposed framework. *Neuroscience & Biobehavioral Reviews*, *139*, 104759.

Mead, G. H. (1913). The social self. *Journal of Philosophy, Psychology and Scientific Methods*, 10(14), 374-380.

Metcalfe, J., & Greene, M. J. (2007). Metacognition of agency. *Journal of Experimental Psychology: General*, 136(2), 184-199.

Miele, D. B., Wager, T. D., Mitchell, J. P., & Metcalfe, J. (2011). Dissociating neural correlates of action monitoring and metacognition of agency. *Journal of Cognitive Neuroscience*, 23(11), 3620-3636.

Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371–378.

Montague, P. R., & Chiu, P. H. (2007). For goodness' sake. *Nature Neuroscience*, *10*(2), 137-138.

Moore, C., Dunham, P. J., & Dunham, P. (2014). *Joint attention: Its origins and role in development*. Psychology Press.

Moore, J. W., & 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., Ruge, D., Wenke, D., Rothwell, J., & Haggard, P. (2012). Disrupting the experience of control in the human brain: Pre-SMA contributes to the sense of agency. *Proceedings of the Royal Society B: Biological Sciences*, 279(1739), 3274-3284.

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

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

Moreton, J., Callan, M. J., & Hughes, G. (2017). How much does emotional valence of action outcomes affect temporal binding?. *Consciousness and cognition*, *49*, 25-34.

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

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), 81-103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78).

Nass, C., Steuer, J., & Tauber, E. R. (1994, April). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78).

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional “trolley problem”. *Emotion*, *12*(2), 364.

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional “trolley problem”. *Emotion*, *12*(2), 364.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of experimental child psychology*, *162*, 31-38.

Northoff, G., Heinzel, A., De Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, *31*(1), 440-457..

Northoff, G., Heinzel, A., De Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, *31*(1), 440-457.

Obhi, S. S., & Hall, P. (2011a). Sense of agency and intentional binding in joint action. *Experimental Brain Research*, 211(3-4), 655-662.

Obhi, S. S., & Hall, P. (2011b). Sense of agency and intentional binding in joint action. *Journal of Experimental Psychology: Human Perception and Performance*, 37(1), 129-139.

Obhi, S. S., Swiderski, K. M., & Brubacher, S. P. (2012). Induced power changes the sense of agency. *Consciousness and cognition*, *21*(3), 1547-1550.

Obhi, S. S., Swiderski, K. M., & Farquhar, R. (2013). The effect of mood on the sense of agency. *Acta Psychologica*, 142(3), 286-293.

Ofcom. (2023). Connected Nations Report. Retrieved March 21,2024, https://www.ofcom.org.uk/research-and-data

Ofcom. (2023). Connected Nations Report. UK Government. Retrieved March 21, 2024, https://www.ofcom.org.uk/research-and-data/multi-sector research/infrastructure-research/connected-nations-2023

Office for National Statistics. (2022). Internet access – households and individuals, Great Britain. Retrieved from https://www.ons.gov.uk

Pacherie, E. (2001). The content of intentions. *Mind & Language*, 16(4), 400-432.

Pacherie, E. (2001). The phenomenology of action: A conceptual framework. *Cognition*, 82(1), 1-32.

Pacherie, E. (2007). The sense of control and the sense of agency. *Psyche*, *13*(1), 1-30.

Pacherie, E. (2011). The phenomenology of joint action: Self-agency vs. joint-agency. In *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 343-389). Cambridge, MA: MIT Press.

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

Pagliari, M., Chambon, V., & Berberian, B. (2021). What is new with Artificial Intelligence? Human-agent interactions through the lens of social agency.

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22-27.

Pascolini, L., Stephenson, L. J., Bayliss, A. P., & Wyer, N. A. (2021). Words of agency: Executed and observed vocal actions induce a temporal binding effect. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(12), 1717.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Peyton, K., Huber, G. A., & Coppock, A. (2022). The generalizability of online experiments conducted during the COVID-19 pandemic. *Journal of Experimental Political Science*, *9*(3), 379-394.

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.

Pfister, R., Pfeuffer, C. U., & Kunde, W. (2014). Perceiving by proxy: Effect-based action control with unperceivable effects. *Cognition*, 132(3), 251-261.

Pfister, R., Tonn, S., Weller, L., Kunde, W., & Schwarz, K. A. (2021). To prevent means to know: Explicit but no implicit agency for prevention behavior. *Cognition*, *206*, 104489.

Pockett, S., & Miller, A. (2007). The rotating spot method of timing subjective events. *Consciousness and cognition*, *16*(2), 241-254.

Pronin, E., Wegner, D. M., McCarthy, K., & Rodriguez, S. (2006). Everyday magical powers: the role of apparent mental causation in the overestimation of personal influence. *Journal of personality and social psychology*, *91*(2), 218.

PubMed. (n.d.). Number of published papers with terms "prolific" or "mturk." Retrieved May 4, 2024, from https://pubmed.ncbi.nlm.nih.gov

Rameson, L. T., Satpute, A. B., & Lieberman, M. D. (2010). The neural correlates of implicit and explicit self-relevant processing. *NeuroImage*, *50*(2), 701-708.

Ratcliffe, M. (2013). Depression and the phenomenology of free will.

Recht, S., & Grynszpan, O. (2019). The sense of social agency in gaze leading. *Journal on Multimodal User Interfaces*, *13*, 19-30.

Reips, U. D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In *Psychological experiments on the Internet* (pp. 89-117). Academic Press.

Reips, U.-D. (2000). The Web Experiment Method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 89-117). Academic Press.

Ritter, K., Dziobek, I., Preißler, S., Rüter, A., Vater, A., Fydrich, T., Lammers, C.-H., Heekeren, H. R., & Roepke, S. (2010). Lack of empathy in patients with narcissistic personality disorder. *Psychiatry Research*, 187(1-2), 241-247.

Robinson, J., Rosenzweig, C., Moss, A. J., & Litman, L. (2019). Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. *PLOS ONE, 14*(12), e0226394.

Rodd, J. (2024). The advantages and disadvantages of browser-based software for online research. *Behaviour Research Methods, 56*(2), 321-337.

Rodd, J. M. (2024). Moving experimental psychology online: How to obtain high quality data when we can’t see our participants. *Journal of Memory and Language*, *134*, 104472.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of personality and social psychology*, *35*(9), 677.

Roselli, C., Ciardo, F., De Tommaso, D., & Wykowska, A. (2022). Human-likeness and attribution of intentionality predict vicarious sense of agency over humanoid robot actions. *Scientific Reports*, *12*(1), 13845.

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years?. *Journal of consulting and clinical psychology*, *58*(5), 646.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature neuroscience*, *4*(5), 546-550.

Sahaï, A., Caspar, E., De Beir, A., Grynszpan, O., Pacherie, E., & Berberian, B. (2023). Modulations of one’s sense of agency during human–machine interactions: a behavioural study using a full humanoid robot. *Quarterly journal of experimental psychology*, *76*(3), 606-620.

Sahaï, A., Caspar, E., De Beir, A., Grynszpan, O., Pacherie, E., & Berberian, B. (2023). Modulations of one’s sense of agency during human–machine interactions: a behavioural study using a full humanoid robot. *Quarterly journal of experimental psychology*, *76*(3), 606-620.

Sahaï, A., Desantis, A., Grynszpan, O., Pacherie, E., & Berberian, B. (2019). Action co-representation and the sense of agency during a joint Simon task: Comparing human and machine co-agents. *Consciousness and Cognition*, *67*, 44-55.

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., & 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.

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

Schmitt, B. H., Gilovich, T., Goore, N., & Joseph, L. (1986). Mere presence and social facilitation: One more time. *Journal of experimental social psychology*, *22*(3), 242-248.

Schurger, A., Pak, J., & Roskies, A. L. (2021). What is the readiness potential?. *Trends in cognitive sciences*, *25*(7), 558-570.

Scott, N. J., Ghanem, M., Beck, B., & Martin, A. K. (2022). Depressive traits are associated with a reduced effect of choice on intentional binding. *Consciousness and Cognition*, *105*, 103412.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*, (3):515-530.

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

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, 88(3), B11-B21.

Sherif, M., & Sherif, C. W. (1953). *Groups in harmony and tension; an integration of studies of intergroup relations.* Harper & Brothers.

Shiraishi, M., & Shimada, S. (2021). Inter-brain synchronization during a cooperative task reflects the sense of joint agency. *Neuropsychologia*, *154*, 107770.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology, 69*, 487-510.

Sidarus, N., & Haggard, P. (2016). Difficult action decisions reduce the sense of agency: A study using the Eriksen flanker task. *Acta psychologica*, *166*, 1-11.

Sidarus, N., Travers, E., Haggard, P., & Beyer, F. (2020). How social contexts affect cognition: mentalizing interferes with sense of agency during voluntary action. *Journal of Experimental Social Psychology*, *89*, 103994.

Sidarus, N., Travers, E., Haggard, P., & Beyer, F. (2020). How social contexts affect cognition: mentalizing interferes with sense of agency during voluntary action. *Journal of Experimental Social Psychology*, *89*, 103994.

Sidarus, N., Vuorre, M., & Haggard, P. (2017). How action selection influences the sense of agency: An ERP study. *NeuroImage*, *150*, 1-13.

Silver, C. A., Tatler, B. W., Chakravarthi, R., & Timmermans, B. (2021). Social agency as a continuum. *Psychonomic Bulletin & Review*, *28*(2), 434-453.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 81(1), 174-176.

Simon, J. R., & Craft, J. L. (1970). Effects of an irrelevant auditory stimulus on visual choice reaction time. *Journal of Experimental Psychology*, 86(3), 272-274.

Simon, J. R., & Small, A. M. Jr. (1969). Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, 53(5), 433-435.

Singer, P., Seymour, B., O'Doherty, J., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439(7075), 466-469.

Singer, P., Seymour, B., O'Doherty, J. P., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661), 1157-1162.

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., & Haggard, P. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nature Neuroscience*, 7(1), 80-84.

Sperduti, M., Delaveau, P., Fossati, P., & Nadel, J. (2011). Different brain structures related to self- and external-agency attribution: A brief review and meta-analysis. *Brain Structure and Function*, 216(2), 151-157.

Sperduti, M., Delaveau, P., Fossati, P., & Nadel, J. (2014). Different brain structures related to self- and external-agency attribution: A brief review and meta-analysis. *Brain Structure and Function*, 219(2), 519-527.

Sperduti, M., Pieron, M., Leboyer, M., & Zalla, T. (2014). Altered pre-reflective sense of agency in autism spectrum disorders as revealed by reduced intentional binding. *Journal of autism and developmental disorders*, *44*, 343-352.

Stagkourakis, S., Spigolon, G., Williams, P., Protzmann, J., Fisone, G., & Broberger, C. (2018). A neural network for intermale aggression to establish social hierarchy. *Nature neuroscience*, *21*(6), 834-842.

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), 1073-1077.

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.

Sui, J., & Humphreys, G. W. (2015). The integrative self: How self-reference integrates perception and memory. *Trends in cognitive sciences*, *19*(12), 719-728.

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human perception and performance*, *38*(5), 1105.

Synofzik, M., & Voss, M. (2010). Disturbances of the sense of agency in schizophrenia. In *Neuropsychology of the sense of agency: From consciousness to action* (pp. 145-155). Milano: Springer Milan.

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., & Newen, A. (2012). I move, therefore I am: A new theoretical framework to investigate agency and ownership. *Consciousness and Cognition*, 21(1), 411-424.

Synofzik, M., Vosgerau, G., & Newen, A. (2012). The feeling of agency: A meta-analysis. *Cognition*, 125(1), 107-114.

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

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.

Tanaka, T., & Kawabata, H. (2021). Sense of agency is modulated by interactions between action choice, outcome valence, and predictability. *Current Psychology*, *40*(4), 1795-1806.

Tankersley, D., Stowe, C. J., & Huettel, S. A. (2007). Altruism is associated with an increased neural response to agency. *Nature neuroscience*, *10*(2), 150-151.

Thomson, J. J. (1984). The trolley problem. *Yale LJ*, *94*, 1395.

Tooley, M. (1980). An irrelevant consideration: Killing versus letting die. *Killing and letting die*, *2*, 103-111.

Tooley, M. (1980). An irrelevant consideration: Killing versus letting die. *Killing and letting die*, *2*, 103-111.

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

Twenge, J. M., & Campbell, W. K. (2003). "Isn't it fun to get the respect that we're going to deserve?" Narcissism, social rejection, and aggression. *Personality and Social Psychology Bulletin*, 29(2), 261-272.

Ulloa, J. L., Vastano, R., George, N., & Brass, M. (2019). The impact of eye contact on the sense of agency. *Consciousness and cognition*, *74*, 102794.

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review, 23*(1), 87-102.

van der Wel, R. P. R. D., & Knoblich, G. (2013). Cues to agency: Time can tell. *Agency and joint attention*, *54*, 256.

Vogel, D. H., Jording, M., Esser, C., Weiss, P. H., & Vogeley, K. (2021). Temporal binding is enhanced in social contexts. *Psychonomic bulletin & review*, *28*(5), 1545-1555.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological science*, *19*(1), 49-54.

Waltemate, T., Gall, D., Roth, D., Botsch, M., & Latoschik, M. E. (2016). The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 27-35.

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.

Watson, P. J., & Morris, R. J. (1991). Narcissism, empathy and social desirability. *Personality and Individual differences*, *12*(6), 575-579.

Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social cognition unbound: Insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science*, *19*(1), 58-62.

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., & Wheatley, T. (1999). Apparent mental causation. Sources of the experience of will. *American Psychologist*, 54(7), 480-492.

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.

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

Weller, L., Schwarz, K. A., Kunde, W., & Pfister, R. (2020). Something from nothing: Agency for deliberate nonactions. *Cognition*, *196*, 104136.

Whalley, H. C., Simonotto, E., Moorhead, W., McIntosh, A., Marshall, I., Ebmeier, K. P., ... & Lawrie, S. M. (2006). Functional imaging as a predictor of schizophrenia. *Biological psychiatry*, *60*(5), 454-462.

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology, 39*(4), 202-217.

Winkler, P., Stiens, P., Rauh, N., Franke, T., & Krems, J. (2020). How latency, action modality and display modality influence the sense of agency: a virtual reality study. *Virtual Reality*, *24*, 411-422.Whalley, H. C., Simonotto, E., Flett, S., Marshall, I., Ebmeier, K. P., Owens, D. G. C., ... & Lawrie, S. M. (2004). fMRI correlates of state and trait effects in subjects at genetically enhanced risk of schizophrenia. *Brain*, *127*(3), 478-490.

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

Yoshie, M., & Haggard, P. (2017). Effects of emotional valence on sense of agency require a predictive model of action outcomes. *Cognition*, 168, 164-173.

Yun, S., Wen, W., An, Q., Hamasaki, S., Yamakawa, H., Tamura, Y., ... & Asama, H. (2018, October). Investigating the relationship between assisted driver's SoA and EEG. In *International Conference on NeuroRehabilitation* (pp. 1039-1043). Cham: Springer International Publishing.

Zajonc, R. B. (1965). Social Facilitation: A solution is suggested for an old unresolved social psychological problem. *Science*, *149*(3681), 269-274.

Zheng, Y. Y., Ma, Y. T., Zhang, J. Y., & Xie, X. (2020). COVID-19 and the cardiovascular system. *Nature Reviews Cardiology, 17*(3), 259-260.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*(4), 493-504.

Zorowitz, S., Solis, J., Niv, Y., & Bennett, D. (2023). Inattentive responding can induce spurious associations between task behaviour and symptom measures. *Nature Human Behaviour*, *7*(10), 1667-1681.

Appendices

Appendix A

Individual Differences in Anthropomorphism Scale from Waytz, Cacioppo & Eply (2010)

We will ask you to rate the extent to which you believe various stimuli (e.g. technological or mechanical items, wild and domestic animals, and natural things) possess certain capacities. On a 0-10 scale (where 0 = “Not at All” and 10 = “Very much”), please rate the extent to which the stimulus possesses the capacity given. Please circle a number to indicate your response.

We will ask you about the extent to which the stimulus has a mind of its own, has free will, has intentions, has consciousness, can experience emotions, is good-looking, is durable, is lethargic, is active, and is useful.

By “has a mind of its own” we mean able to do what it wants.
By “has free will” we mean able to choose and control its own actions.
By “has intentions” we mean has preferences and plans.
By “can experience emotion” we mean it has feelings.
By “has consciousness” we mean able to be aware of itself and its thoughts and feelings.
By “good-looking” we mean attractive.
By “lethargic” we mean moving slowly.
By “active” we mean moving frequently and quickly.
By “useful” we mean able to be used for something.

1. To what extent is the desert lethargic?
2. To what extent is the average computer active?
3. To what extent does technology—devices and machines for manufacturing, entertainment, and productive processes (e.g. cars, computers, television sets)—have intentions.
4. To what extent does the average fish have free will.
5. To what extent is the average cloud good-looking.
6. To what extent are pets useful?
7. To what extent does the average mountain have free will?
8. To what extent is the average amphibian lethargic?
9. To what extent does a television set experience emotions?

10. To what extent is the average robot good-looking?
11. To what extent does the average robot have consciousness.
12. To what extent do cows have intentions?
13. To what extent does a car have free will?
14. To what extent does the ocean have consciousness?
15. To what extent is the average camera lethargic?
16. To what extent is a river useful?
17. To what extent does the average computer have a mind of its own.
18. To what extent is a tree active?
19. To what extent is the average kitchen appliance useful?
20. To what extent does a cheetah experience emotions?
21. To what extent does the environment experience emotions?
22. To what extent does the average insect have a mind of its own?
23. To what extent does a tree have a mind of its own?
24. To what extent is technology—devices and machines for manufacturing, entertainment, and productive processes (e.g. cars, computers, television sets)—durable?
25. To what extent is the average cat active.
26. To what extent does the wind have intentions?
27. To what extent is the forest durable?
28. To what extent is a tortoise durable?
29. To what extent does the average reptile have consciousness?
30. To what extent is the average dog good-looking?