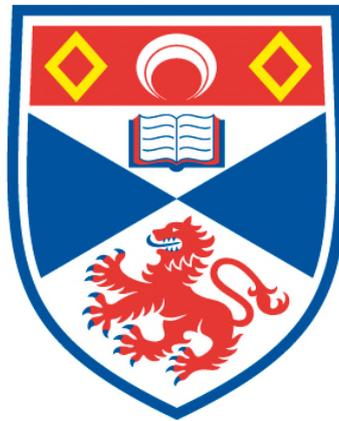


**UNDERSTANDING OBJECT-DIRECTED INTENTIONALITY
IN CAPUCHIN MONKEYS AND HUMANS**

Rouoting Tao

**A Thesis Submitted for the Degree of PhD
at the
University of St Andrews**



2016

**Full metadata for this item is available in
St Andrews Research Repository
at:**

<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/9304>

This item is protected by original copyright

**Understanding object-directed intentionality in Capuchin
monkeys and humans**

Ruoting Tao

1. Candidate's declarations:

I, Ruoting Tao, hereby certify that this thesis, which is approximately 62,000. words in length, has been written by me, and that it is the record of work carried out by me, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in and as a candidate for the degree of Doctorate of Philosophy in September, 2011. This is a record for my study carried out in the University of St Andrews between 2011 and 2015

Date 23/02/2016..... Signature of candidate

2. Supervisor's declarations

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date Signature of supervisor

3. Permission for publication

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:access to printed copy and electronic publication of the thesis through the University of St Andrews

Date 23/02/2016..... Signature of candidate

Date Signature of supervisor

Acknowledgement

The four years was not easy, but was indeed a great journey accompanied by so many great minds and kind hearts. I feel grateful for all the people that have encouraged and supported me on my way to the PhD.

A very special thanks to my family, particularly my parents for their continued love, support and encouragement. You always encourage me to pursue my own interest. I could have never done this without you.

My greatest acknowledgement goes to my first supervisor, Dr Juan-Carlos Gomez. I am grateful for all the discussion, comments, suggestions, patience and encourages through this thesis and my four years.

I would like to thank my examiners, Dr Christine Caldwell and Dr Cat Hobaiter, for the thorough discussion about my thesis and research, and for the very helpful suggestions about the expansion of the thesis.

I would also like to thank all the staff and research students in the School of Psychology and Neuroscience, University of St Andrews, for their help and support. Many thanks to my second supervisor, Dr Amanda Seed, for her helpful advice and great insight. Many thanks to Prof Dick Byrne, Dr Cat Hobaiter, Prof Andy Whiten, Prof Zuberbuhler for great discussions about my studies and about animal cognition. I would also like to thank the colleagues and friends in the Wolfson office and the Lilac room for the encouragements and for the time spent together. A special thanks to Dr Shuang Xia, for the discussions about my and also her study, the mutual encouragements, the companionship in a foreign country, the fun time we spend together in and around St Andrews.

My thanks to all the staff at the Living Links to Human Evolution Research Centre at the Royal Zoological Society of Scotland, Edinburgh Zoo, for granting me the opportunity to conduct my research with the Capuchin Monkeys. Many thanks to the Capuchin Monkeys at the Living Links, and to all the human participants in St Andrews.

Many thanks to the University of St Andrews. I'll always remember the motto AIEN APIΣTEYEIN, ever to excel.

Table of Contents

Acknowledgement	i
Abstract	1
Key word	1
Chapter 1. Introduction	2
1.1 Theory of Mind	3
1.1.1 Belief understanding.....	4
1.1.2 ToM abilities besides belief understanding	6
1.2 Understand intentionality to understand ToM	7
1.2.1 Intentionality and Theory of Mind.....	8
1.2.2 Coding intentionality: to understand the relation between actions and objects.....	11
1.2.3 Understanding intentionality when the object is not directly observable	18
1.3 Aim of the current study	25
Chapter 2. Capuchin Monkey’s inference of objects from object-directed actions	28
Summary.....	28
2.1 Introduction.....	29
2.2 Method	31
2.2.1 Subjects.....	31
2.2.2 Apparatus.....	34
2.2.3 Stimuli	38
2.2.4 Procedure.....	43

2.2.5 Data analysis	43
2.3 Results	45
2.3.1 The effect of action type and outcome type on looking time and number of looks.....	45
2.4 Discussion	54
2.4.1 Capuchin monkeys can distinguish object-directed and non-object- directed actions.....	54
2.4.2 Do Capuchin monkeys infer the presence of the object from the action alone?.....	55
2.4.3 Gaze and grasping as intentional actions	57
2.4.4 From an evolutionary perspective	59

Chapter 3. Understanding of Intentionality in Adult Humans 62

Summary.....	62
3.1 Introduction.....	64
3.2 Experiment 1	66
3.2.1 Method	66
3.2.2 Results.....	76
3.2.3 Discussion	86
3.3 Experiment 2	91
3.3.1 Introduction	91
3.3.2 Method	93
3.3.3 Results.....	100
3.3.4 Discussion of experiment 2.....	106

Chapter 4. Adult human’s automatic encoding of other’s belief 110

Summary.....	110
4.1 Introduction.....	111

4.2 Experiment 1	119
4.2.1 Method	119
4.2.2 Results.....	124
4.2.3 Discussion	127
4.3 Experiment 2	131
4.3.1 Introduction	131
4.3.2 Method	131
4.3.3 Results.....	133
4.3.4 Discussion	137
4.4 Experiment 3	139
4.4.1 Introduction	139
4.4.2 Method	139
4.4.3 Results.....	142
4.4.4 Discussion	144
4.5 Experiment 4	147
4.5.1 Introduction	147
4.5.2 Method	148
4.5.3 Results.....	149
4.5.4 Discussion	151
4.6 Experiment 5	152
4.6.1 Introduction	152
4.6.2 Method	154
4.6.3 Results.....	155
4.6.4 Discussion	157
4.7 General Discussion	159
A summary of the five experiments.....	159

Chapter 5. Capuchin monkey's automatic encoding of other's beliefs 164

Summary.....	164
5.1 Introduction.....	165
5.2 Method.....	169
5.2.1 Participants.....	169
5.2.2 Apparatus and stimuli.....	170
5.2.3 Design and Procedure.....	174
5.2.4 Data analysis.....	174
5.3 Results.....	175
5.3.1 Monkey's attention in the experiment.....	175
5.3.2 Looking time at the attended trials.....	175
5.4 Discussion.....	177
5.4.1 Capuchin monkeys showed no encoding of Belief or Intentionality.....	177
5.4.2 Summary.....	180

Chapter 6. Discussion..... 181

6.1 A review of the rationale & aims.....	181
6.2 Summary of the experiments.....	184
6.2.1 Capuchin monkeys' inference of object from object-directed action (Chapter 2).....	184
6.2.2 Adult human's inference of object from object-directed action (Chapter 3)	185
6.2.3 Summary of chapter 2 and 3.....	186
6.2.4 Adult human's automatic encoding of object-directedness and their implicit understanding of belief (Chapter 4).....	188
6.2.5 Capuchin monkey's automatic encoding of object-directedness and their implicit understanding of belief (Chapter 5).....	189

6.2.6 Summary of Chapter 4 and 5	190
6.2.7 General conclusions & implications about the understanding of intentionality.....	191
6.3 About Implicit ToM.....	193
6.3.1 Implicit ToM of non-human primates.....	193
6.3.2 The Kovács-experiment with adult human is confounded by other factor than belief	Error! Bookmark not defined.
Reference	204
Appendix	215
Appendix I. Ethical approval for research with Capuchin Monkeys	215
Appendix II. test of Monkeys' attention across sessions in chapter 2.....	217
Appendix III. The questionnaire in experiment 1 of Chapter 2.....	220
Appendix IV. Example of the form in chapter 3 and 4	226
Example of forms in chapter 3.....	226
Example of Forms in chapter 4	234
Appendix V. The comparison of the experiments in chapter 5	241
Comparison of experiment three, four and five.....	241
Comparison of the two different procedures.....	243

Abstract

Understanding intentionality, i.e. coding the object directedness of agents towards objects, is a fundamental component of Theory of Mind abilities. Yet it is unclear how it is perceived and coded in different species. In this thesis, we present a series of comparative studies to explore human adults' and Capuchin monkeys' ability to infer intentional objects from actions.

First we studied whether capuchin monkeys and adult humans infer a potential object from observing an object-directed action. With no direct information about the goal-object, neither species inferred the object from the action. However, when the object was revealed, the monkeys retrospectively encoded the directedness of the object-directed action; unexpectedly, in an adapted version of the task adult humans did not show a similar ability.

We then adapted another paradigm, originally designed by Kovács et al (2010), to examine whether the two species implicitly register the intentional relation between an agent and an object. We manipulated an animated agent and the participants' belief about a ball's presence behind a hiding screen. We found no evidence showing that humans or monkeys coded object-directedness or belief. More importantly, we failed to replicate the original results from Kovács et al's study, and through a series of follow up studies, we questioned their conclusions regarding implicit ToM understanding. We suggested that, instead of implicit ToM, results like Kovacs et al's might be interpreted as driven by "sub-mentalizing" processes, as suggested by Heyes (2014).

We conclude that so called 'implicit ToM' may be based upon the computation of intentional relations between perceived agents and objects. But, these computations might present limitations, and some results attributed to implicit ToM may in fact reflect "sub-mentalizing" processes.

Key words: intentionality, object-directedness, implicit ToM, belief,

Theory of Mind, primate

Chapter 1. Introduction

Theory of mind (ToM) describes the ability to represent internal mental states, such as beliefs, desires, intentions, knowledge, and to use these representations to predict the behaviour of others (Premack & Woodruff, 1978). Since 1978 when Premack and Woodruff asked the question 'Does the chimpanzee have a theory of mind?', ample research has focused on both the ontogenetic and evolutionary origins of ToM. The subject has long been investigated within many fields, including comparative and evolutionary psychology (Call & Tomasello, 2008; Herrmann, Call, Hernández-Lloreda, Hare & Tomasello, 2007; Premack & Woodruff, 1978), developmental psychology (e.g., Baillargeon, Scott & He, 2010; Baron-Cohen, Leslie & Frith, 1985; Perner & Wimmer, 1983), cognitive neuroscience (e.g., Gallagher & Frith, 2003; Umiltà et al., 2001) and many other fields. For example, from the comparative perspective, Call and Tomasello (2007) compared chimpanzees, orang-utans and pre-schoolers using a battery of cognitive tasks and proposed a cultural intelligence hypothesis where they argued that humans outperform great apes in understanding social context. From the developmental perspective, Woodward and colleagues conducted a large amount of studies investigating infants' abilities to understand intentional actions (Woodward, 2009b); In addition, the field of neuroscience has added evidence about the neural basis underpinning ToM and intention understanding, e.g. Umiltà and colleagues proposed that the mechanism of intention understanding could be explained by the mirror neuron system (Umiltà et al., 2001). More than 30 years of research have brought us abundant knowledge about the ToM abilities of adults, children, infants, atypical people (e.g. autistic), and non-human animals (hereafter animals). However, we still know little regarding the cognitive mechanisms through which ToM is achieved. In this thesis we propose that understanding intentionality in terms of understanding object-directedness is a fundamental step in ToM abilities. We will examine the abilities of capuchin monkeys and human adults and investigate if they are able to understand object-directedness. In addition, we will examine if the participants are able to infer the object when directedness is perceived from intentional actions. We will mainly focus on the

directedness from an intentional action or an intentional agent to an object. Through these studies, we expect to find evidence supporting our proposal that intentionality is perceived as directed to an object and the ability to infer the object may be a fundamental step in developing theory of mind.

We will start by introducing Theory of Mind: the ability to understand other's mental states, and will then introduce intentionality and propose that understanding intentionality might be important for ToM development. After that, we will review the psychological and neuroscience studies with human infants and non-human primates to show evidence that both humans and primates can understand intentionality by coding behaviour and its relation to a goal as a full configuration. We will propose that the ability to spontaneously code object-directedness is the basic mechanism in the understanding of intentionality and suggest that inferring the goal-object of actions is a key issue to investigate. Though the four studies in this thesis, we will examine this proposal and we will discuss the results of the experiments and their implications.

1.1 Theory of Mind

In 1987, Premack and Woodruff (1978) asked a question: 'Does the chimpanzee have a Theory of Mind?' In their original experiment they presented a female chimpanzee a series of videotaped scenes where a human actor encountered several different problems. Later, the chimpanzee was given several photos to choose from. She consistently chose the one that depicted a potential solution to the problem faced by the human. The authors concluded that this demonstrated her understanding of the actor's purpose. Although several researchers pointed out problems with the methodology and offered other explanations for the result (e.g., Heyes, 1998; Povinelli, Perilloux, Reaux, & Bierschwale, 1998; Savage-Rumbaugh, Rumbaugh, & Boysen, 1978), this paper started the fertile field of Theory of Mind research, where researchers investigate whether individuals have the ability to understand mental states.

1.1.1 Belief understanding

Following on from suggestions that emerged out of the debate generated by Premack and Woodruff's paper (1978), Wimmer and Perner (1983) brought the question of ToM into developmental psychology by testing children's understanding of false belief. Early investigation of theory of mind mainly focused on the understanding of belief. Understanding belief, especially false belief, is of special importance in ToM research because it indicates mental representation, rather than simple behavioural response (Schneider, Slaughter, & Dux, 2014). If one forms a belief about a reality, but does not know then the reality changes afterwards, the original belief becomes false, or in other words, wrong, regarding the current reality. Therefore, to represent this false belief, one cannot depend on the reality or one's own belief to understand another's mind. Thus, understanding false belief indicates the ability to attribute to others' mental representation abilities (Baillargeon et al., 2010; Call & Tomasello, 2008; Sodian, 2009).

In Wimmer and Perner's (1983) experiment, they asked if children could explicitly represent another individual's false belief. To investigate this question, they developed a story where a character, Maxi, put chocolate in one place A, but his mother moved the chocolate to another place B during Maxi's absence. As Maxi was absent when the chocolate was moved, his belief about the chocolate's location was not updated, and so conflicted with both the reality and the subject's belief. If the subject understood Maxi's false belief (FB), when asked where Maxi would look for his chocolate, they should indicate place A; otherwise, if they can only refer to their own belief, place B would be indicated. The task thus tested whether children can explicitly represent others' FB. Typically developing children can reach the milestone of passing the classic FB task at around 4-years-old (see Wellman, Cross & Watson, 2001 for a meta-analysis). Younger children, however, systematically make the mistake of giving a realistic response (indicating place B) in this task (Wellman et al., 2001). The FB task (e.g., Sally-Ann task, Baron-Cohen et al., 1985; Maxi Task, Perner & Wimmer, 1983) has been traditionally considered a key test of ToM, and passing the task is considered a benchmark in the acquisition of explicit ToM, which indicates

the acquisition of the ability to represent other's mental states (Gopnik & Astington, 1988; Wimmer & Weichbold, 1994).

Though abundant evidence indicates that children typically pass the false belief task at around 4 years of age (Wellman et al., 2001; Wellman & Liu, 2004), many recent studies suggest that much younger children and infants may understand false belief to some extent, despite their failure in the classic task (see review in Baillargeon et al., 2010). This ability to succeed in a false belief tasks that does not require explicitly stating the representation of others' belief is described as *implicit* understanding of false belief (Apperly & Butterfill, 2009). Infants in their second year may have established an implicit understanding of false belief. It is demonstrated by several non-verbal FB tasks, e.g., tasks using a violation of expectation paradigm (Onishi & Baillargeon, 1995; Surian, Caldi, & Sperber, 2015; Träuble, Marinović, & Pauen, 2010), anticipatory looking tasks (Clements, Clements, Perner, Clements & Perner, 1994; Southgate, Senju, & Csibra, 2007), active non-verbal helping tasks (Buttelmann, Carpenter, & Tomasello, 2009; Southgate, Chevallier & Csibra, 2010). These findings suggest that humans understand false belief much earlier than 4-years of age and that the ability to represent other's mental states might emerge early in our development.

In addition, similar approaches to the infant studies have also been applied to non-human animals to tackle this topic from an comparative perspective to investigate the evolutionary origins of this capability (Marticorena, Ruiz, Mukerji, Goddu & Santos, 2011). For example, Marticorena and colleagues (2011) followed Onishi and Baillargeon's experiment with infants and used a violation of expectation paradigm. Here when the outcome is contrary to the participants' expectation, they look longer at this unexpected outcome than they would at an expected one. To examine whether rhesus macaques understand another's beliefs, they presented rhesus macaques with situations where an object might hide in one of two places and an agent, in this case, an experimenter, had different beliefs about the object's location. When the agent searched in one of the two locations for the object, the monkey's looking time was recorded to indicate if the location being searched conformed with the monkey subject's expectation. For instance, when the object

was hidden in location A but the experimenter held a false belief (the object was in location B), the monkey, if it can understand false belief, should be surprised and would look longer when the experimenter searched at the location with the object, instead of the place where the experimenter believed the object should be. Their results, however, suggested no evidence that rhesus macaques understood the experimenter's false belief (Marticorena et al., 2011). This experiment, as with many other FB tasks applied to non-human primates, suggested that non-human primates may not be able to understand other's belief (e.g., Call & Tomasello, 1999; Kaminski, Call, & Tomasello, 2008).

Nevertheless, it doesn't follow that non-human primates cannot understand other's mental states at all. In Marticorena et al's experiment, the monkeys were surprised when the experimenter searched at the wrong location while she had the right knowledge about the location of the object. It indicated that the monkeys understood the experimenter's knowledge and the monkeys have some capability of recognizing other's mental state.

To briefly summarize, human beings have some ability to represent false belief from a very early age. Non-human primates cannot understand belief well, but they are able to understand some mental states.

1.1.2 ToM abilities besides belief understanding

Although much emphasis has been placed on false belief, ToM ability is not limited to belief understanding. From its original definition: the ability to understand other's mental states, ToM also involves the understanding of emotions, desires, intentions, and knowledge, etc. (Call & Tomasello, 2008; Premack & Woodruff, 1978; Wellman & Liu, 2004). And abundant literature has reported evidence in pre-schoolers, infants, and non-human primates regarding their understanding of different aspects of mental states besides belief.

For example, infants as young as 14-months have already adopted a fundamental understanding that other people have goals and intentions (Gergely, Bekkering, Király, Damme, & Wilson, 2002; Meltzoff, 1995); infants' ability to follow an other's gaze, especially following the gaze geometrically (behind visual obstacles),

provides evidence that at the end of their first year infants can understand others' perception and attention (Moll & Tomasello, 2004; Shepherd, 2010).

Not only children and infants have these ToM abilities, non-human primates also show evidence of understanding some mental states, such as other's intentions and goals (e.g., rational imitation: Buttelmann, Carpenter, Call, & Tomasello, 2007; distinguishing intentional and accidental actions: Call, Hare, Carpenter, & Tomasello, 2004; Phillips, Barnes, Mahajan, Yamaguchi, & Santos, 2009), and perception and knowledge (e.g., Bräuer, Call, & Tomasello, 2007; Flombaum, Santos, & Haven, 2005; Hare, Call, Agnetta, & Tomasello, 2000; Hare & Tomasello, 2004). Thus a range of studies suggest that, apart from belief understanding, there are many other early developed ToM abilities in human infants and shared ToM abilities in human and primates in dealing with mental states.

1.2 Understand intentionality to understand ToM

With the ample facts at hand, the main research question gradually shifts from whether infants and animals (especially non-human primates), can understand the meaning of other's behaviour and mental states, to how this understanding is achieved (Apperly & Butterfill, 2009; Baillargeon et al., 2010; Baron-Cohen, 1991, 2001; Gómez, 2009; Leslie, 1987, 1994; MacLean et al., 2012). Many researchers have tried to tackle this issue by different ways. For example, Apperly and Butterfill (2009) suggested that implicit ToM might have emerged early in evolution and infant development to help deal with ToM relevant problems before it can be mentally represented; Gallagher and Frith (2003) used an fMRI approach to investigate the brain areas related to the ToM tasks; Gergely and Csibra (2003) argued that mental representation is not always involved, the perception of the other's intention is more crucial. As we are particularly interested in a proposal that focuses on the understanding of intentionality and suggests that it is likely the foundation of ToM abilities (Gómez, 2008, 2009), in this section, we will introduce intentionality and why it might be important for ToM development.

1.2.1 Intentionality and Theory of Mind

In philosophy, the term intentionality refers to ‘the power of minds to be about, to represent, or to stand for, things, properties and states of affairs’ (Jacob, 2010). It initially emerged from the Latin word *intendere*, which means being directed to an object (Jacob, 2010). The term was reintroduced by the German philosopher Franz Brentano, where it was often summarised as ‘aboutness’. It refers to the directedness from mental acts to the external world (Brentano, 1874; Gomez, 2009). For example, if I say ‘He looks at the dog’, the mental act is looking and the external world, or in this case physical object, is the dog. Intentionality in this case is directed from his perception to the object. In another example, when someone reaches for a bar of chocolate, the intentionality is directed from his desire to the goal. Or when someone introduces a theory or an idea, the intentionality is directed from his mind to a mental object. In all of these examples, the mental ability connects the physical phenomena, e.g., looking and a dog, reaching and chocolate, introducing and a theory. Moreover, it always points from the subject to the object, and the directedness from the subject to the object is more than just the physical phenomena, but can have meaning beneath, which can be perceived or interpreted cognitively (Gomez, 2009).

Intentionality is involved in various mental activities, which always act upon an object. Note that ‘object’ here is to the contrary of the term ‘subject’. It can indicate a specific thing but not necessarily so. For example, if I say ‘I know the dog’, the dog can be a specific individual but it can also be the concept of dog. Either way, it is the object of my knowing. Or if someone desires something, the desire has to be directed to a desired object. Similarly, one has to intend what is intended, perceive what is perceived, know what is known (Brentano, 1874).

Note, however, that intentionality should be distinguished from the term intention (Gómez, 2009; Jacob, 2010). As introduced above, intentionality in our studies means the directedness to an object and will be interchangeable with ‘object-directedness’ in this thesis. Intention, on the other hand, often refers to purposefulness (Meltzoff, 1995) or willingness (Call, Hare, et al., 2004) in psychological studies. For example, in an experiment that showed that 18-month

infants can imitate a demonstrator rationally, based on the demonstrator's intention instead of his action, the term intention refers to being purposeful. To be specific, the infants in the experiment understand that the purpose of the demonstrator was to turn on the light; therefore when the demonstrator's hands were occupied and he used his head to turn on the light, the infant turned the light on by the usual and more efficient approach, i.e., by hand. In the control group, where the demonstrator used his head to turn on the light while his hands were free, the infants imitated the head action too since they understood that the demonstrator purposefully didn't use his hand (Meltzoff, 1995). This experiment was about infants' understanding of another's intention in terms of purpose, but was not about the intentionality that we will talk about in the current thesis. For another example that explores intention in terms of willingness but not intentionality, infants and primates can differentiate an experimenter's willingness. When someone failed to hand over a toy, infants as young as 9-months old could differentiate whether the agent accidentally dropped the toy or uncooperatively teased the infant (Marsh, Stavropoulos, Nienhuis, & Legerstee, 2010; Tomasello, Carpenter, Call, Behne, & Moll, 2005) and, when tested under similar conditions, non-human primates also behave similarly to infants (Call, Hare, et al., 2004; Phillips et al., 2009). In the latter case intention refers to willingness, which again should not be confused with intentionality.

Now the meaning of intentionality is made clear, we would like to bring in Theory of Mind, which makes things more interesting. Theory of mind involves the ability to understand other's mental states (Premack & Woodruff, 1978). Although the debate regarding whether primates and infants are truly capable of understanding other's mental states continues (Call & Tomasello, 2008; Lurz & Krachun, 2011; Povinelli & Vonk, 2004), ample empirical evidence has demonstrated that they are at least able to adjust their behaviour according to their observation of others' behaviour. For example, infants and many primate species can follow other's gazing direction (Rosati & Hare, 2009) and chimpanzees even move around a barrier, which blocks the line of gaze, to look for an object of interest (e.g., Tomasello, Hare, & Agnetta, 1999). Some researchers concluded from this that chimpanzees understand others' perceptions and intentions, and this was the reason that they

were not satisfied at finding only the barrier, but not something more interesting, along the gaze line (Tomasello, Hare, & Agnetta, 1999; Itakura, 2004; Maclean & Hare, 2012); but other researchers insist that gaze-following doesn't necessarily involve understanding the mental experience of seeing or attending to an object (Heyes, 1998; Lurz & Krachun, 2011; Povinelli & Eddy, 1997). However, it is worth noting that there is never an actual line of gaze, no physical connection between the agent (or the agent's eye) and the object: the line of gaze does not exist physically and cannot be observed directly. Primates' ability to follow the non-existing gaze line and their understanding of the connection between the eye and the object of interest itself already indicates an understanding beyond the observable behaviour (Gómez, 2008, 2009). This understanding of the directedness from the gaze to the object has been suggested to be achieved through the perception of intentionality (Gómez, 2008, 2009).

To perceive intentionality has been suggested as a possible corner stone for ToM abilities (Gomez, 2009). For example, by perception of intentionality, one can understand others' perception and knowledge, as has been shown in the case above about gaze. Visual perception involves seeing an object, and acoustic hearing. By coding the intentionality from seeing and hearing to the object, the ability of understanding others' perception and knowledge is enabled.

In some other cases the object of intentionality is not a physical object, but an object that only exists mentally, or, in other words, 'intentional inexistence' or an 'intentional object' (Brentano, 1874; Gomez, 2009). As Brentano suggested:

'this intentional inexistence is characteristic exclusively of mental phenomena... We can, therefore, define mental phenomena by saying that they are those phenomena which contain an object intentionally within themselves.'

Since all mental activities involve intentionality (Brentano, 1874), ToM abilities are highly likely to relate to, and perhaps be built on, the ability to perceive intentionality (Gómez, 2009). Abundant empirical evidence that examines ToM

related abilities indeed indicates human and non-human primates' ability in understand intentionality, as will be discussed below.

1.2.2 Coding intentionality: to understand the relation between actions and objects

In this section, we would like to review some empirical evidence that indicates infants' and primates' understanding and coding of intentionality.

1.2.2.1 Evidence in preverbal infants

A series of experimental findings indicate that young infants can understand intentionality, i.e., they can understand object-directed actions in the sense of coding a relation between an agent and an object (e.g., Hamlin, Hallinan, & Woodward, 2008; Woodward, 2009). In their first year, infants have already developed the ability to code object-directed actions, e.g., pointing and grasping, as directed towards a certain goal-object (Sodian & Thoermer, 2004; Sommerville & Woodward, 2005; Woodward, & Guajardo, 2002; Woodward, 1998b, 1999).

For example, in 1999, Woodward reported an experiment using a violation of expectancy paradigm (Woodward, 1999). Infants at the age of 9 months were habituated to an event in which an actor acted on one of two toys that sat side by side on a small stage. In the test phase, the toys changed position and the actor acted either towards the new toy by the old path (New Toy) or the old toy by the new path (New Path). Grasp was used as the object-directed action, in contrast to the non-object-directed action of flop (an accidental drop of the back-of-the-hand). Results showed that in the grasp condition, infants looked longer at the new toy than at the new path trials, indicating that they expected the action to be related to the previous goal object instead of the previous location. However, infants did not differentiate the two test trial types when the motion was a flop, suggesting that they discriminated goal-directed action from actions performed without a goal, and they did not relate non-object-directed action to an object. Younger infants at five to seven months olds also showed the same pattern of response as the 9-month-olds did (Woodward & Guajardo, 2003). Subsequent experiments explored a broader range of goal-directed actions including gazing (Paulus, 2011; Phillips, Wellman, &

Spelke, 2002; Sodian & Thoermer, 2004; Woodward & Guajardo, 2003) and pointing (Woodward et al., 2002). Using the same paradigm researchers showed that by 12-month of age, infants related others' gaze (Woodward & Guajardo, 2003) and pointing gestures (Woodward et al., 2002) to a goal object in the same way as they did with grasping. These authors concluded that infants under one year of age distinguish between object-directed and apparently non-object-directed behaviour. They did not perceive others' object-directed actions as simple movements, but coded a relation between the action and the goal-object (Woodward, 1999, 2009b).

In 2012, Cannon and Woodward further followed up with an eye tracking experiment and suggested that 11-month old infants could predict the goal of an agent's intentional action (Cannon & Woodward, 2012). In the experiment, they habituated the participants to videos where a grasping action was directed toward one of two objects. In the test events the location of the objects were swapped (as in the previous experiment), but instead of reaching for an object, the grasping action was incomplete and stopped between the objects. Eye tracking data revealed that infants looked from the hand to the object that was familiarized in the habituation videos significantly more often than at the other object. As a comparison, when the grasping action was replaced by a wooden claw, the infants did not look to the familiarized object more often than to the alternative one. The authors concluded that infants can 'actively use goal analysis to generate on-line prediction of an agent's next action' (Cannon & Woodward, 2012).

From the perspective of the understanding of intentionality, we suggest, however, that these experiments show that infants can perceive intentionality. Firstly, they can distinguish object-directedness from accidental actions, suggesting that they are able to identify intentionality. Secondly, when they perceive intentionality, they expect the object-directed action to be directed at its goal, suggesting that they code intentionality as a relationship from the subject to the object.

In addition, it has been suggested that infants' early understanding of intentional actions might be essential for attributing mental states, and might be a

precursor of later ToM development (Aschersleben, Hofer, & Jovanovic, 2008; Wellman & Phillips, 2001; Woodward, 1998b). Empirical work has indicated a correlation between infants' attention to object-directed actions and pre-schooler' ToM ability and was suggested as evidence (Aschersleben et al., 2008; Wellman et al., 2004). For example, Aschersleben and colleagues (2008) presented a longitudinal experiment and claimed that infants' understanding of intentionality predicted their ToM ability when they were around 4 years old. They first tested infants' understanding of intentionality following the Woodward-paradigm (Woodward, 1999) and then called the children back when they reached the age of 4 years to test them with a ToM scale, which included two false belief tasks (Wellman & Liu, 2004). They found that children's performance at the false belief tasks correlated with the decrement of attention from the habituation to the test trial in the Woodward-test. They concluded that the correlation indicated a link between the understanding of intentional actions and ToM abilities and they further suggested that infants' early understanding of intentionality was likely a precursor to later ToM development (Aschersleben et al., 2008).

To briefly summarize, infants can identify and understand intentionality, and this ability relates to later development of ToM.

1.2.2.2 Evidence in primates

The ability to understand intentionality by appreciating the relationship between agent and object not only develops early ontogenetically, but also developed early evolutionarily, as it is not limited to human infants, but is found in non-human primates as well (Kano & Call, 2014; Rochat, Serra, Fadiga, & Gallese, 2008; Umiltà et al., 2001).

For example, great apes were tested in a similar eye-tracking paradigm to that used with the infant's (Cannon & Woodward, 2012; Kano & Call, 2014). Bonobos, chimpanzees and orang-utans also showed an understanding of intentionality from the action to the object (Kano & Call, 2014). In the experiment, great apes were habituated to the event of a human grasping one of two objects. Then, as in the infant study, the positions of the two objects were swapped and the hand made an

incomplete reaching action between the objects. Eye-tracking was used to monitor great apes' expectation of object-directed actions and it revealed that they predictively looked at the familiarized object instead of the alternative one. This finding indicated that great apes can also understand intentionality in the way that infants do. And primates, again like infants, expect object-directed action to be directed at its goal-object.

Experiments with non-human primates using an eye tracking technique or looking time as the measurement are relatively sparse compared to those with infants (Kano & Call, 2014). Nevertheless, findings on their understanding of intentionality are not entirely absent. For example, numerous experiments of the so called "object choice paradigm" provide some relevant information regarding primates' understanding of intentionality. This paradigm has been widely used in primates to study their understanding of informative actions such as grasping, gazing, and pointing (Anderson, Montant, & Schmitt, 1996; Anderson, Sallaberry, & Barbier, 1995; Barth, Reaux, & Povinelli, 2005; Bräuer, Kaminski, Riedel, Call, & Tomasello, 2006; Call, Hare, & Tomasello, 1998; Itakura, Agnetta, Hare, & Tomasello, 1999; Mikl, Soproni, & Miklósi, 2006; Mulcahy & Call, 2009; Wood, Glynn, Phillips, & Hauser, 2007; Wood & Hauser, 2011). In this paradigm, a food reward is first shown to the subject and then hidden (the process of baiting is not shown to the subject) in one of two occluded locations (usually in containers or behind occluders). Then the experimenter produces some informative cue (e.g., a pointing gesture) directed at the baited location, following which the animals are allowed to choose one location and retrieve the reward. After watching the action of reaching towards a location, apes and monkeys are able to follow the direction of the action and retrieve the food (Call, Tomasello, et al., 2004; Wood et al., 2007; Wood & Hauser, 2011), indicating that they can understand the relationship between the reaching action and the hidden object. In comparison, the accidental gesture flop (Wood, Glynn, & Hauser, 2008) or the communicative gesture pointing (Call, Tomasello, et al., 2004) is not regarded as intentional by the participants, indicating that they can differentiate different actions in terms of object directedness. For the hand-flop, the gesture itself is not designed to be object-directed (Woodward, 1999) and accordingly is not seen

as intentional by rhesus macaques (Wood et al., 2008). Non-human primates generally fail to use communicative cues, pointing for instance, from a human experimenter (see reviews, Mikl et al., 2006; Mulcahy & Hedge, 2012; Rosati & Hare, 2009) and are sometimes described as ‘performing poorly’ in the object choice task in contrast to human infants (e.g., Bräuer et al., 2006; Itakura & Tanaka, 1998). However, this could be due to different motivations, as suggested by Hare and Tomasello (2004). Primates may have difficulties in understanding communicative intentions and perform better in competitive than in communicative contexts. And their failure in utilizing the communicative gesture could also be due to a difference in their ability to understand different types of intentional behaviour. After all, pointing is not a natural action in the behavioural repertoire of primates but arguably exclusively human. Therefore we conclude from the above results that non-human primates can distinguish intentionality from non-intentionality.

To summarize, the above evidence from non-human primates showed that non-human primates can perceive intentionality too. As human infants, they can also distinguish object-directedness from non-intentional actions, although intentionality might be different in their perspective and the human perspective. Secondly, when they perceive intentionality, they also expect the object-directed action to be directed at its goal, suggesting that they can also code intentionality as a relationship from the subject to the object. In addition, since this ability is shared by primates and infants, and that of infants is likely a precursor of ToM, it might also play an important role in primates’ ToM abilities.

1.2.2.3 Physiological evidence

More than behavioural evidence, physiological studies about the mirror neuron system (MNS, see review, Rizzolatti & Craighero, 2004) support the proposal that agents appreciate the relationship between action and goal-object (Gallese, 2007; Umiltà et al., 2001; Wood & Hauser, 2008). Mirror neurons, which are a type of neuron first discovered in the macaque pre-motor cortex, activate both while the same subject is performing or observing actions such as grasping, tearing, holding or manipulating objects (Gallese & Goldman, 1998; Rizzolatti et al., 1988). The MNS therefore matches others’ actions with the observer’s own motor representation

(Csibra, 2003; Gergely & Csibra, 2003; Umiltà et al., 2001). But what is more important for our topic of intentionality is that cell responses in MNS may reflect a coding of the interaction between an action and an object. These cells do not fire either when the subject just observes an action without a target object or observes the goal object without the action; only when the action is directed to the object, are the corresponding mirror neurons activated and therefore the alleged intentionality or object-directedness identified (Umiltà et al., 2001).

This property indicates that mirror neurons may play a crucial role in coding the relationship between object-directed action and the goal-object. This has been illustrated by several studies (see reviews, Fogassi, 2011; Gallese & Goldman, 1998; Ocampo & Kritikos, 2011), among which a single-cell recording study by Umiltà et al (2001) is of special interest. Here, they showed macaques a grasping action directed towards an object either in full vision or with the final goal-object hidden behind an occluder. In the occluded-cases, the monkey was always informed in advance whether the object existed or not, with the occluder only introduced afterwards. The researchers recorded that mirror neurons were activated with the same pattern when the object was directly observable and when it was behind the screen and therefore the actual grasping could only occur out of sight. It appeared that the neurons automatically reconstructed the missing goal of the action. However, the same neurons failed to respond in the control conditions when the monkey had been shown that there was no object behind the screen prior to the grasping action. Interestingly, the only difference in the two hidden conditions was the monkey's awareness of the object's existence based upon his previous ability to see the object or the absence of an object. Therefore, the authors concluded that the MNS does not just code the observation of the grasping action, but also the feature of object-directedness (Csibra, 2003; Gergely & Csibra, 2003; Umiltà et al., 2001). Or in other words, the MNS may be the biological foundation of the ability to understand intentionality.

More importantly, this study suggested that object-directedness may be attributed to the action alone, without a full visual presentation of the goal-object, although in the current case it needs an appropriate supporting context, (previous

direct perception of the existence of the object). Building on this finding, we made a hypothesis: when coding intentionality, the subject, the object, and their relationship of object-directedness might be regarded not as independent pieces, but as a full configuration of intentionality. When part of the components, e.g., the object-directed action and its goal-object, are observed, the implicit full configuration, i.e., the relation of object-directedness from the action to the object, can be automatically completed. It is similar to the gestalt perception where observable pieces are automatically perceived as a full configuration. For example, in Figure 1 - 1, when we observe the fragments of lines, we can actually perceive a square instead of merely four corners. The action and the object perceived are (to some extent) like the corners, and intentionality is like the full configuration of the square. Infants and primates in the Woodward-experiments, as well as in the other behavioural experiments, may have perceived the complete configuration of intentionality and therefore, related the action to its goal-object. The macaque monkey in the mirror-neuron-experiment knew about the object beforehand, and, upon observing the incomplete action, its mirror neuron system detected the gestalt intentionality. In this way, the directedness from the action to the object does not necessarily involve mental representation, such as intention.



Figure 1 - 1 the Gestalt square

1.2.3 Understanding intentionality when the object is not directly observable

The studies of infants and primates reviewed above are consistent with the proposal that intentionality is coded by the relationship between the action and the goal, which are perceived as a whole instead of being detected as two separate elements related by the attribution of an unobservable mental state (Gómez, 2009).

What would be more interesting, however, is if the object is not directly observable and nor is it known about beforehand, will it always be assumed to exist? It is interesting because, in this case, the object is one that does not have to exist physically or in reality, but can be an object that only exists in mind, or the so called 'intentional object' (Brentano, 1874; Gomez, 2009). In this case, it is no longer putting the action (or the subject) and the object together through an intentional relation, but uses the perception of intentionality to infer about the existence of the object. In other words, as intentionality is object-directed, there must be an object to which the action is directed to, even if none is or was immediately visible. Once the presence of the object is assumed, to interpret about the object would further complete the understanding of a social scenario. For instance, the simplest scenario is when the object is actually a physically one, (e.g., a food reward is expected in the object-choice task); an object of interest is found in the gaze following study. The interpretation of the object can also involve mental representation. For example, a monkey can believe that the sound from the bush indicates that a predator is hidden behind, and as a consequence the monkey runs away; when in fact it is another monkey playing and making the bush move. The physical object that moves the bush is the second monkey, but the first monkey reacts to the intentional object it has imagined: a non-existent predator. By means of interpreting the perceived intentionality, ToM ability might therefore be achieved and the perception of intentionality might thus be a first step of ToM. Nevertheless, what needs to be determined first is whether intentionality can be perceived when the object is not directly known about (Gomez, 2009). It would serve as a crucial link between coding intentional relations between agents and real objects, and coding intentional relations between agents and inferred object.

None of the literature has addressed this point directly, but some existing studies may be relevant for addressing the wider issue.

1.2.3.1 Inferring objects from gaze

A number of gaze following studies provide suggestive evidence that young human infants and some non-human primates may have the ability to infer previously unseen targets from others' gaze (see review of general gaze following, Rosati & Hare, 2009). We mentioned earlier that non-human primates can follow another's gaze around barriers (Tomasello et al., 1999) and that this ability is shared by human infants too (Moll & Tomasello, 2004). Moll and Tomasello (2004) reported that both 12-month and 18-month-old infants followed an adult's gaze line around a barrier, and checked back at the adult's gazing direction, as if making sure that they were looking at the same thing. Note that the infant had not seen behind the barrier beforehand and did not know if there was any object of interest. The authors reported that infants made efforts to go around the barrier, indicating that they expected the experimenter's gaze to be directed not only to the directly observable barrier, but towards something unseen at the other side (Moll & Tomasello, 2004). As a result they suggested that when infants detected an agent looking at something, they have the understanding that other people are seeing something different from what they initially see on their side of the barrier. Other researchers argue that gaze following is merely a simple behavioural reflex, where as long as the subjects detect an other's looking direction, they automatically follow it (see argument for simple behavioural reflex, Povinelli & Eddy, 1996, 1997). But if this was the case, the infants would not look for more when they've found the barrier in their gaze line. As they have no previous knowledge about what is behind the barrier, their expectation of the object of interest is likely to depend on their inference of an intentional object. The ability to follow gaze around barriers and past distractors is shared by human infants (Dunphy-Lelii & Wellman, 2004; Moll & Tomasello, 2004), great apes (Bräuer, Call, & Tomasello, 2005; Okamoto-Barth, Call, & Tomasello, 2007; Rosati & Hare, 2009; Tomasello et al., 1999) and monkeys (Amici, Aureli, Visalberghi, & Call, 2009; Burkart & Heschl, 2006; Emery, 2000). It is often interpreted as evidence of understanding 'seeing' in a ToM sense – understanding that others must be seeing

an object and therefore, understand others' perception and knowledge (see review, Call & Tomasello, 2008).

From the perspective of understanding Brentanian intentionality, the subjects do not have to understand another's perception as an internal, unobservable mental state, but may just perceive the other's gazing as intentional in the Brentanian sense. Or in other words, they have established the intentional relation between the agent and a potential object (Gómez, 2009; Scerif, Gómez, Byrne, & Gomez, 2004). Moreover, it is important to note that before following the gaze behind barriers, the subject had not seen the potential object of attention in advance. The fact opens up the possibility that they may have been displaying the capability to infer or imagine the existence of unseen targets when they perceive intentionality from others' gaze and attention (Csibra, 2003; Gergely & Csibra, 2003; Gómez, 2009). This suggestion is supported by a study of attention understanding in Diana monkeys (Scerif et al., 2004). In this study, researchers showed Diana monkeys photographs of familiar conspecifics orienting towards one of two locations covered with screens, followed by the display of a toy, which was hidden behind one of the locations, appearing when the screens were lowered. The position of the toy either corresponded to the gaze direction or not. The researchers found that monkeys' first inspection, total duration of looking, and number of looks were more likely to be directed to the position that the monkey in the photograph was orienting towards. In addition they checked back at the conspecific's looking direction more often when the toy appeared in the opposite position. The authors concluded that the monkeys followed the attention in a 'relatively automatic' way, but expected the presence of a target to be linked to the looker's attention before they actually saw the target. This would suggest that the monkeys expected a relationship of intentionality between the agent and the object of attention, which would be a different process from simple reflexive gaze following behaviour, and different from an understanding of others' mental representations of intentions and goals (Scerif et al., 2004).

From these studies, we suggest that infants and primates may have the ability to infer a goal from the gaze of others even when none is immediately available. But before closing the case, there are still some shortcomings and limitations to discuss.

Firstly, in the geometrical gaze following studies, the researchers suggested that when the apes checked back at the gazer, it provided evidence of expecting an object of interest when they found no object along the gaze line (Bräuer et al., 2005). However, this interpretation remains somewhat subjective and speculative. Maybe the subjects that failed to find an interesting object were just looking back at the face of the other because in the absence of a competing stimulus, the face itself is an interesting focus to observe; the 'gaze direction' is just a rich interpretation. The checking back behaviour only provides an indirect evidence of the expectation of an object aligned with the direction of gaze. Secondly, spider monkeys and capuchin monkeys (Amici et al., 2009) followed the gaze around barriers, but never checked back. It is difficult to tell if they had any expectation of an object or if they just followed the direction of the gaze in a reflexive way. To better examine whether the apes and monkeys understand gaze or other actions in the Brentanian sense, we need better paradigms that address more directly the issue of whether monkeys are or not inferring the existence of an object from the intentional behaviour of the agent.

In addition, although the study of Diana monkeys (Scerif et al., 2004) adopted a new paradigm with the potential for addressing the two points above, the researchers only tested 6 monkeys and found a significant result in only 4 of them, which was a small number of subjects and did not provide strong statistical power. In addition, in this study, the gazing cues were displayed by photos of two different monkeys, i.e., a dominant one and a young female. The significant effect only occurred in the case of the dominant. It is possible that the dominant's gaze is more significant for group members to attend to, but the number of comparisons was too limited and it remains possible that the effects were caused by other confounding factors.

Taking these limitations into consideration, we conclude from the above evidence that infants and primates may be able to infer the object from an intentional action, gazing, but that more direct and robust evidence is needed to support this hypothesis. Moreover, and more importantly, gazing is one specific type of action that might indicate intentionality, but alone it is not sufficient to argue for

the Brentanian understanding of intentionality. Intentional actions in general and in a wider range would provide more convincing evidence.

1.2.3.2 Inferring the object from intentional actions

The literature directly addressing the issue of inferring an object from intentional actions is very sparse. In most of cases, the existence of the object is known in advance and the inference of the object from actions besides gazing is not addressed. Nevertheless, some existing literature may provide some relevant information.

Firstly, the object-choice task, which we introduced above, may provide some relevant information regarding the ability to understand others' intentional actions and intentionality. We discussed that primates can differentiate intentional from non-intentional action earlier. However, when the subject observed the action directed to one of two occluded locations, the object was not directly observable, but the subjects could still understand the directedness. It suggested that from observing the actions alone, without direct access to the object, intentionality might be perceived. Note however, in most of the tasks, the object was shown in advance (e.g., Anderson, Sallaberry, & Barbier, 1995; Itakura & Tanaka, 1998) and therefore, the subjects did not necessarily have to infer about the presence of the object. In contrast to showing the object beforehand, in the mirror neuron experiment, the subject was shown the absence of the object (Umiltà et al., 2001). In this case, mirror neurons were not activated when the grasping action was observed, showing that the presence of the object was essential in coding intentionality. However, the experiment did not show whether intentionality could be perceived when the presence of an object was uncertain.

From the commonly used paradigms investigating intentional action and its goal-object, direct evidence for the question of whether primates are able to perceive intentionality and infer possible potential targets from intentional actions remains lacking. To address the issue of whether the object can be inferred when intentionality is perceived, a new paradigm should be able to present the subject

with intentional actions and a situation in which there may, but not necessarily, exist a target of the action.

A preliminary study using a violation of expectation looking-time design with Capuchin monkeys (Kersken, Rollins, & Gómez, n.d.) suggested a possible experimental paradigm. The violation of expectancy paradigm is commonly used with preverbal human infants to study their understanding of various physical and social scenarios (e.g., Baillargeon, 1995; Wynn, 1998; Xu & Carey, 1996). It has been adapted by researchers in comparative psychology for use with non-human primates (Anderson, Kuroshima, Kuwahata, & Fujita, 2004; Santos & Hauser, 1999). In this paradigm, subjects are presented with different events. If they detect an event that is unexpected (i.e. it is in violation of physical or psychological principles), they look longer at it than at a control event. Thus, duration of looking can be used as a measure of expectation (Baillargeon, 1995).

In the study by Kersken et al., monkeys faced two screens and an experimenter sitting at the centre. As the screens were in a closed position at the beginning of each trial, the monkeys were not informed about the existence of an object situated behind one of the screens. The monkeys were presented with a grasping action directed to behind one of the screens, such that its final part was occluded by the chosen screens. All that was visible to the monkeys was the two screens, the initial hand action and the direction of grasping; but the object, its location, and the final part of the hand action were occluded and had to be inferred from the available information. In this experiment, after being shown a grasping hand disappearing behind one of the two screens, the screen was opened and the monkeys were confronted with one of two outcomes. In one, upon opening the screens, the object appeared in the corresponding location to the action direction. In the other, the hand had reached to the screen behind which there was no object. The six monkeys tested looked longer as a group when no object was found behind the screen corresponding to the action direction, which indicated a violation of their expectation. This was interpreted as evidence that the monkeys might have inferred the existence of an object from the grasping action alone. In a control condition, the grasping action was replaced by a hand-flop action, as in Wood et al's experiment

(Wood et al., 2007), and the monkeys' looking behaviour suggested that they did not expect the object to appear at the corresponding location of the action.

In this paradigm, the subjects were never informed of the existence of the object, but only watched the intentional actions. Nonetheless, they predicted the presence of an object as the target of the action. From the results we can conclude that monkeys might be able to imagine objects when they detect an intentional action (e.g., grasping), but not a non-intentional action (e.g., flop), without having known the object before, or without the objects actually existing. Thus, the subjects may have perceived the intentionality of the grasping action in the sense of directedness to an object, with the object being completely 'intentional', not previously anchored in reality.

In this new paradigm, the subjects are presented with a situation in which no object necessarily exists. This is different from Woodward's (1998, 1999) paradigm where both the action and the objects are presented to the subjects in advance. It is also different from Umilta's (2001) experiment where the monkey was informed of the presence or absence of the object, and therefore there was no need to infer the object's existence. Woodward's and Umilta's design may reveal the coding of an intentional relationship between existing agents and objects, while this new paradigm may indicate the ability to infer or imagine an object from the intentional relationship, which is crucial to the Brentanian intentionality understanding hypothesis. In neither the gaze following experiments (Moll & Tomasello, 2004) nor the object task ones (e.g., Bräuer et al., 2006) did the researchers find direct evidence of the expectation of an object upon perceiving intentional actions. In comparison, the current paradigm directly addressed this issue.

However, the Kersken et al's experiment had a number of limitations. Firstly, although they found some positive evidence, their study only examined 6 capuchin monkeys, which is a limited sample size for a behavioural study. In addition, they also tested 4 Diana monkeys (*Cercopithecus diana*), but the result was not significant and indicated no support for the inference of the object from intentional actions.

The small sample size and inconsistent results between species questioned the robustness of the evidence from the capuchin monkeys.

Secondly, when the screens were opened the monkeys were able to observe that behind the screen, either the hand grasping the object in the congruent conditions, or the hand in grasping position but without an object in the incongruent one. The authors interpreted that the longer looking time at the incongruent condition indicated that the monkeys' expectation was violated since the grasping action was not directed to an object. However, an alternative interpretation could also plausibly explain the results: the grasping action at an empty void looked unusual and strange. As a comparison, for the flop action, the hand rested on the platform in a relaxed and resting way; however, the grasping hand shape was still in tense and was not in a comfortable resting appearance. Therefore, the monkey might have regarded the grasping action as bizarre and wanted to observe it more carefully.

Thirdly, in Kersken et al's experiment, when the experimenter conducted the action, her gaze always accompanied the action. Thus the hand action and the gazing cue were a combined signal and it was unclear whether it was from the hand action or the gazing cue that the expectation was generated.

Despite of these limitations, this study provided some positive evidence for the assumption that an object is always expected when intentionality is perceived, even when none is immediately observable, which, as discussed above, is a key point in testing whether intentionality is perceived in the Brentanian sense.

1.3 Aim of the current study

From the current literature, we reach several conclusions. Firstly, there seems to be very good evidence that infants and non-human primates code intentional relations between objects and agents. As we have demonstrated above, infants can relate various actions, e.g. grasping, gazing, pointing, to their specific goal-object, and non-human primates have a similar ability for some actions such as grasping.

Secondly, some evidence suggests that when coding such relations, primates may not necessarily need to observe the action and the object simultaneously. When the object was shown beforehand, they may be able to 'remember' or represent it from the intentional actions to generate their representation of the intentional relation. It might be like the gestalt perception where a partially observed configuration would automatically reveal the full configuration.

In addition, when the object was not shown beforehand, primates and infants seem to be able to infer about the existence of the object from others' gaze. It suggests the possibility that the intentional relation can be coded from perceiving only the action, when no direct evidence of the potential object exists. However, gazing is not fully representative for intentional actions in general and there is no convincing evidence, other than the pilot study by Kersken et al, that investigates other intentional actions regarding the inference of the potential object.

The aim of the current thesis is to build upon Kersken et al's experiment and try to address the issue of understanding intentionality in the Brentanian sense, especially regarding the inference of object. We propose a hypothesis based on the reviewed evidence about human infants and non-human primates: **intentionality is coded as a relation between intentional agents and objects, and the observation of certain actions will trigger the perception of intentionality.** Thus, when perceiving an action as intentional, the observer automatically infers a potential object, even in the absence of any independent evidence about the existence of the object. This ability may be a fundamental step for ToM development.

A way of testing whether behaviour is understood as intentional is to test if an observer expects an object when they perceive an intentional action. So the first two studies (chapter 2 and 3) aim to establish whether the presence of an unseen object can be inferred from an intentional action and whether this is only the case for actions with intentions, but not for actions without intentions. We will investigate whether the subjects will infer the existence of a goal object from a grasping action by a violation of expectation paradigm. We will examine two species, a non-human primate species, capuchin monkeys, and humans, in order to compare their

performance and investigate the possible evolutionary origin of this ability. Another two studies (chapter 3 and 4) will aim to investigate capuchin monkey and human adult understanding of intentionality. These two studies will not be focused on the inference of the object, but on coding the agent and the object together as a whole configuration of intentionality.

Chapter 2. Capuchin Monkey's inference of objects from object-directed actions

Summary

The ability to understand intentionality is potentially important for the development of theory of mind. Ample evidence has suggested that human infants can code intentionality by relating an intentional action and an object together (e.g., Woodward, 2009). But evidence in non-human primates is sparse (Kano & Call, 2014). In addition, it remains unclear whether non-human primates can infer the potential object from object-directed actions, if the object has not been shown beforehand. To investigate these issues, we used an adaptation of the violation of expectation (VoE) paradigm to test if Capuchin monkeys possess such abilities.

We presented 20 Capuchin monkeys videos of an object-directed (grasping), and a non-object-directed action (flop) action directed behind one of two screens. These were followed by one of three outcomes, there was: no object (Object-absent condition), an object present behind the congruent screen (congruent condition), or an object present behind the incongruent screen (incongruent condition). The monkeys' looking time at the three outcomes was coded as a measurement of their expectation.

The results suggested that the monkeys can differentiate the actions by their object-directed intentionality. They had no specific expectation about the object for the non-object-directed action flop. For grasping, however, the longer looking time at the incongruent than the congruent outcome indicated that the monkeys expected the object-directed action to be directed at an object in a particular location. However, when no object was shown behind the screens, the monkeys did not look longer in the grasping condition. This suggests that their object expectation may be activated only retrospectively after seeing an actual object. The results suggest that Capuchin monkeys can understand intentionality as object-directedness

by coding the goal-object from the object-directed action, but that there are limitations when the objects are not visible from the beginning.

2.1 Introduction

As described in the general introduction, ToM has long been a focal point in developmental and comparative psychology (e.g., Baillargeon, Scott, & He, 2010; Call & Tomasello, 2008; Gallagher & Frith, 2003; Perner & Wimmer, 1983; Premack & Woodruff, 1978; Umiltà et al., 2001). The main research question has gradually shifted from whether infants and animals, especially non-human primates, can understand the meaning of others' behaviours and mental states, to how this understanding is achieved or, in other words, the different mechanisms of ToM (Apperly & Butterfill, 2009; Baillargeon et al., 2010; Baron-Cohen, 1991; Gómez, 2009a; MacLean et al., 2012; Sodian, 2009). We proposed that understanding intentionality, in the sense of directedness to an object, might be important in the development and evolution of ToM (see chapter 1 and also, Aschersleben, Hofer, & Jovanovic, 2008; Umiltà et al., 2001; Wellman et al., 2004; Wellman, Lopez-Duran, LaBounty, & Hamilton, 2008) and therefore the ability to infer about an object from intentional actions is a key point to investigate (see chapter 1 and also Gómez, 2008, 2009).

The first aim of the current experiment is to examine whether a non-human primate species, the Capuchin monkey, can understand the relation between object-directed action and a goal-object. Understanding intentionality by coding the relation between an action and an object develops early in human infancy (Woodward et al., 2002; Woodward & Guajardo, 2003; Woodward, 1998a, 1999). At as young as 6-months, human infants are already able to distinguish object-directed actions from other actions and to understand the goal-directed action together with its goal-object (Woodward, 1998a). Some findings shows that non-human primates can understand intentionality in a similar way too, but empirical work on this issue is very sparse (Kano & Call, 2014).

In addition to perceiving the relation between actions and goals when presented together, we further suggested that the inference of an object from

intentional actions may happen automatically when interpreting object-directed actions whose object is not visible (see chapter 1 and also Gómez, 2008, 2009). Studies about geometrically gaze following (e.g., Amici, Aureli, Visalberghi, & Call, 2009; Bräuer et al., 2005; Moll & Tomasello, 2004) provide some support for this proposal; however, direct empirical evidence, (except for a pilot study, Kersken et al., n.d.), addressing this issue with intentional actions other than gaze is still lacking.

The aim of this experiment is to investigate whether Capuchin monkey can infer an object from an object-directed action when the presence of the object is not shown in advance. This ability would be a fundamental component in understanding intentionality in a Brentanian sense (Gómez, 2008).

We will use the violation of expectation paradigm to examine Capuchin monkey's expectation of an object after watching an object-directed or a non-goal directed action. The design is based on the preliminary study by Kersken et al reviewed in chapter 1.

The violation of expectation paradigm has been successfully used in pre-verbal infants and non-human primates in investigating their understanding of various physical and social scenarios (Anderson et al., 2004; Onishi & Baillargeon, 2005; Santos & Hauser, 1999; Woodward, 1999; Xu & Carey, 1996). In this paradigm, the subjects are shown two events, one of which is in conflict with what one would expect (for example, an object that remains standing in mid air when its support is removed). In our experiment, we will present Capuchin monkeys with different T-test events and will use their looking time as an indication of their expectation. After watching an object-directed action directed to one of two hiding spots, if the monkeys infer a potential object from the action, they should look for longer if no object appears in the corresponding location. However, if they do not infer an object from the intentional action, their looking times would not be different in the two conditions.

We will present the Capuchin monkeys two different actions, one object-directed (hand in grasping movement) and one non-object-directed (hand in flop movement). We expect the monkeys to only infer the object from the object-

directed action. Grasping is chosen as the object-directed action because it is a very common intentional action in most primate species. In their everyday life, primates use their hands to grasp food, branches, and other objects, and to groom and interact with each other in play and aggressive interactions. Grasping hands are one of the defining adaptations of primates (Gómez, 2004). Thus, we regard this action as the most salient and universal goal-directed action related to objects. We will also present the subjects a non-object-directed action, an action of hand-flop, as the control condition (Wood et al., 2008, 2007; Wood & Hauser, 2011; Woodward, 1999). By presenting these two actions, we expect to be able to compare the monkeys' expectations of intentional and non-intentional actions, in the sense of object-directed (grasping is always about an object) vs non-object directed (flop is not about an object). This method will allow us to determine whether the subjects can infer the existence of an unseen object from the mere contemplation of an intentional action.

To briefly summarise, in the current chapter we are concerned with the question: can monkeys infer objects from intentional (in the sense of object-directed) actions from observing the intentional actions alone when the object is not shown in advance?

2.2 Method

2.2.1 Subjects

We tested 20 brown Capuchin monkeys (*Cebus apella apella*) at the Living Links to Human Evolution Research Centre at Edinburgh Zoo (see Table 2 - 1 for more information). Eleven of them were included in the final data analysis; the other nine were excluded for various reasons. Six monkeys participated in only one or two sessions because they did not come back to finish the all the conditions of the experiment. One, Carlos, stayed in the cubicle during the test but did not engage with the current experiment; the last two, Flojo and Sylvie, did not manage to attend to at least two trials in each condition.

**Table 2 - 1 Monkey's profile and participation in the experiment.
(Monkey's ages were when the experiments finished. The first 11
monkeys were included in the reported results and statistical analyses)**

Monkey	Age (Yr:m)	Gender	Finished sessions	Included in stats
Chico	3:8	M	9	Y
Ruben	2:7	M	9	Y
Junon	12:5	F	9	Y
Anita	15:2	F	9	Y
Inti	3:6	M	9	Y
Figo	6:7	M	9	Y
Diego	10:5	M	9	Y
Diablo	41:?	M	8 (Dead after 8th)	Y
Meako	4:9	M	9	Y
Ximo	2:9	M	9	Y
Torris	1:11	M	9	Y
Flojo	1:7	M	8 (did not want to be separated)	N, failed to finish at least 2 trials in each condition
Sylvie	9:5	F	9	N, failed to finish at least 2 trials in each condition
Carlos	6:6	M	7	N, not attending to the current T-test
Rufo	3:3	M	6	N, Too anxious in cubicles
Toka	9:1	M	4	N, Did not come back to be separated
Alba	1:4	F	1	N, Did not come back
Luna	1:8	F	1	N, Too anxious to be separated
Lindo	1:5	F	2	N, Did not come back
Lana	17:3	F	1	N, Did not come back

The Capuchins monkeys were from two social groups, East group and West group, housed in two separate but similar enclosures, each with an indoor section, 32.5m²*6m, and an outdoor section, approximately 900m² (Griffey, 2011; Macdonald & Whiten, 2011). The monkeys were free to move from one section to the other through several entrances. Apart from these entrances, monkeys could

also enter or exit their enclosures via two rows of four transparent (Perspex) testing cubicles (each 0.5m*0.5m*0.5m), which were along the side of a research room beside their indoor section. Figure 2 - 1 indicates the enclosures and the research room (Griffey, 2011; Macdonald & Whiten, 2011). In addition, the Capuchins were housed in mixed species group with squirrel monkeys, which had their own indoor section and shared the outdoor section with the Capuchins.



A



B

Figure 2 - 1 The housing conditions and research environment (A. Living links from the air. Each inside enclosure includes an indoor section for Capuchins, a research room and an indoor section for squirrel monkeys. Photo: Stephen Evans; taken from Macdonald & Whiten, 2011; B. The east research room and testing cubicles. Photo: Mark Bowler; Taken from Griffey, 2011)

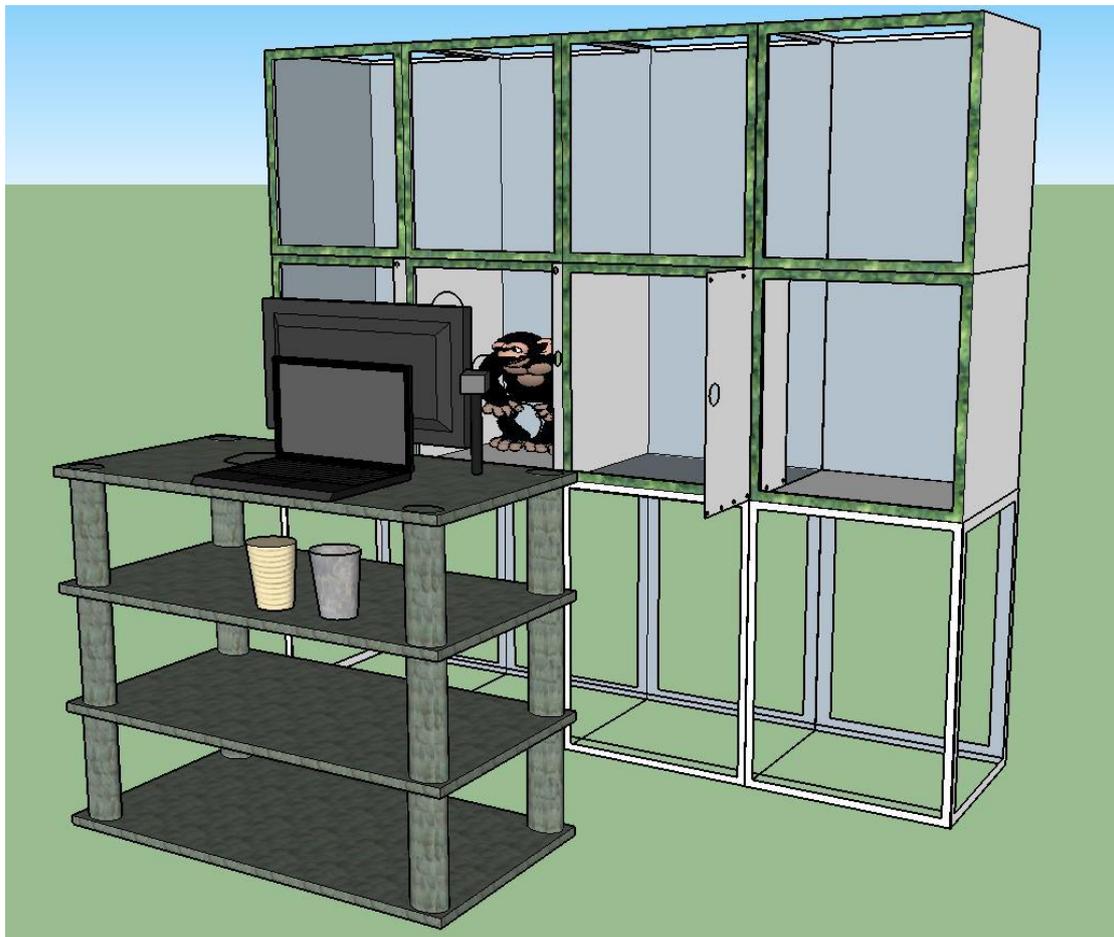
Testing was conducted on each group once a day for two hours between 11:00 and 16:00. Monkeys were fed three times daily with fresh fruits and vegetables and water was available *ad libitum*. The experiment was conducted in the research room where the cubicles were located (see Figure 2 - 1).

The authorization for this experiment was provided by the Living Links Centre at Edinburgh zoo and the ethical permission to conduct experiments with the animals has been obtained from the Ethical Committee of University of St Andrews (see Appendix I).

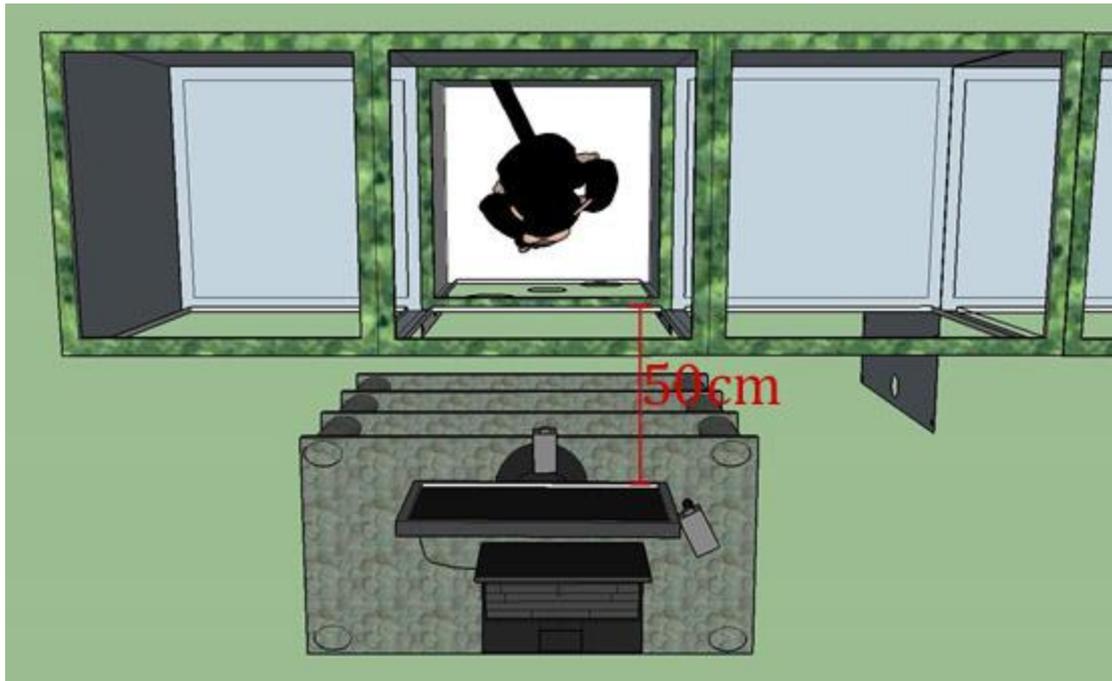
2.2.2 Apparatus

The monkeys were separated and tested individually in the cubicles (see Figure 2 - 2). They have been habituated to stay in the cubicles for research sessions by the Living Links Centre before the current experiment. The cubicles can be closed by

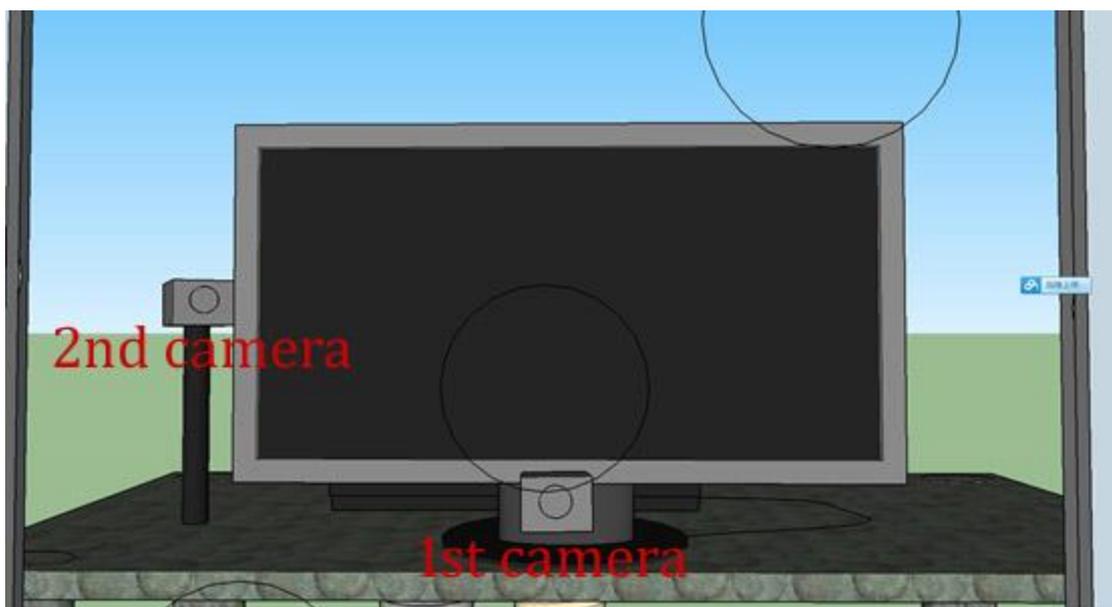
slides and the monkeys are allowed to go back to their group if they put their hands on the slides. A 17-inch computer monitor was located on a platform at equal height to the base of the cubicles and 0.5m directly in front of the front panel. Display of stimuli was controlled by a computer (Lenovo ThinkPad X220i) behind the monitor. Subjects' visual behaviour was recorded by two digital cameras (Panasonic SDR-S26), one placed at the centre directly below the monitor and the other at either the right or left side of the monitor. Monkeys' looking time data were extracted from videos and were analysed frame by frame using Coral VideoStudio Pro X4.



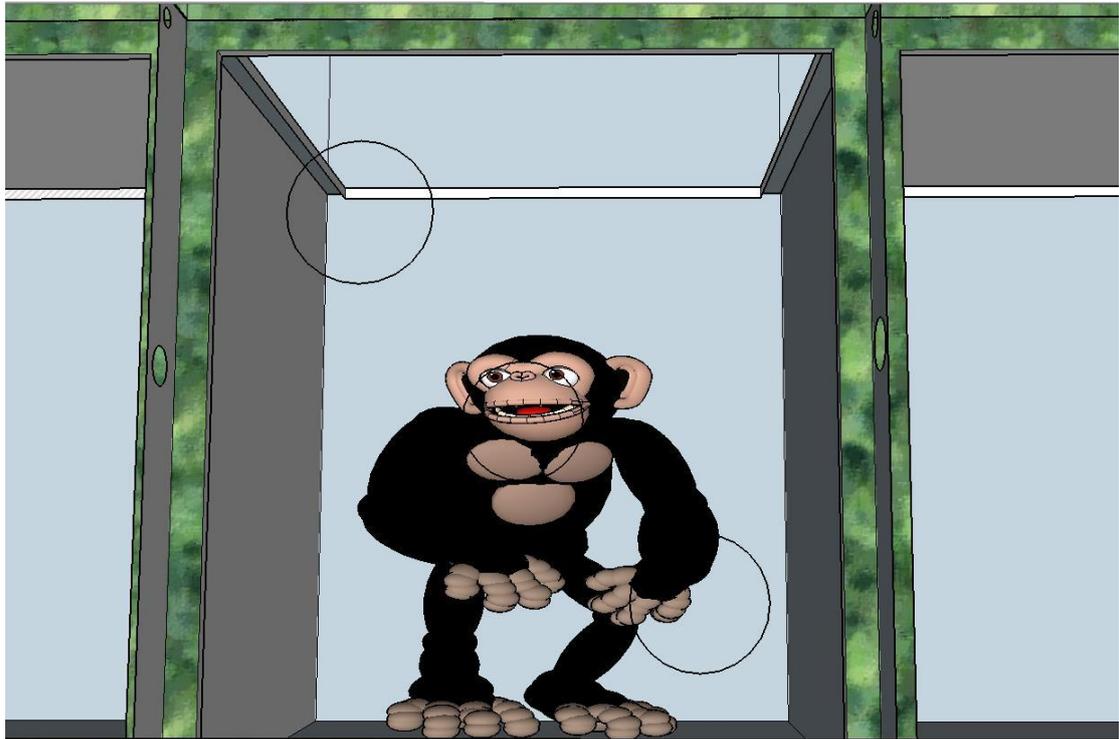
- A. The experimental setting. A platform with the monitor stands in front of one of the middle cubicles of the bottom layer. The monitor is controlled by the laptop behind it. A monkey is separated from its group and locked in the cubicle for the experiment session.



B. Top view. The monitor is 50 centimetres away from the front panel of the cubicle.



C. The monkey's view. The videos are displayed on the monitor. The 1st camera is located at the centre bottom of the monitor and the second camera at the middle of either the left or right side of the monitor. The experimenter is hidden behind the monitor during each trial.



C. The 1st camera's view. The 1st camera records the monkey's behaviour from the front panel.



D. The 2nd camera's view. The second camera records the monkey's behaviour in supplement to the 1st camera.

Figure 2 - 2 The experimental settings

2.2.3 Stimuli

We created an introductory video clip and 12 different experimental clips for the monkeys to watch. All the videos were recorded in a quiet room in the School of Psychology and Neuroscience, at the University of St Andrews and were edited using Window Movie Maker.

The videos showed two hiding screens (which could be opened simultaneously), an agent, who directed an action towards behind one screen, a curtain (which could occlude the agent's face to eliminate gazing information) and three different outcomes regarding an object's appearance.

The introductory video introduced the general settings of the experimental videos (see Figure 2-3). In this video, the agent sat at the centre with her hands resting at the centre of two closed hiding-screens. Then the agent looked left and right behind the screens, indicating that she could see and she knew what lay behind the screens. After that, the curtain rolled down until just above the two hiding-screens, occluding the agent's face so that the gazing cue would not affect the results. The introductory clip lasted for 13 seconds.



Figure 2 - 3 Frames from the introductory video



Phase 1



Phase 2



Phase 3



Phase 4



Figure 2 - 4 frames from the experimental clips. The left and right showed the flop congruent and grasping incongruent condition respectively.

All experimental clips had four phases (see Figure 2-4). In the first phase, the screens were closed and the agent directed an action from the centre to behind one screen using the cross-lateral hand (so that the grasping action looks more natural) until the whole hand was occluded. This phase was crucial for monkeys to observe the action. In the second phase, the screens opened simultaneously and the agent drew the hand back to the centre. In the third phase, the screens remained open and the hands stayed resting at the centre. Monkeys could see the outcome (presence and location of object) in these two phases and their looking time during these two phases was used to indicate their expectation. In the last phase (after stopping the looking time recording), the agent picked up the object with both her hands, stood up, and left the scene; if the object was absent, the agent simply stood up and left the scene. The fourth phase was designed to inform the subject that the object was moved away and there would not necessarily be an object in the next trial. The four phases lasted for 4s, 2s, 8s and 4s (3s if the object was absent), respectively. And therefore, the whole video clips of the congruent and incongruent conditions lasted for 18 seconds and the videos of Object-absent conditions lasted for 17 seconds.

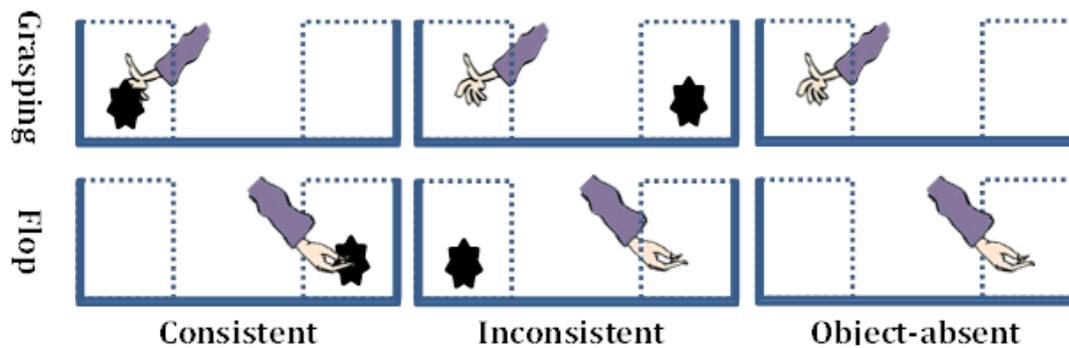


Figure 2 - 5 The 6 experiment conditions (2 Action types X 3 Outcomes)

In the experimental clips, there were two types of actions (i.e., Grasping, the object-directed action, and Flop, the non-object-directed action). Each action was directed towards one of two locations (right/left) which were paired with three different outcomes regarding the position of the object (i.e., Object-absent, Congruent, Incongruent). The combinations resulted in 12 different video clips. The

two crucial factors were the Actions and the Outcomes, leading to 6 different conditions (see Figure 2-5).

If the monkeys can infer the object from the object-directed action, then they would look for a shorter time in the Congruent conditions than the Incongruent and the Object-absent condition. Note that if the monkeys anticipated the presence of a target object for the grasping action, they should be 'surprised' when none appeared at all and so look longer. If this is only the case for the object-directed action but not for the non-object-directed action, the looking times to the three outcomes would not be different for the non-object-directed action.

In the Grasping condition, the agent raised a hand, formed the grasping gesture and moved towards behind a screen (see Figure 2 – 4). Then the hand stopped behind the screen (therefore invisible to the monkey) grasping the object (or mimicking grasping if there was no object. In the Flop condition, the agent raised and rotated the hand, formed the flop gesture and moved the hand behind a screen. The hand (now invisible to the monkey) stopped resting on the platform (beside the object if there was one) completely hidden behind the screen. When the screens opened simultaneously, the hand returned to the centre and rested at its initial position.

In the congruent outcome, the object would appear behind the screen that the action was directed to. In the in-congruent outcome, the object was placed at the opposite position. In the Object-absent outcome, no object was present.

Note that in the second phase, when the hiding screen opened, the hand was returning to the centre. The reason for doing this was to overcome some potential confounding factors from the Kersken study (Kersken et al., n.d.). In the Kersken study (see chapter one), after the action, the hand did not return but stayed at the location behind the occlude, either grasping the object or lying on the platform in grasping position but without an object (incongruent condition), or, in the flop condition, the hand laid besides the object or just on the platform (incongruent condition).

In addition, for the flop action, which was designed as an accidental hand drop, the hand was resting on the platform in a relaxed way; however, the grasping hand shape was still in its grasping shape. Therefore, the monkey might have regarded the stationary grasping position without an object as bizarre on its own and observed it for longer but without having predicted the presence of an object in advance, just as a reaction to an unusual configuration. We therefore decided to make the hand return to its original location as a way of ensuring that any differential reaction to the experimental conditions was to the expectation generated by the monkeys, and not to the nature of the existing configuration.

One potential further advantage of moving the hand back to the centre is that the monkeys' attention may be captured by the returning hand, and therefore return to the centre, thereby equalizing the attentional scanning necessary to find the object in both congruent and incongruent conditions.

2.2.4 Procedure

Each subject was presented with 9 test sessions on different days. Each session contained an introductory trial followed by six experimental trials, including the two action conditions paired with the three outcomes in a balanced order. The direction of action was randomized within sessions and balanced between sessions. The subjects were rewarded three raisins between trials to keep them motivated to stay in the cubicles regardless of their performance in the task. After the experimental trials, the subjects were rewarded a small handful of raisins for participating. At the end of each session, a peanut was put at the top centre, the bottom centre, the left centre and the right centre of the monitor to calibrate the monkey's looking direction in the videotape. While coding for the monkeys' looking time, these four points were used as reference of the monkeys' looking direction, i.e., if the monkey was looking no wider than the four points, the duration of look was included in the looking time.

2.2.5 Data analysis

All looking behaviours were coded offline by rewatching the videos. The footage from the 1st camera served as the premier data source and the footage from the 2nd

camera was used as supplementary when a monkey's gaze direction was not clear from the 1st camera. A look was defined as at least 4-frames (1/25 second per frames) of looking within the range of the monitor. Four out of 25 frames (25 frames = 1 second) was chosen because in published looking time experiments with non-human primates, 5 out of 30 frames (30 frames = 1 second) are defined as a look (Marticorena, Ruiz, Mukerji, Goddu, & Santos, 2011; Martin & Santos, 2014). Our camera and video coding software could only divide a second into 25 frames and 4 out of 25 (0.16s) was similar to 5 out of 30 (0.17s). We added the frames of all looks during the 10-second-looking-time-window (which starts from when half of the object was revealed, including the second and third phases) to calculate the monkeys' looking time.

Two periods in each trial were crucial for data coding. The first period was in the first phase from 0.5 seconds, when the hand had formed the shape of the action, until 2-seconds, when half of the hand was occluded by the screen. We coded whether the monkeys looked or not during this period as an indication of their attention to the stimulus, i.e., whether they noticed the action and its direction. If they looked at least once for 4 frames during this period, we recorded the trial as being attended to and included it in further analysis.

The second crucial period was the second and third phases. The screens opened and the outcome was revealed during the 10 seconds of the two consecutive phases, but the scenes were not the same. In the 2-seconds of the second phase, with the opening of the screens, the hand was still moving towards the centre of the platform; in the 8 seconds of the third phase, the hand was resting at the centre and the scene was still. Considering that the two phases were both showing the outcome, we added the two together and analysed the monkeys' looking time and the number of looks during the 10 seconds for the main statistical tests. In addition, we also examined monkeys' looking time in the two seconds of the second phase and the 8 seconds in the third phase separately to determine if the still moving hand during phase 2 may have any effect on the monkeys' attention.

We mainly used 2*3 two-way repeated-measure ANOVAs with Action (grasping/flop) and Outcome (Congruent, Incongruent and Object-absent) as within subject factors to assess the effect of the two factors. But because the Object-absent condition was not directly equal to the other two, we also examined the two factors using 2*2 two-way repeated-measure ANOVAs, excluding all the conditions where the object was not present.

2.3 Results

Firstly, we excluded all trials where the monkey was not paying attention, i.e., did not have at least one 4-frames look, to the initial action. It left us with 350 attended trials from the 567 tested trials (62%) for looking time analyses. In addition, of the 9 sessions conducted, we decided to only include the first 6 sessions in further data analyses considering the drop of interest in the last three sessions (see detailed analysis in appendix II). Unless specified, the statistical analyses in the following sections only included the first 6 sessions.

2.3.1 The effect of action type and outcome type on looking time and number of looks

We examined the monkeys' total looking time and number of looks using a two-way repeated-measure ANOVA with Action (grasping/flop) and Outcome (Congruent, Incongruent and Object-absent) as within subject factors to assess whether the monkeys' looking time was affected by different actions and different outcomes about the goal object.

Looking time in the 10 seconds after object location is revealed

Firstly, we analysed the individuals' mean looking time during the 10-second window (see Table 2-4 and Figure 2-8). Mauchly's test indicated that the assumption of sphericity had not been violated for the main effect of Outcome, ($\chi^2(2) = 4.28, p = .12$), or the interaction between the two factors, ($\chi^2(2) = .32, p = .85$), and therefore, all the reported tests were not adjusted for sphericity and the degrees of freedom required no correction. A significant main effect for Outcome was detected, ($F(2,20) = 3.84, p = .04$), but no significant main effect of Actions was found, ($F(1, 10)$

= 1.44, $p = .26$). More importantly, we found a significant interaction between Action and Outcome, ($F(2, 20) = 4.43, p = .03$), which indicated that the type of action had different effects on monkey's looking time depending on the object's state.

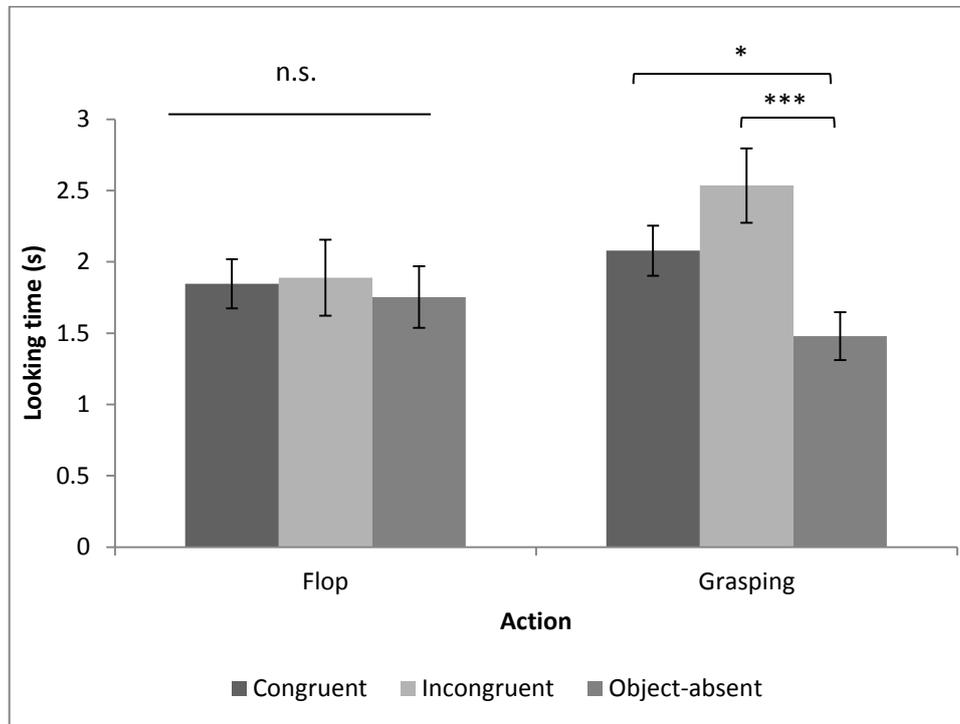


Figure 2 - 6 Monkeys' average looking time in the 10-second window in each condition

To further analyse this interaction, we used a one-way repeated ANOVA to compare the three outcomes for each action. The results revealed a significant difference for Grasping ($F(2,20) = 8.28, p = .002$, Partial *Eta Squared* = .45), but not for Flop ($F(2,20) = .13, p = .88$, Partial *Eta Squared* = .013), Watching a grasping action significantly affected the monkeys' looking time on the three different outcomes of congruency, while flop did not.

Post hoc tests (Turkey test with Bonferroni adjustment) in the grasping condition revealed that the monkeys looked longer at the incongruent ($t(10) = 3.26, p = .009$) and congruent ($t(10) = 3.09, p = .011$) outcomes than the no-object outcome. There was no difference in looking time between the congruent and incongruent conditions ($t(10) = 1.85, p = .283$). In addition, because the Object-absent condition was not directly equal to the other two, where an object was presented, and we expected the participants to look longer at the incongruent

condition than the congruent one, we conducted a planned paired T-test for the congruent and incongruent outcomes of grasping. The T-test revealed difference between the two conditions, ($t(10) = -1.848, p = .047$), one-tailed, suggesting longer looking time at the incongruent outcomes than at the congruent outcome. In comparison, no difference was found for Flop by ANOVA, ($F(2,20) = .13, p = .88$), suggesting that the monkeys had no specific expectation about the object for the action flop.

We also compared the monkeys' looking time between the two action-conditions at each outcome. Paired T-tests showed significant longer looking time at the incongruent outcome if the action was grasping than if it was flop, ($t(10) = -2.26, p = .047$), but showed no significant difference in looking time between the actions at the congruent outcome, ($t(10) = -1.07, p = .31$), or the Object-absent outcome, ($t(10) = 1.20, p = .26$).

Table 2 - 2 Monkeys' mean looking time and number of looks in each condition

		Outcome				
Action			Congruent	Incongruent	Object-absent	Total
Looking time (s)	Grasping	Mean	2.08	2.54	1.48	2.03
		SD	0.59	0.87	0.56	0.80
During full 10 seconds	Flop	Mean	1.85	1.89	1.75	1.83
		SD	0.57	0.88	0.72	0.71
Total		Mean	1.96	2.21	1.62	
		SD	0.58	0.92	0.64	
Number of looks during full 10 seconds	Grasping	Mean	3.01	3.26	2.30	2.86
		SD	0.97	0.88	0.78	0.94
Number of looks during full 10 seconds	Flop	Mean	2.78	2.56	2.58	2.64
		SD	0.55	0.71	0.73	0.66
Total		Mean	2.90	2.91	2.44	
		SD	0.78	0.86	0.75	
Looking time (s) during the	Grasping	Mean	0.52	0.49	0.51	0.50
		SD	0.38	0.23	0.32	0.31
Looking time (s) during the	Flop	Mean	0.40	0.38	0.34	0.37
		SD				

first	2		SD	0.27	0.20	0.14	0.20
		seconds	Total	Mean	0.46	0.43	0.42
			SD	0.32	0.22	0.26	
Looking	(s)	Graspin	Mean	1.56	2.05	0.97	1.53
		g	SD	0.60	0.92	0.46	0.80
during the	8	Flop	Mean	1.44	1.51	1.42	1.46
			SD	0.51	0.82	0.67	0.66
last	seconds	Total	Mean	1.50	1.78	1.19	
			SD	0.54	0.89	0.61	

If we included all the 9 sessions in the analyses, the only significant difference lay between the looking time during the 10-second window after they watched the grasping action and the flop action, when the factor Outcome was combined together, ($F(2,20) = 3.92, p = .04$). The monkeys looked longer at the grasping action in general than the flop action. Although the results from all the 9 sessions were different from that from the first 6 session, the trend was similar. However, we believe that the data about the drop in attention to the actions during the first phase fully justify reliance on the results of the first 6 blocks.

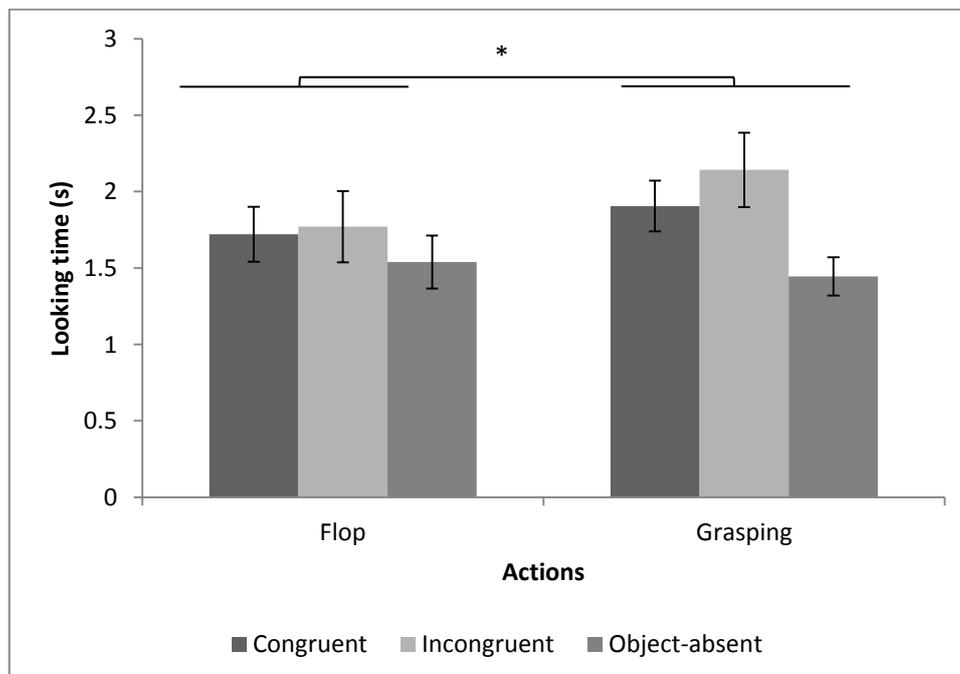


Figure 2 - 7 Monkey's average looking time in the 10-second window in each condition when all the 9 sessions are included

Number of looks during the 10 seconds window

Some of the monkeys' looks were very quick checks and although they did not have very long looking time in total, their number of looks might indicate their attention to the scenario. Although the number of looks was not an independent measurement from looking time, it provides an additional way of exploring the results. We examined individuals' mean number of looks (see Table 2-2 and Figure 2-8) during the 10-second window as an additional test.

Mauchly's test indicated that the assumption of sphericity had not been violated for the main effect of Outcome ($\chi^2(2) = 0.82, p = 0.66$) or the interaction between the two factors ($\chi^2(2) = 2.53, p = 0.28$) and, therefore, all the reported tests were not adjusted for sphericity and the degrees of freedom requires no correction. There was a marginally significant main effect of Outcome ($F(2,20) = 2.95, p = 0.08$) but no significant main effect of Actions ($F(1, 10) = 2.19, p = 0.17$). Again we found a significant interaction between the type of action and the type of outcome ($F(2, 20) = 4.32, p = 0.02$), indicating that the type of actions had different effects on number of looks depending on the outcome of the object's position.

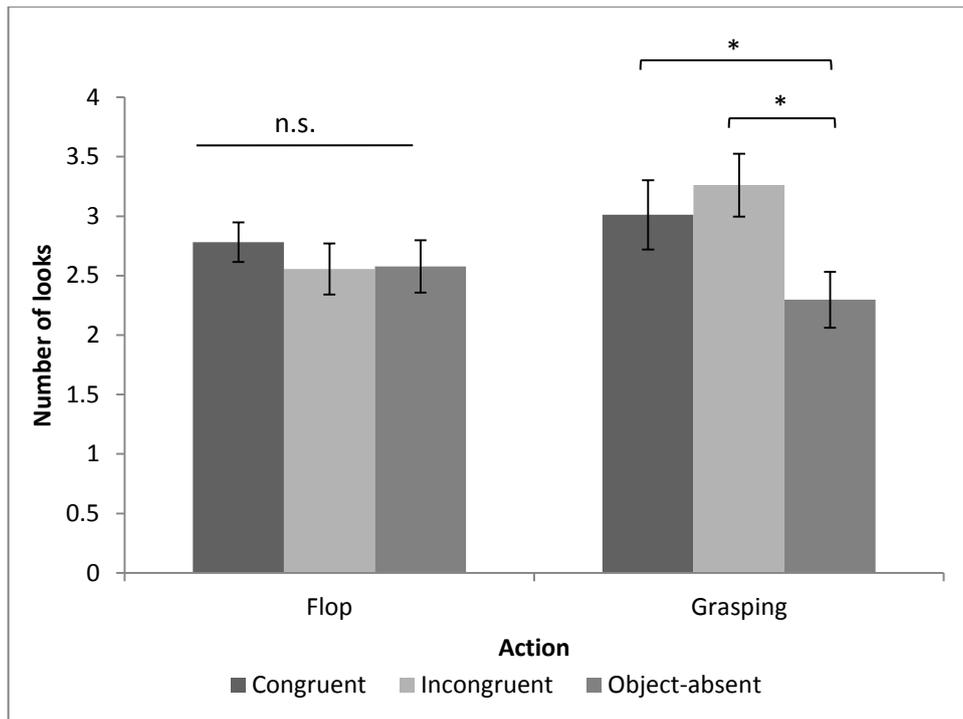


Figure 2 - 8 Monkeys' average number of looks in the 10-second window in each condition

To break down the interaction, we again conducted a one-way repeated ANOVA comparing the three outcomes for each action. The number of looks at different outcomes was significantly different for Grasping, ($F(2,20) = 5.21, p = .015$) but not for Flop, ($F(2,20) = .54, p = .59$) reflecting that watching the grasping action significantly affected monkey's number of looks at the three different outcomes of the object, while watching the flop action did not. Post hoc test (Turkey test with Bonferroni adjustment) for grasping revealed that the monkeys looked more often at the incongruent outcome than the Object-absent one ($t(10) = 2.98, p = .041$) and they tended to look more often at the congruent outcome than the Object-absent one ($t(10) = 2.60, p = .076$), but the difference between the congruent and incongruent conditions was not significant ($t(10) = .75, p = 1.00$). We again conducted a planned paired T-test for congruent and incongruent outcomes of grasping and with this measurement the T-test revealed no significant difference between the two conditions, $t(10) = -.75, p = .47$, suggesting no difference for the number of looks at the incongruent and the congruent outcome. This finding means that the significant difference found in the previous analyses of total amount of

attention, must be due to a longer duration of the fixations during the incongruent outcomes in the grasping condition.

We also compared the monkeys' looking time at each outcome between the two action-conditions. Paired T-tests showed marginally significant longer looking time at the incongruent outcome if the action was grasping than if it was flop, ($t(10) = -2.20, p = .052$), but showed no significant difference of looking time between the actions at the congruent outcome ($t(10) = -1.20, p = .26$), or the Object-absent outcome ($t(10) = 1.44, p = .18$).

These results almost repeated the results using looking time as the measurement, and, therefore, supported the findings. The only difference was that the monkeys did not look more often at the incongruent outcome than the congruent one for grasping. This could be due to individual differences in inspection patterns, i.e., monkeys' average number of looks (regardless of trial type) ranged widely from 2.00 times per trial to 3.61 per trial, perhaps making the variance too big to reach statistical significance.

On the other hand, however, this also suggest that the overall longer looking time found in the previous analyses is due not to an increase in number of fixations, but to their longer duration during the incongruent grasping condition, which was also suggested by Wass and Smith (2014).

Looking time in the first 2-seconds and the last 8 seconds

Thirdly, we used the same analysis to examine the monkeys' looking time during the first 2-seconds of the outcome, when the hand was moving back from the screen to the centre, and the last 8-seconds, when the hands were both resting at the centre and the scene remained still (see Table 2-4 for details). We wanted to know if the two periods were equally crucial in reflecting the monkeys' expectation of the object. For example, phase two might contribute more to the looking time since the moving hand could still show the direction of the action; but phase three might also be the main influence as the hand was resting and no longer a distraction for the monkey who could now observe the outcome scene on its own.

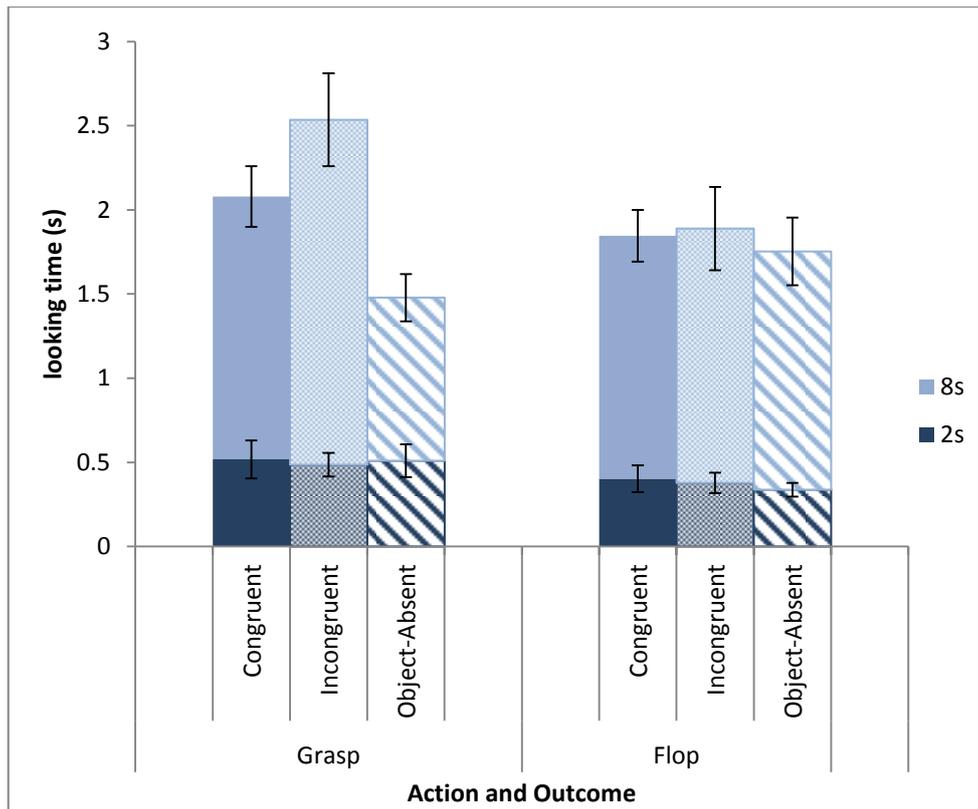


Figure 2 - 9 Monkey's mean looking time in the 2- and 8-second window

For the first 2-seconds, Mauchly's test indicated that the assumption of sphericity had not been violated for the main effect of Outcome ($\chi^2(2) = .60, p = .74$), or the interaction between the two factors ($\chi^2(2) = 4.44, p = .11$), and, therefore, the tests in the two-way repeated measure ANOVA were not adjusted for sphericity. We found a marginally significant main effect for Actions, ($F(1,10) = 4.28, p = .06$), which indicated possible longer looking time at grasping than at flop. But the results revealed no significant effect for the other main factor, Outcome, ($F(2,20) = .12, p = .89$), or for the interaction, ($F(2,20) = .007, p = .91$).

For the last 8-seconds, we first conducted the Mauchly's test. The results indicated that the assumption of sphericity had not been violated for the interaction ($\chi^2(2) = 1.03, p = .60$), and, therefore we did not adjust for sphericity when reporting the interaction; however, the sphericity had been violated for the main effect of Outcome, ($\chi^2(2) = 6.51, p = .039$), so we reported the results of ANOVA using Greenhouse-Geisser corrected degree of freedom. There was a marginally significant main effect for Outcome ($F(1.32,13.2) = 4.02, p = .057$). But no significant

main effect of Actions was found ($F(1, 10) = .18, p = .68$). Similarly to what we found for the whole 10-seconds period, we found a significant interaction between the type of action and the type of outcome ($F(2, 20) = 4.23, p = .029$), indicating that the type of actions had different effects on the monkey's looking time depending on the outcome of the object's position. To break down this interaction, we first performed the one-way repeated ANOVA comparing the three outcomes for each action. The result revealed significant difference for Grasping, ($F(2,20) = 8.04, p = .03$), but not for Flop ($F(2,20) = .067, p = .94$), reflecting that watching the grasping action significantly affected the monkey's looking time at the three different outcomes of object, while flop did not. Post hoc test (Turkey test with Bonferroni adjustment) for grasping revealed that the monkeys looked longer at the incongruent ($t(10) = 3.74, p = .012$) and congruent ($t(10) = 3.00, p = .040$) outcome than the no-object outcome, whereas the difference of looking time between the congruent and incongruent conditions was not significant ($t(10) = 1.57, p = .44$). We also conducted the planned paired T-test for congruent and incongruent outcomes of grasping. The T-test revealed no difference between the two conditions, ($t(10) = -1.57, p = .15$). Again we compared the monkeys' looking time at each outcome between the two action-conditions and no significant difference was found between any pairs, congruent, ($t(10) = -.63, p = .54$), incongruent, ($t(10) = -1.71, p = .12$), Object-absent, ($t(10) = 1.80, p = .10$).

To briefly summarize, for phase 2, the monkeys looked longer at the grasping condition than the flop condition when the hand was drawing back to the centre, but no significant difference between the three outcomes was found. The 8-seconds in phase 3 was more similar to what we found for the whole 10-seconds period: the monkeys' looking pattern of the three outcomes was different for the two actions. They are more interested in, i.e., looked longer at, the condition where an object appears than otherwise, but only when the initial action was grasping. However, when the outcome phase is split into the 2- and 8-seconds phases, none of the phases reached significance in the crucial distinction between congruent and incongruent grasping conditions, which in the 10-second analysis is significant.

2.4 Discussion

In this experiment, we tested if Capuchin monkeys understand the intentionality of actions in the sense of object-directedness by showing an ability to infer the existence of an object from an object-directed action (grasping), without previous information about the object. We compared the expectations of the monkeys (as reflected by their looking time) in response to the outcomes of two different actions: an object-directed action (grasping) and a non-object directed action (flop). We expected the monkeys to show longer looking times when the object directed action was revealed to have been directed at an incongruent or empty location with no object.

2.4.1 Capuchin monkeys can distinguish object-directed and non-object-directed actions

Firstly, the monkeys distinguished the two different actions in terms of object-directedness, i.e. their pattern of looking time at the three outcomes was different for the two actions. For the non-object-directed action (flop) there was no looking time difference between the three outcomes (congruent object, incongruent object, no object). This suggested that the monkeys had no specific expectation about the existence of an object for the non-object-directed action. In comparison, the different looking times at the three outcomes for the object-directed action (grasp) suggested that the monkey did expect an object linked to the grasping action. Therefore, we argue that Capuchin monkeys are capable of detecting the difference between grasping and flop in terms of their intentional relation to an object even when the object is not immediately perceivable and it appears only after the action has ceased (the screens were lowered to reveal the outcome only after the hand had started to return to its original location). The longer looking time at the incongruent outcome than the congruent outcome for the object-directed action indicates “surprise” (violation of expectation) and is consistent with the interpretation that the monkeys expected an object as the goal of the object-directed action.

Our findings therefore provide evidence that Capuchin monkeys understand object-directedness by coding the action as directed to a goal-object. This is in

principle comparable to the findings with pre-verbal infants and findings with other primate species suggesting that certain actions are perceived as directed to a goal object (Cannon & Woodward, 2012; Kano & Call, 2014; Woodward et al., 2002; Woodward & Guajardo, 2003; Woodward, 1999).

However, our findings go further suggesting that the object-directedness of grasping is not only perceived when both object and grasping are visible simultaneously, but also *anticipated*, when only the action is initially visible.

This would be consistent with the findings of Umiltá et al. (2001) suggesting that monkeys can interpret an action as object grasping even if they cannot see the grasping of the object itself. It is enough for the monkeys to have seen the existence of an object behind a screen for their mirror neurons for grasping to fire when they see a hand in grasping position disappearing behind the screen. However, there is no mirror neuron firing for grasping if the monkeys previously saw that there was no object behind the screen.

Our study goes beyond this finding by showing that the monkeys act as if they expect an object even when they have not been previously informed about the presence or absence of the object behind the screen. Our study suggests that monkeys can not only remember if there was an object and connect it to a subsequent grasping action, but they may *anticipate or imagine* the presence of the object from the contemplation of the grasping action alone.

2.4.2 Do Capuchin monkeys infer the presence of the object from the action alone?

There is, however, a problem with the interpretation that Capuchin monkeys understand the existence of the object from the grasping action alone. Thus far, we've mostly discussed about the congruent and incongruent conditions above, but the Object-absent outcome is also crucial for our study. In the condition where no object appeared in the end, the monkeys were not surprised as we had expected. We had predicted that if the monkeys anticipated the presence of a target object for the grasping action, they should be 'surprised' when none appeared at all. However, the no object condition produced no evidence of prolonged looking time: it was

instead the condition that provoked the least attention from the monkeys. This suggests that the Capuchin monkeys did not, in fact, spontaneously infer the presence of a potential object from observing the object-directed action of grasping. This seems to contradict the finding of the contrast between grasp and flop that suggested an expectation that there would be an object congruent with the grasping action, but not the flop.

Note that the contrast between the flop and grasping conditions precludes a simple explanation of the results in terms of the appearance of an object logically requiring more attention than the appearance of no object. The nature of the action performed before the screens were removed did affect the attention patterns of the monkeys.

One possibility is that this result, together with the mirror neuron study that we introduced earlier (Umiltà et al., 2001), suggests that monkeys might only represent the object-directedness of grasping once the presence of the object is known. In the mirror neuron experiment, although the object was occluded and could not be observed when the grasping action was conducted, the rhesus monkeys still understood that the grasping was directed at the object. As the monkeys had known about the presence of the object beforehand, they might remember its presence. However, in our experiment, the object was not displayed before the action was conducted and the Capuchin monkeys had no information about the presence of the object at all.

One possible way of reconciling these apparently contradictory results could be that the monkeys cannot generate an object representation from watching the action only, or that any representation generated is overridden immediately by reality. Instead, the revealing of the object, in the congruent and incongruent condition, may retrospectively evoke the monkeys' memory of the intentionality of grasping. Once they know that an object was present, their memory of the grasping action, still active, makes them assign the object to the intentional direction of the grasping, which led to the longer looking time in the non-matching incongruent outcome. This suggests an intriguing intermediate possible answer to the main

question of the experiment: although the Capuchin monkeys generate something like an empty “object slot” when seeing the grasping action, this slot needs to be filled by reality. It is not enough to create the expectation of an object, and in the absence of an object the action is not interpreted as object-directed and therefore generates no violation of expectation.

If in Umiltá et al.’s (2001) experiment the monkeys needed a memory of the object to understand the subsequent action of grasping as object-directed (as indicated by a physiological measure of neurone firing), in our experiment the monkeys needed the subsequent presence of an object for their memory of the action to be interpreted as object-directed (as indicated by the behavioural measure of looking time). Thus, we could conclude that the monkeys did not infer the presence of an object from the grasping action alone, but did code the difference between intentional and non-intentional actions (in the sense of being object-directed) because they reacted differently to the eventual presence of a congruent versus incongruent object depending upon the nature of the object.

These results are compatible with a Gestalt interpretation of the perception and coding of intentional relations between agents and objects (Gómez, 2008), but one in which both the agent’s action and the object of the action must be perceived, although not necessarily at the same time. The monkeys may put together and code as intentionally related actions and objects that are perceived in succession. We found, however, no evidence that they could infer an as yet unobserved object from the action alone.

2.4.3 Gaze and grasping as intentional actions

In the gazing following experiments where great apes and monkeys tried to find an object of interest when they noticed others’ gaze (Amici et al., 2009; Bräuer et al., 2005; Bugnyar, Stöwe, & Heinrich, 2004; Burkart & Heschl, 2006; Dunphy-Lelii & Wellman, 2004; Moll & Tomasello, 2004; Tomasello et al., 1999), the object was not known beforehand (as in our experiment), yet the subjects seemed to understand intentionality from merely seeing the other’s gaze direction. We suggest that two reasons are likely to have resulted in the different results. In the current

experiment, we eliminated gaze cues to avoid the confounding of cues and therefore the impossibility of disentangling them. It is possible that non-human-primates can infer the object of intentionality only from gazing information, and maybe from the combination of gaze and action, but not from hand action alone. Indeed, the natural behaviour of primates, human beings included, would typically consist of gaze and action cues combined together. For example, if a monkey is grooming another one, it will be looking at where the hand is active. This opens up interesting questions about which cues of intentionality are necessary to identify object-directedness.

The different methodologies of the experiments may also play a role in the different results. For example, the type of measurement and experimental environment in our looking time experiment may lead to less salient behaviour than in the gaze following experiments. In the geometric gaze following task, the participants can move in a larger arena and actively go around barriers to trace the gazing direction and, therefore, display more explicit behavioural evidence of expecting an object. In our experiment, the monkeys are confined in a much smaller cubicle, which won't allow such locomotion to show their expectations. In our experiments, some monkeys, e.g., Chico and Ruben, tried to stand up and get down in an apparent attempt to look behind the screens in the initial phase, when the hand was reaching behind the screen. This might indicate that the monkeys had inferred that there might be something behind that screen the actor was reaching to. This was, however, not done by all the monkeys and it would be difficult to tell if they were simply trying to look behind the screen primed by the actor's action without any specific object expectation.

Using looking time as the main measurement may have potential problems too. Looking time (at least, when estimated from videotape recordings) may not be sensitive enough to detect subtle differences. For example, during the 2-seconds of phase two, when the screen was open but the hand was still moving back, the monkeys looked longer at the object-directed than the non-object-directed actions. Specifically, during those 2-seconds in the object absent condition grasp did not generate more attention than flop. But instead of coding the monkeys' looking time and number of looked from the video, if we could record their visual behaviour in a

more subtle way, for example, record the fixations and saccades using eye tracking device to make sure where specifically they are looking at, it might be able to show their initial expectation of an object (maybe what we tentatively called 'an object slot' above), but this was immediately overridden by reality.

Finally, the gaze cue in the gaze following experiment was given by a real person and the screens and experimental scenario consisted of real objects, whereas the stimuli in our experiment were all presented by video, which might be weaker in drawing the monkey participants' attention and therefore result in a weaker experimental effect.

Eye-tracking techniques might help in providing more detailed information than the looking time measurement alone. They would help in recording the looking time more accurately and would help to identify the monkeys' saccades and exact attention of focus during the study. For example, if the monkeys check the empty locations more often in the incongruent condition than the congruent condition, it might indicate that the monkeys remembered the direction of the action. But by our coding of looking time only, it is hard to tell where exactly the monkey was looking at.

2.4.4 From an evolutionary perspective

As mentioned earlier in the first chapter, our experiment was developed from a preliminary study with Capuchin monkeys and Diana monkeys using a violation of expectation design by Kersken et al (Kersken et al., n.d.). But the current experiment was different from their original one, mainly in that 1) we included the additional Object-absent condition in addition to their congruent and incongruent conditions, and 2) we used videos to conduct the experiment rather than human actors acting live onsite. In spite of the last difference, our results from the congruent and incongruent conditions are consistent with their results from the six tested Capuchin monkeys. With these consistent findings, we are able to more confidently conclude that Capuchin monkeys are able to understand intentionality in the sense of object-directedness even when the action and the object are not perceived simultaneously.

However, the results of our additional condition (object absent) indicate that their suggested interpretation (Capuchin monkeys can infer or imagine the existence

of an object from the grasping action alone) may not be correct. The monkeys showed no surprise to the complete absence of the supposedly inferred object, but just surprise to the fact that, if there eventually was an object, its location was incongruent with the direction of the action. This may appear to be a subtle difference, but it could reflect a crucial cognitive difference between humans and non-human primates: the ability to infer imaginary intentional objects versus the ability to code intentionally only real objects and actions, even if they need not be perceived simultaneously.

In sum, our findings reveal a sophisticated intentional ability in Capuchin monkeys: they can code intentionally actions even if they are never perceived in conjunction with their objects. This supports behaviourally the physiological findings of Umlilt et al (2001) with another monkey species, but also expands upon their study by showing that previous knowledge of the object is not necessary to trigger an intentional coding of an action. However, our findings also suggest that Capuchin monkeys may not be able to generate a representation of an intentional object before obtaining direct evidence of its existence.

Although available evidence suggests that the ability to understand intentionality when both action and object are simultaneously visible is shared by several monkey species, great apes and human infants, the ability to code intentionality when objects and actions are presented sequentially, targeted by our paradigm, remains largely unexplored. In their pilot study, Kersken et al also tested Diana monkeys and these did not look differently at the congruent and incongruent conditions, which suggests that they did not interpret the actions intentionally. These different findings for the two species raise the question of the phylogeny of this ability, and the need to test more primate species with our paradigm, both for intentional coding of sequentially presented action and objects, and crucially, our original aim of testing the ability to infer objects from the action alone.

In the next chapter, we will test human adults with a similar experiment to explore if in human adults, using the same test, we find consistent evidence of

inference of the intentional object not dependent upon the direct perception of the actual object.

Chapter 3. Understanding of Intentionality in Adult Humans

Summary

Understanding object-directedness from actions is a key socio-cognitive skill; however, we still know little about how object-directedness is understood. In this chapter, we investigated whether adult humans could infer a potential object from object-directed actions performed without a visible object with the same stimuli used with monkeys in the previous chapter.

In the first experiment, we presented adult humans with a reaction time task (measurement used instead of looking time), where they were required to press one of three keys to indicate the presence or location of an object that they had not seen beforehand. If they could infer the object from an object-directed action, we'd expect their reaction time to be shorter when an object appeared at the congruent location, than at the incongruent location, or when no object was present. The results, however, showed that the 20 tested adults' reaction times in the congruent and the incongruent conditions were not different, indicating that they had formed no expectation about the object's presence or location from the actions. Although the participants' reaction time was always longer when the object was absent, it was not necessarily due to their expectation of the object's presence. A follow up experiment where no action was conducted, showed that participants still responded more slowly to the object's absence. In an additional control condition, when the participants knew about the object's presence and location before watching the actions, the type of action did not affect their final response.

We concluded that adult humans in the current paradigm failed to show an ability to automatically infer the potential object from the object-directed action. But

whether they truly lack this ability, or the experimental conditions failed to register it, requires further investigation.

3.1 Introduction

In the previous chapter, we used looking time (LT) as a measurement to investigate whether Capuchin monkeys can infer potential objects from object-directed actions as opposed to non-object-directed actions.

We had two major findings. Firstly, the monkeys could differentiate the two actions on intentionality: they expected the object-directed action, but not the non-object directed one, to be congruent with the object location. However, they showed no ‘surprise’ when no object appeared behind the screens. From these results, we concluded that the monkeys had the ability to code an intentional relation without the need to perceive action and object at the same time. However, they lacked the capability to infer the presence of an object merely from watching the grasping or flop actions without *previous* or *subsequent* knowledge of the existence of the object. We suggested the interpretation that the presence of the object may trigger a memory of the object-directedness of the grasping action, or that the intentional, object-directed action generated some sort of “object slot” that was not activated unless an object was subsequently perceived.

In the current chapter, we aimed to adapt the method explored with the Capuchin monkeys, to be used with adult humans. We asked adult participants to take part in a series of reaction-time tasks in which they watched object-directed and non-object-directed actions performed towards behind one of two barriers, without knowing what was hidden at each location. The position of an object was then made visible and the participants were required to respond as quickly and as accurately as possible to indicate the position of the object. We took their reaction time as the measurement of their expectation of the potential object, predicting that if participants anticipated the existence and position of an object from the object-directed action, their reaction times would be faster..

In the first experiment, we asked human participants to watch a short video clip similar to the one that we had shown to the Capuchin monkeys, i.e., a video containing an action directed to behind one of two hiding-screens, followed by three outcomes, i.e., congruent, incongruent and Object-absent. The participants needed

to press a matching key to decide if there was an object and which screen it was as quickly as possible when the hiding-screens were opened. In contrast to the monkey experiment, we recorded the participants' reaction time as measurement of the participants' expectation of the existence and the position of the object. Reaction time was used instead of looking time because the latter does not work well with adults, whereas RT has been successfully used as a measurement for implicit mentalistic reasoning generating similar results as looking time in human infants in a variety of tasks (Kovács, Téglás, & Endress, 2010; Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Cohen & German, 2009, 2010). Two factors were crucial in this experiment. The first was the type of action: object-directed or non-object-directed; the second was target-congruency, object position congruent to the direction of the action, object position incongruent to the direction of the action or no object at either location. According to our hypothesis that the agents would expect an object as the target of the object-directed actions, the incongruent and object absent conditions should be unexpected for the participants who watched the grasping action and, therefore, we predict that they would react more slowly in these two conditions than in the grasping-congruent condition. Given that the flop action was not object-directed, the participants should not have any specific expectation about object and, therefore, we predicted that the reaction time after the flop-action video would not be affected by the different reaction targets.

The second experiment was a follow up study to the first experiment. It consisted of three sessions to answer the questions that arose from the first experiment, as will be discussed later. It consisted of fewer repeated trials so that the results would not be largely affected by a practice effect. In addition, it also included a condition where the object would be shown beforehand so that the participants would not need to infer about the presence of the object, but could directly encode the intentional relationship between the actions and the object.

It should be noted that the direction of the actions, even if it was not object directed, might bias the participants' attention to one direction, resulting in faster reaction to the target if the object is at the congruent location. However, we predicted that this effect should be the same for both actions, with the hypothesized

object expectation effect resulting in an additional shortening of the RT for the object-directed action.

In sum, with the conditions in this task, we would be able to investigate (1) whether the participants differentiate object-directed actions and non-object-directed action by expecting the object-directed action, but not the non-object-directed action, to be always congruent with the object location, as the Capuchin monkeys did in our last experiment. And (2), whether the human participants would show evidence of inferring the presence of an object from the object-directed action when no object was shown at all, which the monkeys were not capable of.

3.2 Experiment 1

3.2.1 Method

3.2.1.1 Participants

A total of 20 participants (Female : Male = 12:8; average age = 21.0) were paid £5 each to participate in the current experiment. They were recruited from the participant-pool in the University of St Andrews. All participants were required to have normal or corrected-to-normal vision. None of them had participated in similar experiments beforehand. They were invited to a quiet room in School of Psychology and Neuroscience to participate in the experiment. Ethical approval was granted by the ethical committee of University of St Andrews.

3.2.1.2 Apparatus and general procedure

The experiment was presented on an 18 inch flat-square CRT monitor (IIYAMA, LS902UT, Resolution: 1280*1204, Refresh Rate: 85Hz). The participants sat approximately 50 cm from the monitor, facing the centre of the screen (see Figure 3 - 1). Their response was made on the keyboard and was recorded by E-Prime. The experimental procedure was controlled by E-Prime as well.

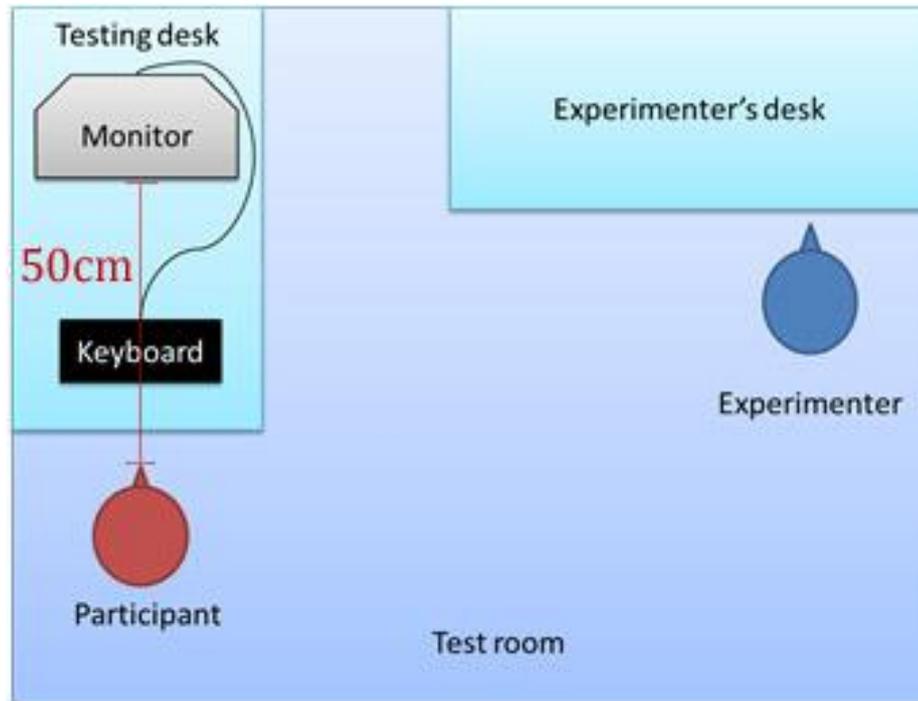


Figure 3 - 1 Experimental environment

At the beginning of the experiment, the experimenter welcomed the participant to the room and read them the instructions for the task. The task was introduced as a reaction-time task but object-directedness per se was not mentioned.

After signing the consent form, the participants first watched an introductory video clip where they were presented with the general scenario of the task. The introductory trial was exactly the same as that shown to the monkeys in the previous study (see Figure 3 – 2).

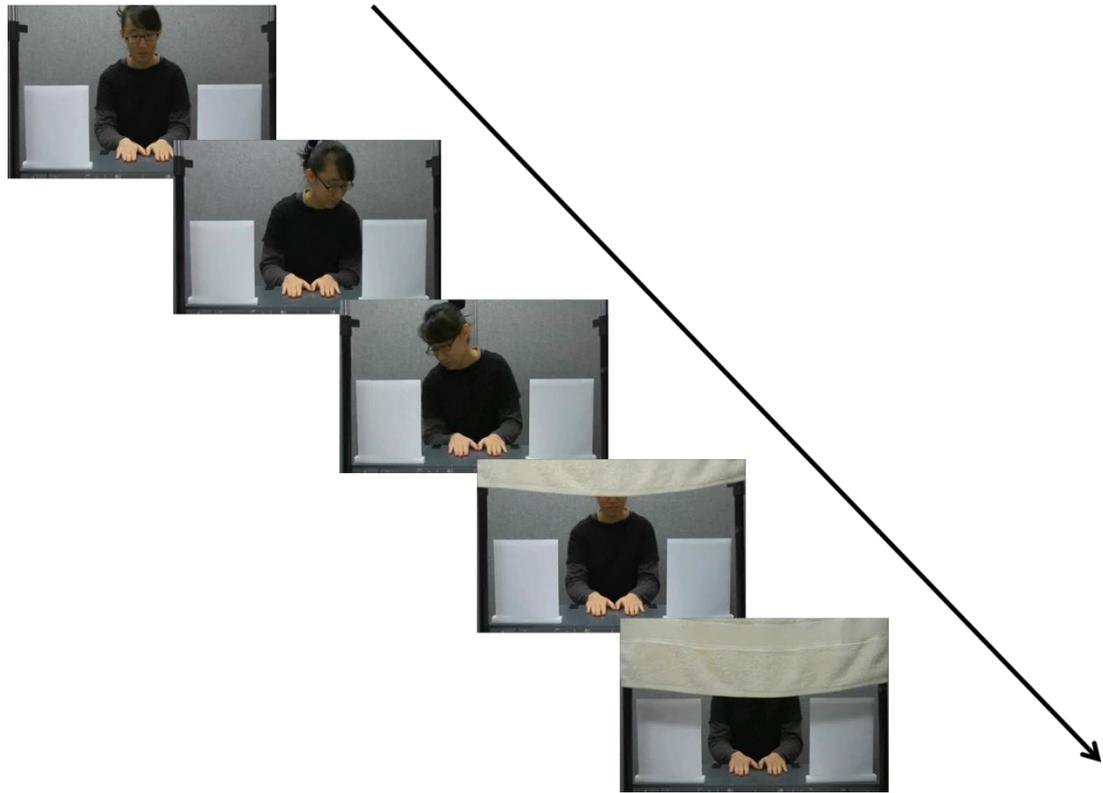


Figure 3 - 2 Procedure of the introductory clip

The participants then watched the experimental video clips where an action was directed towards behind one of two hiding screens (see Figure 3 - 3). When the hiding screens opened simultaneously, they were required to press a key on the keyboard as fast as possible to indicate the presence/location of the object. Depending on the presence/position of the object, the participants should respond as quickly and accurately as possible by pressing the key “left”, “down” or “right” to inform about the object as behind the left screen, absent, or behind the right screen using their index finger, middle finger, or ring finger, respectively (see Figure 3 - 4). The videos were exactly the same as those used in the monkey experiment before the opening of the screens. But there was a slight modification after the screens opened so that it better fitted the reaction time task, as will be discussed in detail later.

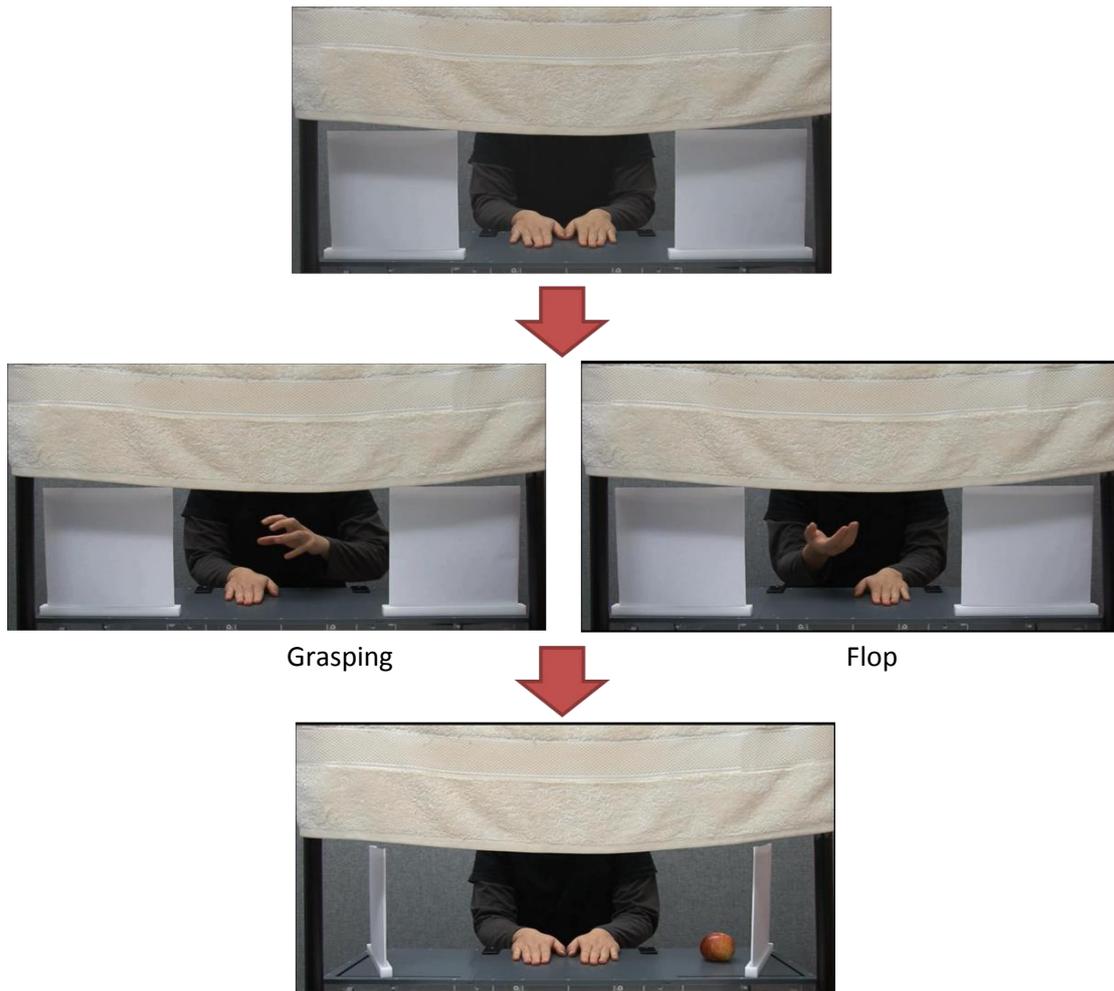


Figure 3 - 3 Procedure in each experimental video clip

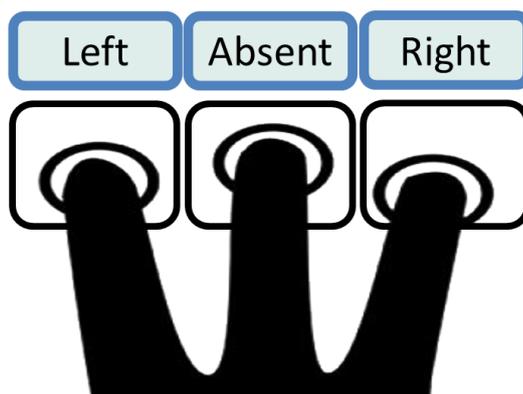


Figure 3 - 4 The response keys

After finishing the reaction-time task on the computer, participants were asked to fill in a brief questionnaire. The questionnaire was to collect the participants' explicit views regarding the task and the two actions. For example, we asked if they noticed the different gestures in the videos and if they tried to ignore the movement of the hands. We also asked if they regarded either action as object-directed, intentional, or whether they felt that it affected their reaction to the object (see Appendix III for the complete questionnaire).

At the end, the experimenter read the debriefing to the participant and described the purpose of the study, explaining that it was about understanding object-directedness. The experimenter remained seated quietly in the same room through the experiment but did not monitor the participant during the experiment.

3.2.1.3 Design

The experiment was a 2*3 with-in subject design (see Figure 3 - 5), the same as that in the monkey study. Again the two factors were Action and Outcome. For the first factor, Action, we again used the object-directed action, grasping, and the non-object-directed action, flop, as in the last chapter. If the participants inferred the goal only from the object-directed action, we would expect a difference in reaction time pattern to the two actions. The control action, flop, also helped to control for confounding factors such as the trajectory of movement and the direction of attention etc.

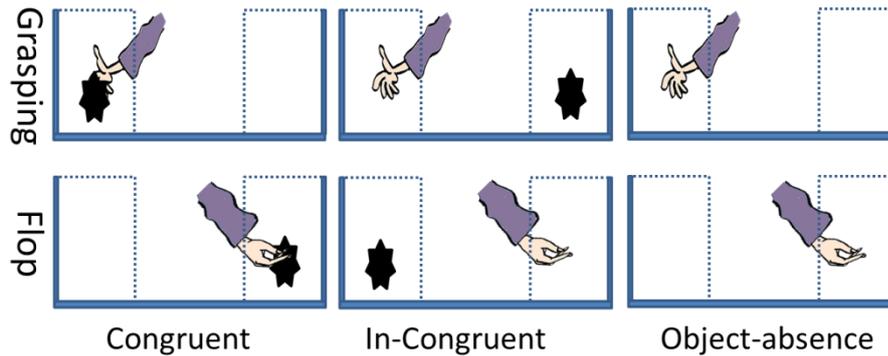


Figure 3 - 5 An illustration of the experimental conditions.

The second factor was Outcome. It included three different conditions: in the congruent condition, an object appeared at the congruent location to where the action had been directed; in the incongruent condition, it appeared at the incongruent location to the action direction; and in the Object-absence condition, no object was present when the two screens opened. If the participants inferred the object from the object-directed action, we would expect them to respond more quickly in the congruent condition than the other two. But since the non-object-directed action was not supposed to activate the participants' expectation of the object, their reaction time was expected to be similar in the three conditions, unless there was attention cueing to one side, in which case there could be some bias of faster reaction time to the congruent condition. But in the latter case, we would expect the bias to appear for both actions and a larger difference between the congruent and incongruent condition would indicate the participants' expectation about the object.

It must be noted that, as in the monkey experiment, when the screens opened, the hand returned to the centre so that when the object was revealed, the position of the hand would not bias the attention of the observer to one side. It also avoided the strange gesture of grasping at nothing or the hand resting by an ungrasped object. As in the monkey experiment, this served to ensure that the participants were responding to the action itself rather than just the final scene.

During the 30-minutes reaction time task, each participant first had a practice block of 12 trials to get familiar with the reaction keys and the experimental scenario.

They then took 4 test blocks of 36-trials each and after each block they were allowed a brief rest if they wanted to. The 36 trials in a block included equiprobable factorial combinations of the actions (grasping or flop), cue direction (left or right) and object position (left, right or object absent). Each condition was repeated twice in a block. Within each block, the trials were presented to participants in a different randomized order. Feedback for correct and incorrect responses was given in the practice block but not in the test blocks.

In addition, three probe questions were randomly presented in between clips in each block, asking which action was presented in the clip before. This was to ensure that the participants paid attention to the difference between the actions as opposed to simply waiting until the screen opens.

3.2.1.4 Stimuli

Reaction time task

The stimuli displayed included an introductory clip and six experimental clips. They were essentially the same videos used in testing the Capuchin monkeys.

In the introductory clip (see Figure 3 - 2), both hiding-screens were closed and an actress sat at the centre with both of her hands resting at the centre of the two screens. After she had looked left and right behind the screens, a curtain rolled down and occluded her face to eliminate the gazing cue. The introductory clip lasted for 13 seconds and was exactly the same as in last chapter.

Figure 3 - 6 illustrates the experimental clips. The experimental clips consisted of two phases: before and after the screens opened. The first phase lasted for 4 seconds before the screen opens, presenting an action directed towards behind one of the two screens. At the beginning of the clip, the actress sat with both her hands resting at the centre of the platform. Then she performed an action, either the grasping or the flop, to behind either screen. The hiding-screens were closed all the time and the curtain occluded her face all through the video. This part was exactly the same as the first phase of the videos in the monkey study. But in the 4th second, when the hiding-screens opened simultaneously to show one of the three outcomes,

the video was slightly different to the monkey study. Instead of returning slowly to the centre, the hand was seen immediately back at centre after the screens opened. This adaptation had to be done due to the use of RT (instead of looking time) as the dependent variable. The participants had to respond as soon as possible after the opening of the screens, and if the hand took as long as 2-seconds to return to the centre, we could not eliminate the confounding factor of the hand being at one side at the time of response. After the opening of the screens, therefore, a still image remained showing the final outcome (congruent, incongruent, or object absent) for 3-seconds. The object in this experiment was an apple, as was used in the monkey study.

A second difference with the monkey experiment was that, at the end of each video, we did not present the participants with the removal of the object, but instead, an instruction screen appeared, asking the participants to press a key to start the next trial. The removal of the object was presented in the monkey experiment to show that the presence of the object was not necessary in the next trials. But we decided to abandon this design, because the adult humans had 12 practise trials before the test and they should be easily able to realize that the presence of the object was not relevant to its present or location in the previous trials. And given that the adults had many more trials (144 test trials) than the monkeys (54 trials if all 9 sessions were included), if all the trials included the removal of the object, the experiment would have reached an excessive duration.

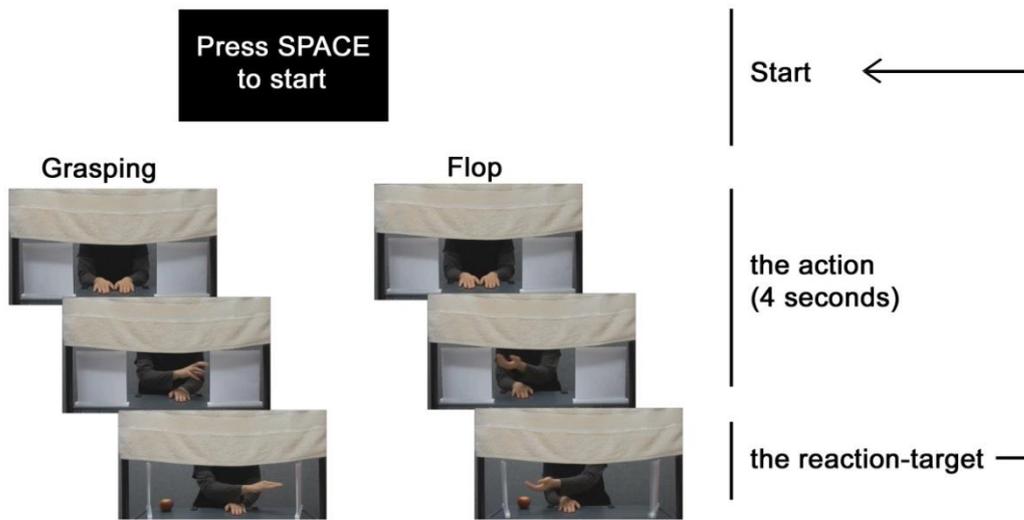


Figure 3 - 6 Experimental procedure

Questionnaire

The questionnaire administered after completion of the tasks included two sections. In both sections, the participants were required to rate statements on a 7-point scale (see Table 3 - 1).

Table 3 - 1 the 6-point scale of statement rating

Score	Description
1	Strongly disagree
2	Disagree
3	Somewhat disagree
4	neutral
5	Somewhat agree
6	Agree
7	Strongly agree

In the first section, the participants were asked to rate five statements about the task (see Table 3 - 2). From these questions, we planned to investigate the participants' general feeling about the task. In the second section, we planned to investigate the participants' explicit view about the two actions. The participants were presented with the video of an action directed to behind one of two hiding screens, followed by 12 statements regarding this action and then the other action

followed by the 12 statements again. The 12 statements were categorized into three categories: 1) if the action was object-directed, 2) if the actress knew about the object, and 3) if they found the action relevant to the object, to examine different aspects of the participants' explicit views about the agent, the action, and the object. The order of the actions presented was counterbalanced between participants. The order of the questions was randomized for each action and participant.

Table 3 - 2 The statements in each section

Section of questions	Aim of the question	Questions
Section 1. Statements about the task	Participants' general feeling about the task	The task is easy.
		I made some mistakes in the task.
		I noticed the different gestures in the videos.
		The video before the screens opens was distractive.
		I tried to ignore the movements in the video and focused on the moment when the screen open.
Section 2. Statements about each action	Object-directedness of the action	This action indicated the location of the object.
		This action had nothing to do with the object.
		This action did not inform about the object.
		The action was aimed to the object.
	The agent's intention about the object	The actress in the video knew if there's an object.
		The actress did not know the position of the object.
		The actress tried to reach the object.
		I tried to figure out if the actress knew about the object.
	The general strategy in the task	This action distracted me from the task.
		I ignored this video when making the judgement.
		This video was helpful for me to succeed in the task.
		This action looked random and irrelevant to the task.

3.2.1.5 Data analysis

To examine the effect of the two actions and the three outcomes on the participants' reaction time, I analysed the data of the correct responses in all four experimental blocks combined, (excluding the practice blocks). We used a 2*3 repeated measure ANOVA with Actions (Flop or Grasping), and Outcome (congruent, incongruent, no object) as the two factors. When Mauchly's test indicated violation of assumption of sphericity, the degrees of freedom were adjusted by the Greenhouse-Geisser approach.

In addition, as in the monkey study, to explore the potential difference between the response to the object's location and its absence we also conducted a 2*2 repeated measure ANOVA with Actions (Flop or Grasping), and Congruency (congruent or incongruent) as the two factors.

3.2.2 Results

3.2.2.1 The Incorrect trials

Before the main test, we first examined the rate of correct responses and the distribution of the incorrect trials in each condition. The participants responded incorrectly in 2.19% of all the test blocks (see Figure 3 – 7 and Table 3 - 3). A Chi-square test showed that the number of incorrect responses was not distributed evenly across conditions ($\chi^2(2) = 15.57, p = .008$). Mistakes were found more often when reporting the presence of an object in a particular location in the object absent conditions than in the congruent or incongruent conditions ($\chi^2(2) = 12.10, p = .002$). The number of mistakes made for the two actions was not significantly different ($\chi^2(1) = 1.92, p = .17$).

We also excluded all the Object-absent conditions to examine if the location of the appearance of the object would affect the participants' error rate differently for the two actions. Chi-square test revealed no significant difference, between the two actions and the two outcomes ($\chi^2(1) = 2.59, p = .11$).

Table 3 - 3 Number of errors in each condition of the testing blocks

Condition	Congruent	Incongruent	Object-absent
Flop	10	7	20
Grasping	4	8	14

Note: 480 trials in total for each condition

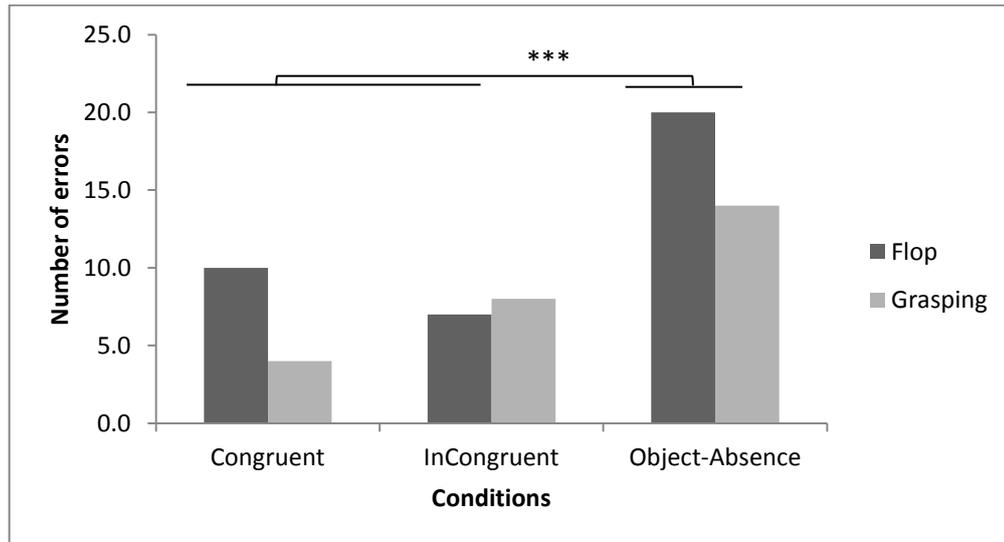


Figure 3 - 7 Number of errors in each condition of the testing blocks

In the following results, we excluded all the incorrect trials when reporting the mean RT for each participant.

3.2.2.2 The effect of Action and Outcome on reaction time

We first examined the participants' RT to each condition in the 4 testing blocks. . Contrary to our expectations, the results of the 2*3 ANOVA (See Table 3 - 4 and Figure 3 - 8) showed no interaction effects, $F(1.2,23.6) = .23$, $p = .79$. However, both main effects were significant. The reaction time was significantly higher for flop than grasping although the difference was only a few milliseconds ($F(1, 19) = 5.78$, $p = .027$). The main effect was significant for Outcome ($F(1.04,19.79) = 35.75$, $p < .001$), too. Paired T-test (with Bonferroni adjustment) showed that the subject responded significantly more quickly at both the congruent ($t(39) = 7.98$, $p < .001$), and the incongruent ($t(39) = 8.20$, $p < .001$), outcomes than in the Object-absent conditions.

But no difference was revealed between the congruent and incongruent conditions ($t(39) = .020, p = 1.00$).

We also conducted a planned 2*2 2-way ANOVA excluding the object-absence condition and assessed the effect of Action and Congruency. We found a significant main effect of Action ($F(1,19) = 4.85, p = .040$), where again reaction to Flop was slower than RT to Grasping. However, no main effect of Congruency was found ($F(1,19) = .00, p = .99$), indicating that the congruent and incongruent conditions did not have a different effect on the participants' RT. Again, there was no interaction ($F(1,19) = .080, p = .78$).

Table 3 - 4 Reaction Time in each condition (ms) of the testing blocks

		Outcomes		
		Congruent	Incongruent	Object-absent
Flop	Mean	401.6	400.9	496.2
	SD	56.5	56.8	81.6
Grasping	Mean	392.9	393.6	482.7
	SD	66.4	62.0	94.0

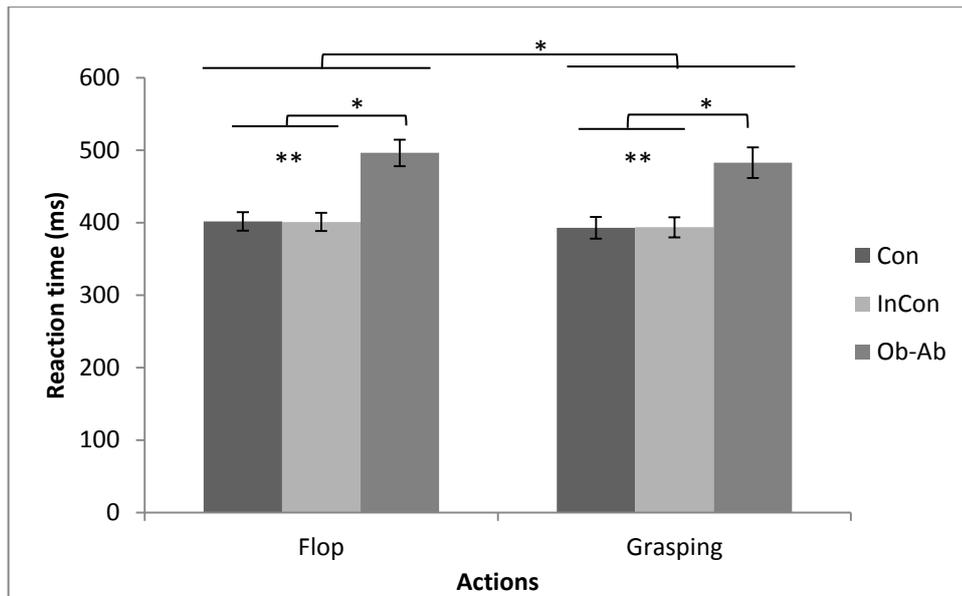
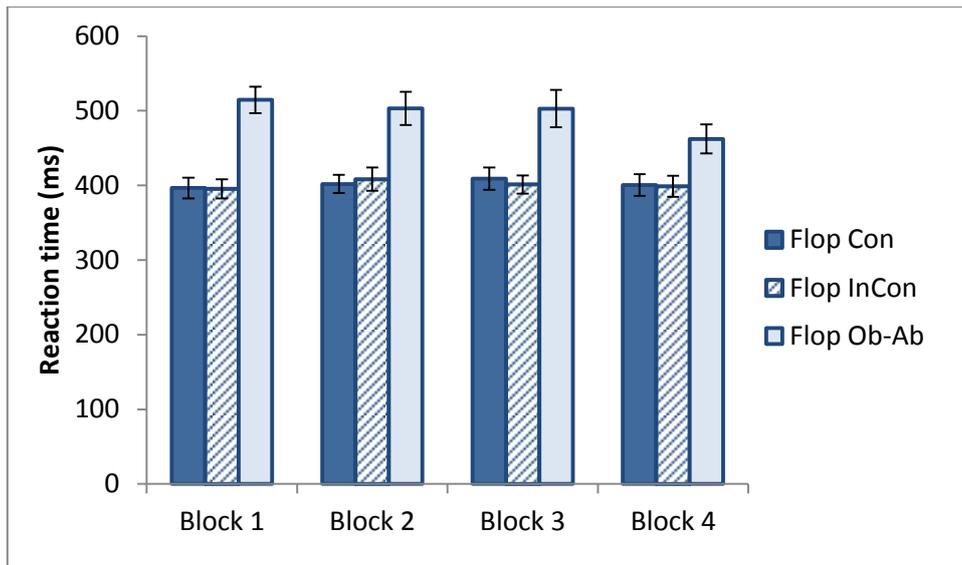


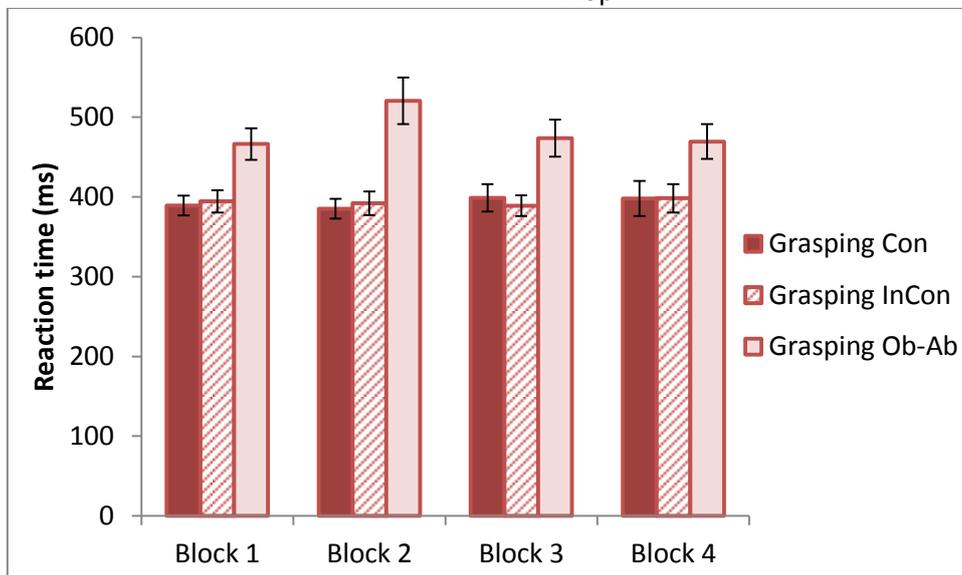
Figure 3 - 8 Mean (SE) reaction time (ms) in each condition (Note: Con, InCon and Ob-Ab represents Congruent, Incongruent, Object-absent, respectively.)

3.2.2.3 The effect of Action, Outcome and Block order on reaction time

There was the possibility that the participants learned to ignore the actions during the repeated conditions of the test thereby suppressing any expectation of the object. Therefore, we examined if the block order (one, two three, four) had any effect or interaction with the two relevant factors, Action and Outcome. We used a three-way repeated measure ANOVA, adding in block order as the third factor, to investigate this issue. Figure 3 - 9 illustrated mean RT in different conditions in each block. Again we found significant main effects of Outcomes ($F(1.0,19.8) = 35.48, p < .001$), and Actions ($F(1,19) = 5.92, p = .025$), but no interaction for these two factors ($F(1.24, 23.62) = .18, p = .84$). But notably, although the results revealed no main effect of Block order ($F(3, 57) = 1.27, p = .30$), it showed a significant two-way interaction between Block and Outcome ($F(6, 114) = 3.62, p = .003$) and a significant three-way interaction of the three factors ($F(3.5, 66.4) = 2.94, p = .032$).



A. Flop



B. Grasping

Figure 3 - 9 Mean RT (SE) at different Outcomes across the testing blocks for each action (A. Mean RT after watching flop, B. Mean RT after watching Grasping)

To look further into how the reaction time varied over the testing blocks, we examined the effect of Action and Outcome in each block separately by with a 2*3 ANOVA (Action * Outcome). Across all four blocks, the reaction time was significantly longer in the Object-absent conditions than the congruent or incongruent conditions, as shown in Figure 3 - 10 and Table 3 - 5. The interaction of Action and Outcomes,

however, was different across the 4 blocks. It was only significant in the first block (see Table 3 - 5).

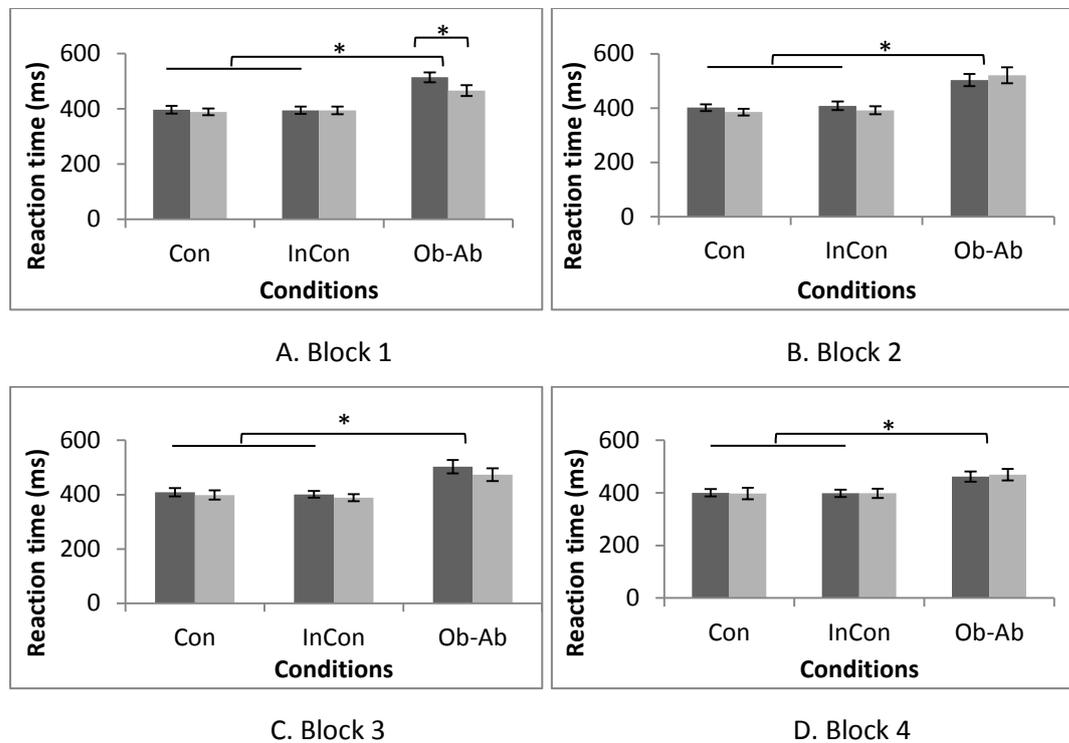


Figure 3 - 10 Mean RT in each condition in each block. (Dar and light grey represent flop and grasping respectively).

Table 3 - 5 ANOVA test results for each block.

Two-way ANOVA						
Block No.	Action		Outcome		Action*Outcome	
	F	Sig.	F	Sig.	F	Sig.
1	4.6	0.046*	46.60	< .001***	5.07	0.011*
2	0.4	0.53	29.21	< .001***	2.30	0.14
3	11.0	0.004**	22.20	< .001***	0.79	0.43
4	0.059	0.81	15.22	.001**	0.18	0.74

*p < .05, **p < .01, ***p < .001

We examined the first block more closely, in order to investigate the difference between conditions. We first compared the RT of the two actions for each outcome. When the outcome was Object-absent, the tests revealed significantly longer RT for flop than grasping ($t(19) = 2.58, p = 0.019$), but no difference was found between the

two actions when the outcome was congruent ($t(19) = 1.07, p = 0.39$) or incongruent ($t(19) = 0.08, p = 0.93$). We also looked at the RT between the outcomes for each action, the reaction time was significantly longer in the Object-absent condition than the other two conditions for both actions (see Table 3-6 for t and p values). But, in contrast to what we had predicted, no difference was detected between the congruent and incongruent outcomes for either action, indicating no specific expectation regarding the presence of an object by the adults.

Table 3 - 6 Paired T-tests for the outcomes for each action in the first block

Action	Outcome pair	t(19)	p
Grasping	Con - InCon	0.76	.46
	Con - Ob-Ab	0.46	.000***
	InCon - Ob-Ab	8.42	.000***
Flop	Con - InCon	0.15	.88
	Con - Ob-Ab	4.94	.000***
	InCon - Ob-Ab	3.98	.001***

3.2.2.4 The practice block

As mentioned above, the participants might have learnt from the task that the position of the object was random across the task and that the action was irrelevant to the task. They could have learnt this not only from the experimental blocks, but also from the practice blocks. Although the practice blocks were not originally designed as the main test, we decided to examine this very first experience of the participants in this task.

The participants had fully understood the tasks in the four testing blocks, but in the practice block some of them were not familiar with the rules. As a consequence some participants did not respond as soon as the screens opened, but waited and observed the outcomes for longer; others guessed about the results and tried to respond before they actually saw the target. Therefore we added an additional criteria for the data in the practice block. Only RTs that fell in the range of 150ms to

1250ms were regarded as valid data for this analysis. With these criteria, we were left with 221 trials in the practise block.

We conducted a two-way ANOVA with the remaining data (See Figure 3 - 11 and Table 3 - 7). There was a significant main effect of Outcome ($F(1.39, 26.35) = 48.31, p < .001$). But neither the main effect of Action ($F(1, 19) = .38, p = .55$), nor the interaction between the two factors ($F(1.28, 24.31) = .90, p = .38$) was significant.

The data from the practise trials confirmed the results of the four blocks of the main test, suggesting that the lack of the predicted effect was not due to a practice effect.

Table 3 - 7 Reaction Time in each condition (ms) of the practise block

		Outcomes		
		Congruent	Incongruent	Object-absent
Flop	Mean	395.6	455.9	603.0
	SD	84.1	111.2	151.3
Grasping	Mean	411.2	412.8	600.9
	SD	60.7	115.0	139.7

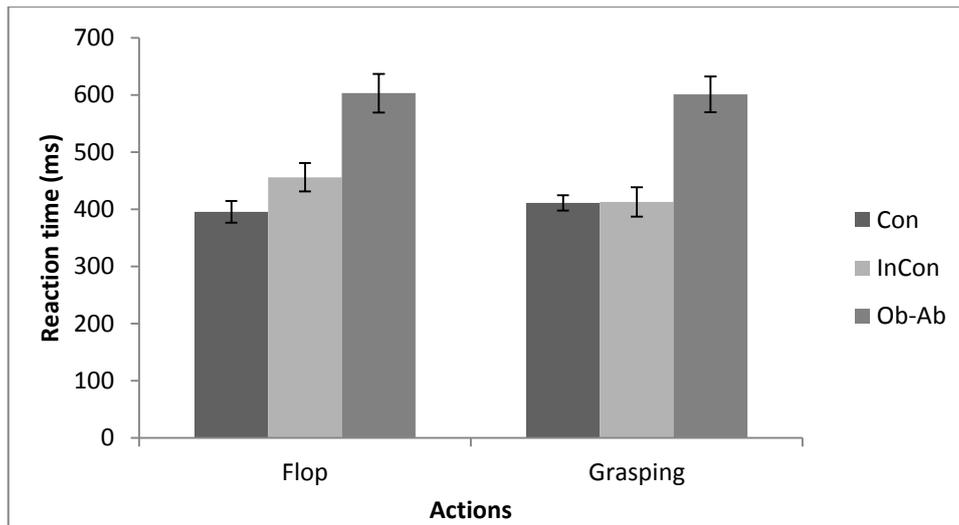


Figure 3 - 11 Mean (SE) reaction time (ms) in each condition of the practise block

3.2.2.5 The probe questions

For the probe questions, which were implanted among the test trials randomly, the participants made mistakes in 49 (21.5%) out of 228 trials in total. The error rate was significantly lower than chance level ($\chi^2(1) = 74.12, p < 0.001$), indicating that the participants noticed the difference between the two actions. Fewer mistakes were made in the first block (see Table 3 - 8).

The tendency of the error rate in the probe questions to increase over blocks might indicate the participants progressively tended to pay less attention to the type of action. However, the level of correct judgment remained very high and, in any case, this only indicated the participants’ explicit awareness of the actions. If the actions had an implicit effect (which was the target of our study), it would not matter much if they could not explicitly make the correct explicit judgement.

Table 3 - 8 Error rate of the probe question in each block

Block number	1	2	3	4	Total
Error rate	14%	19%	26%	26%	21.5%

3.2.2.6 Questionnaire

Table 3 - 9 shows the participants’ average score (from raw scores) on each question. We transformed the score of negative questions to a standard score and

grouped the questions into four categories. The first category involves the participants' general opinion about the task; the second the directedness of the action to the object; the third the agent's knowledge about the object; the fourth the role of the action in the task.

Results suggest that the participants found the task easy. They noticed the difference between the two actions and were aware of the mistakes they made. They did not find the actions distracting for them to respond to the object and did not make efforts to ignore the actions to focus on the opening of the screens.

Table 3 - 10 showed the participants' average standardized score in the last three categories for the two actions. The participants did not in general find that the actions were directed to the object. In spite of this, a paired T-test revealed a significant higher score for Grasping than Flop regarding the four questions about whether the action was directed to the object ($t(19) = 2.28, p = .03$). It suggested that the participants *explicitly* regarded the grasping action as more object-directed than flop, despite the implicit measures revealing no such distinction. But for the four questions about the agent's intention about the object, the two actions did not have different influence on the participants' answer ($t(19) = .24, p = .81$). Yet when asked whether the agent knew about the position of the object, and if she was trying to reach the object, the participants were more likely to agree when the action was grasping than flop. From the last four questions, we found that the participants did not find the actions very relevant to the position of the object or to help them make the right response. The two actions were not regarded differently for the last four question ($t(19) = 1.34, p = .20$).

Table 3 - 9 Mean score on each question on the questionnaire (Score 1-6, strongly disagree, disagree, somewhat disagree, somewhat agree, agree, strongly agree.)

Questions	Flop	Grasping	About the Task
About the task			
I made some mistakes in the task.			5.1
I noticed the different gestures in the videos.			5.0

I tried to ignore the movements in the video and focused on the open of the screens.		3.8
The task is easy.		4.7
The video before the screens opens was distracting.		3.7
The action's directedness to the object		
The action was aimed to the object.	2.3	2.7
This action did not inform about the object.	4.9	4.3
This action had nothing to do with the object.	4.5	4.3
This action indicated the location of the object.	2.5	2.4
The agent's knowledge about the object		
I tried to figure out if the actress knew about the object.	2.9	2.9
The actress did not know the position of the object	3.2	3.8
The actress in the video knew if there's an object.	3.4	3.4
The actress tried to reach the object.	2.3	2.9
The distraction/relevance of the actions to the task		
I ignored this video when making the judgement.	4.1	4.3
This action distracted me from the task.	3.5	3.7
This action looked random and irrelevant to the task.	3.6	3.9
This video was helpful for me to succeed in the task.	2.7	2.5

Table 3 - 10 Standard score for the last three category of questions

	Flop	Grasping
The action and the object (goal-directedness)	2.3	2.6
The agent and the object (intention)	3.1	3.1
The distraction/relevance of the actions to the task	3.1	3.0

3.2.3 Discussion

The main aim of this experiment was to compare the performance of adult humans with that of Capuchin monkeys in a task of inferring the presence of an object from an object-directed action using the same stimuli, but using reaction time in the place of looking time.

In contrast to the Capuchin monkeys, human adults failed to show any evidence of distinguishing between grasp and flop in terms of object directedness. They did not expect the object location to be congruent with the type of action

3.2.3.1 The lack of difference between the congruent and incongruent outcome

The first and the most important result that we expected to find was longer RT at the incongruent outcome for the object-directed action, grasping, as only the object-directed action was predicted to generate an expectation of an object based on the features of the action. However, we found no difference between the congruent and the incongruent outcome for grasping. The result was the same for the non-object directed action, flop. The human adult participants simply did not expect the object's location to be congruent with the direction of the actions at all, thereby suggesting they may not have implicitly inferred the presence of an object linked to the performance of a typical object-directed action. There are two possible explanations that could have led to the lack of difference regarding congruency.

Practice effect?

One possibility is that the participants had too many repeated trials learning that the object's location was not related to what happened before the screen opens at all. We may have trained them to ignore the events before the screens' opening and just focus on detecting the object itself once the screens opened. However, when we examined the order of the blocks, and even the practise block, we found no practice effect.

Posner effect?

Another possibility is that the paradigm failed to detect the implicit understanding of intentionality. The actions might have triggered the participants' implicit perception of intentionality and object expectation, but this might have decayed during the time interval between the initial perception of the action and the screens opening, with RT being an indication of voluntary behaviour, rather than of an implicit cognitive event (Shepherd, 2010). This might be plausible because in the

Posner paradigm, which is an experimental procedure very similar to the one used in the current experiment, findings indicate that when a cue, such as an arrow or a pointing gesture, is shown in a particular direction, adult humans will respond more quickly to targets appearing at the corresponding location only if the interval between the cue and the target is less than about 200 ms (Posner, 1980). With intervals longer than 400ms, the cue no longer affects the participants' responses. In our experiment, the shape and direction of the hand could be recognized at around 0.5 seconds into the trial; but the screen would not open until the end of the fourth second. In this case, the interval between the initial presentation of the cue (which continued until the hand was occluded by the screen at around 3-seconds) and the target was far longer than 400ms. Thus, if our grasping action provoked an object expectation similar to the Posner effect bias, its effect would no longer be visible at the time of response in our experiment. We will come back to this issue in the general discussion of this chapter.

The longer reaction time to the absence of the object

We predicted that if the participants expected an object from the grasping action, their reaction time should be slower when the object was absent than when it appeared at the congruent location for the goal-directed action. Although we indeed find results consistent with this prediction, there was every indication that the longer reaction time was not due to the expectation that an object should be present because the control action (flop) also produced longer RTs in the object absent condition. As only the grasping action was designed as object-directed, the control action flop should not have triggered the participants' expectation of the object. But, in the experiment, the reaction time for grasping and flop was exactly the same: longer for the absent object than for the object appearing at the congruent or incongruent location to the direction of the action.

This suggests two possibilities. Firstly, the control action might actually have been regarded by the participants as an object-directed action, and their longer reaction time to the absence of the object would indicate the participants' inference of the object. Some participants indeed suggested after the experiment that the

action flop looked like requesting, which is a very commonly used gesture and typically object-directed. A way to address this problem would be to compare the two actions with another control action that is more neutral in terms of object-directedness than either flop or grasping. This was done in a follow up experiment presented next in this chapter.

The second possibility is that responding to the absence of the object may simply require more scanning time to make sure that there is no object at *either* location, whereas the response to object presence can occur as soon as it is detected in one location without having to check the second. In this case, the longer reaction time at the Object-absent condition should not be attributed to the violation of expectation to the inference of an object, but be an effect of the search task unrelated to the previously shown actions. In the follow up experiment, we will further examine if, when no previous action is shown, the participants still respond more slowly to the absence than to the presence of an object.

3.2.3.2 The two actions and object-directedness

Did the participants pay attention to and differentiate between the two actions? The pattern of reaction time to the three outcomes looked almost identical for the two actions and it seemed that only the outcome affected the participants' reaction time. Did the participants focus only on detecting the object, failing to pay attention to the two actions? It didn't seem so according to the probe questions. The participants gave the right answer to the majority of the probe questions (80%), suggesting they explicitly noticed the difference between them.

Moreover, they scored a little higher for grasping than for flop at object-directedness in the questionnaire. This may indicate that they explicitly regarded grasping to be more object-directed than flop. Nonetheless, the difference was small and it may also be possible that at a more implicit level (the one targeted by the reaction time measure) the two actions were coded similarly in terms of object-directedness. Rather than jumping to any conclusions, we suggest that more empirical work would be helpful in answering this question. Again, to compare a more neutral action with the current two might assist in clarifying this issue.

3.2.3.3 Comparison with capuchin monkeys

We expected a similar pattern of results for the Capuchin monkeys and adult humans, with the humans showing in addition evidence of expecting an object in the object absent condition. The results, however, turned out to be very different for the two species.

Firstly, we expected the RT (Reaction time) and LT (looking time) to be longer at the incongruent outcome than the congruent outcome only for the object-directed action. If the participants expected an object by perceiving object-directedness, they would find the incongruent outcome violated their expectation and thereby, would look longer or respond more slowly in this condition only for the object-directed action. This was true of Capuchin monkeys, (i.e., longer LT at the incongruent than the congruent outcome for grasp, but similar LT for congruent and incongruent flop). In contrast, however, human participants' reaction times to the congruent and incongruent conditions were not different for either grasp or flop. At first sight, this might seem to indicate that capuchin monkeys could understand object-directedness, while adult human couldn't.

However, we suggested that there are a number of reasons why humans failed to show their understanding of intentionality in our experiment. We tried to present the Capuchin monkeys and adult human with as similar a task as possible. So we presented them with the same videos. But the measurements were different. For the monkeys, we used looking time, as typically done in the violation of expectancy paradigms for infants, and for the humans we used reaction time as a comparable measurement, following the example of previous studies (for example, Kovács et al., 2010). Nonetheless, the use of different measurements may bring in other potentially confounding factors.

For example, when the screens were opened, the monkeys could still observe the outcome for 10 seconds, which might allow them to retrospectively code the relationship between the object and the previously seen action. But the adults had to make their response immediately after the object was shown, which might preclude any retrospective coding of an action-object relation. The adults' reaction

time could only indicate their expectation about the object from watching the actions. They would not have the opportunity to generate the sort of retrospective relation that we suggested the monkeys could be encoding after seeing the object (see chapter 2). It remains undetermined whether adult humans would also code object-directedness from the action to the object once both of them were shown. Therefore, in the follow up experiment, we will examine if the action would affect the adults' reaction time to the object, when they knew about the object in advance.

Secondly, although the human adults did show longer RT to the object absent condition (consistent with the expectancy of an object), whereas the monkeys showed shorter LT (inconsistent with object expectation), we cannot conclude that this demonstrates the humans did infer an object from the action. On the one hand, as discussed earlier, the non-object-directed action was as effective as the object-directed action in provoking this effect, and on the other, the effect could be explained as being due to the need to scan both locations before being able to respond that there was no object, which would always lead to a longer RT.

It is difficult to draw any conclusions from the comparison between capuchin monkeys and adult humans on their ability to understand intentionality for now. We will come back to this issue in the general discussion of this chapter, after a follow up experiment with adult humans in which we will try to address some of the potential shortcomings identified in the experiment.

3.3 Experiment 2

3.3.1 Introduction

In this follow up experiment we mainly aim to examine the potential shortcomings identified in the last experiment.

The first is whether the longer reaction time to the absence of the object should be attributed to an inference of the object from the actions or to the fact that it is simply cognitively harder to respond to 'nothing' than to 'something' (e.g., the scanning needs to be more thorough to conclude that there is no object). Therefore in the first session of the current experiment, we eliminated the preliminary actions

and required the participants to respond directly to the three outcomes. If the participants again responded slower to the absence of the object, it would suggest that it took longer to respond to the absence than to the presence and location of the object. Otherwise, it would suggest that the actions in the last experiment had evoked the participants' expectation of an object to be present.

Secondly, it remains undetermined whether both grasping and flop were regarded as object-directed and whether a neutral more clearly non-object-directed action would be perceived differently. To determine this, we added another action, tapping at the centre of the platform between the screens. It was not directed to either location and never reached the object. By adding this control action, we sought to investigate whether grasping and flop were both regarded as object-directed actions. In addition, this will also act as a control for the possibility that the actions directed to either screen would bias laterally the participants' attention. Thus the three actions, grasping, flop and tapping were presented in the second session of the current experiment before the simultaneous opening of the two screens.

In addition, the number of trials was reduced so that any practice effect would be minimised. In the current experiment, we presented the participants with only two repeated trials in each condition.

The next question was whether previous knowledge about the object would affect the reaction time. To address this, in the third session of the current experiment, we showed the participants if the object was present or not and, if so, where it was. In this case, they always knew about the object before they watched the action, thereby eliminating the requirement of inferring the potential object from the action. Especially in the conditions where the object was present, revealing in advance the presence of the object might facilitate the processing of the subsequent action as object-directed. Specifically, we expected that the adults' RT would be longer when the object location suggested by the action was incongruent with the previous knowledge.

Finally, we made some modifications to the video clips to increase the proportion of efficient area on the computer screen. In the video clips of the last

experiment, large areas of the screen were not providing information regarding the action or object-directedness. For example, the hiding curtain that occluded the actress' facial area occupied almost half of the screen but provided no information about intentionality; while the space between the two hiding screens, where the action could be seen, only occupied about 1/6 of the whole screen. The small area of effective information might not have drawn enough attention from the participants to the actions. In the current experiment, we modified the videos to increase the proportion of efficient area (see method for details).

3.3.2 Method

3.3.2.1 Participants

A total of 30 (Female : Male = 23:7, average age = 27.7) participants were paid £2 each to participate in the current experiment. They were recruited from the participant-pool in University of St Andrews. All participants were required to have normal or corrected-to-normal vision. None of them had participated in similar experiments, including our last experiment. They were invited to a quiet room in School of Psychology and Neuroscience to participate in the experiment.

3.3.2.2 Apparatus and general procedure

The apparatus was the same as in the last experiment.

The participants were invited to the lab in the School of Psychology and Neuroscience at University of St Andrews. The experimenter first read them the instructions and they signed the consent form. At the beginning of the reaction time task, they were presented with an introductory video clip. Afterwards, they finished three sessions of reaction time tasks. In between every two they were allowed to have a short break if they wished to. At the end of the reaction time tasks, they were shown the debriefing form and the experimenter explained the purpose of the study.

3.3.2.3 Design

As described in the introduction, the experiment was divided into three sessions. In all the three sessions, the participants were required to respond by

pressing one of three keys to indicate the presence and location of an object, when the two hiding screens opened. The object in this experiment was a small toy lion.

In the first session, the participants were presented with 6 trials without actions. This session had a single factor, the outcome when the screens were open. It included three conditions and was a within subject design. The three conditions were, object presented behind left screen, object presented behind right screen and object absent (See Figure 3 - 12).

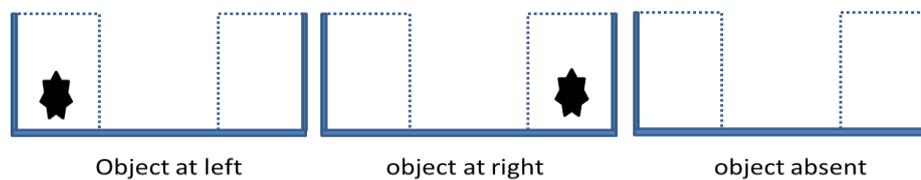


Figure 3 - 12 the three conditions in the first session

In the second session, we used a 3*3 with-in subject design (see Figure 3 - 13). The two factors were Action and Outcome. Besides grasping and flop, we added a third action, tapping at the centre, as a neutral action. The three outcomes were again congruent, incongruent and object-absent. Note that here, although the action tapping was directed to the centre and therefore does not truly involve a clearly congruent or incongruent direction to the location of the object; for all the three actions, in the congruent condition, the arm and hand at the congruent side to the object were used to produce the action. Therefore, we kept the factor, Congruency, for this action as well.

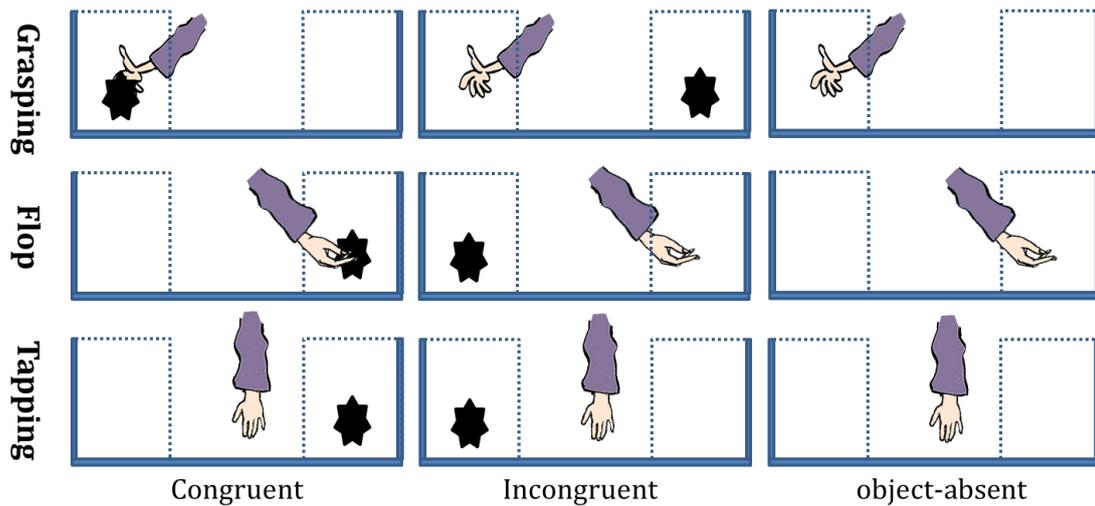


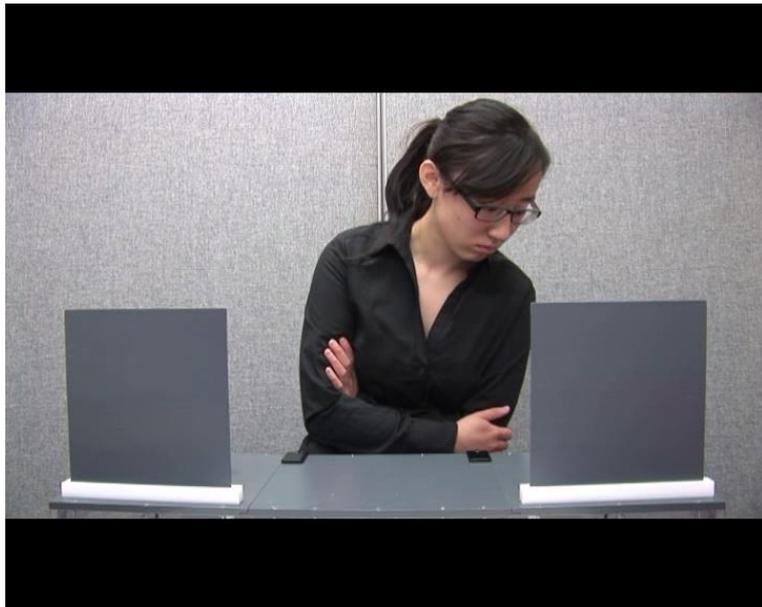
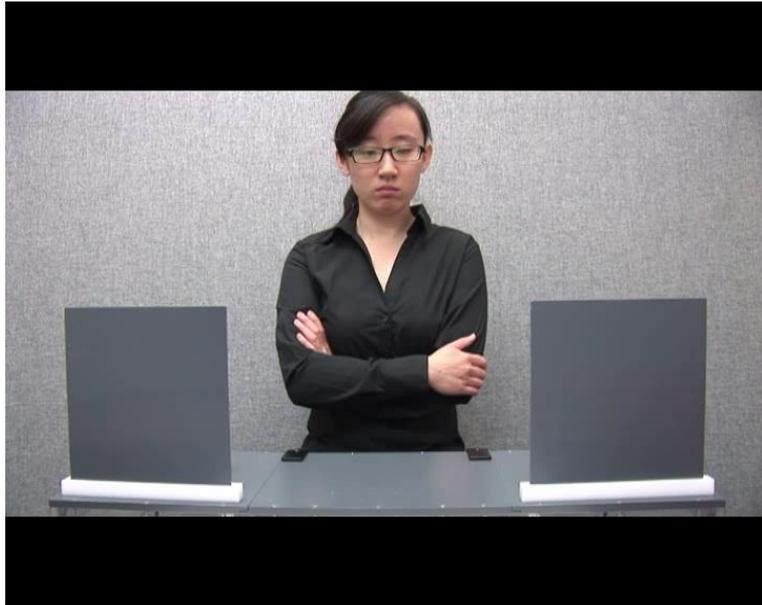
Figure 3 - 13 the 9 conditions in the second session

In the third session, the design was as used in the second session, but with the addition that the participants were shown the outcome before the actions (followed again by the outcome) to investigate whether the participants were affected by the actions when they knew about the outcomes in advance.

3.3.2.4 Stimuli

The stimuli displayed were videos similar to those used in the last experiment, but with the following adjustments to facilitate participants' attention at the action area.

The introductory clip lasted for 11 seconds with a similar procedure to the last experiment (see Figure 3 - 14) . However, instead of resting both hands at the centre of the screen, the actress sat with her arms crossed to reduce the participant's initial attention to the hands. And, as the curtain was eliminated in the test sessions to enhance the participants' attention to the critical area, i.e., the two screens and the space between where the action was visible, no curtain fell at the end of the introductory clip.



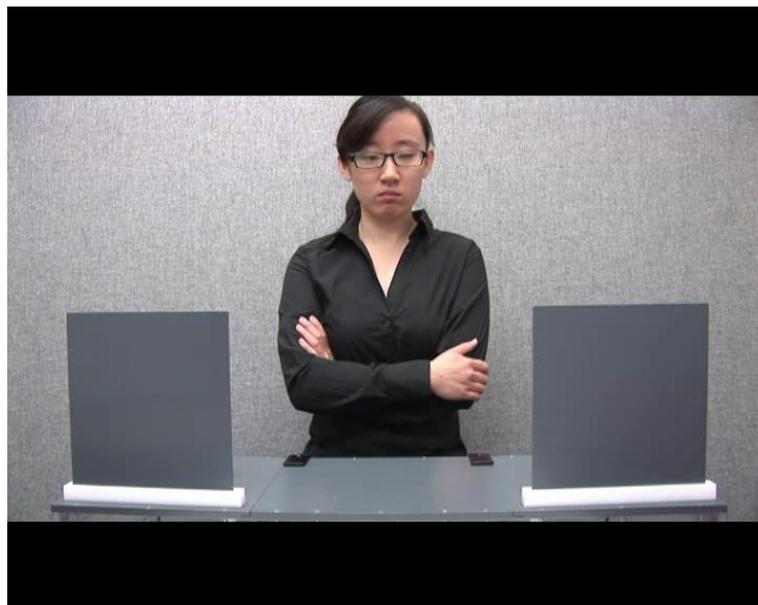
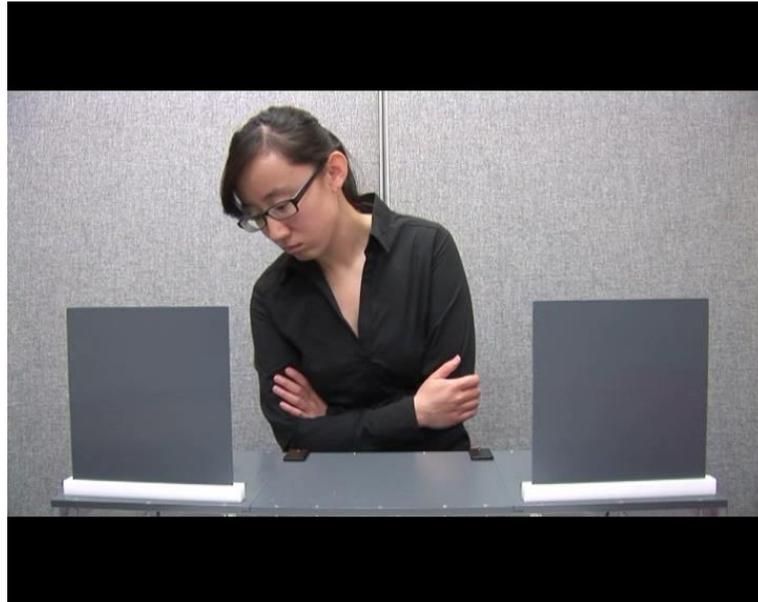


Figure 3 - 14 Procedure of the introductory clip

Instead, in the experimental clips, the videos were cropped to exclude the face above the chin of the actress so that any information from the eyes or head orientation was eliminated (see Figure 3 - 14, 3 – 15). The video always began with the hiding-screens closed and the actress sitting at the centre with her arms crossed. At the end of the video, the hiding-screens opened simultaneously and one of three

outcomes was revealed: an object behind the left screen, an object behind the right screen, or an absent object. The participants were required to press a matching key as quickly and accurately as possible when the screens opened.

For the first session each video clip lasted 2-seconds before the screen opened to reveal the outcomes. The actress sat still with her arms crossed without moving.

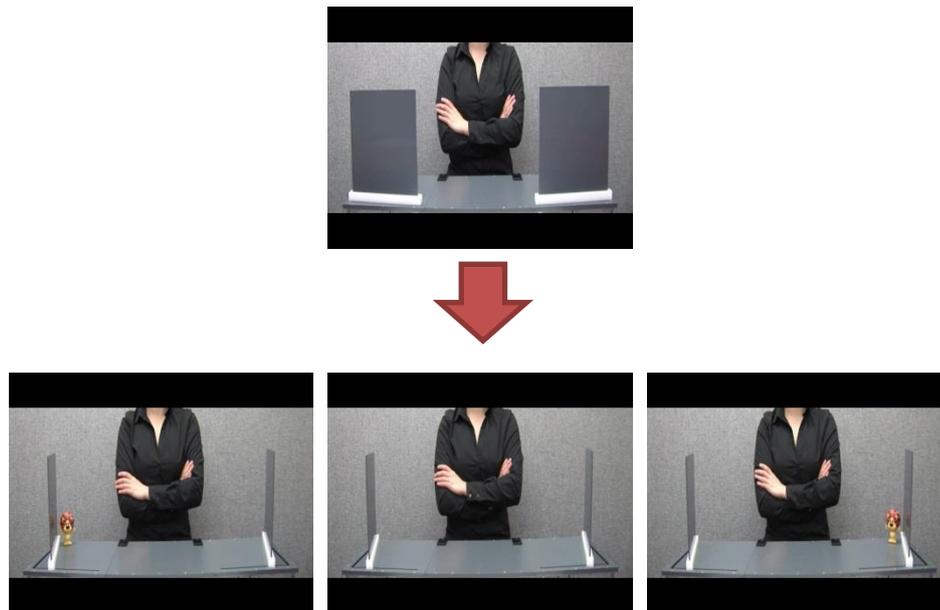


Figure 3 - 15 Session one, no action

For the second session, each clip lasted for 4-seconds before the screen opened, presenting an action directed towards behind one of the two screens. After 1-second of sitting still with crossed arms, the actress performed an action: grasping, flopping, or tapping the centre (see Figure 3 - 16). At the end of the 4th second the hiding-screens opened simultaneously while the hand was drawing back.



Figure 3 - 16 the three actions in session two

In the third session, the video clips were the same as in the last session except that the target scene (where the screens were open and the presence/position of the object was visible) was presented for 1-second before each trial with the aim of making the participants aware of the object position from the beginning.

3.3.2.5 Data analysis

We first examined the error rate for all the sessions. Then we used a 3*3 repeated ANOVA to examine each session separately. Action and Outcome were again our two main factors. In addition, for the second and third sessions, we conducted an analysis excluding all the conditions where the object was absent and compared the congruent and the incongruent conditions. This was because the

longer reaction times in the first session indicated different cognitive processes when responding to the presence and absence of the object and thereby the reaction time at the absence of the object would not be directly comparable with the congruent and incongruent conditions. In addition, we compared the second and the third session to examine if knowledge of the object beforehand affected the participants' behaviour in the experiment. Finally, we also compared the second session with the last experiment to investigate whether the participants inferred the object from the object-directed action when the effect of practice was reduced.

3.3.3 Results

3.3.3.1 The incorrect trials

The participants responded incorrectly for 2.54% of all the test trials (Table 3 - 11 and Table 3 - 12). The number of incorrect trials was too small for further analysis.

Table 3 - 11 Number and percentage of incorrect trials in session 1

Session		Left	Right	Object-absent
1	No.	3	2	3
	%	5	3.33	5

Table 3 - 12 Number and percentage of incorrect trials in session 2 and 3

Session	Condition		Congruent	Incongruent	Object-absent
2	Flop	No.	1	4	2
		%	1.67	6.67	3.33
	Grasping	No.	2	2	3
		%	3.33	3.33	5
	Tapping	No.	1	2	1
		%	1.67	3.33	1.67
3	Flop	No.	0	0	2
		%	0	0	3.33
	Grasping	No.	1	0	3
		%	1.67	0	5
	Tapping	No.	0	0	0

In the following tests we excluded all the incorrect trials from analysis. In addition, following standard practice in RT experiments, we excluded the trials with reaction times shorter than 150ms (1.19%), in which participants were likely to have responded when the screen moved rather than when they found the object, or longer than 1250ms (0.8%), in which they might not have responded as quickly as possible. We calculated the mean RT for each participant in each condition and then the reported the mean in each condition as the average across participants.

3.3.3.2 Reaction time to different positions of the object (Session one)

We first examined the participants' RT to different locations of the object in the first session, i.e., with no cuing actions (Figure 3 - 17 and Table 3 - 13). A repeated measure ANOVA with object position (Left, Right, Absent) as the factor revealed a significantly different RT ($F(1.54, 43.04) = 37.78, p < .001$) between the three conditions. Paired T-tests with Bonferroni adjustment showed that RT to Object-absent was significantly longer than to object-left ($t(28) = 5.60, p < .001$), or object-right ($t(28) = 6.75, p < .001$). But no difference was detected between object-left and object-right conditions ($t(29) = 1.61, p = .14$). The results suggest that the participants responded much more slowly when the object was absent than when it was present even when no cue action was presented. This indicated that the longer reaction time to the absence of the object was not likely to be caused by the inference of the object from the intentional action, but by an inherent difficulty in detecting 'nothing' vs 'something'.

Table 3 - 13 Reaction time to different position/presence of the object in the 1st session (ms)

	Left	Right	Object-absent
Mean	493.8	456.0	695.5
SD	125.6	128.5	207.4

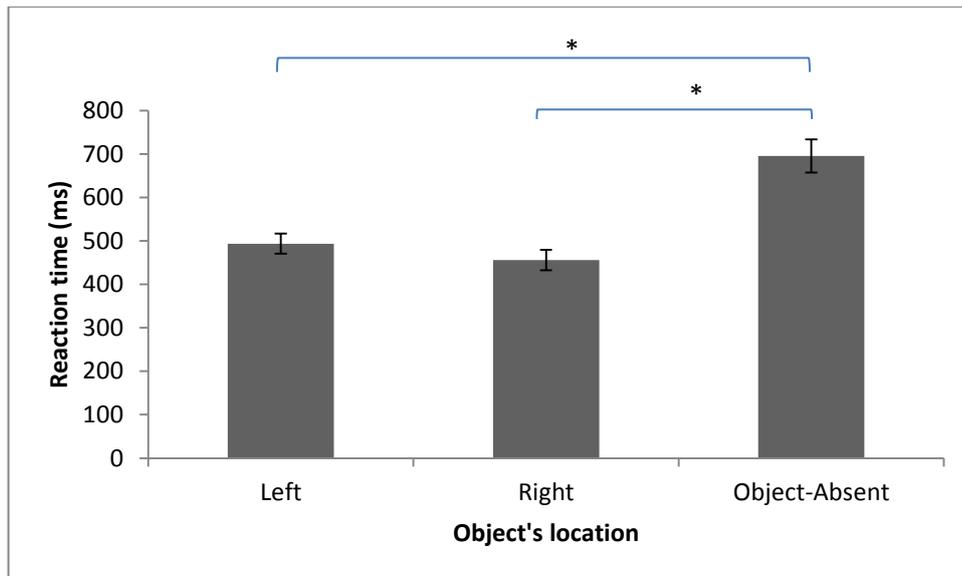


Figure 3 - 17 Reaction Time to different position/presence of the object in the 1st session (ms)

3.3.3.3. The effect of Actions Congruency on reaction time without knowledge of the object beforehand (session two)

We then examined the participants' RT in the second session (See Table 3 - 14 and Figure 3 - 18) to test whether the actions and their congruency had any effect on the participants' reaction time. We conducted a two way repeated measure ANOVA using Actions (Grasping, Flop and Tapping) and Outcome (Congruent, Incongruent and Object-absent) as the two factors.

The 3*3 ANOVA revealed a significant main effect of Action ($F(2,56) = 3.12, p = .048$) and Congruency, ($F(1.2,34.6) = 41.17, p < .001$). But the interaction was not significant ($F(2.9, 82.4) = 1.49, p = .22$). To examine the simple main effect of Congruency and Action, we used Paired T-tests with Bonferroni adjustment. We found a trend of longer reaction time to Flop than to Grasping ($t(29) = 2.48, p = .062$), but the differences between Tap and Flop, $t(29) = .68, p = 1.00$ or Tap and Grasping ($t(29) = 1.72, p = .33$), were not significant. For Congruency, the results showed that the participants responded more slowly to the Object-absent condition than the Congruent ($t(29) = 6.85, p < .001$), and the Incongruent condition ($t(29) = 6.23, p < .001$). No difference between the Congruent and Incongruent conditions ($t(29) = .84, p = 1.00$) was found.

Table 3 - 14 Mean and SD of reaction time in the 2nd and 3rd session (ms)

Session	Condition		Congruent	Incongruent	Object-absent
Session 2 No Previous knowledge	Flop	Mean	410.4	432.3	560.0
		SD	75.4	75.2	151.3
	Grasping	Mean	408.4	413.9	513.7
		SD	83.1	72.6	116.4
	Tapping	Mean	417.7	411.5	556.5
		SD	90.4	72.2	159.5
Session 3 Previous knowledge	Flop	Mean	344.8	369.8	411.4
		SD	85.1	101.4	154.7
	Grasping	Mean	347.9	368.6	439.0
		SD	111.3	143.2	183.9
	Tapping	Mean	353.0	329.7	414.7
		SD	110.8	96.1	148.8

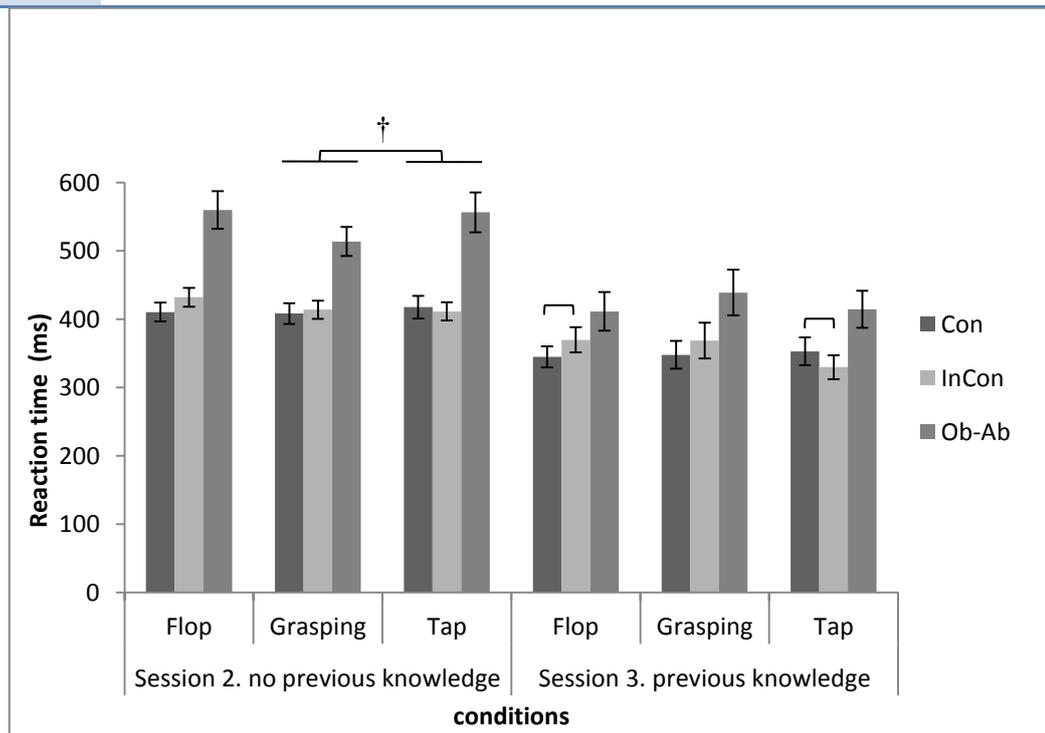


Figure 3 - 18 Mean reaction time (ms) of each condition in the 2nd and 3rd session (SE)

In addition, because the first session revealed a significant difference for the outcome of object-absence, even if no action was presented at all, we conducted an additional analysis of session 2 excluding the Object-absent trials. We conducted a 3*2 two-way ANOVA with Actions (Flop, Grasping and Tap) and Congruency (Congruent and Incongruent) as the two factors but found no significant main effect of either factor, Action ($F(2,56) = 1.20, p = .31$), Congruency ($F(1,28) = .33, p = .57$) or the interaction ($F(1.5,42.0) = 1.35, p = .26$).

3.3.3.4 The effect of Actions and Congruency on reaction time with knowledge of the object beforehand (Session three)

We then examined the participants' RT in the third session (See Table 3 - 14 and Figure 3 - 18) to test if the actions and the congruency had any effect on the participants' reaction when they knew from the beginning the location or absence of the object. We conducted a two way repeated measures ANOVA using Actions (Grasping, Flop and Tapping) and Outcome (Congruent, Incongruent and Object-absent) as the two factors.

The ANOVA revealed a significant main effect of the Outcome ($F(1.4, 41.1) = 15.41, p < .001$). But neither the main effect of Action ($F(1.6, 46.2) = 1.17, p = .31$), nor the interaction ($F(2.9, 81.6) = .96, p = .41$), were significant. The participant response was slower to the Object-absent condition than the Congruent ($t(29) = 4.56, p < .001$), and the Incongruent condition ($t(29) = 3.81, p = .002$). No difference was detected between the Congruent and Incongruent conditions ($t(29) = .85, p = .42$).

Again, we excluded the Object-absent trials and examined whether the action and congruency of the object location affected the participants' reaction time. The 3*2 ANOVA showed no significant main effect of either factor, Action ($F(1.4, 41.4) = 1.42, p = .25$), Congruency ($F(1,29) = .72, p = .40$), although the interaction was approaching significance ($F(2,58) = 2.89, p = .064$). To further examine the interaction, we conducted Paired-T-tests for each action and found that the participants showed a trend to react more slowly to the incongruent than the Congruent outcome for flop ($t(29) = 1.94, p = .062$), but no difference was found for grasping ($t(29) = .96, p = .34$). The reaction time to the congruent outcome of

tapping, however, was significantly longer than that to the incongruent outcome ($t(29) = 2.34, p = .026$).

3.3.3.5 The effect of knowledge (RT Decrease)

The only difference between the 3rd and the 2nd session was that in session 3 the participants had knowledge of the location of the object before the screens opened. So the time difference between the 2nd and the 3rd session should be caused by this knowledge. We calculated the time decrease from session 2 to session 3 by subtracting the RT of session 2 from session 3 (see Figure 3 - 19) to examine if the factors Action and Congruency affected the decrease of the reaction time in a different way. We examined the decrease of reaction time using a two-way ANOVA. The test showed a significant main effect of Action ($F(1.6, 46.0) = 4.20, p = .029$) and Congruency ($F(1.3, 37.9) = 7.14, p = .007$), but no interaction ($F(3.1, 88.7) = 1.07, p = .37$). The decrease of RT in the object absent condition was significantly more salient than in the Congruent ($t(29) = 2.95, p = .025$), and Incongruent ($t(29) = 2.10, p = .028$), outcomes. But the difference between the Congruent and the Incongruent conditions was not significant ($t(29) = .96, p = 1.00$).

The decrease of reaction time was marginally significantly more salient for Tapping than for Grasping ($t(29) = 2.48, p = .058$), but no difference was found between Flop and Tapping ($t(29) = .36, p = 1.00$), or grasping and flop ($t(29) = 2.03, p = .16$). In addition, when the outcome was Object-absent, the decrease of reaction time showed a tendency to be larger for flop than for grasping ($t(29) = 1.94, p = .62$).

Knowing that there was no object behind the screen was marginally less helpful when the 'distractor' action was grasping than when it was tap or flop. This suggested that grasping did convey a slightly stronger 'suggestion' of there being an object that may interfere with the knowledge that actually there is none. Interestingly, this supposedly stronger object cueing was only manifest when there was a conflict with previous knowledge, i.e., when the grasping action was known to be counterfactual in advance.

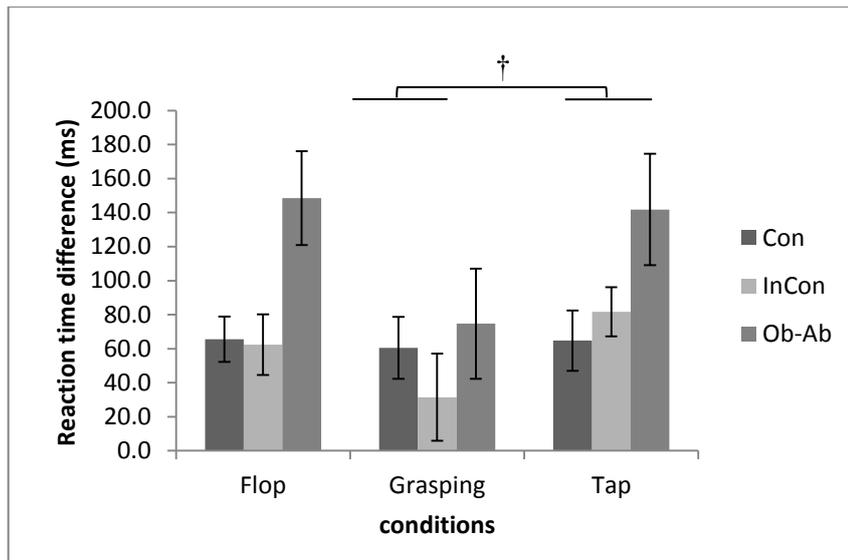


Figure 3 - 19 Mean reaction time difference (ms) to each condition between the 2nd and 3rd session (SE)

3.3.4 Discussion of experiment 2

In the current experiment, we tested human adults' reaction time (RT) to three different outcomes (Left/right position or absence of an object) when they watched no cueing actions to examine if responding to 'nothing' is cognitively harder than responding to 'something'. We also replicated experiment 1 with an additional, more clearly non-object-directed action, tapping, as a control condition. Finally we tested the influence of having previous knowledge of the outcome to examine the potential interference of counterfactual object-directedness upon RT.

When no cueing action was conducted, it still took the participants longer to respond to 'nothing' than to the position of the object, regardless of whether it was behind the left or right screen. The longer RT to the object's absence is therefore not caused by an expectation of an object from perceiving the intentionality of preceding actions. Responding to the absence of an object seems to be cognitively more difficult than responding to the presence and location of something. When the participants were making the judgement about the object, if the object was present, its appearance on the computer screen might have quickly attracted the participants' attention and provided them with the information that they needed to respond. But

when the object was not present, the participants had to scan both locations to make sure of its absence before making the move to respond. Therefore it always took longer to respond when the object was absent. As the longer reaction time to the object's absence is not affected by the actions, it seems reasonable to exclude the Object-absent condition from further analyse. In the following discussion, we will focus on the comparison of congruent and the incongruent outcomes.

3.3.4.1 Adults do not distinguish object-directed action from non-object directed action

The second goal of this experiment was to examine if the participants show sensitivity to object directed actions when a more clearly non-object directed action is used as control. The action flop in the previous chapter was potentially ambiguous (e.g., it could be interpreted as a request gesture), so in the current experiment we added the action of tapping on the centre of the platform: this was not an object-directed action (beyond the apparatus itself) and its localised direction was always directed to the centre and therefore not cueing any of the potential object locations. However, contrary to our predictions, the reaction time pattern to the object's locations after watching the three actions was not different from each other. And the pattern was consistent with the first session where there was no cueing action and with the results reported in chapter 2 for flop and grasping. The three actions were not coded differently and it seems that no action was regarded as indicative of the location of the object at all, as if the participants were not engaging in an intentional, object-directed coding of the action.

Moreover, this could not be attributed to a practice effect (learning that the actions were not predictive of the objects), because we used a significantly shortened number of trials in the second session of this experiment and the results were still consistent with those in the first experiment. Therefore we conclude that the failure to find a significant effect of the object-directed action was not due to multiple trials, or in other word, practice effect.

3.3.4.2 When knowing about the object in advance, adults still did not seem to differentiate the three actions, but grasping might have indicated some evidence of object-directedness.

In the third session of the current experiment, we provided adult humans with information about the object *before* presenting the actions. In this case, they had seen both the object and the object-directed actions before their final response. Through this manipulation, we expected to remove the requirement of imagining the object from the action and predicted that, if they expected the object-directed action to be directed at the object, their RT would be longer when the cue contradicted the previous knowledge about the object.

In contrast to our prediction, however, the object-directed action (grasping), didn't affect the participants' reaction time at the objects' location as we had expected. The reaction time to the incongruent condition was not longer than the congruent condition for grasping or the other actions. We therefore failed to find evidence of coding object-directedness even when both information about the object and the action had been directly given.

The only evidence suggestive of some degree of coding object-directedness came from an additional analysis. Knowing about the object in advance in session 3 reduced RTs in general in relation to the previous sessions, but this was especially effective for reporting the absence of the object (despite this decision still being significantly slower, but much less so). Interestingly, the reduction in RT to report object absence was almost significantly smaller for the grasping action, as if this action produce a possibly stronger suggestion of the presence of an object than flop or tap. Moreover, the decrease of reaction time to the contradictory cue of grasping, say, to the right when the object had been seen to be on the left, showed a trend to be smaller than that of flop too, again indicating that counterfactual grasping might have produced a slightly stronger interference with the previous knowledge about the object than flop or tapping.

3.3.4.3 Conclusion

To sum up, using the current paradigm we did not find direct evidence supporting the hypothesis that human adults generate an implicit expectation of a relevant object when they perceive an object-directed action alone. However, we did find some evidence that they were sensitive to the counterfactual nature of a goal-directed action that was in conflict with previously established knowledge about object presence and location.

Chapter 4. Adult human's automatic encoding of other's belief

Summary

In our last study with adult humans we found little evidence of the participants' understanding of intentionality (in the sense of object-directedness, Chapter 3) in relation to an object that has not yet been perceived but could be inferred from an object-directed action. In comparison, however, a very similar task with a non-human primate species, showed some understanding of intentionality in Capuchin monkeys, although only after the object was perceived (Chapter 2). To further explore the potential limitations of adult humans' ability to code intentionality automatically, this chapter aims primarily to re-examine this issue with another paradigm originally designed by Kovács and colleagues (Kovács et al., 2010), which showed evidence of complex automatic ToM abilities in both human adults and preverbal human infants. With a modified version of this paradigm, including conditions that required responding to absent objects, we again found limited evidence for adults' ability to encode intentionality in relation to objects that are not directly perceived. Having, in addition, failed to replicate some of Kovács et al's results, we further investigate the paradigm and suggest that their positive results might only occur under certain limited experimental conditions, and question if they truly reflect an implicit detection of belief about objects. To conclude, we suggest that whether adult humans can automatically encode intentionality about non-perceived objects in the sense of object directedness remains an open question.

4.1 Introduction

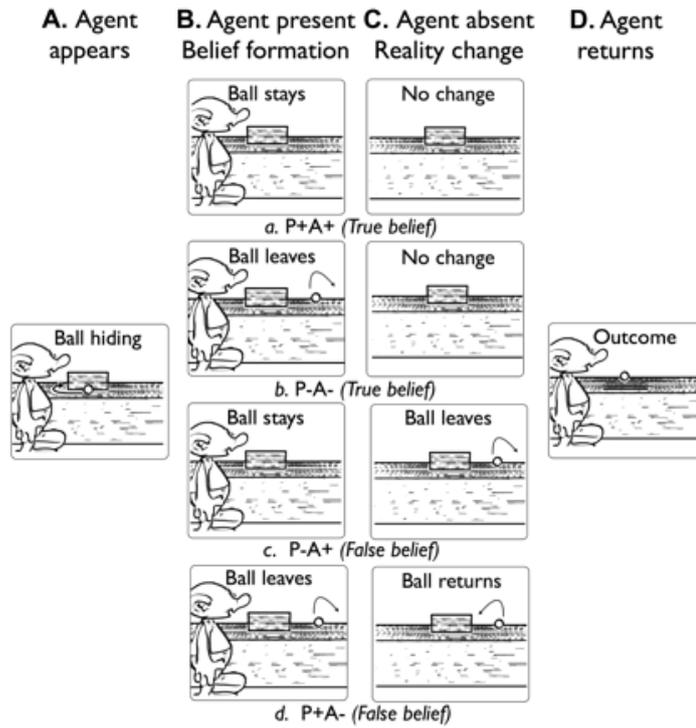
As discussed in the last chapter, we found conflicting results in Capuchin monkeys and adult humans. The experiment with Capuchin monkeys showed evidence that the monkeys could encode the intentionality of an action performed when its object was not visible as long as the object became visible (in some sort of retrospective encoding), but that they showed no evidence of having ‘inferred’ an object from the action alone. However, the results with human adults provided little evidence of a similar effect. The apparently better performance of Capuchin monkeys over adult humans was unexpected. The ability to code intentionality (in the sense of object-directedness) automatically might be an important component of ToM and humans are typically much more proficient in ToM abilities (Herrmann et al., 2007). Therefore we expected adult humans to at least equal with Capuchin monkeys, if not outperform them, in understanding intentionality about non-perceived objects.

As discussed in the previous chapter, one possible explanation of our negative results with humans might be limitations in the design of the paradigm. Specifically we discussed the possibility that the measurement of reaction time (instead of looking time, as with the Capuchin monkeys) might have failed to capture the actual implicit capabilities of the humans. Therefore in the current chapter, we attempt to use another paradigm to re-examine adults’ implicit understanding of ToM, one in which positive results have been reported with the use of RT with adults as an equivalent measure to Looking Time with non-verbal participants (in this case, human infants).

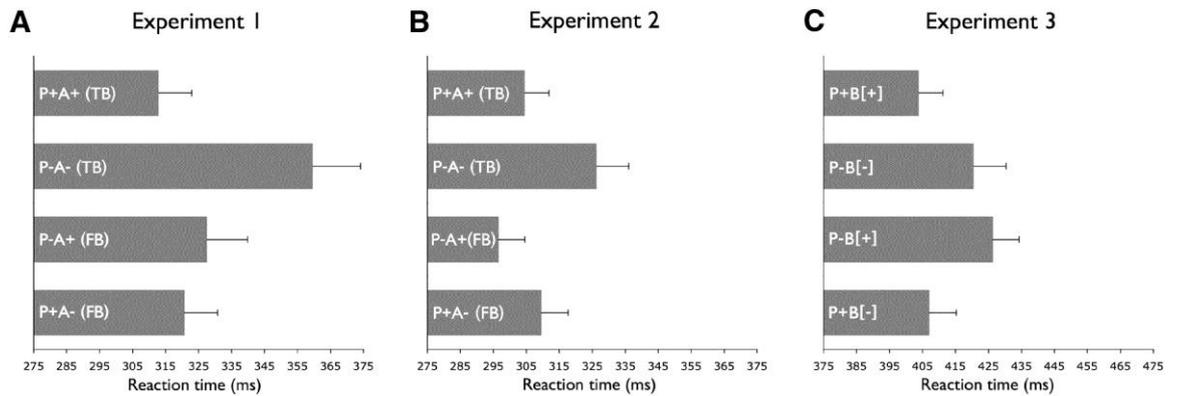
The task was initially designed by Kovács et al (2010), where they reported automatic computation of other’s belief about object location and its interaction with the participants’ own beliefs in modulating the reaction times in an object detection task in human adults. These results were equivalent to those found with 7-month human infants using Looking Time instead of RT as the dependent variable. In their task (see Figure 4 - 1), human adults are required to watch videos of an object hiding behind a screen and then, when the screen opens, they are required to press

a key if they detect the object. In their video, there is also an irrelevant agent (A) who believes that the object is either present (+, + stands for believing the ball hides behind the screen) or absent (-, - stands for believing the ball is not behind the screen) behind the screen.

They found that only when neither the participant (P) nor the agent (A) believed that the ball was present (P-A- condition), adults respond more slowly (see Figure 4 - 1.II), which indicating a violation of their expectation. But when at least one of them (either P or A) believed that the ball was behind the screen (P+A-, P-A+ and P+A+), they respond more quickly. Especially intriguingly is that, in the P-A+ condition, the participants seem to ignore their own true belief about the absence of the object, and follow the irrelevant agent's false belief about its presence. This key finding led to their conclusion that the participants' automatically compute the beliefs of an agent even when the agent's belief is erroneous and irrelevant to performing the task. The effect does not occur if instead of an agent, the video displays an inanimate entity such as a stack of boxes. This experiment provided evidence that adult behaviour is affected by an automatic registration of others' belief attributable to an implicit ToM system (Kovács et al., 2010).



I



II

Figure 4 - 1 Procedure and results of Kovács et al's task (Kovács et al., 2010). A. Experiment procedure in the four conditions. Note: P and A represents the participant and the agent in the video respectively, + and - represents believing the ball is behind the screen and believing the ball is not behind the screen, respectively. B. Results of Kovács et al's task in adult human. Note: In experiment 1, the agent stands beside the screen when the screen opens. In experiment 2, the agent was not present when the screen opens. In experiment 3, a pile of boxes, instead of the agent, acted as the agent following the procedure in experiment 1.

One difference between Kovacs' et al. task and ours is that in their displays no object-directed action is performed by the agent other than introducing the object and putting it down on the stage. The smurf figurine just enters or leaves the scene by sliding in or out, but does not subsequently act or specifically look in the direction of the target object. It just remains present possibly contemplating what is going on. In contrast, in our task we had a real agent conducting actions that were directed to an initially non-perceived object.

However, the coding of intentionality relationship between agents and objects does not require the performance of particular actions performed upon objects. For example, an observer can code whether an agent has "witnessed" or not something, and therefore whether the agent has entered into an intentional relation (in the Brentanian sense of object-directedness or aboutness) with that object. So that the object now is intentionally "available" to the agent, or, if the object changes location during the agent's absence, the object becomes intentionally "unavailable" (see Gómez, 2008, for an elaboration of the idea of "intentional availability"; and see also Apperly and Butterfield, 2009, for their similar notion of 'registration' of events).

It is therefore possible to try to re-interpret Kovacs et al. study from an object-directed intentionality perspective.

To explain Kovács et al's results from this perspective, we suggest that object-directedness to the target from either the agent or the participant is automatically registered by the participants when the hiding process of the ball is witnessed. The intentional relation featuring the hidden object is accessible when the hiding-screen opens and it helps enhance the participants' response to the presence of the object. We emphasise the importance of registering the intentionality to an object, rather than the understanding of the other's belief. All conditions start with the coding of an intentional relation to the ball by both the observer and the agent. In the P-A- condition, the intentional relation is broken for both the agent and the observer when in the final part of the display, the ball leaves the scene, and therefore the appearance of the object at the end is more unexpected and the reaction time is the longest. But in Kovács et al's (2010) two key conditions (P+A- and P-A+), where the

participant's and the agent's exposure to the events conflict, one intentional relation with the object remains active — between the participant and the object in P+A- and between the agent and the object in P-A+, which evokes the participants' expectation of the object and leads to a similarly faster response in detecting the object. And, of course, in the P+A+ condition both the participant's own relation and the agent's relation to the object remain active and accelerate RT. The key finding is that when the participant knows that the object is gone, the fact that the agent did not witness this, makes the participant react as quickly as when he knows himself that the object is there. Our proposal is that this happens because the agent-object relation has not been updated in the absence of the agent, and it guarantees that a representation of the object behind the screen is still active in the mind of the participant. Under this hypothesis, the crucial factor is whether the events witnessed by the participants have led them to keep active in their record of the situations an intentional relation featuring the object. We will refer to this explanation as the object-directedness interpretation.

The alternative interpretation proposed by Kovács et al's focuses on the match of the participant's and the agent's belief and the outcome to respond to: when either the participant's or the agent's belief matches the outcome, i.e. the object's presence in the Kovács-experiment, it makes the presence of the object expected and therefore results in a shorter reaction time. To explain in more detail, in the P-A- condition, neither the agent's nor the participant's belief about the object is the same as the outcome, and therefore the participants' reaction time to the outcome is longer than the other three conditions. In the P-A+ condition, the agent's belief about the object being present shortens the reaction time compared with the P-A- condition; in the P+A- condition, the participants' belief about the object's presence shortens the reaction time in relation to P-A-; and in the P+A+ condition, both the participant's and the agent's correct and congruent beliefs produce the shortest RT. As long as one party, either the agent or the participant, holds a belief that is consistent with the final outcome when the hiding-screen opens, the participants' reaction time is shorter than when the outcome violates all the beliefs. We will refer to this hypothesis as the belief-outcome congruency interpretation.

Both the belief-outcome congruency interpretation and the object-directedness interpretation can explain the results. Kovács et al (2010)'s experiment cannot distinguish between these two interpretations. This is due to the fact that in their design the participants were only required to respond to the presence of the object, but never to the outcome of the object not being present. What was measured was their RT to the *presence* of the object when it was or not expected. As a result, when intentionality about the object is encoded, the belief also matches the outcome that the object is present. For example, in the P-A+ condition, the agent and the object form an intentional relation and at the same time, the agent's belief conforms with the reality of the object being present. In the P-A- condition, no object-directedness is coded, which leads to longer reaction time to the object's presence, but at the same time, the agent's and the participant's belief both contradict the outcome of the object's presence, which might also result in the same effect. Hence, when the target response is about the presence of the object, the two interpretations will predict the same results.

To disentangle the two possible interpretations, it is vital to design an experiment where the participants will behave differently according to the two hypotheses. A simple modification of Kovács et al's (2010) experiment can achieve this. By requiring the participants to respond to the object's absence, and not just to its presence, we expect different effects according to the two proposals (Figure 4 - 2). Firstly, the P+A+ and P-A- conditions will act as the two baseline conditions and the two mechanisms predict the same results that it will take longer to respond to the P+A+ than the P-A- condition. For the P-A- condition, as no object is expected from coding object-directedness and neither subject's belief contradicts the outcome, the participants' reaction time would be short since nothing violates their expectation. For the P+A+ condition, both parties had coded a relation to an object and both parties beliefs contradict the outcome, so the reaction time is long. Following the two different interpretations, the difference emerges between the P+A- and the P-A+ condition where the agent's and the participant's intentionality coding and beliefs conflict. When one of the subjects believes that the object is present (A+ or P+), although one of subject's belief conflicts with the outcome that the object is

absent, the other subject's belief matches the outcome. Therefore, according to the belief-outcome congruency interpretation, since one subject's belief is consistent with the outcome (A- or P-), this will lead to a shorter reaction time than the P+A+ condition (Figure 4-2.II.B), mirroring the results of the object present conditions. However, according to the object-directedness interpretation, the expectation of the object either by A+ or P+ will prolong the participants' reaction time, making it longer than the P-A- condition (Figure 4-2.II.C).

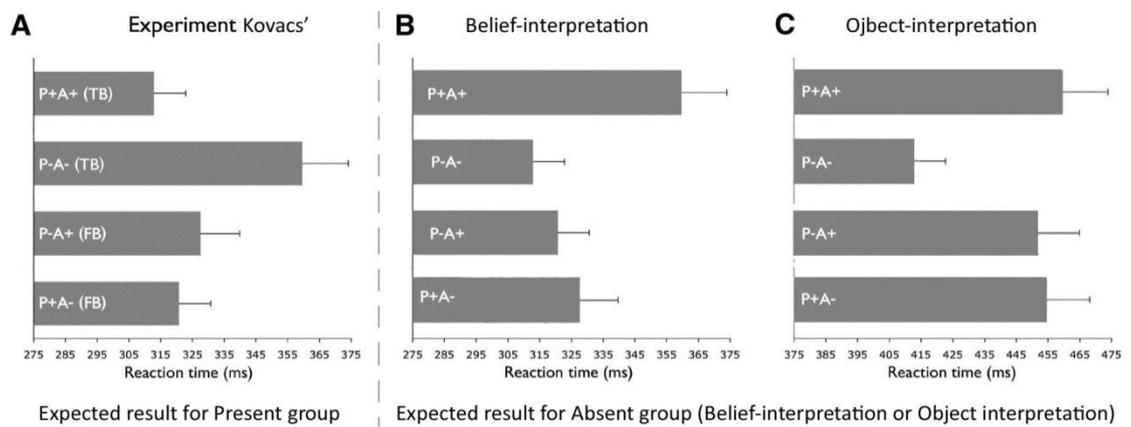


Figure 4 - 2 The no-object-target version and the two predicted results.
 Note: A. Kovács' original results in experiment 1. B. the predicted results of the no-object-target version task if coding belief explains their result better. C. the predicted results of the no-object-target version task if coding object explains their result better.

In addition, we suggest that the object-directedness interpretation is consistent with the results of the additional experiment with infants reported by Kovács et al's (2010). The authors used the same video presentations with young infants but a different measurement, looking time. In contrast to the adult study, in the infant experiments, they always showed the outcome where the object was absent, the one that we suggest could help in distinguishing the two interpretations. The infants' looking time in the tested conditions fits the object-directedness hypothesis (Figure 4-3. A, B). The looking time of infants was longer at P+A+ condition than the P-A- condition, as would be expected. More importantly, the looking time in the P-A+ condition is longer than that in the P-A- condition. In a control experiment where no outcome is shown, the difference between the P-A+ and P-A- condition is eliminated. The infant's performance in the experiment suggest

a promising direction that object-directedness might be the better interpretation of their results with adults than the belief-outcome congruency hypothesis. However, Kovács and colleagues never tested the infants in all four conditions, and therefore the contrast that we propose for each interpretation can not be tested with their published results. We believe that a direct comparison of all the four conditions is necessary to differentiate the two hypotheses.

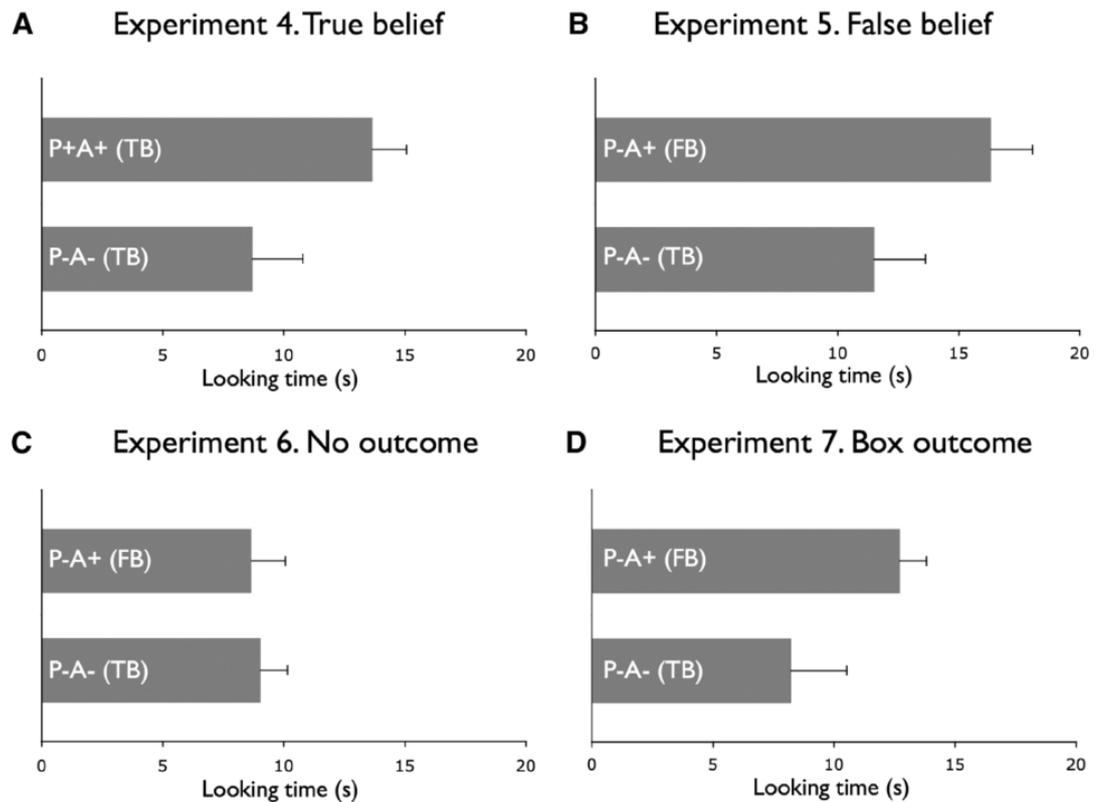


Figure 4 - 3 Results of Kovács et al's task with infants (Kovács, Téglás & Endress, 2010).

Therefore, based on the general idea and procedure of Kovács et al's experiment, we conducted the current study with adult humans, first, to replicate their positive results in a task that seemingly requires the coding of agents' relations with absent objects, and second to test whether the belief-outcome congruency interpretation or the object-directedness interpretation better explain the results.

4.2 Experiment 1

4.2.1 Method

4.2.1.1 Participants

A total of 48 adults (Female : Male = 39:9, average age = 21.8) were paid £3 each to participate in the current experiment. They were recruited from the participant-pool in the University of St Andrews. All participants were required to have normal or corrected-to-normal vision. None of them had participated in similar experiments beforehand. They were invited to a quiet room in the School of Psychology and Neuroscience to participate in the experiment. Ethical approval was granted by the ethical committee of the University of St Andrews.

They were randomly assigned to two groups, resulting in 24 participants in the group responding to the absence of the object and the other 24 in the other group responding to the presence of the object. Five participants were excluded from each group due to large number of incorrect responses, i.e., more than 10 incorrect responses (25%) in the test. The remaining 38 participants' (average age: 20.9; 7 males and 31 females) data were included in the data analysis.

4.2.1.2 Apparatus and Materials

They watched animated videos on an 18 inch flat-square CRT monitor (IIYAMA, LS902UT, Resolution: 1280*1204, Refresh Rate: 85Hz). The videos included an agent, a ball, a hiding-screen on a table and a bin under the table (see Figure 4 - 4) . The participants were required to press a key when certain events occurred.

The participants were shown videos on the computer monitor. All videos lasted for 16.8 second and had four phases (see Figure 4 - 4). In the first phase, the agent came into the scene and the ball rolled and hid behind the screen. Phase I lasted for 4.4 seconds. In the second phase, the agent stayed and watched the ball rolling in the scene. The final position of the balls was either behind the screen or in the bin. This phase lasted for 3 seconds. The agent left the scene at the beginning of the third phase. In his absence, the ball rolled again. The final position of the ball was either behind the screen or in the bin. The third phase was 4.6 seconds in total. The agent

returned at the beginning of the fourth phase and then the screen vanished, showing either the presence or the absence of the ball behind it. After the screen vanished, the scene stayed for 2.4 seconds before the trial's end. The final phase was 4.8 seconds in total. In addition, the ball could last be seen at 11.68s.

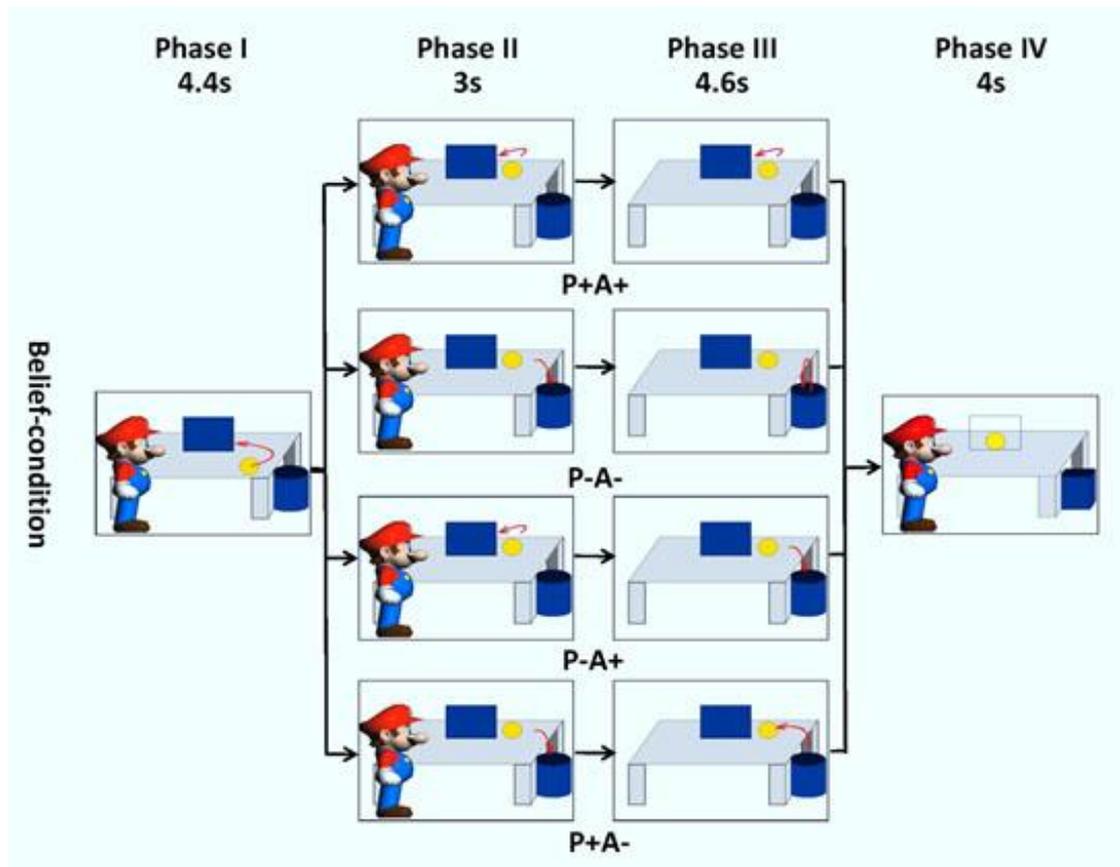


Figure 4 - 4 The 4 phases of videos and the 4 conditions. (P: participant, A: agent, +: believe the ball is behind the screen, -: believe the ball is absent behind the screen. The time (seconds) in the brackets after each Phase indicates the duration of Mario's movement and the ball/screen's movement).

The videos followed Kovács et al's (2010) general methods in manipulating the participant's and the agent's belief about the presence of the ball behind the hiding-screen. The first and the last phase were exactly the same in all conditions, but the second and the third phase varied determining the agent's and the participant's beliefs regarding the presence/absence of the ball behind the screen. Different combination of the participant's and the agent's beliefs resulted in four different conditions.

In Phase II of the P+A+ condition, the ball rolled out from behind the screen and then rolled back behind the screen. At the end of the phase, both the participant and the agent believed that the ball was behind the screen. When the agent was away in Phase III, the ball repeated those motions and stopped behind the screen. Therefore the reality did not change when the agent was absent and both the participant (P) and the Agent (A) still believed the ball was hidden behind the screen.

In Phase II of P-A- condition, the ball rolled into the bin in the second phase when the agent was watching, leading to the result that both the agent and the participant believed the ball was not behind the screen. In phase III, it rolled out and back into the bin when the agent was away. Thus in this condition, both the agent and the participant still believed the ball was not behind the screen.

In Phase II of P+A- condition, the ball rolled into the bin in the second phase and then rolled back from the bin to behind the screen in the third phase. In this condition, the agent's initial belief that the ball is behind the screen was updated to the ball not behind the screen; meanwhile, the participant's initial belief that the ball was not behind the screen was updated to the ball being behind the screen.

In Phase II of P-A+ condition, the ball first rolled out from behind the screen and then hid again behind the screen during the agent's presence, making the participant believe the ball to be behind the screen while the agent believed otherwise. Then the ball rolled into the bin when the agent was absent, leading to the agent's belief of the ball being behind the screen and the participant's belief of the ball not being behind the screen.

Though the videos were similar to Kovács et al's (2010), we made some modifications to ensure that the participants watched the same number of movements of the ball in each video. By doing this we were hoping to eliminate the possible difference of cognitive load for the participants if they watched the different number of events between conditions. For example, in Kovács et al's original experiment, the ball did not move at all in the P+A+ condition after the first phase, but in comparison, it moved twice, i.e., leaves the scene and returns to behind the

screen, in the P+A- condition. In our experiment, the ball always moved twice, once in phase II and once in phase III in all conditions.

4.2.1.3 Design and Procedure

The experiment was a 2 (Agent's Belief, AB hereafter, of object's presence/absence behind the screen) * 2 (Participant's belief, PB hereafter, of object's presence/absence behind the screen)* 2 (Group: respond to absence/presence) mixed design. AB and PB are two within subject factors and Group was a between subject factor.

We decided to use a between subject design, instead of a with-in subject block design, for three main reasons. Firstly, the switch of tasks between blocks of responding to the absence and presence might affect the participants' reaction time after switching, especially over the first few trials and possibly over the whole block. The between subject design would avoid this problem. Secondly, we wanted to avoid too many repeated trials and keep the participants free from any practise effect. Lastly, the group of participants responding to the presence of the object could act as a replication of Kovács et al's original experiment.

Our experimental procedure was based on Kovács et al's (2010) study. The participants were invited to the lab and were given verbal instructions by the experimenter (see appendix IV). Half of the participants were required to press the right key on the keyboard with their right hand when the screen vanished and they detected the ball behind the screen (presence group); the other half were instructed to press the same key when the screen vanished and they detected the absence of the ball (absence group). They were asked to respond as quickly as possible and were aware that their reaction time would be recorded. Again following Kovacs et al., all participants were also required to press a left key with their left hand when they detected the agent leaving the scene in the video, to ensure that they paid attention to this key event.

The task started with 4 practise trials. After the practise trials, all participants were presented with 5 blocks of experimental trials. Each block contained one trial of each condition with randomized order. Therefore the participants each had 5

repeats of the four different belief-conditions paired with two outcomes, leading to 40 trials in total in the experiment. The experimental process was controlled by E-prime.

4.2.1.4 Data analysis

E-prime recorded the participants' reaction time in each condition. The recorded time was relative to the start of the videos. We then subtracted the time from the beginning of videos to the moment the hiding-screen opens to calculate the reaction time to the key event data. We excluded all trials that were incorrect (4.2% for the present group and 3.4% for the absent group) and then calculated the participants' average reaction time in each condition.

For statistical analysis, we first examined the effect of Groups on the participants' reaction time by a mixed factor three-way repeated measure ANOVA (GLMM), with Group (Group: Absence/Presence) as the between subject factor, Agent's belief (AB: A+/A-) and Participant's belief (PB: P+/P-) as two within subject factors.

Since the task for the two groups was different, we then examined the two groups separately. Using two two-way repeated measure ANOVAs, with Agent's belief (AB: A+/A-) and Participant's belief (PB: P+/P-) as two within subject factors, we were hoping to examine the effect of the participant's own belief and the agent's belief on their reaction time.

We then used a one-way ANOVA with the four conditions as the only factor, to compare the four conditions directly. Paired T-test with Bonferroni adjustment were performed to compare every pair of conditions.

In addition, hoping to replicate Kovács et al's (2010) finding, we adopted their method of data analysis and examined the two groups of participants separately again. Following their method, we regarded the P-A- condition, where neither believed the ball was behind the screen, as the baseline condition and used paired T-tests to compare with the other conditions for the presence group. For the absent

group, except for using P-A- as the baseline, we also used P+A+ as the baseline condition to distinguish the interpretation of belief from object-directness.

4.2.2 Results

Figure 4 - 5 and Table 4 - 1 illustrate the participants' reaction time to different conditions .

Table 4 - 1 Two groups' average reaction time (ms) to the four conditions in experiment one

Group		P-A-	P+A-	P-A+	P+A+
Absent	Mean	389.2	380.8	361.4	361.4
	SD	77.1	84.0	78.3	66.7
Present	Mean	361.1	336.2	350.1	329.2
	SD	92.8	94.4	80.0	81.8

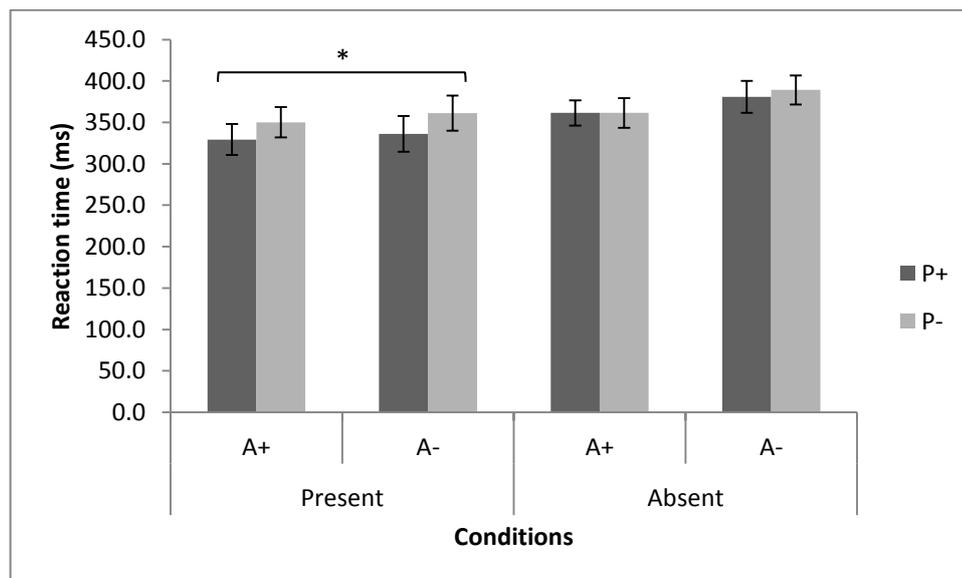


Figure 4 - 5 Two groups' average reaction time (ms) to the four conditions in experiment one (error bar: SE)

4.2.2.1 Responding to the presence/absence of the object

The mixed factor three-way repeated measure ANOVA indicated that the participants were equally good at detecting the presence or the absence of the object: Group ($F(1,36) = 1.52, p = .22$). No interaction was revealed, suggesting that reacting to the presence or absence of the object did not have a different effect in

the belief-conditions: Group*AB*PB ($F(1,36) = .017, p = .90$); AB*PB ($F(1,36) = .15, p = .70$); AB*Group ($F(1,36) = 1.34, p = .26$); PB*Group ($F(1,36) = 1.59, p = .22$).

Surprisingly, when the two groups were combined together, if the agent believed the ball was present, the participants' response was faster regardless of whether they were responding to the presence or absence of the ball ($F(1,36) = 6.69, p = .014$). The main effect of participant's belief was approaching significant too: PB ($F(1,36) = 3.34, p = .076$), indicating a longer reaction time when the participant's believed the ball was absent.

4.2.2.2 Effect of Participant and agents' Belief in each group

As the tasks for the two groups were actually different, although we found no interaction between Group and other factors, we examined the two groups separately. Each group was tested using two-way ANOVAs with AB (Agent belief) and PB (Participant belief) as the two factors.

For the presence group, the participants responded significantly faster when they believed the object was behind the screen ($F(1,18) = 9.13, p = .007$). But the agent's belief did not affect the participants' reaction time ($F(1,18) = 1.27, p = .27$) and there was no interaction between the two factors ($F(1,18) = .033, p = .86$).

For the absence group, however, the participant's own belief did not affect their response to the absence of the object ($F(1,18) = .11, p = .75$). But surprisingly, if the agent believed the object was not behind the screen, the participants' reaction time was longer at the absence of the object than when it believed it was present ($F(1,18) = 5.85, p = .026$). Again, no interaction was found ($F(1,18) = .13, p = .73$).

4.2.2.3 Comparison of the four conditions

To directly compare the four conditions, instead of using AB and PB as two different factors, we used a one-way ANOVA to examine the two groups separately.

In the group responding to the presence of the object, the difference of reaction time in the four conditions was approaching significance ($F(3, 54) = 2.46, p = .072$). Therefore we compared each pairs of conditions to investigate where the potential difference was. The t ($df=18$) values and p values with Bonferroni

adjustment are shown in Table 4 - 2. The results indicated that the reaction time in the P-A- condition was significantly longer than that in the P+A+ condition ($t(18) = 3.06, p = .040$) suggesting that the belief of the participant and the agent combined together may have affected the participants' reaction time, as we had expected.

In the Absence group, the difference between conditions was not significant ($F(3.54) = 1.51, p = .22$), and no RTs between pairs of conditions were significantly different (Table 4 - 2).

Table 4 - 2 Results of the paired T-tests in experiment one

	Pair of conditions	t(18)	p(2-tailed)	
			Bonferroni adjustment	no adjustment
Presence	P-A- vs P+A-	1.54	.84	.14
	P-A- vs P-A+	.78	1.00	.45
	P-A- vs P+A+	3.06	.040*	.007***
	P+A- vs P-A+	-1.20	1.00	.25
	P+A- vs P+A+	.52	1.00	.61
	P-A+ vs P+A+	2.02	.35	.058 ⁺
Absence	P-A- vs P+A-	.43	1.00	.68
	P-A- vs P-A+	2.18	.26	.043*
	P-A- vs P+A+	2.25	.22	.037*
	P+A- vs P-A+	1.02	1.00	.32
	P+A- vs P+A+	1.12	1.00	.28
	P-A+ vs P+A+	.004	1.00	1.00

4.2.2.4 Following Kovács' approach

We followed Kovács et al's approach and examined the group responding to the presence of the object using paired T-test. According to Kovács et al's (2010) results, the participants in this group would respond slower to the three conditions (P+A+, P-A+ and P+A-) where at least either the participant or the agent believed that ball was behind the screen than to the baseline condition (P-A-), where neither believed the ball was behind the screen. In our experiment, paired T-tests with no

adjustment (Table 4 - 2) showed that the participant's RT was longer in P-A- than P+A+ condition ($t(18) = 3.06, p = .007$) which conforms with our expectation and Kovács et al's results. But crucially no difference was found for the other conditions compared with the baseline: P-A+ ($t(18) = .78, p = .45$); P+A- ($t(18) = 1.54, p = .14$). In addition, the reaction time in P-A+ condition was marginally significantly longer than that in the P+A+ condition ($t(18) = 2.02, p = .058$). It indicated that the true belief of the participant, which contradicts the outcome of the object's presence, might have prolonged the participants' response.

For the Group responding to the object's absence, the P+A+ condition is the more surprising condition from the belief interpretation (Kovács et al, 2010). In this condition the agent and the participant's beliefs both conflict with the reality and, therefore, it should prolong the RTs. Using RT in P+A+ as the baseline, the belief interpretation would predict shorter reaction time in the other conditions, where at least one subject's belief was congruent with reality. Paired T-tests revealed that in fact the participants responded more quickly in the P+A+ condition ($t(18) = 2.25, p = .037$), than P-A-, which contradicts the expectation by the belief-interpretation. No difference was found between P+A- ($t(18) = .43, p = .68$) or P-A+ ($t(18) = .004, p = 1.00$) and the baseline.

Following the object-directed interpretation, on the other hand, the registration of the relationship between an agent and the object affects the participant's reaction. In the case of responding to the absence of the object, the P-A- condition becomes the baseline since only in this condition no registration was available, whereas for the other three conditions we expected a slowing down in the participants' response. Paired T-tests revealed a significantly shorter RT to the P+A+ ($t(18) = 2.25, p = .037$), and P-A ($t(18) = 2.18, p = .043$) condition than the baseline, which are both in the opposite direction to the prediction of the object-directed interpretation. RT to P+A- was not different from to P-A- ($t(18) = .43, p = .68$).

4.2.3 Discussion

In this experiment, we again tried to investigate the question whether adult human understand object-directedness implicitly. By modifying Kovács et al's (2010)

experiment, we expected to differentiate whether a belief-outcome congruency interpretation and an object-directedness interpretation. However, the results did not support either the object-directedness interpretation or the belief-outcome congruency interpretation; moreover, we failed to replicate the original results found by Kovács and colleagues.

4.2.3.1 Interpretation of the results

The group responding to the presence of the object was a replication of Kovács et al's (2010) experiment, but we were able to replicate only part of their results: the participants' behaviour was only affected when the agent's and the participant's beliefs were combined together, but one single party's belief was not strong enough to affect the participant's reaction time. When we followed Kovács et al's (2010) statistical analysis more closely and conducted multiple paired T-tests without adjustment, we discovered that the participant's own belief affected their reaction time, as found by Kovács et al. But the agent's belief alone did not have the same effect, which provides the most important evidence that the participant can automatically understand other's belief or register intentionality. We therefore failed to replicate these authors' main findings.

The results from the group responding to the absence of the object were even more surprising. The lack of difference between the two baseline conditions, the P-A- and the P+A+, did not allow us to build the benchmarks of reaction time for the other two conditions, and the overall lack of difference across the four conditions did not support either the belief or the object-directed interpretations. Counter-intuitively, the participant's own belief does not affect their RT at all in this task. When the participants know that the object has left the hiding place, we expected them to respond more quickly as the outcome should match their expectation. The lack of effect may indicate that they failed to engage in the experimental scenario, i.e., they might not have paid enough attention to the movements and location of the object, thereby failing to register the location of the object.

To briefly conclude, although we followed the general procedure of Kovács et al's (2010), we obtain different results and failed to reach their conclusion that

other's belief automatically affect one's own belief. The results from the object-absence group are also unexpected and cannot help in distinguishing the belief-outcome congruency interpretation from the object-directedness interpretation.

4.2.3.2. Comparison of the methodology with the Kovács experiment

The most significant difference between our video clip and the original clip used by Kovacs et al. is the stronger relationship between the agent and the object in Kovács et al's. The so-called irrelevant agent in their experiment could in fact be described as the "owner" of the object, because he drops the object on the table in the scenario in the first place. This initial relationship to the object may have drawn more attention to the agent's intentional relation to the object in the original study. In comparison, the agent in our experiment is truly irrelevant to the object, i.e., he only acts as a bystander and never interacts with the object directly besides watching in the direction of the hiding screen. Our scenario was targeting an intentional relation of witnessing or somehow perceiving the whereabouts of the object. However, it is possible that only when a strong relation between the agent and the object involving an action (even if it is one of dropping or letting go) is observed, will participants code object-directedness and expect the object to appear together with the agent.

Secondly, our method to open the hiding-screen is different from Kovács et al's (2010). In our experiment, the hiding-screen vanishes, leaving the frame of the screen still visible to indicate its previous location (see Figure 4 - 4); while in Kovács et al's videos, the whole screen falls forward. The remaining frame in our experiment may have retained the participants' attention in the object-area and led to their ignoring of the agent and its perspective.

Thirdly, the procedures of our video presentation in the four conditions are different from Kovács et al's (2010). To ensure that the participants experience the same number of movements of the ball, the ball always moved once when the agent is present and once again when the agent is absent, whereas in Kovacs et al. the agent leaves the scenario at different time points in relation to its return at the end in each condition. In addition, the agent is designed to leave at the same time point

in all conditions in our experiment. These two modifications were aimed to improve on the original study by standardizing the procedure and events in the videos of the four conditions and reducing the possibility that the participants are reacting to these secondary differences rather than the experimental manipulations. It could be argued that the original design has its pros, i.e., the movement of the object is more continuous, e.g., in P+A+ and P-A- condition, the ball's movement is not interrupted by the agent leaving. The continuous movement of the ball might have provided the participants with a better understanding of the ball's position; while the two movements, once before the agent's leaving and once after, in our experiment might have made the participants confused about the position of the object. The increased cognitive load might disrupt implicit ToM processing (Schneider, Lam, Bayliss, & Dux, 2012). But the cons of this design is that the differences in the video presentation in different conditions might bring in potential confounds that may affect the participants' performance

In addition, we used a different agent from Kovács et al's (2010) experiment. However, they are both animated characters and should not have affected the results if belief or object-directedness is the crucial factor that influences the reaction time.

To examine the difference between our experiment and Kovács et al's original design that might have led to the different results, we conducted four follow-up experiments with adult humans. Before pursuing our exploration of the belief-outcome congruency interpretation versus the object-directedness interpretation in their paradigm we need to be able to replicate Kovács et al's (2010) results. Therefore in the follow up experiments, we only test participants' response to the presence of the object, as in the original study, while trying different variations in the experimental design to pinpoint which one is the crucial one in order to replicate Kovacs et al original studies.

4.3 Experiment 2

4.3.1 Introduction

In the current experiment, we sought to replicate Kovács et al's method (2010) more closely by strengthening the relationship between the agent and the object. In a sense, the lack of effect with the purely observational version of the clips in which the agent just witnesses the object movements, without a more direct involvement with the object, may be taken to favour an object-directedness interpretation over a Belief computation interpretation. It is because more clear signs of there being a relation of the agent to the object would be required. To test this possibility, we edited the videos from the last experiment by adding to the beginning a sequence where the agent puts the ball on the table before it starts rolling, so that the agent had the "ownership" to the object as in Kovács et al's original design. We expected the stronger relation of the agent and the object would enhance the participants' attention to the experimental manipulations and lead to a replication of Kovacs et al results.

We also modified the opening method of the hiding-screen to make it closer to Kovács et al's original design. When revealing the presence/absence of the object, we changed the screen from vanishing to falling forward, and also eliminated the remaining frame that suggested the position of the screen in the last experiment. With these modifications, we expected to avoid any undesired retention of the participants' attention within the area of the hiding screen allowing them to pay closer attention to the whole scenario.

4.3.2 Method

4.3.2.1 Participants

A total of 31 adults (Female : Male = 18:13, average age = 21.1) was paid £3 each to participate in the current experiment. They were recruited from the participant-pool in the University of St Andrews. All participants were required to have normal or corrected-to-normal vision. None of them had participated in similar experiments beforehand. They were invited to a quiet room in the School of

Psychology and Neuroscience to participate in the experiment. Ethical approval was granted by the ethical committee of University of St Andrews.

Two of the participants were excluded from the data analysis: one participant due to program failure and another due to a large number of incorrect responses, i.e., more than 10 incorrect responses (25%) in the test. The remaining 29 participants, including 13 males and 16 females, at the average age of 21.2 years, were included in the data analysis.

4.3.2.2 Apparatus and Materials

The apparatus was the same as in the last experiment. The materials were short animation videos, identical to those in the last experiment, except that at the beginning of each video, the agent came in with the ball in his hands and put it on the table; and at the end of each video, the screen fell forward instead of vanishing to a frame.

4.3.2.3 Design and Procedure

Aiming to replicate Kovács et al's (2010) experiment, we only tested one group of participants, all responding to the presence of the object. The design was a 2 (Agent's Belief: object presence/absence behind the screen) by 2 (Participant's belief: object presence/absence behind the screen) factorial design, resulting in 4 conditions (P+A+, P-A+, P+A-, P-A-) in total.

The procedure was identical to the last experiment, and in addition, participants were asked afterwards whether they knew the purpose of this experiment.

4.3.2.4 Data analysis

All the incorrect trials (2.1%) were excluded from the data analysis. Reaction time was calculated in the same way as in experiment one.

For statistical analysis, we first conducted a two-way repeated ANOVA to examine whether the participant's belief and the agent's belief had any effect on participants' reaction time to the object's presence. Then we compared the four conditions directly using one-way repeated ANOVA. Lastly, we followed Kovács et

al's (2010) approach and used t-tests without correction to compare the baseline, P-A-, to the other conditions.

In addition, we compared the current experiment results and the the results of the presence-group in experiment one to better determine the effect of the "ownership" relation. We conducted a two-way mixed ANOVA using the four conditions as with-in subject factors and Ownership as a between subject factor.

4.3.3 Results

4.3.3.1 Comparison of the four conditions

Figure 4 - 6 and Table 4 - 3 show the participants' average reaction time to the different conditions. The two-way ANOVA reveals no main effect of PB ($F(1,28) = 2.68, p = .11$), nor of AB ($F(1,28) = 2.05, p = .16$). But the interaction is significant ($F(1,28) = 4.28, p = .048$), indicating AB affects the participant's response differently with the two types of PB.

Table 4 - 3 Participants' average reaction time (ms) to the four conditions in experiment two

	P-A-	P+A-	P-A+	P+A+
Mean	382.5	351.2	355.9	362.1
SD	73.0	64.3	61.0	106.4

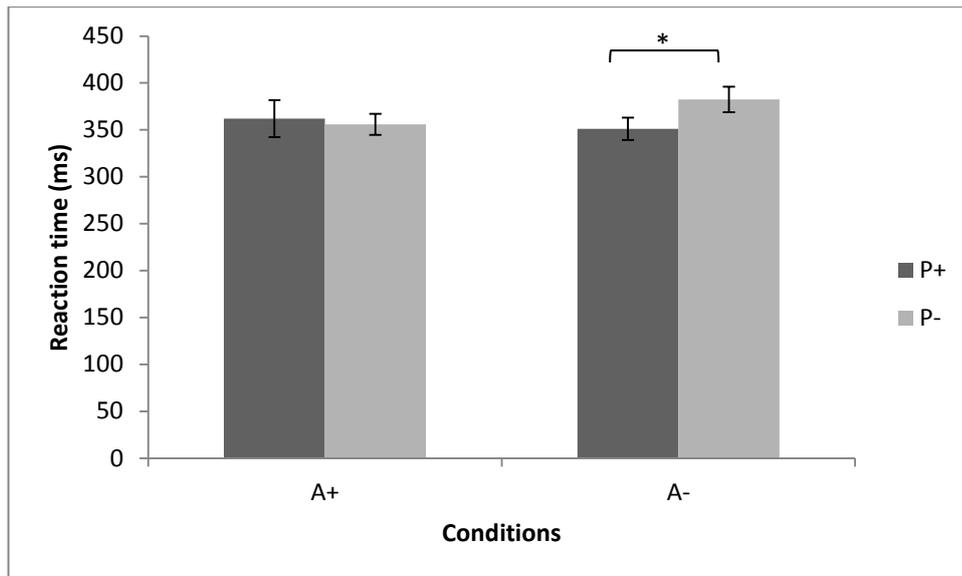


Figure 4 - 6 Participants' average reaction time (ms) to the four conditions in experiment two (error bar: SE)

To examine the interaction in detail, we conducted a one-way ANOVA using the four experimental conditions as the only factor. Mauchly's test reveals that the assumption of sphericity is violated ($\chi^2(5) = 24.25, p < .001$). Therefore we report the p-values with Greenhouse-Geisser adjustment here. The reaction time in the four conditions are significantly different ($F(3,84) = 3.35, p = .044$). We then used paired T-test with Bonferroni adjustment to examine the difference between each pair of conditions (see Table 4 - 4). The tests reveal that RT in P-A- condition is significantly longer than that in P+A- condition ($t(28) = 4.03, p = .002$). Although RT in P-A- seems longer than that in P-A+ ($t(28) = 2.64, p = .081$), and P+A ($t(28) = 1.82, p = .47$), neither is statistically significant. No significant difference is detected between P+A- and P-A+ either.

Table 4 - 4 Results of the paired T-tests in experiment two

Pair of conditions	t(28)	P (2-tailed)	
		Bonferroni adjustment	No adjustment
P-A- vs P+A-	4.03	.002*	.000***
P-A- vs P-A+	2.64	.081	.014*
P-A- vs P+A+	1.82	.47	.080

P+A- vs P+A+	-0.99	1.00	.33
P-A+ vs P+A+	-0.42	1.00	.68
P+A- vs P-A+	-0.66	1.00	.52

4.3.3.2 Following Kovács' approach

Again we followed Kovács et al's method (2010) and used P-A- as the baseline condition. We compared the other conditions to the baseline using paired T-test without adjustment (see Table 4 - 4). Reaction time in P+A- condition and P-A+ condition are both significantly shorter than the baseline, P+A- vs P-A- ($t(28) = 4.03$, $p < .001$); P-A+ vs P-A ($t(28) = 2.64$, $p = .014$), however, surprisingly, the difference between P-A- and P+A+ is approaching significance ($t(28) = 1.82$, $p = .08$).

4.3.3.3 Comparison of the two experiments

To investigate whether the ownership and the method of opening the hiding screen had any effect on the participants' reaction time to the four conditions, we conducted a two-ways mixed ANOVA using the four conditions as a with-in subject factor and the two experiments as a between-subject factor. Mauchly's test shows the assumption of Sphericity is not met ($\chi^2(5) = 13.14$, $p = .022$), so we report the ANOVA with Greenhouser-Geisser adjustment. The mixed ANOVA reveals a significant main effect of conditions ($F(3,138) = 4.67$, $p = .007$). But the interaction between experiment and conditions is not significant ($F(3,138) = .93$, $p = .44$), indicating the manipulation between experiments did not bring about any effect.

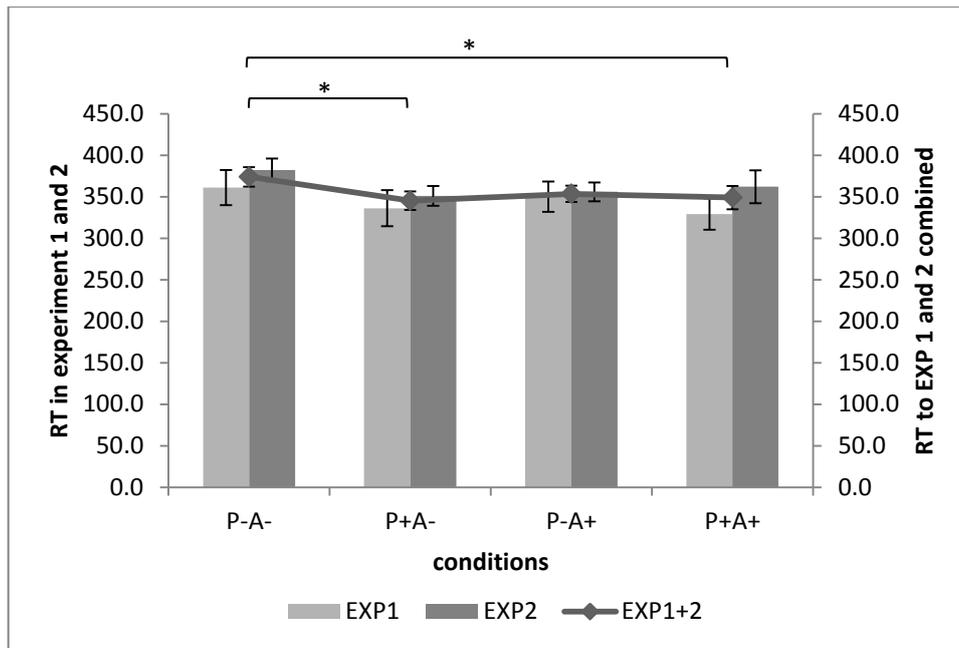


Figure 4 - 7 Participants' average reaction time (ms) to the four conditions in experiment one and two combined (error bar: SE)

Therefore we combined the data in the two experiments (see Figure 4 - 7 and Table 4 - 5) and analysed the effect of conditions using a one-way repeated ANOVA. Mauchly's test suggests the assumption of sphericity is not met; therefore the following analysis is adjusted using Greenhouse-Geisser adjustment. The reaction times in the four conditions are significantly different, $F(3,141) = 4.89, p = .005$.

Paired T-tests with Bonferroni adjustment show significant longer reaction time in P-A- than P+A- ($t(47) = 3.67, p = .004$), and P+A+ ($t(47) = 3.16, p = .016$), but not than that in P-A+ ($t(47) = 2.47, p = .10$). No difference is found between other pairs of conditions (see Table 4 - 6).

When no adjustment is made, following Kovács et al (2010), the results are the same as Kovács et al 's findings, i.e., the reaction time in P-A- is significantly longer than in the other three conditions (see Table 4 - 6).

Table 4 - 5 Participants' average reaction time (ms) to the four conditions in experiment one and two combined

	P-A-	P+A-	P-A+	P+A+
Mean	374.0	345.3	353.6	349.1
SD	81.2	77.0	68.4	97.8

Table 4 - 6 Results of the paired T-tests for experiment one and two combined

Pair of conditions	t(47)	P (2-tailed)	
		Bonferroni adjustment	No adjustment
P-A- vs P+A-	3.67	.004*	.001***
P-A- vs P-A+	2.47	.10	.017*
P-A- vs P+A+	3.16	.016*	.003*
P+A- vs P+A+	-.45	1.00	.66
P-A+ vs P+A+	.46	1.00	.65
P+A- vs P-A+	.46	1.00	.19

4.3.4 Discussion

4.3.4.1 Ownership and opening of screen are not the key factor

By adding the agent putting the ball on the table, we sought to strengthen the relation between the agent and the object and replicate Kovacs et al findings. However, the strengthened relation did not generate the expected effect, i.e., the participants did not automatically register intentionality or the agent's belief, in instead confirming the negative results found in experiment 1. Changing the opening method of the hiding screen did not change the results from experiment 1 either. When comparing the findings of the current experiment with the presence-group in experiment 1, the reaction time pattern to the four conditions was the same. Therefore we exclude the possibility that the failure to replicate Kovács et al's experiment in experiment 1 was due to these two factors in the video stimuli.

However, the results in the current experiment better replicated Kovács et al's results when exactly the same statistical tests were used (i.e., comparisons without corrections), we found consistent results with Kovács et al's findings: when either

the participant or the agent believed the ball was present (P+A- and P-A+), the participants responded faster than when both of them believed the ball was absent. The replication, however, was not perfect because the reaction time was not faster in the P+A+ condition (the one where both subjects hold the correct belief) than the baseline P-A- (both subjects hold the incorrect belief), questioning the interpretation that the participants are automatically encoding the belief of the agent.

4.3.4.2 The two experiments together replicates Kovács et al's findings

Given that the modifications in stimuli between the two experiments did not change the pattern of reaction time to the four conditions, we combined the data from both experiments to enlarge the sample size and found, with 48 participants in total, the same results as Kovács et al's (2010), but only if we follow their approach of statistical analysis. As Kovács et al's findings, the three non-baseline conditions (P+A+, P+A-, P-A+) affected the participants' behaviour by enhancing their reaction time, indicating that the agent and the participants' beliefs both modulated their reaction time to the presence of the object.

We, however, have some doubts about Kovács et al's approach to data analysis. Kovács et al's approach of statistical testing might be problematic in not adjusting the significance level for multiple-paired T-test. The difference between P-A- and P-A+ was only significant when no adjustments were made. However, the lack of adjustment may have increased the possibility of detecting a significant difference by chance. In comparison, when a Bonferroni adjustment for multi-paired T-test were made, the significance-level dropped from significance ($p = 0.017$) to non-significance ($p = 0.10$), suggesting that the agent's belief did not affect the participant's reaction time. Although there was still a trend suggesting that the reaction time in P-A- was longer than P-A+, notably the sample size was twice the size of that as in Kovács et al's experiment while the effect of condition was not as robust. It indicates the possibility that the agent's belief alone was not comparable to the participant's belief in affecting the participants' behaviour.

4.3.4.3 Summary

To conclude, the current experiment alone again failed to replicate Kovács et al's findings. Although the first two experiments put together replicated Kovács et al's findings, this was only possible when using their approach of data analysis without Bonferroni corrections, which casts some doubt over the validity of their interpretation, and at the very least suggests the effects are extremely fragile.

4.4 Experiment 3

4.4.1 Introduction

Our previous two experiments both replicate part of Kovács et al's (2010) results but neither successfully replicated all. Moreover, when combined, there still seems to be only some evidence supporting their findings, and the results were not as robust as those originally reported. Considering that they obtained robust results from three experiments using the same method, we have to be cautious before reaching a different conclusion. As discussed after experiment 1, the procedure for each condition in our experiments is not exactly the same as that in their experiments and it might have led to the different results. In the current experiment, therefore, we try to conduct the closest replication of the procedure and examine again whether we could replicate their results. We modified the videos in experiment 2 and created a set of stimulus that had the same procedure as Kovács et al's (2010) experiment.

4.4.2 Method

4.4.2.1 Participants

A total of 32 adults (Female : Male = 24:8, average age = 21.2) recruited from the participant-pool in the University of St Andrews took part in the experiment. One participant was excluded due to program failure and another was excluded due to pre-knowledge of the purpose. Thirty of them (average age = 21.0; Female : Male = 22:8) were included in the data analysis.

All participants reported normal or corrected-to-normal vision. They were naive to the purpose of the study and had never participated in our previous

experiments. And after the experiment, they all reported not knowing the purpose of this experiment. The participants each received £3 compensation for their participation. Ethical approval was granted by the ethical committee of the University of St Andrews.

4.4.2.2 Apparatus and Materials

The apparatus were the same as in the last experiment. The materials were short animation videos, which used the same agent from the last experiment. But the table is longer, lying across the scenario and the bin was removed as in Kovács et al's experiment (see Figure 4 - 8). The procedure in each video was also different from the last experiment and was the same as Kovács et al's (2010) experiment (Figure 4 - 8).

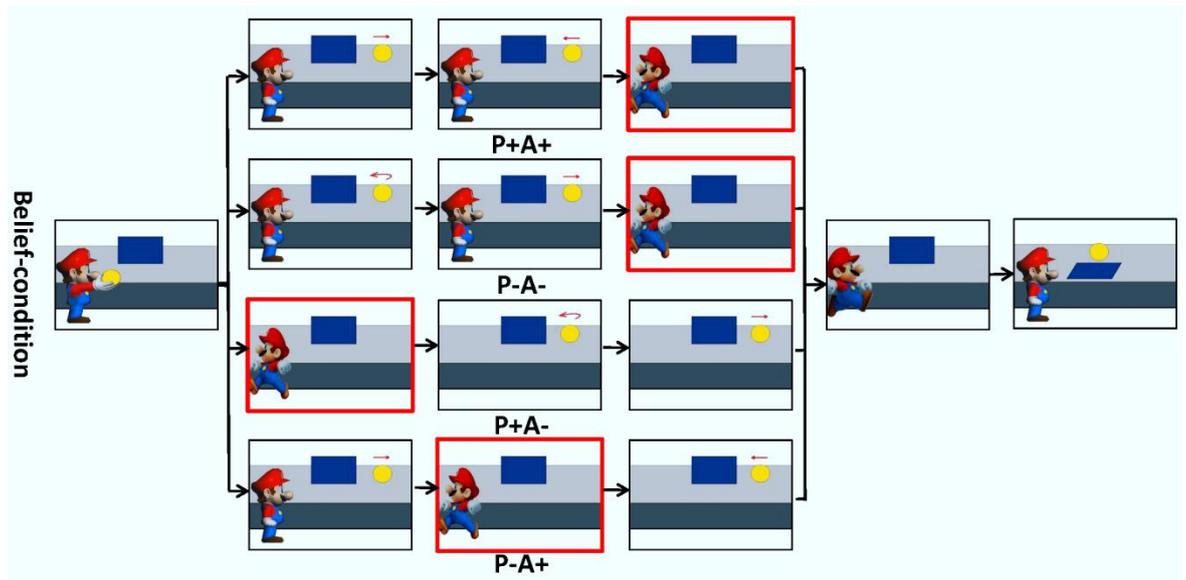


Figure 4 - 8 Procedure of the videos in each condition in experiment three (Note: the red frame shows when the agent leaves the scenario)

In the first phase, the agent put the ball on the table and the ball hid behind the hiding screen. Then the different events regarding the ball and the different attendance of the agent created the four conditions, described below.

In the P+A+ condition, the ball rolled out of the scenario and then rolled back behind the hiding-screen. During the process, the agent stayed and was watching all the time. When the ball stopped moving, the agent left the scene. Therefore both

the agent and the participant knew the ball should be present behind the hiding-screen.

In the P-A- condition, the ball rolled out of the screen and rolled back, then it came out of the hiding-screen again and rolled all the way out of the scenario. The agent was present during the events and then left the scene when the ball had stopped moving. Thus in this condition, the participant and the agent both witnessed the ball leaving the scenario, i.e., not present behind the hiding-screen.

In the P-A+ condition, the ball's movement was the same as that in the P-A- condition. But the agent left before the ball rolled out of the screen and was not present for any of the events after the initial hiding in the first phase. In this case, the participant has seen all the events and knew the ball was not present behind the screen; however, the agent only saw the ball's initial hide and should believe the ball is present behind the hiding screen.

In the P+A- condition, the ball moved the same way as that in the P+A+ condition. The agent was present when the ball first rolled out of the scenario and then left before the ball rolled back. Therefore, missing the ball's final move, the agent believed the ball was absent; witnessing all the moves, the participant believed the ball was present behind the screen.

At last, in all conditions the agent returned before the hiding-screen opened. When the hiding-screen opened, the participants were required to respond to the presence of the object.

4.4.2.3 Design and Procedure

Aiming to replicate Kovács et al's (2010) experiment first, we again only tested one group of participants, all responding to the presence of the object. The design was the same as in the last experiment. The procedure was equivalent to the last experiment.

4.4.2.4 Data analysis

Participant's reaction time was calculated as that in experiment one. Statistical analysis was similar as that in experiment two. The incorrect responses (2.2%) were excluded from the data analysis.

4.4.3 Results

4.4.3.1 Comparison of the four conditions

Figure 4 - 9 and Table 4 - 7 show the participants' average reaction time to the different conditions. A two-way ANOVA revealed that the effect of the participant's belief was approaching significance ($F(1,29) = 3.60, p = .068$) i.e., the participants responded faster if they themselves believed the object was present behind the screen. But no significant effect of the agent's belief was found ($F(1,29) = 1.68, p = .21$). More importantly, the agent's belief affected participant's reaction time differently when the participants' own belief was different ($F(1,29) = 34.15, p < .001$).

Table 4 - 7 Participants' average reaction time (ms) to the four conditions in experiment three

	P-A-	P+A-	P-A+	P+A+
Mean	463.0	403.0	404.1	430.6
SD	121.6	113.3	93.8	100.7

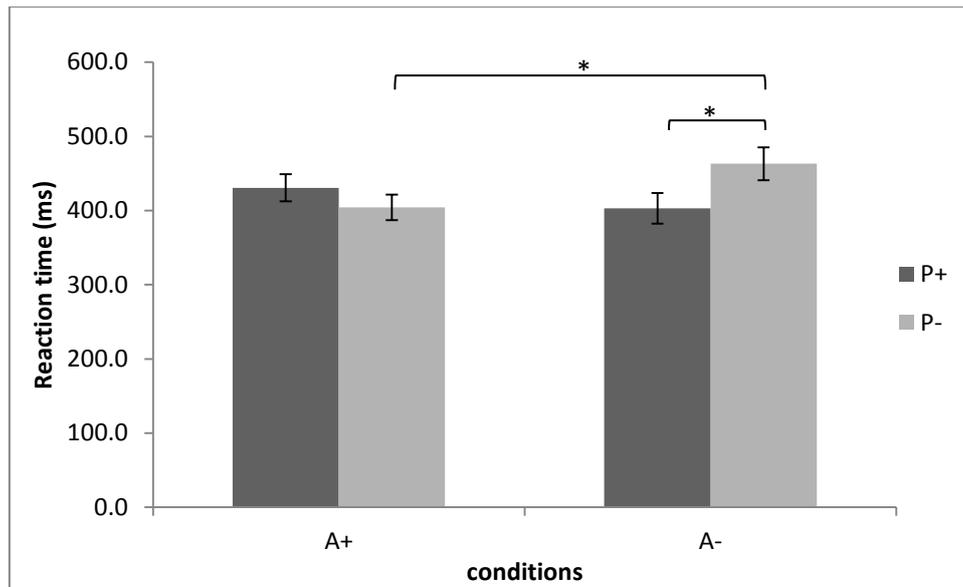


Figure 4 - 9 Participants' average reaction time (ms) to the four conditions in experiment three (error bar: SE)

Table 4 - 8 Results of the paired T-tests in experiment three

Pair of conditions	t(29)	P (2-tailed)	
		Bonferroni adjustment	No adjustment
P-A- vs P+A-	4.73	<.001***	<.001***
P-A- vs P-A+	3.44	.011*	.002*
P-A- vs P+A+	1.86	.44	.073
P+A- vs P+A+	-2.67	.074	.012*
P-A+ vs P+A+	-2.59	.089	.015*
P+A- vs P-A+	-.097	1.00	.92

4.4.3.2 Following Kovács' approach

We followed Kovács et al's method (2010) and conducted paired T-tests without adjustment using P-A- as the baseline condition. We compared the other conditions to the baseline and found that RT in the P+A- condition and the P-A+ condition were both significantly shorter than the baseline, P+A- vs P-A- ($t(29) = 4.73$, $p < .001$); P-A+ vs P-A- ($t(29) = 3.44$, $p = .002$), however, the RT in P-A- only had a trend to be longer than in the P+A+ condition ($t(29) = 1.86$, $p = .073$).

4.4.4 Discussion

4.4.4.1 *Partial replication of Kovács et al's results*

In the current experiment, we closely replicated Kovács et al's (2010) method and found some similar results. The participant's reaction seems to be affected by both their own belief and the agent's belief. When the participants or the agent believed the object was present, their reaction time was shorter than if neither of them believed so.

However, when both the agent and the participant believed the ball was present, the reaction time was not reduced. Although direct comparison using paired T-test (without adjustment for multiple comparison) showed a trend of longer reaction time in the P-A- than the P+A+, the difference did not reach significance. Moreover, when comparing multiple pairs, increasing numbers of separate comparison increases the chance of finding a significant difference that may only be attributed to chance, and therefore, making an adjustment is essential. But when Bonferroni adjustment is made, the difference between the reaction time in the P-A- and the P+A+ condition is far from significant ($p = 0.44$). These results are the same as those in experiment 2 when no adjustment for multiple comparisons is made for T-test in the last experiment. As has been discussed following experiment 2, due to the lack of difference between the P+A+ and P-A- condition, we argue that we should be cautious in reaching the conclusion that the positive results in the P+A- and P-A+ conditions indicate automatic registration of other's belief. Longer reaction time in the P-A- than the P+A+ is a crucial precondition for reaching the above conclusion because in these two conditions, the participant's and the agent's belief are not conflicting and the difference between both is the most basic assumption if beliefs affect the participants' behaviour.

The lack of robust difference between these two conditions when the proper statistical tests are conducted makes it confusing as to why the reaction times in P+A- and P-A+ condition are both shorter than that in the P+A+ condition. In our expectation, when both the participant and the agent believed the ball is present, their reaction time should at least be equivalent, if not faster than when one party

believes so. Therefore we again suggest that whether the participant's and the agent's belief is the key factor that affects participants' reaction time as suggested by Kovács et al (2010), needs further investigation.

To conclude the comparison of the current experiment to Kovács et al's, we replicated all their results when we follow their statistical analysis. But when a proper correction for multiple comparisons is applied, we are not able to replicate some of their results. The lack of difference between the two key conditions, P-A- and P+A+, suggests that we be cautious in reaching any conclusion regarding automatic belief registration.

4.4.4.2 The real factor that may have affected participants' reaction time

Looking back to our three experiments, the current one generated the closest result to Kovács et al's (2010) study. In comparison, the other two replicated fewer of Kovács et al's (2010) results, i.e., in experiment 1 the participants reacted only faster in P+A+ than P-A- and in experiment two participants responded only faster in P+A- than P-A-.

When examining the difference between the three experiments, the most significant modification is the change of the procedure of videos in the current experiment. It brings forward a question: is belief the direct reason to explain the difference of reaction time between conditions in the current experiment? Or could it be the specific procedure of the videos in the four conditions that led to the different reaction times?

Therefore we examine the different procedures closely here. In the first two experiments, i.e., experiment 1 and experiment 2, the order of events in the four conditions are identical. The ball always moves once when the agent is present and moved again after the agent leaves in each condition. And the agent always leaves at the same time point relative to his return. In these, the participant's reaction time is a little longer in P-A- condition than the other conditions.

In the current experiment (and in Kovács et al' original study), however, the agent leaves at different points in different conditions, which could potentially bring

about a difference in focus of attention considering the participants are also required to respond when the agent leaves. In P+A+ and P-A- condition, the ball has completed all movement before the agent leaves. So the participants' response to the agent's leaving is directly followed by the return of the agent and the opening of the screen. When the screen opens, the participant's attention might still remain at the agent, and in order to perform the reaction task about the object, they have to shift their attention from the agent to the screen area. The time to shift their attention may be the reason why the participants responded more slowly in the P-A- condition. And there also a trend that it takes longer for them to make the response in the P+A+ condition. Firstly, T-tests without adjustment indicated longer reaction times in the P+A+ condition than the P+A- and the P-A+ condition. Secondly, although the majority of the participants (18/30) responded slowest in P-A- condition, for the rest, two thirds (8/12) participants' RT is the longest in P+A+ condition (see Table 4 - 9). In comparison, in P-A+ condition, the ball completes all movements after the agent has left. Thus, the participants have a long enough interval between responding to the leaving of the agent and to the object when the screen opens so that they could have shifted their attention from the agent back to the ball or screen area before the final response. No participant in the current experiment responds slowest in this condition (see Table 4 - 9). In the P+A- condition, the ball moves both before and after the agent's leaving and the interval between the two reactions was shorter than P+A+ and P-A-, but longer than P+A-. Four participants' reaction times are the slowest in this condition.

We therefore suggest that the procedure of the conditions, rather than belief per se, might be a more direct factor that affects the participants' reaction time in the current experiment. A follow up experiment using the identical procedure of video presentations will help in confirming this suggestion. Hence in the next experiment, we will use a different scenario and agent, but the same experimental design to investigate this issue.

Table 4 - 9 Number of participants that responded slowest at the condition in experiment three

Condition	P-A-	P+A-	P-A+	P+A+
No. of participant	18	0	4	8

4.4.4.3 Summary

To sum up, whether adult humans can automatically register intentionality or other's belief still remains unclear. We have concerns that, in the current experiment, the specific procedure of the task rather than intentionality or belief is a more direct factor in affecting the participants' response.

4.5 Experiment 4

4.5.1 Introduction

As discussed in the last experiment, the longer reaction time in P-A- and P+A+ condition might have been caused by the specific procedures for the conditions. But before reaching this conclusion from a single experiment, we suggest further testing to confirm the results. In the current experiment, we will use the same procedure to replicate the last experiment.

Instead of conducting an exact replication, however, we will use a different agent in a different environment, but provide a similar scenario: the background and the table in the last experiment are replaced by green grassland; the hiding-screen is replaced by a wooden board; the agent by a capuchin monkey's profile picture and the ball by an orange (see Figure 4 - 10). This is for two main purposes. In the first place, it is a preparation for a future study (See Chapter 5) with monkeys as a comparison to the series of experiments in the current chapter. These stimuli are deemed to be more monkey-friendly and the current experiment will act as a better matched comparison to interpret the behaviour of the monkeys, as will be discussed in chapter five. On the other hand, it is included here because it also serves as a further attempt at replication of Kovács et al's results, and as a measure of the effect of the procedure difference in Kovács et al's paradigm.

4.5.2 Method

4.5.2.1 Participants

A total of 24 adults took part in the experiment (Female : Male = 23:1, average age = 21.2). They were recruited from the participant-pool in the University of St Andrews. All participants reported normal or corrected-to-normal vision. They were naive to the purpose of the study and had never participated in our previous experiments. They were invited to a quiet laboratory room in the School of Psychology and Neuroscience to participate in the experiment. After the experiment, they each received £3 compensation for their participation. Ethical approval was granted by the ethical committee of the University of St Andrews.

4.5.2.2 Apparatus and Materials

The apparatus were the same as in the last experiment.

The materials were short animation videos, which followed the method of Kovács et al's (2010) experiment and were similar to experiment 3. But as described, the agent and the scenario were replaced by a monkey-friendly version (Figure 4 - 10). In addition, the orange was initially emerged from one end of the scene on the grass, instead of putting down by the monkey.



Figure 4 - 10 Examples of the scenario and agent in experiment three

4.5.2.3 Design and Procedure

The design and procedure were equivalent to the last experiment. The participants had to identify the presence of the target object as soon as possible after watching the same set of manipulations as in the previous experiment.

4.5.2.4 Data analysis

Data analyses were similar to that in the last two experiments. All incorrect responses (3.6%) were excluded from the data analyse.

In addition, we examined the number of participants that responded the slowest in the four conditions as we analysed in the discussion of the last experiment in order to examine if the procedure in the videos affected the participants' reaction time.

4.5.3 Results

4.5.3.1 Comparison of the four conditions

Table 4 - 10 and Figure 4 - 11 show the participants' average reaction time to different conditions. A two-way ANOVA revealed that the participants' belief had a significant main effect ($F(1,23) = 7.14, p = .014$) and the main effect of the agent's belief was approaching, but did not reach, significance ($F(1,23) = 3.34, p = .81$). More importantly, there was an interaction between PB and AB: the agent's belief affected participant's reaction time differently when the participants' own belief was different ($F(1,23) = 18.89, p < .001$).

To examine the interaction, we compared the four conditions directly using repeated one-way ANOVA and the test showed a significant difference between conditions ($F(3, 69) = 10.88, p < .001$). Paired T-test with Bonferroni adjustment (see Table 4 - 11) revealed significant longer reaction time at the P-A- condition than the other three: P+A- ($t(23) = 4.65, p = .001$), P-A ($t(23) = 4.01, p = .013$), and P+A ($t(23) = 3.04, p = .035$). In addition, RT in P+A+ had a trend to be longer than that in P+A- ($t(23) = 2.80, p = .061$). No difference was found between P+A- and P-A+ or P+A+ and P-A+.

Table 4 - 10 Participants' average reaction time (ms) to the four conditions in experiment four

	P-A-	P+A-	P-A+	P+A+
Mean	428.6	365.5	376.8	392.2
SD	71.8	58.9	45.4	72.1

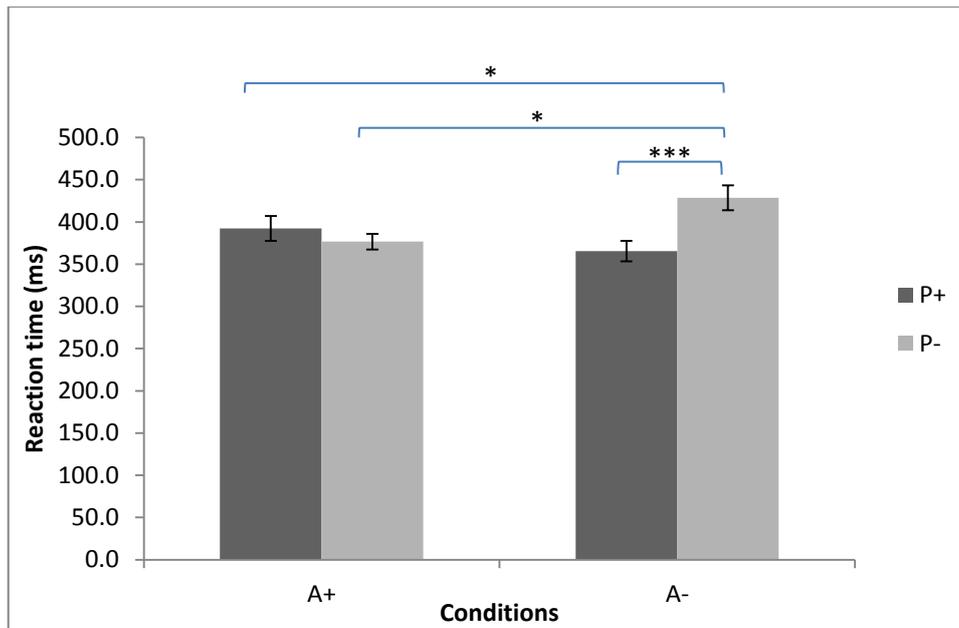


Figure 4 - 11 Participants' average reaction time (ms) to the four conditions in experiment four (error bar: SE)

Table 4 - 11 Results of the paired T-tests in experiment four

Pair of conditions	t(23)	P (2-tailed)	
		Bonferroni adjustment	No adjustment
P-A- vs P+A-	4.65	.001***	<.001***
P-A- vs P-A+	4.01	.003*	.001***
P-A- vs P+A+	3.04	.035*	.006*
P+A- vs P+A+	-2.80	.061	.010*
P-A+ vs P+A+	-1.32	1.00	.20
P+A- vs P-A+	-1.07	1.00	.29

4.5.3.2 Following Kovács' approach

We followed Kovács et al's method (2010) and conducted paired T-tests without adjustment using P-A- as the baseline condition. We compared the other conditions to the baseline and found that RT in P-A- was significantly longer than in the other three conditions (see Table 4 - 11). In addition, with this analysis, participants responded significantly more quickly in the P+A- than P+A+ condition ($t(23) = 1.32, p = .010$).

4.5.3.3 Each participant's RT

In addition, for each condition we counted the number of participants who responded most slowly in that condition (see Table 4 - 12) and found that, as we found in the last experiment, more than half, 14 out of 24, of the participants in the current experiment responded most slowly in P-A- condition. For the rest, half (5/10) participants' RT were the longest in P+A+ condition. Only five participants altogether took the longest time in either the P-A+ (3 of the 5) or the P+A- condition (2 of the 5).

Table 4 - 12 Number of participants that responded slowest at the condition in experiment four

Condition	P-A-	P+A-	P-A+	P+A+
No. of participant	14	3	2	5

4.5.4 Discussion

4.5.4.1 The procedure of video presentation, rather than intentionality or belief is the key factor

When the order of events remains unchanged and the scenario, i.e., the background, the agent and etc., is altered, we find similar results to our last experiment: besides the shorter reaction time to the P-A+ and P+A- than to the P-A- condition, the reaction time in the P+A+ condition is also shorter than the P-A- condition. Following Kovács et al's (2010) suggestion, these results indicate that the agent's belief is automatically registered by the participant. But looking back to the four experiments, we suggest an alternative explanation of these results. The change of procedure in experiment 3 and 4 has led to a robust pattern of reaction times to the four conditions, i.e., longest reaction time in P-A- condition, shortest reaction time in P+A- and P-A+, and an intermediary the reaction time in the P+A+ condition. The longer reaction time in the P+A+ condition than the P+A- and P-A+ condition is hard to explain by the registration of intentionality or belief: if intentionality or belief registration assists the participants to respond faster in this object-detection task, P+A+ (where both the agent's and the participant's beliefs match the outcome) this should show reduced reaction time with an equivalent, if not a larger, effect than the asymmetrical belief registration where only one subject, either the participant's or

the agent's belief matches the outcome. That was what Kovacs et al. reported in their original study. Our suggestion is that the RT effects found in experiments 3 and 4, as well as Kovács et al's original experiment might reflect differences in the procedure of video presentation, rather than in the registration of intentionality or belief.

In addition, at the initial stage of the current experiment, the orange was not placed on the grass by the monkey, but emerges from one side of the computer screen. That is to say, the ownership of monkey to the orange does not exist in the current experiment as in the third experiment. Hence the relation between the subject and the object is not as strong as in experiment 3. The weakened relation does not decrease the effect of different conditions in this experiment, suggesting that the ownership between the subject and the object is not essential, as has been suggested by experiment 1 and 2. It raises a further question: if the strength of relationship between the subject and the object is not important, is the intentional relation important to the effect at all? In other words, if the agent is not an intentional agent who does not have any belief at all, will the results turn out to be the same as in the current experiment? If so, the evidence would strongly argue that the procedure of the video presentation and its effect upon the participants' attention, instead of the registration of intentionality or belief, is the real cause of Kovács et al's findings (2010). In the next experiment, we will replace the agent with a stack of boxes to eliminate any attribution of intentionality and belief and investigate this issue.

4.6 Experiment 5

4.6.1 Introduction

Aiming to establish whether it is the structure of events shown in videos or belief computation that leads to the results in experiments 3 and 4, we modified experiment 3 and replaced the agent with a stack of boxes to eliminate the effect, if there is any, of the agent's belief.

In Kovács et al's original study, they also used a stack of boxes (B for the boxes) to replace the agent as a control condition. They reported that the participants' reaction time was not affected by any "belief" of the boxes, but was only affected by the participants' own belief (see Figure 4 - 12, Experiment 3). In contrast to their two experiments containing an agent, the pattern of reaction time in the P+B+, P-B- and P+B- conditions in the control experiment was identical to that of P+A+, P-A- and P+A-. But the reaction time in the P-B+ was not shorter than in P-B- like that between P-A+ and P-A-. Therefore Kovács and colleagues concluded that in the B+ condition, the participants did not attribute a belief to the boxes', as one would expect given that they are not agents, and therefore their presence or absence during key events was not able to affect the participants' reaction time. But the longer reaction time at the two P- conditions than the two P+ conditions still indicated that the participants' own belief affected their reaction time in the task.

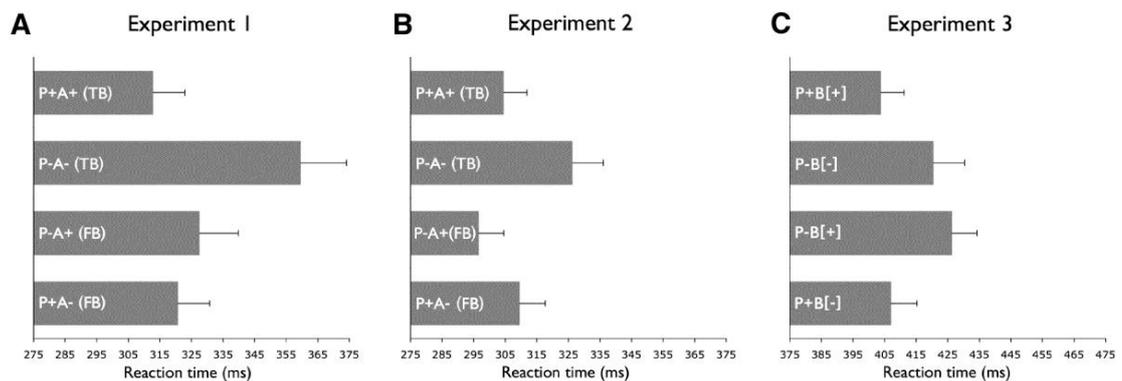


Figure 4 - 12 Results of Kovács et al's task in adult human Note: In experiment 1 and 2, the agent was an animated figure. In experiment 3, a pile of boxes, replaced the agent and acted as the agent. (Kovács et al, 2010)

In the current experiment, we followed this approach. The participants could still watch all the events about the object and hold their own beliefs about the presence or absence of the ball, but now there was no agent present in the scene that could form beliefs about the events depicted on the screen. If the differences of reaction time in the previous experiments were the result of belief computations, we expect differences to emerge only from the participants' own belief, but not the boxes' "belief". In contrast, if the structure of events in the videos was the real factor

that affected the reaction time in the last experiment, we expect to find the same pattern of results as in experiment 4.

4.6.2 Method

4.6.2.1 Participants

A total of 24 adults took part in the experiment (Male: Female = 8:16, average age = 22.4). They were recruited from the participant-pool in the University of St Andrews. All participants reported normal or corrected-to-normal vision. They were naive to the purpose of the study and had never participated in our previous experiments. They were invited to a quiet laboratory room in the School of Psychology and Neuroscience to participate in the experiment. After the experiment, they each received £3 compensation for their participation. Ethical approval was granted by the ethical committee of the University of St Andrews.

Two participants were excluded due to self-report of pre-knowledge about the purpose of the experiment. The remaining 22 (Male: Female = 8:14, average age = 22.5)) were included in the data analysis.

4.6.2.2 Apparatus and Materials

The apparatus was the same as the last experiment. The materials were short animation videos, which were the same as in experiment three, except that the agent was replaced by a stack of boxes (see Figure 4 - 13).

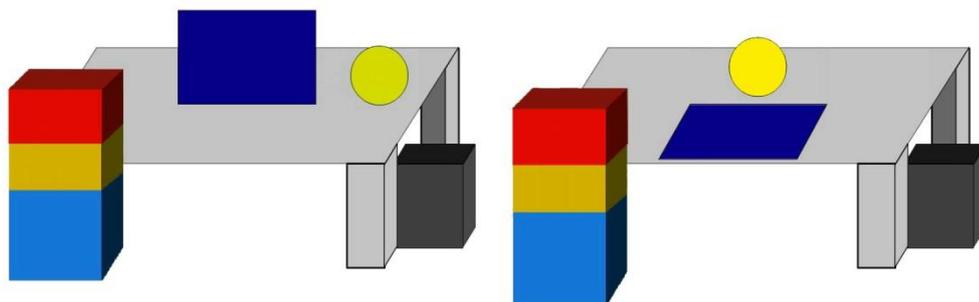


Figure 4 - 13 Examples of the scenario in experiment 5

4.6.2.3 Design and Procedure

The design and experimental procedure were the same as in the last experiment.

4.6.2.4 Data analysis

Data analysis was similar to that in the last three experiments, except that the factor Agent's belief was replaced by the factor Boxes' Belief. All the incorrect responses (3%) were excluded from the data analysis.

4.6.3 Results

4.6.3.1 Comparison of the four conditions

To examine the interaction, we compared the four conditions directly using repeated one-way ANOVA. Mauchly's test revealed that the assumption of sphericity was violated ($\chi^2(5) = 14.73, p = .012$). Therefore we reported the p-values with Greenhouse-Geisser adjustment. The ANOVA showed significant difference between conditions ($F(3, 63) = 10.32, p < .00$). Paired T-test with Bonferroni adjustment (see Table 4 - 14) revealed significant longer reaction time at P-B- condition than P+B- ($t(21) = 4.24, p = .002$), and P-B+ ($t(21) = 4.13, p = .003$), but not than P+B+ ($t(21) = 2.52, p = .12$). No other difference was found.

Table 4 - 13 and Figure 4 - 14 show the participants' average reaction time to different conditions. The two-way ANOVA revealed that the participants' belief, ($F(1,21) = 5.28, p = .032$), and the boxes' 'belief' ($F(1,21) = 5.18, p = .033$) both had a significant main effect. More importantly, the interaction was also significant, $F(1,21) = 10.00, p < .001$.

To examine the interaction, we compared the four conditions directly using repeated one-way ANOVA. Mauchly's test revealed that the assumption of sphericity was violated ($\chi^2(5) = 14.73, p = .012$). Therefore we reported the p-values with Greenhouse-Geisser adjustment. The ANOVA showed significant difference between conditions ($F(3, 63) = 10.32, p < .00$). Paired T-test with Bonferroni adjustment (see Table 4 - 14) revealed significant longer reaction time at P-B- condition than P+B-

($t(21) = 4.24, p = .002$), and P-B+ ($t(21) = 4.13, p = .003$), but not than P+B+ ($t(21) = 2.52, p = .12$). No other difference was found.

Table 4 - 13 Participants' average reaction time (ms) to the four conditions in experiment five

	P-A-	P+A-	P-A+	P+A+
Mean	434.9	364.9	370.0	391.0
SD	83.8	58.6	58.9	69.5

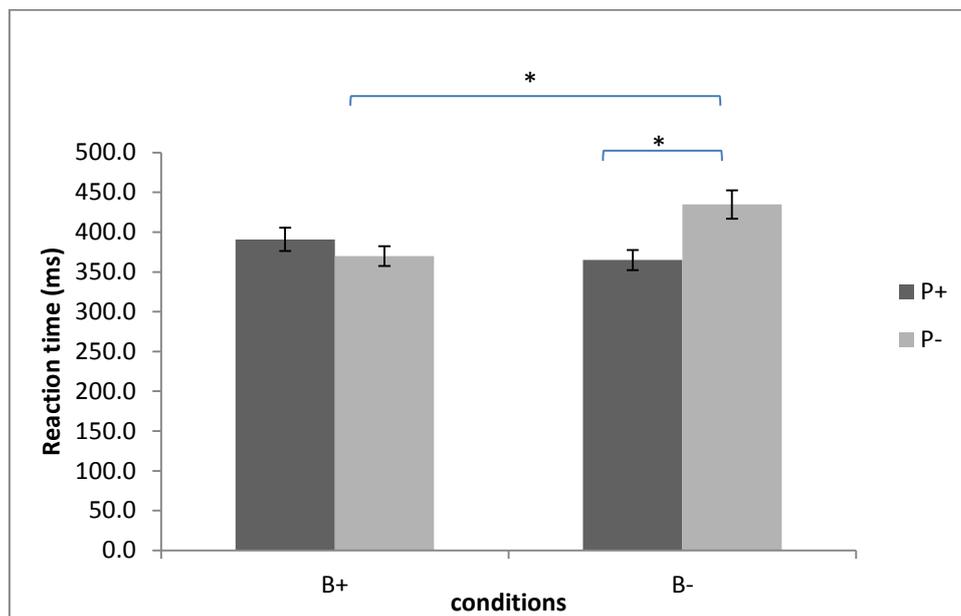


Figure 4 - 14 Participants' average reaction time (ms) to the four conditions in experiment five (error bar: SE)

Table 4 - 14 Results of the paired T-tests in experiment five

Pair of conditions	t(21)	P (2-tailed)	
		Bonferroni adjustment	No adjustment
P-B- vs P+B-	4.24	.002***	<.001***
P-B- vs P-B+	4.13	.003***	<.001***
P-B- vs P+B+	2.52	.12	.020*
P+B- vs P+B+	-2.42	.15	.025*
P-B+ vs P+B+	-1.60	.75	.12
P+B- vs P-B+	-.61	1.00	.55

4.6.3.2 Following Kovács' approach

Using P-B- as the baseline condition, we used paired T-tests without adjustment and compared the other conditions to the baseline. The results showed that RT in P-B- was significantly longer than in the other three conditions (see Table 4 - 14). In addition, participants responded significantly more quickly in P+B- than P+B+ condition ($t(21) = 2.42, p = 0.025$). These results are identical to Kovács et al's experiment using the animated character as the agent, but different from that using the stack of boxes as the agent.

4.6.3.3 Each participant's RT

In addition, for each condition we counted the number of participants who responded slowest in that condition (see Table 4 - 14) and found that, similar to what we found in the last experiment, more than half (13/22) of the participants in the current experiment responded slowest in P-B- condition. And for the rest, 4 participants' RT were the longest in P+B+ condition, 3 in P-B+ and 2 in the P+B- condition.

Table 4 - 15 Number of participants that responded slowest at the condition in experiment five

Condition	P-B-	P+B-	P-B+	P+B+
No. of participant	13	3	2	4

4.6.4 Discussion

In the current experiment, we aimed to eliminate the agent's intentionality, or belief, but keep all the other features of the experiment, e.g., the trajectory of the ball and the movement pattern of elements, in order to check for the possibility that other factors, rather than intentionality or belief, determined the results in the previous experiments. Therefore, following the lead from one of Kovacs et al. controls, we used a stack of boxes, which has no perspective or belief, to replace the agent in the experiment. The prediction was that only the participants' belief should affect RTs.

However, we again found longer reaction time in P-B- condition than the other three conditions, and the pattern of results is the same as in experiments 3 and 4, where either an animated agent or a monkey photo acted as intentional agents with potential perspective and belief. The similar results in the current experiment and the previous two experiments (experiment 3 and 4) suggest that either the stack of boxes does not successfully eliminate belief computation, or belief is not the main factor that has affected the participants' reaction time in the last three experiments. We suggest that the latter is more likely the case for the following two reasons. Firstly, the stack of boxes, as suggested by Kovacs et al., eliminates any cues of belief or intentionality. Although the boxes exactly follow the trajectory of the agent in experiment 3, they do not look like intentional agents. For example, it does not have a front or back and hence does not face to or look at the hiding-screen as the agent does, and it does not turn around when it leaves the scenario. With these manipulations, the stack of boxes delivers no cue of an animated being or sign of intentionality. As a consequence, any belief computation of the agent should have been eliminated.

Secondly, if it is the belief that determines the participants' reaction time, in the current experiment only the participants' own belief would affect the results. However, the results indicate that not even the Participants' own belief is not the only relevant factor: if it were, RT in the two P+ conditions would be faster than that in the two P- conditions since although the agent's belief is eliminated, the participants still have their own belief. This is what Kovács and colleagues reported in their control experiment with a pile of boxes instead of the agent. Our results, however, do not follow this pattern and do not show the effect of the participant's own belief. A potential explanation for this difference might be that, the pile of boxes in Kovács et al's control experiment was stationary and "was present in all the movies during their entire duration" while the stack of boxes in our experiment mimicked the agent's movement in the videos. Note that the movement does not reproduce the kinematics of an animated being: it is a simple displacement that could be entirely mechanical. Therefore we argue that it is the movement of the boxes and the agents (in the previous two experiments) that had an effect on RT,

perhaps by introducing a potential distraction to the participants' attention that might have affected their response in the task.

Moreover, the comparison of the last three experiments with the first two experiments suggests that the change in experimental procedure indeed determines the participants' pattern of reaction time to the four conditions. It remains unchanged that the reaction time in the P-A- condition is the longest among all the conditions. But the major difference is that the reaction time in P+A+ is longer than in the P+A- and P-A+ condition. This finding is in accord with what we have discussed after experiments 3 and 4 that the specific procedures of the video presentation and their effect on the participants' attention, instead of any belief or intentionality computation, is the real cause of the findings in our last three experiments, as well as for Kovács et al's findings (2010).

4.7 General Discussion

The studies reported in this chapter emerged as an attempt to address the surprising results found in Chapters 2 and 3, suggesting that Capuchin monkeys did better than adult humans in coding the object-directed intentionality of actions when their objects are not directly shown.

To address the possibility that the negative results with humans were due to an inability of our experimental procedure to capture their true skills, we decided to use another paradigm, a visual detection task, which was originally applied successfully in human adults (with RT measurements) and infants (with LT measurements) to access their implicit belief understanding (Kovács et al, 2010), to examine again if adults can automatically register object-directedness. Our initial suggestion was that the results of Kovacs et al. could actually be due to the ability to code object-directedness, rather than belief computation, and we modified their design to test the alternative interpretations

A summary of the five experiments

Building upon Kovács et al's visual detection task, we conducted a series of five experiments with adult humans.

In experiment 1, we tried, first, to replicate their original experiment based on detecting object presence, and second, to distinguish between predictions made by their Belief interpretation and our Object-directedness interpretation using a novel condition involving the detection of object absence. Unexpectedly, we failed to replicate Kovács et al's results in the object present group, the one that reproduced their original study. And the result for the Object-absent group did not support either the belief or the object-directed interpretations. Given that the results in the presence group were not consistent with Kovács et al's original experiment, we focused the next experiments on understanding the failure to replicate their results.

In the second experiment, we strengthened the relationship between the agent and the object, hoping to replicate their results with a closer reproduction of their stimuli, which might also emphasise the coding of object-directedness. However, again the effect of the agent's and the participants' beliefs did not match Kovács et al. original results, suggesting that the relation of 'ownership' of the agent to the object was not the key factor in causing our different results.

As the pattern of reaction time in the four conditions was similar between experiment 1 and 2, we combined the data and this revealed the same results as Kovács et al's, but only when we exactly followed their approach of data analysis. However, this approach of data analysis might be problematic because it does not adjust for multiple comparisons, which may lead to a higher probability of detecting a difference by chance. When proper adjustments were made, we could not completely replicate their results because of a lack of effect in the condition with asymmetrical beliefs between the agent and the participant (P-A+). This indicates that the agent's belief alone is not effective in affecting the participants' reaction, which is key to Kovács et al. claim that there is automatic encoding of others' beliefs.

Given the results of the first two experiments, we felt it was not possible to conclude that adults can automatically register object-directedness or belief as has been suggested by Kovács et al (2010).

In an attempt to disentangle the factors that control for the effect they report, in the next experiments we tried to exactly replicate Kovács et al's method,

abandoning our attempt to ensure that the video presentations in the four experimental conditions were structurally equivalent and detailed description in the method section of experiment one), by exactly following Kovács et al's original design and detailed description in the method section of experiment three). Despite this, in experiment three, we again only partially replicated Kovács et al's results, and failed to reproduce the difference between two crucial baseline conditions: P-A- and P+A+, without which the results cannot be interpreted as demonstrating automatic computation of agents beliefs. In experiment four, where we used a living being, i.e., a Capuchin monkey's profile picture, instead of the animated character, we finally replicated Kovács et al's findings in the agent condition, but unfortunately we found exactly the same results in a control condition using a stack of boxes instead of an animated being, This finding suggests that it is not Belief (or object-directedness) that produces the pattern of results, but the specific procedures of video presentation that affects the participants' attention differently in the different conditions (see detailed discussion in the discussion section of experiment 3).

Kovacs et al had used a "pile of boxes" control condition in which they reported only an effect of the participants' own belief upon RT, but no effect of the boxes "belief". However, Kovács et al's control experiment did not follow the same procedure of video presentation used with the agent. We did use the exact same presentation for the Agent and the Object, and found the same results for Agent and Boxes. Thus our results suggest that the structure of events in the video presentation, rather than any Belief computation might explain the results (see more discussions in the discussion of experiment 5). This finding would support a recent critical re-assessment of the infant and adult implicit false belief literature by Heyes (2014) suggesting that, rather than implicit mindreading skills, the findings could reflect "submentalizing" processes based upon domain-general cognitive mechanisms such as automatic attentional orientation, or the automatic coding of spatial locations of objects.

To summarize, given the failure to replicate Kovács et al's original study, we were not able to pursue our original aim of further exploring the ability to code object-directedness in human adults by using a "proven" paradigm that might go

beyond any procedural limitations of our paradigm in Chapters 2 and 3. Instead we found consistent evidence that challenges Kovacs et al.'s much heralded findings of the automatic encoding of other's beliefs by human adults, which, added to our previous failure to find evidence of automatic encoding of object directedness to unperceived objects, casts doubts on the assumed abilities of humans to implicitly code mental states. This finding is in strong contrast with the increasing evidence of implicit mindreading in adults (Apperly et al., 2006; Back & Apperly, 2010; Butterfill & Apperly, 2013; Kova et al., 2014; Kovács et al., 2010), but fits with recent critical interpretations of the literature suggesting that some of the purported effects may be due to domain general processing constraints (Heyes, 2014).

Although we failed to replicate the neat results reported by Kovacs et al., we nonetheless found some potentially interesting effects. For example, in the first two experiments of this chapter, in which we improved Kovacs et al. design by ensuring that the event structure of the video sequences was fully comparable across conditions, whenever a significant difference was detected in the object present conditions ($p < 0.05$), this involved the reaction time being longer in the P-A- condition than other conditions (P+A+, P+A- or P-A+). For example, when the two experiments were put together, the participants' reaction time was significantly longer in the P-A- condition than in the P+A- and the P+A+ condition; in experiment 1, significantly longer in the P-A- than the P+A+ condition; experiment 2, significantly longer in the P-A- than the P+A-, and a trend of longer reaction time in the P-A- than the P-A+ ($p = 0.08$). The consistently longer reaction time in the P-A- condition may indicate that the presence of another agent in scenarios with conflicting events (e.g. ball appearing in one, disappearing in another), may have some effect on the way in which such scenarios are processed, even if this influence is not due to the computation of false beliefs or intentional relations with objects

The group of participants responding to the absence of the object in experiment 1 was initially designed to distinguish the belief-outcome congruency hypothesis and the object-directedness hypothesis, but it generated very unexpected results that could not be explained by either hypothesis. The reaction times in the four conditions are not significantly different from each other, indicating

that neither intentionality nor belief had any effect on the task when the response required is to recognise the absence, rather than the presence of an object. This is reminiscent of the results of chapter 3, using the object-directed actions paradigm, in which the participants always responded much more slowly to the absence than the presence of an object. It might be that in both paradigms responding to “nothing” requires more effort or a different type of processing than responding to “something”. and this may mask any effect of the experimental conditions upon the participants’ reaction time in this task.

Be that it as it may, in Chapter 2, with the object-directed action paradigm, we unexpectedly found more evidence of sensitivity to the object-directedness of the actions in the monkeys than in the humans. In the next and final experimental chapter, we will test if the same happens with the current design based on Kovacs et al, and whether Capuchin monkeys show a different sensitivity to the conditions displayed in the same stimuli.

Chapter 5. Capuchin monkey's automatic encoding of other's beliefs

Summary

This chapter aims to employ the Participant/Agent Belief paradigm developed with adult humans in the previous chapter to investigate whether Capuchin monkeys can automatically register others' beliefs or intentional relations to objects. A total of 18 monkeys were presented with the same videos used in Experiment 5 in the previous chapter where the monkeys' own belief and an agent's belief (a picture of another monkey) about an object's location were manipulated. The result showed that the monkeys' looking times were not influenced either by their own belief or the other agent's belief about the object's location. The negative results suggested that the monkeys don't have the capability of belief attribution even in an implicit ToM task and any coding of intentionality had no effect on their visual behaviour in this task. In comparison with human performance, the monkeys showed no evidence of retroactive interference, which suggests that they attended to the videos in a different way to the humans.

5.1 Introduction

In the last chapter, we tried to adapt an object-detection task originally developed to test implicit coding of false beliefs to seek for evidence of adults' ability to understand object directed intentionality. However, we failed to replicate the original results reported by the creators of the task and ended up concluding that this paradigm, when applied with appropriate controls and correct statistical analyses, shows little evidence of either Belief or object-directed intentionality coding. Our results seem instead to support recent proposals that the original findings can be better explained through "sub-mentalizing" processes.

The finding could be considered to be consistent with what we detected in chapter 3 using another reaction time paradigm, where adult humans failed to show the ability to code object-directedness when the object of the action was not directly shown. In comparison however, a primate species, the Capuchin monkeys, showed some signs of coding object-directedness in the experiment presented in chapter 2 using the same stimuli as those used in chapter 3 with adults. In this chapter we sought to test the Capuchin monkeys using the same task and stimuli used in the last chapter (experiment 4, chapter 4) to complete our comparative investigation of monkey and human performance in implicit mentalistic tasks.

As we have discussed in the introduction of chapter 4, the experimental paradigm was developed from an object-detection task originally designed by Kovács and colleagues to assess humans' implicit understanding of others' belief (Kovács et al., 2010), with a suggestion that it could also provide evidence for the ability to code object-directed intentionality. In Kovács et al.'s original experiment, they present the participants videos in which the belief of the participant and an animated agent are manipulated to create four conditions with different beliefs about an object's presence behind a hiding-screen. After watching these videos, the participants are required to respond to the presence of the object when the hiding-screen opens. If the participants' expectation is affected by their own belief and/or the agent's belief, their reaction time will be prolonged by the violation of the belief content. From their results, Kovács et al. concluded that humans can automatically register others'

beliefs (Kovács et al., 2014; Kovács et al., 2010). In comparison, however, from our five experiments, which all closely reproduced the experimental conditions of Kovács et al.'s study, we suggested that the results with adult humans should not be attributed to the implicit understanding of belief, but instead, to reactions to differences in the video presentations across the four conditions, which would support the “sub-mentalising” interpretation recently proposed by Heyes (2014)

As in the object-directed actions task presented in Chapters 2, with the Capuchin monkeys we will be using a looking time paradigm, instead of the RT paradigm used with adult humans. One possibility is that the different structures of the video presentations do not affect the results of the experiment as much in the looking time task. In the reaction time task, the participants are required to press a key as quickly as possible to decide if the object is present when the hiding-screen opens and the outcome is shown. Given that, in some conditions the participants are also required to respond to the agent's leaving (as a way of ensuring that they have noticed this key event affecting the agent's Belief), this might have differently affected the participants' attention to the hiding-screen-area when the object was revealed. The shift of attention to the location of the object might have resulted in a longer reaction time in the P+A+ and the P-A- conditions due to this procedural difference (see experiment 4 in chapter 4 for details). However, when looking time is used as the measurement of the participants' expectation, the participants have a relatively long ten second period to observe the outcome, and they are not required to respond to any intervening event.

As with reaction time, the prediction is that the looking time will also be prolonged by any violation of expectation and will therefore indicate whether any belief or object-directedness is encoded.

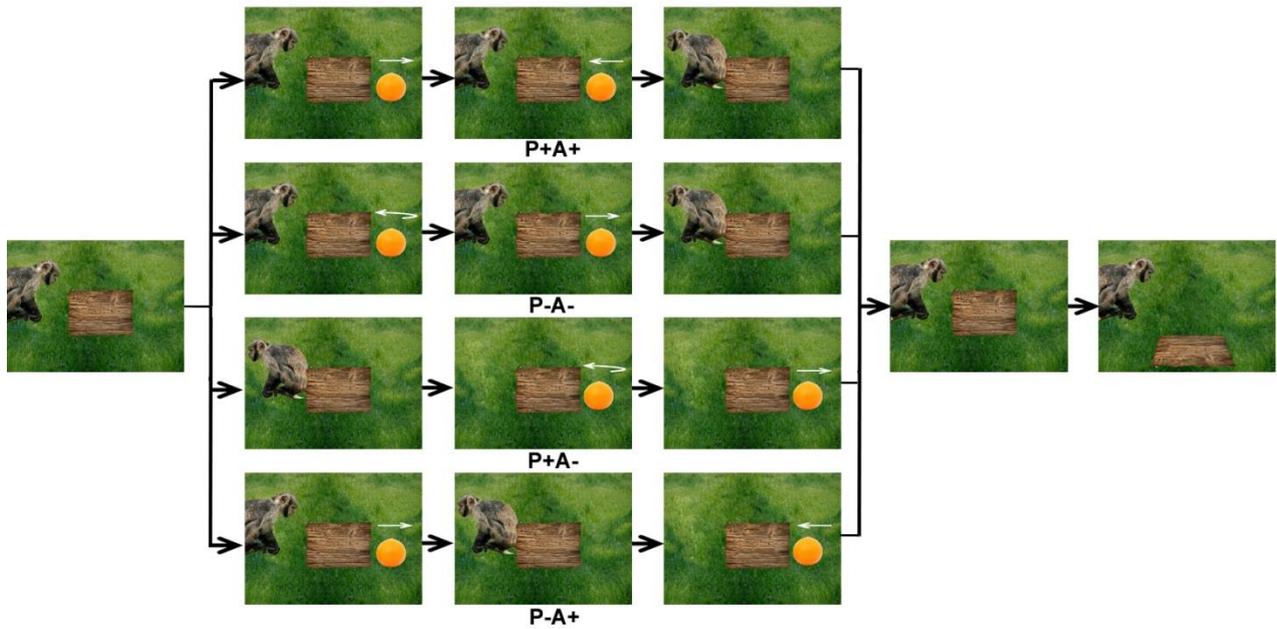


Figure 5 - 1 The procedure of video presentation in the four conditions

The looking time paradigm has been used with both non-verbal human infants and non-human primates (Kovács et al., 2010; Martin & Santos, 2014). As an equivalent of the reaction time task with adults, Kovács et al (2010) used looking time as the measurement to test 7-month-old infants with exactly the same stimuli as the adults, reporting positive evidence for infants' implicit registration of others' beliefs (see details of the infant study in the introduction of chapter 4). After we had completed our own study presented in this chapter, Martin and Santos (2014) published a new experiment in which they adapted Kovács et al. (2010) paradigm to non-human primates using the same looking time paradigm to investigate rhesus macaques' implicit understanding of other's belief. However, although they found that the monkeys' looking time was indeed affected by their own belief in the task (i.e., they looked for longer if no object was revealed behind the screen when they thought there was one), the monkeys failed to show any evidence of encoding the belief of another agent spontaneously (see Figure 5 - 2, Martin & Santos, 2014): i.e., the agent's belief that the object was present or absent did not affect the monkeys' looking time at all. The authors concluded that rhesus macaques, in contrast to human adults and babies, do not understand the false beliefs of others.

Our study therefore can be seen as an extension of Kovacs et al (2010) paradigm to a new species of primate, Capuchin monkeys. However, in addition, it offers a direct comparison with the performance of humans in exactly the same tasks (Martin & Santos only tested monkeys), and one in which the humans did not perform as predicted by Kovacs et al's interpretation.

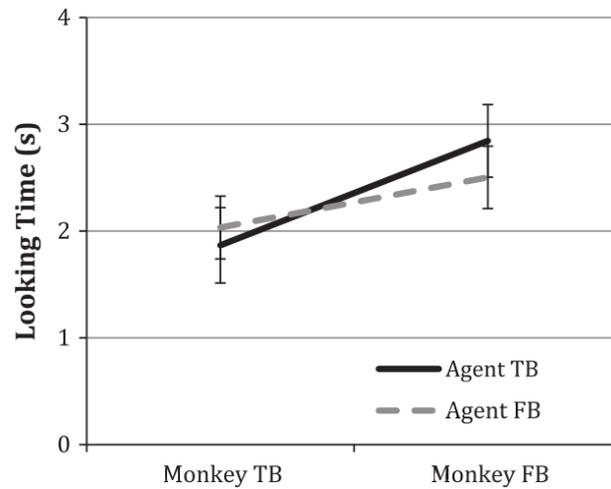


Fig. 2. Results. Mean looking time (in s) \pm SEM across monkeys in each condition: Monkey Belief (True or False) and Agent Belief (True or False).

Figure 5 - 2 Rhesus macaques' average looking time in the task (Martin & Santos, 2014)

Both Kovacs et al. looking time experiment with infants, and Martin and Santos's looking time study with monkeys were based upon the presentation of a different outcome from the original adult experiment, i.e., in looking time the object was always absent when the hiding-screen opened. In the current study, we will follow suit and present the Capuchin monkeys with the absence of the object as the outcome. With this outcome, P+A+ and P-A- still act as the two baseline conditions, i.e., we expect to find longer looking time in the P+A+ than the P-A- because of violation of the expectation that the object will appear behind the screen. However, returning to our two alternative hypotheses (the object-directedness hypothesis and the belief-outcome congruency hypothesis) presented in the last chapter, we predict different results for the P+A- and P-A+ conditions. If monkeys encode object-directed

intentionality, we expect their looking time to be longer in both the P+A- and P-A+ condition than the P-A- condition, because the coding of intentionality suggests the existence of the object, which would violate the outcome of the object's absence. On the other hand, if the monkeys code beliefs, the looking time should be shorter in the P+A- and P-A+ condition than in the P+A+ condition, because either the agent's or the participants' belief that there is no object conforms to the outcome and will shorten the reaction time. See the introduction of chapter 4 for more details on the rationale for these alternative predictions.

5.2 Method

5.2.1 Participants

A total of 18 Capuchin Monkeys (*Cebus apella apella*) took part in the experiment. All monkeys were housed at the Living Links to human evolution research centre at Edinburgh Zoo (see Table 5 - 1 and chapter 2 for more details).

Table 5 - 1 Monkey's profile.

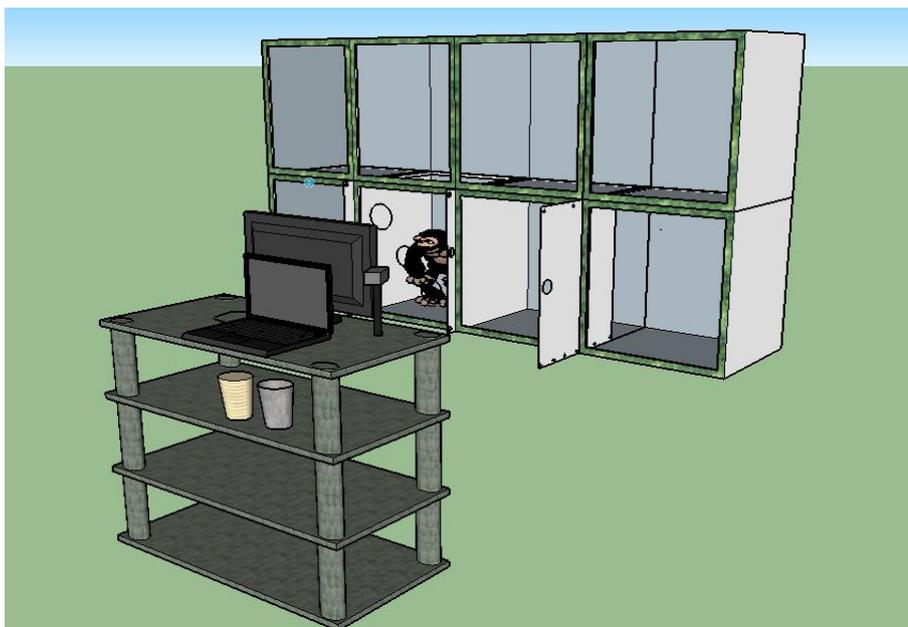
(The age shows the monkeys' age when the experiment finished)

Monkey	Age (Yr:m)	Sex
Chico	05:00	M
Junon	13:10	F
Penelope	08:05	F
Flojo	02:10	M
Ruben	03:10	M
Anita	16:03	F
Carlos	07:09	M
Toka	09:04	M
Figo	07:10	M
Diego	11:09	M
Ximo	04:00	M
Alba	02:07	F
Inti	04:07	M

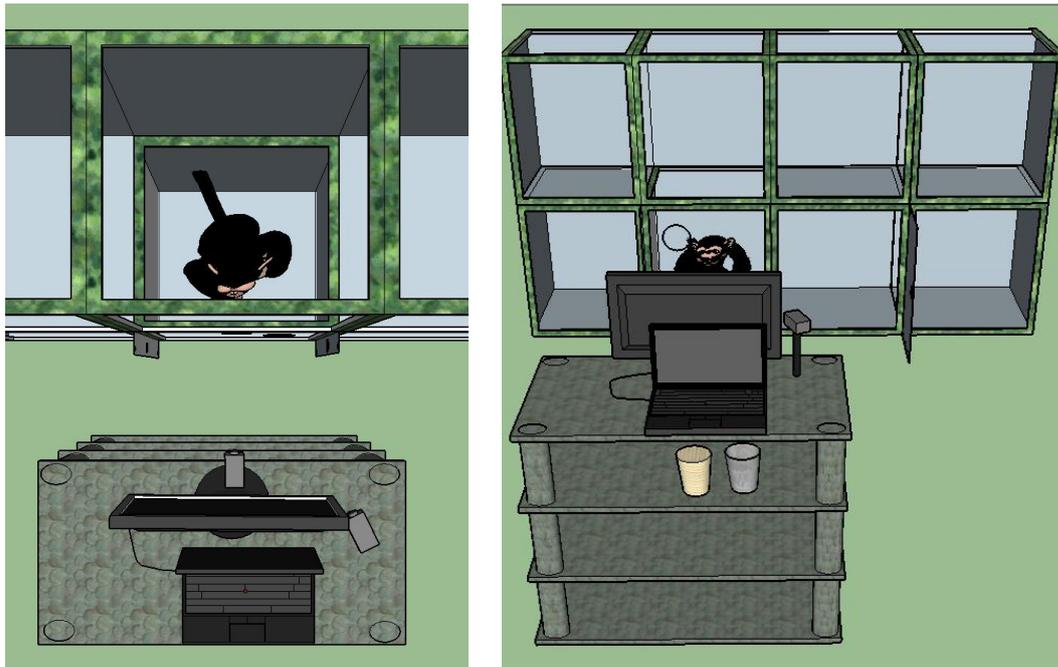
Rufo	04:06	M
Luna	02:11	F
Pedra	06:02	F
Torres	03:03	M
Sylvie	10:08	F

5.2.2 Apparatus and stimuli

Figure 5 - 3 illustrates the setup in the experiment (see also Figure 2 – 1 and Figure 2 – 2 in chapter 2). Monkeys were separated and tested individually in cubicles to which they had been habituated and were motivated to stay for research sessions. A computer monitor (Dell™ E198WFP Flat Panel Monitor, 19 inch) was located on a platform at an equal height to the base of the cubicles and 0.5m directly in front of the front panel. Display of stimuli was controlled by Microsoft Powerpoint on a laptop (Lenovo ThinkPad X220i) behind the monitor. Subjects' visual behaviour was filmed by two digital cameras (Panasonic SDR-S26) for offline coding. One camera was placed at the centre directly below the monitor and the other at either right or left side of the monitor (Figure 5 - 3. B & D). Monkeys' looking time data were coded offline frame by frame using Coral VideoStudio Pro X4.



A



B

C



D

Figure 5 - 3 The experiment setup. (A. the whole view; B. Top view; C. Experimenter's view; D. Monkey's view)

The stimuli were the same short animation videos used in Experiment 4 of Chapter 4 with human adults. They involved an orange (the object), a wooden hiding board and a profile photo of a capuchin monkey (the agent, A) on the background of a grassland (see Figure 5 - 1 & Figure 5 - 4). We modified Kovács et al's (2010) method and created four testing conditions to manipulate the participant's (P) and

the agent's perspective about the presence of the orange behind the hiding board (Figure 5 - 1). The stimuli also included an introductory video clip (all the videos were created by 3D flash animation (Version 4.9.8.7).

In the introductory clip, the orange moved from the left to behind the wooden board and then the board opened at 2.8 seconds, revealing the orange. The video then stopped and remained still for 10 seconds as in the last picture in Figure 5 - 4. During the 10-seconds, the participant's looking time was recorded. In the introductory trial, the outcome was consistent with both the agent's and the participant's belief and it served as an introduction to the scenario.



Figure 5 - 4 The introductory trial

The four testing conditions involved a similar scenario to the introductory clip (Figure 5 - 1). The orange first moved to behind the board. But then it kept moving after the initial hiding process. At a certain point in different conditions, the agent left the scenario and therefore was ignorant as to any movement of the orange if it moved again, leading to the four belief conditions (P+A+, P-A-, P+A-, P-A+). At the end, the agent returned to the scenario, sat beside the board and looked at the location behind the board when the board opened at 7.4 seconds. In all experimental conditions, the orange failed to appear behind the board or anywhere else on the screen. The final outcome scene lasted for 10 seconds.

In the P+A+ condition, after the first hiding behind the board, the orange rolled out of the screen but then returned and rolled behind the board. The agent watched all this movement and then left the scenario. Therefore, both the participant and the agent believed the orange was behind the board. When the agent returned to the

scenario, the board opened, revealing no orange. Thus the outcome contradicted both the agent's and the participant's belief, or their object-directed encoding. As a consequence, the participant should experience a violation of expectation and be surprised and look longer at the outcome.

In the P-A- condition, after the initial hiding, the orange emerged from behind the board without leaving the scene and moved back to behind the board. Then it rolled out of the board again and left the scene. The agent was present all through the process, resulting in both the agent and the participant believing the orange should not be resting behind the screen. Then the agent left and came back. When the board opened, no orange was there, which was consistent with both parties' beliefs, or with both parties lack of encoding of object-directedness given the disappearance of the object, and therefore, the monkey's looking time was expected to be the shortest among all conditions.

In the P-A+ condition, the movement of the orange was the same as in P-A-, but the agent left before the orange emerged from behind the board. In this case, the agent held a false belief that the orange was behind the board while the participant had the true belief that the orange had left. The outcome did not contradict the participant's belief but, from Kovacs et al. automatic Belief coding perspective, the participants would look longer if they had registered the agent's belief to the object, which contradicted the outcome. From the object-directedness perspective, we predicted that, since the object had been seen to disappear from the scene, no object remained to be encoded, and therefore looking time should be comparable to condition P-A-.

In the P+A- condition, the orange left the scene from behind the board in the agent's presence but then, while the agent was away, the orange rolled back and hid behind the occluder. The agent, therefore, believed the orange to be away while the participant believed the orange to be behind the board. When the board opens, the outcome is opposite to the participant's own belief, but consistent with the agent's belief. The Belief interpretation would therefore predict a similar looking time to the P-A- condition. In contrast, from the object-directedness perspective, we predicted

that, since the participant had last encoded the object to be present, looking time should be longer as this constituted a violation of the expectation.

5.2.3 Design and Procedure

The experiment was a 2 (Agent's Belief, of the object's presence/absence behind the board) * 2 (Participant's belief, of object's presence/absence behind the board) factorial design, resulting in 4 conditions (P+A+, P-A+, P+A-, P-A-) within subjects.

Each participant had 4 experimental sessions in total, resulting in 4 repeated trials for each condition. In each session, all the participants watched 2 introductory videos at the very beginning. Then they were exposed to two testing clips before another introductory clip, followed by the other two testing conditions. A session contained the four conditions each and the order of conditions was counter balanced between sessions. A session started when a monkey was separated and sat comfortably in a cubicle. After each trial, the monkeys had short breaks with 2-4 raisins as rewards. After each session, they again had some rewards before being allowed to go back to their social group. The interval between sessions was at least one day.

5.2.4 Data analysis

As in chapter 2, we coded the monkey's looking time offline frame by frame (25 frames = 1 second). As in chapter 2, only looks longer than 4 frames were included in the looking time data. The looking frames were added up in the 10-second-looking-window to calculate a monkey's total looking time in each condition.

In addition, we examined the monkey's attention at the videos before the hiding-board opened and categorized the trials into two categories: attended, where the monkey watched at least for 4 frames when the orange was visible; not attended, where the monkey did not look at the screen area at all.

5.3 Results

5.3.1 Monkeys' attention in the experiment

First, we assessed monkeys' attention to the videos displayed (see Table 5 - 2). The monkeys did not pay any attention to 64 out of the 286 test trials (22.4%) and 63 of the 215 introductory trials (29.3%). We excluded these trials in the following analysis. This left us with 222 test trials and 147 introductory trials that were attended to. In the following analysis, we report all the data from the attended trials.

Table 5 - 2 Monkey's attendance to the videos (Note: H1, H2, H3 represent the first, second and third introductory trial.)

Attendance	P-A+	P+A+	P-A-	P+A-	H1	H2	H3	Grand Total
Not	16	12	19	17	21	24	23	132
Attended	55	60	53	54	50	48	49	369
Total	71	72	72	71	71	72	72	501

5.3.2 Looking time at the attended trials

Table 5 - 3 and **Error! Reference source not found.** show the monkeys' average looking time in different conditions. And Table 5 - 4 shows the data of each individual. We first averaged the looking time in each condition for individual monkeys and then calculated the monkeys' average looking time in the four conditions.

Table 5 - 3 Monkeys' average looking time to different conditions of the attended trials(s)

	H1	H2	H3	P-A+	P+A+	P-A-	P+A-
Mean	1.70	1.28	1.15	1.40	1.41	1.50	1.53
SD	0.94	0.81	0.84	1.02	0.86	0.77	1.01

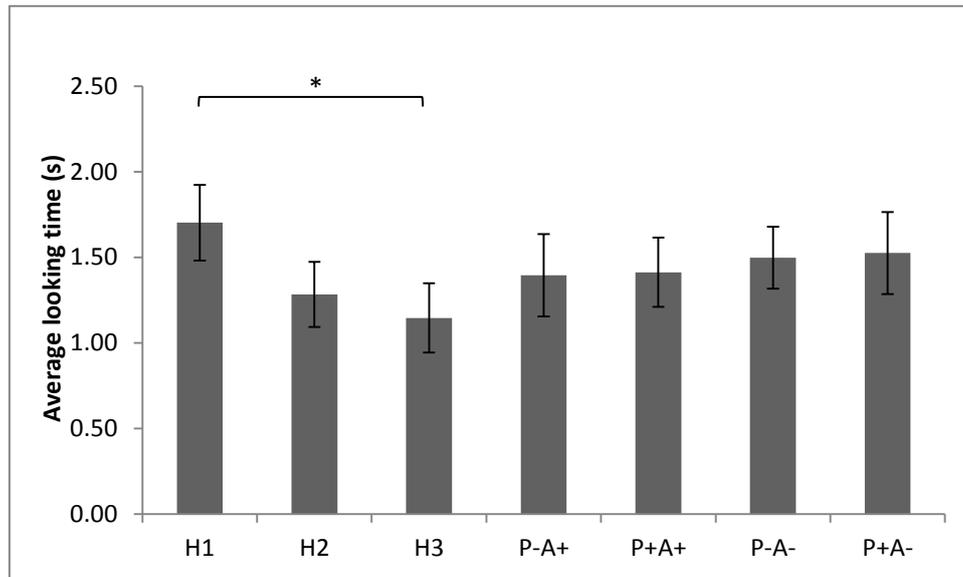


Figure 5 - 5 Monkeys' average looking time in different conditions of the attended trials (error bar: SE)

Table 5 - 4 Individuals' average looking time in each condition of the attended trials(s)

	P-A+	P+A+	P-A-	P+A-	H1	H2	H3
Chico	2.82	2.41	2.80	4.46	4.01	2.56	2.46
Junon	1.19	2.87	1.61	1.86	1.09	2.36	0.66
Penelope	0.72	0.83	0.88	1.12	0.80	0.52	1.17
Flojo	2.99	1.88	2.59	2.49	2.68	1.66	1.74
Ruben	2.95	2.91	2.36	2.77	2.49	2.01	2.33
Anita	2.84	1.55	1.24	1.45	1.73	1.97	1.42
Carlos	2.92	2.65	2.29	2.44	2.22	2.21	1.87
Toka	0.70	1.26	0.63	1.36	0.80	0.22	0.88
Figo	1.00	0.97	0.74	0.83	3.16	0.62	0.12
Diego	1.02	0.42	1.24	0.46	0.78	0.72	0.56
Ximo	0.44	0.68	2.08	0.79	1.57	0.42	0.22
Alba	0.68	1.49	2.09	1.26	1.84	2.32	2.52
Inti	0.88	1.19	0.00	1.17	1.68	1.33	0.00
Rufo	0.38	0.05	0.94	0.10	1.20	0.25	0.15

Luna	0.47	0.85	1.57	0.78	1.12	0.90	0.70
Pedra	1.86	1.46	0.75	1.45	1.41	1.19	1.53
Torres	0.37	0.37	1.75	1.68	0.28	1.39	
Sylvie	0.89	1.59	1.40	0.99	1.78	0.44	1.16

A one way repeated measure ANOVA revealed a significant difference across the three introductory trials ($F(2,32) = 5.83, p = .007$). T-test with Bonferroni adjustment showed a trend of drop of looking time in the second introductory trial from the first ($t(17) = 1.92, p = .08$), but no difference was found between the second and third introductory ($t(16) = .470, p = .81$). In addition, the looking time at the third introductory was significantly shorter than the first ($t(16) = 2.91, p = .03$).

For the four testing conditions, a two-way repeated ANOVA using the Agent's Belief and the Participant's Belief as the two main factors revealed no main effect of either factor: AB ($F(1,17) = .58, p = .46$), PB ($F(1,17) = .036, p = .85$). The interaction was not significant either ($F(1, 17) = .005, p = .94$). In additions, we compared the four conditions directly using a one-way repeated ANOVA. Again, the results revealed no significant difference between conditions ($F(3,51) = .24, p = .87$).

Following Kovács et al's (2010) method of analysis with their adult participants, we also used paired T-test to compare P-A- with other conditions. The looking time was not significantly different between any pair: P+A+ ($t(17) = .44, p = .67$), P+A- ($t(17) = .174, p = .86$), and P-A ($t(17) = .46, p = .65$).

5.4 Discussion

5.4.1 Capuchin monkeys showed no encoding of Belief or Intentionality

The results show that the monkeys did not look differently at the four testing conditions, suggesting that they did not automatically encode either other agents beliefs or their intentional relation to objects. This finding is consistent with the findings reported by Martin and Santos in their own adaptation of the paradigm. However, in contrast to their study, we unexpectedly found that the Capuchin

monkeys also did not react to their own belief about the presence of the object: they showed no prolonged looking time when the object they last saw disappear behind the screen failed to re-appear. This could be taken to indicate that they failed to engage with video presentation of the task or failed to encode the relevant aspects of the scenes.

The most surprising finding is indeed the lack of difference between the two baseline conditions P-A- and P+A+. This suggests that even when the agent's and the monkey's beliefs are consistent to each other and both violate the outcome, the Capuchin monkeys did not find the outcome surprising. Even if the Capuchin monkeys were not able to understand the other's belief, they should at least be affected by their own belief or encoding of the object location.

One possible explanation of these unexpected findings might be that the Capuchins failed to understand the videos. Martin and Santos experiment was conducted with real people acting out the scenes, and maybe the videos failed to engage the Capuchins attention. However, our positive results in Ch 2 suggest that Capuchins do in principle have the ability to perceive what is going on in the videos and react differently to different outcomes and conditions.

The videos in the current experiment may, however, have been particularly difficult to follow for the monkeys. The actions displayed in chapter 2 had physical contact with the object, while the agent in the current experiment only stood and watched the movements of the object; the object was moved by the agent in chapter 2 while it moved itself in the current experiment; the agent was a human experimenter acting out in chapter 2, while the agent in the current experiment was only a profile photo of a Capuchin monkey.

The fact that the only significant effect that we found was a steady decline in the amount of attention payed to the outcome of the introductory trials suggests that indeed the monkeys progressively lost interest in what was being presented to them in the videos.

There are some additional reasons why the performance of both 7-month-old human infants in Kovács et al. (2010) and rhesus macaques in Martin and Santos (2014) might have been better. In Kovács et al's experiment with infants, the video presentation is an infant-controlled presentation, i.e., when the infants look away, the video pauses until the infants look back at the screen. As a consequence, it ensures that the infants are more likely to perceive all the relevant details of the experimental scenarios. In comparison, in our experiment the videos were presented continuously although we were able to exclude the trials where the monkeys had not looked at the video at all during the presentation. However, this post hoc check could not guarantee that the monkeys were as fully engaged in watching the relevant part of the stimuli as the infants were.

Martin and Santos' (2014) experiment with rhesus macaques did not use a macaque-controlled presentation, but, as mentioned, instead of using videos to present the four conditions, the human experimenters performed the conditions live. Their experiment was conducted by an experimenter using a foam core stage and a plastic apple, while ours was displayed through animation videos and the object was an image of an orange. Monkeys' attention might be better when both the scenario and the object look more realistic, as in Martin and Santos' experiment. Therefore the rhesus monkeys might track the 'real' apple's location better than the capuchin monkeys tracked the orange's on the videos. But the fact that in our experiment the capuchin monkeys' own beliefs did not affect their looking time suggests that they may have failed to understand the scenarios in the videos altogether.

Secondly, to ensure that the Capuchin monkeys would attend to at least some of the presentations of the scenarios, in our experiment each Capuchin monkey had more trials than the infants and the rhesus macaques. The Capuchins had three experimental trials and four test trials in one session and they all had four experimental in total. In comparison, Kovacs et al' infants had only two introductory trials and two test trials each, and Martin and Santos' rhesus macaques had only one test trial after the two introductory trials. Maybe the shorter experimental sessions and fewer trials maintained a higher level of curiosity and attention in the infants and the Rhesus. In the case of our Capuchin monkeys, although using more trials

achieved the goal of getting at least one attended trial in each condition for each Capuchin monkey, the longer experiment and more repeated trials might have made them bored with the experiment very soon and perhaps contributed to their lack of understanding of the videos.

In future studies, we suggest that it is vital to enhance the monkeys' attention to the stimuli. This could be achieved by conducting the experiment live with actors, as in Martin and Santos, instead of displaying videos and by reducing the number of trials for each participant but increasing the number of participants in total. In addition, if eye tracking devices can be used in the experiment, a monkey-controlled presentation of videos might achieve a better effect in delivering the experimental conditions.

5.4.2 Summary

In conclusion, we found that the monkeys performance in this this experiment gave no evidence of automatic coding of Belief of Intentionality, but it also gave no evidence of coding guided by the differential properties of the video displays, as suggested for the performance of the human adults in Experiments 4 and 5. The lack of any differential effect, except for the introductory trials, suggests that either the monkeys did not pay enough attention to the video displays to be able to react to the different presented contents, or they were not able to perceive the intended differences.

Chapter 6. Discussion

6.1 A review of the rationale & aims

Theory of Mind (ToM), the ability to understand others' mental states (Premack & Woodruff, 1978), is an important ability for highly socialized species like human beings. An increasing body of study has focused on the different mechanisms of ToM and their development and evolution, with a particular focus on so called "implicit" or automatic mechanisms (Apperly & Butterfill, 2009; Aschersleben et al., 2008; Call & Tomasello, 2008; Leslie, 1987a; Onishi & Baillargeon, 2005; Sodian, 2009; Woodward, 2009a). Among them, the ability to understand intentionality, i.e., the object-directedness from intentional actions to an object, attracts our attention because of its early ontogenetic emergence in humans (Woodward, 2009a), early phylogenetic emergence in non-human primates (Wood et al., 2007) and a positive correlation with other ToM abilities (Aschersleben et al., 2008), which indicate that it is likely an important precursor and foundation of ToM abilities in general (Gómez, 2009). Therefore in this thesis, we have mainly focused on human and non-human primates' ability to understand intentionality.

The term 'intentionality' is used in slightly different senses in psychology and philosophy. In psychology it is largely a synonym of 'goal-directedness', whereas in philosophy it has a wider meaning: it refers to a key property of all mental states — in that they are always *directed* to or *about* something different to themselves. One always thinks about something, wants something, sees something, etc. Thus, intentional or goal-directed action in the psychological sense would just be a type of intentionality in the philosophical sense (Gómez, 2008, 2009).

Both developmental and comparative studies show that non-human primates and human beings understand the relation between intentional actions and their intentional objects (e.g., Aschersleben et al., 2008; Hauser & Wood, 2010; Kano & Call, 2014; Woodward, 1999). For example, studies with infants suggest that they distinguish object-directed actions from non-object-directed actions and recognize its feature of directedness to an object that is the goal of the action (e.g., Biro &

Leslie, 2007; Phillips & Wellman, 2005; Woodward, 2009a). Non-human primates share this ability as well (Kano & Call, 2014).

It has been suggested that applying the wider philosophical notion of intentionality to this type of study may offer an alternative means to of understanding the problem of the evolution of Theory of mind, one that goes beyond the traditional disputes contrasting *Behaviour-reading* (animals just code the observable aspects of behaviour) with *Mind-reading* (animals code behaviour and the unobservable mental states beyond behaviour) (Gómez, 2009). The proposal is that, if intentionality in the philosophical sense is the key defining feature of mental states, then what an organism needs to do to be a 'mentalist' is to code the directionality of behaviour to external objects, and this can be done through some type of Gestalt perception where a relation is perceived between other organisms and their objects of action or attention. Thus, the object-directed actions, the object and the directedness from the action to the object are perceived as a whole. The above cases of perceiving the goal-directedness of actions to objects would be examples of this, but also abilities such as following the gaze of others to the objects they are looking at (Gomez, 2008, 2009). Findings in favour of this ability to code intentionality can be found in studies of gaze following, attention understanding and object-choice (see more discussion of these studies in chapter 1).

This Gestalt perspective on intentional relations predicts that when only part of a relation is observable, the whole relation can still be perceived or coded from the available elements. Some physiological studies, for example, suggest that even after an object has been covered behind a screen, object-directedness can still be triggered by the observation of the object-directed action (Umiltà et al., 2001). It indicates the possibility that the hidden object is remembered and when an appropriate action is performed the coding of object-directedness is triggered by combining the currently perceived action with the remembered object (Gómez, 2009).

This result is particularly interesting because it suggests an ability to attribute "intentionality" (both in the psychological sense of goal-directedness and the

philosophical sense of object-directedness) when the object of an action is not directly perceived. Some of the key ToM skills, that have frequently been considered to be exclusive human, are the ability to engage in pretend play (for example, acting upon imaginary objects) and the ability to attribute false beliefs (for example, predicting that someone will try to find an object where it actually is not). These two abilities involve the attribution of objects that can not be directly perceived (because they are imaginary or no longer exist) to actions.

However, the physiological experiments discussed above demonstrated the ability to 'imagine' the object of the action only after the object had been previously shown to the observer (Umiltà et al., 2001). So, it was more a matter of "remembering" rather than imagining or inferring the object. Could object-directedness be triggered and the object be imagined from only the observation of object-directed action, without previous information about whether or not there is or not an object? According to Gómez (2008), this would be a more important ability to understand the origins of the skills of Pretence and False belief attribution.

The ability to infer an object from the perceived intentionality of an action might be a crucial component of ToM development and evolution. However, no direct experimental evidence exists about this ability. Gaze following experiments suggest that both human and non-human primates might expect an object when they see someone looking in a particular direction (Rosati & Hare, 2009). However, no studies have specifically and systematically targeted the ability to infer objects from actions when no previous information has been given about the existence of an object.

Therefore, this thesis sought to comparatively examine the coding of object-directedness from the perception of actions without previous knowledge of their objects in both human adults and non-human primates. In the following sections of this chapter, we will recap the findings of our studies and discuss their implications for our knowledge of how humans and primates understand object-directedness and the notion of implicit Theory of Mind abilities.

6.2 Summary of the experiments

We conducted two comparative studies using two different paradigms in this thesis: first, a study of capuchin monkeys (chapter 2) and human adults (chapter 3) about their inference of a potential object from object-directed and non-object directed actions; and second, a study of capuchin monkeys (chapter 5) and human adults (chapter 4) about their automatic understanding of the intentional relation between an agent and an object in a paradigm initially designed to assess implicit understanding of other's belief.

6.2.1 Capuchin monkeys' inference of object from object-directed action (Chapter 2)

The results of the first experiment showed two main findings. Firstly, when an object is revealed after the actions, the capuchin monkeys can differentiate the object-directed action and the non-object directed action by expecting the object-directed action to be directed at the goal-object. This was shown by participants' longer looking time, (i.e., 'surprise'), at the outcome when the object-directed action, grasp (but not the non-object directed action, flop) had been directed to the alternative location instead of where the object was revealed. We suggest from these results that capuchin monkeys showed an understanding of object-directedness that goes beyond that shown in the findings with pre-verbal infants (Cannon & Woodward, 2012; Woodward et al., 2002; Woodward, 1999) and non-human primates (Kano & Call, 2014). Our findings suggest that the intentional relation of object directedness can be coded even if the action and the object have never been seen at the same time.

Secondly, the experiment indicated that if no object was revealed after the action, capuchin monkeys were apparently not able to imagine and infer the object, as indicated by their lack of longer looking time when no object was revealed either behind the congruent or the incongruent location. This suggested that, although capuchin monkeys understood object-directedness from the action without having to see the object at the same time, this was not due to the inference of an object behind the screen from the action alone. We suggested that a possible

interpretation was that the object-directed action might provoke the coding of an “object-slot”, but in the absence of a real object appearing shortly afterwards, this was not enough to generate a violation of expectation. It is as if the reality of there being no object overrides any suggestion of object-directedness generated by the contemplation of the action.

This would fit with the idea that non-human primates may have a cognitive system geared towards suppressing counterfactual representations (Gómez, 2008, 2009). Human children engage in pretend play from an early age, for example, pretend that a banana is a phone. But non-human primates do not seem to do so, perhaps due to the rooted ability to understand intentionality. A key step in the evolution of pretence and false belief understanding might be the inhibition of such suppression, or the intentionality to a certain realistic object, enabling the evolution of mental representation.

6.2.2 Adult humans’ inference of object from object-directed action (Chapter 3)

In the next chapter we aimed to compare adult humans in a similar paradigm using the same stimuli as with the monkeys, but with reaction time measures instead of looking time as the dependent variable, to investigate whether human adults could infer an object from object-directed actions.

The results, to our surprise, showed no evidence that adult humans understood object-directedness. Unlike the capuchin monkeys (chapter 2), adult humans didn’t distinguish the object-directed from the non-object-directed action when the object was not shown at the same time as the action, and seemed to regard all the actions as non-object-directed. Thus, when the object appeared at the incongruent location to the direction of the action, they didn’t show any effect in their reaction times, suggesting that they did not expect the object directed action to be congruent with the object location. In this respect, their performance was worse than that of capuchin monkeys (chapter 2). Neither did they show either any ability to infer a potential object from object-directed action when no object was revealed behind any of the screens, because they showed longer reaction times in the

absence of the object to both the object and the non-object directed actions. We discussed the possibility that these negative results were due to methodological aspects of the design, specifically the relatively long interval between the beginning of the action with the object cue (grasp vs flop) and the opportunity to respond to the outcome, and the 50% validity rate (percentage of trials where the action was congruent with the object), both of which might have given a better opportunity to the human participants to override any implicit, automatic ToM computations and control their response more explicitly.

However, if this were the case, then this would open up the need to question the adaptive value of the hypothetical implicit ToM systems. Apperly & Butterfill (2009) suggested that humans have an implicit ToM system, which deals with ToM-like problems but does not involve mental representation. This system provides us an automatic and fast way to problem solve in social situations. However, if an interval above 1 sec between the cueing action and the outcome is enough to make disappear any automatic computation of intentionality, one wonders if such an alleged implicit system would have any adaptive effect in real life situations, beyond artificial laboratory conditions where drastically abridged stimuli and outcomes (e.g., photos of actions instead of video clips) are shown.

On the other hand, we found no evidence of a decreased effect of the intentionality cues over blocks of trials, which would have been consistent with the idea that the adults were learning to ignore the cues because of their low predictive value.

The results from this chapter suggested that adult humans showed no effect of automatically encoding object-directedness in this task. This result was all the more surprising as Capuchin monkeys did show some ability to code object-directed intentionality with exactly the same stimuli when tested with a looking time measurement.

6.2.3 Summary of chapter 2 and 3

These first two empirical chapters (chapter 2 and 3) were designed to study intentionality by examining if the participants can infer the object from object-

directed actions when the object was not known beforehand. The comparative study expanded our knowledge of the ability to understand object-directedness from human infants to non-human primates and directly compared the performance of the two species using the same stimuli.

Despite this positive result, neither the capuchin monkeys nor the adult humans succeeded in inferring the object from only observing the object-directed action. The Capuchins showed no evidence of 'surprise' when no object at all appeared after the grasping action, and the humans adults, although clearly showed longer RTs to reporting the absence of the object, did so for both object-directed and non-object directed actions, including one in which the action did not refer at all to the screens (tap at the centre).

Therefore, it would appear that, at best, the Capuchins were capable of coding an intentional relation only if they had been exposed to separate direct evidence of both the action and the object, but not from the action alone. In this respect, our behavioural findings extend the physiological findings of Umiltà et al. (2001). Here a non-human primate species, macaque monkeys, was capable of coding an action alone as object-directed if the presence of the object had been shown in advance. We show that Capuchins can code object-directedness even if the object is shown *after* the completion of the action. This suggests a flexible and sophisticated ability to code object-directed intentionality as a relation between agents and objects that is not dependent upon direct simultaneous perception of both components of the relation. However, it falls short of showing the ability we were targeting — inference of an object expectation from the action alone, without any direct evidence of the object. The observation of object-directed actions alone is not sufficient in triggering the participants' inference of an object, or at least any inferred representation of an object is not strong enough to resist the perception of counterfactual reality. We suggested that it could, for example, consist of something like an "object slot" associated to a specific spatial location that would only be activated if an actual object appears, congruent or incongruent with that location.

Given the different results between capuchin monkeys and adult humans, especially the negative results from humans, we raised two doubts. Firstly, is the paradigm in chapter 3 adequate for detecting the adults' implicit ability? Secondly, is it valid to compare the two species using two different measurements of their reaction to the same stimuli, i.e. looking time for capuchin monkeys and reaction time for adult humans? Therefore in the next two experiments, we used a different paradigm, which had purportedly succeeded in comparing adult humans with RT measures with non-verbal human infants using Looking time measures, to re-examine capuchin monkeys' and adult human adults' understandings of object-directedness.

6.2.4 Adult humans' automatic encoding of object-directedness and their implicit understanding of belief (Chapter 4)

In the third empirical chapter (chapter 4) we modified a published experimental paradigm (Kovács, Téglás, & Endress, 2010) which originally succeeded in investigating implicit Theory of Mind ability in adult and infant humans. The authors interpreted their results as providing evidence of automatic belief computation, but we suggested that they could be interpreted through the automatic encoding of object-directedness in terms of the intentional relation between the agent and the object. We extended the original paradigm (entirely based on reactions to outcomes where the object was expectedly or unexpectedly present) by adding conditions where the object was absent in the final outcome, and reasoned that the predictions in the absent conditions would be different for a belief and an object-directedness interpretation. In addition, given that this paradigm successfully found comparable results with adult humans using reaction time and infants using looking time as measurements, we hoped it could be used in comparing adult human and capuchin monkeys (chapter 5) in the same task with the same stimuli.

The results were unexpected. They did not support either the Belief or the Object-directed interpretation, because to begin with we failed in the attempt to replicate Kovács et al's original results in the relevant conditions of our experiments

1 & 2. To try to determine if the source of our failure to replicate was slight changes we had introduced to the experimental procedure (aiming at a better control of variables such as the structure of the events shown in the videos corresponding to the different conditions) we conducted a number of additional experiments. The results from these experiments (experiment 3, 4 & 5) suggested that, rather than belief understanding or object-directedness, our results, and those of Kovacs et al (2010) could be better explained by superficial differences in the specific presentation of the stimuli that might actually have affected the participant's reaction time. Specifically, we found evidence that could support the alternative explanation recently proposed by Heyes (2014; paper published after we had completed our experiments) that the RT effects might have been caused by retroactive inhibition in the key conditions. Heyes suggested that due to the procedural difference between the conditions in Kovács et al's experiment, the events might interfere with the participants' memory and thereby affect their response in the test. This suggestion is very similar to our earlier discussion about procedural differences. Our experimental manipulations effectively (and unwittingly) controlled for retroactive inhibition by making all conditions comparable in this respect.

In sum, the absence of evidence provided no support for adults' ability in understanding either automatic belief computation or object-directedness, and they support recent alternative interpretations suggesting that instead of implicit ToM, results like Kovacs et al's might be interpreted as driven by other factors, "sub-mentalizing" processes (i.e., domain general cognitive processes) for example (Heyes, 2014).

6.2.5 Capuchin monkeys' automatic encoding of object-directedness and their implicit understanding of belief (Chapter 5)

The aim of the final empirical chapter was initially to expand Kovács et al's experimental paradigm (either in its Belief or its intentional object-directedness interpretations) to non-humans. Despite the negative results obtained with human

adults, we reasoned that the Capuchin monkeys might react differently to the stimuli, and we wanted to investigate whether by using a different measurement, looking time, the experimental design could work for Capuchin monkeys.

The results, however, indicated no encoding of either object-directedness or Belief by capuchin monkeys, as had been found with the adults. Moreover, the capuchin monkeys were indifferent to the four experimental conditions, indicating no effect of their own perspective (i.e., having last seen the object present or absent). After completing our experiment, a new paper by Martin and Santos (2014) reported a similar study using an adaptation of Kovacs et al, paradigm with rhesus macaques. The authors reported no Belief sensitivity in their monkeys. Although they had no control human group using the same adapted stimuli, they reported that the monkeys looking time was affected by their previous own knowledge of the object, i.e., they looked for longer if no objects appeared where they had just seen one. The complete lack of differentiation among conditions that we found in our Capuchin subjects led us to conclude that they might have failed to attend to or understand properly the experimental videos. Note that the Martin and Santos study with rhesus macaques was conducted using real people and with live manipulation of actual objects as opposed to our virtual ones.

6.2.6 Summary of Chapter 4 and 5

Chapters 4 & 5 used similar paradigms with identical stimuli to compare adult humans and capuchin monkeys in a task aimed at distinguishing the encoding of object-directedness in terms of encoding the intentional relationship between an agent and an object and the implicit computation of others' beliefs. Unfortunately, the results of the baseline conditions failed to replicate the original findings of Kovács et al (2010), thereby rendering impossible the exploration of the alternative interpretations. However, our exploration of this failure to replicate the original results with successive modifications in the experimental procedure led us to conclude that the reported effects might have been due to uncontrolled differences in the superficial structure of the experimental videos, which would support the sub-mentalizing' interpretation of implicit false belief recently proposed by Heyes (2014).

6.2.7 General conclusions & implications about the understanding of intentionality

The main aim of the thesis was to determine if Capuchin monkeys could infer (or imagine) the existence of an object that they have not seen from seeing an object-directed action such as grasping performed towards behind a screen. Following Gómez (2008), it was assumed that human adults would be capable of this. But we did not know whether a nonhuman primate would be. It was known that nonhuman primates can code intentional relations between the action of grasping and the object of the action when both terms of the relation are visible. Physiological evidence further suggested that grasping could be coded when the object of the grasp was no longer visible (but had been seen before), but not if the monkeys had been shown that there was no object behind the screen. What remained to be determined, according to Gomez, was whether monkeys would be able to infer an object when they did not know if there was anything behind the screen (an ability that was assumed human adults would undoubtedly possess).

What we found was that the monkeys showed an intermediate ability: they were not able to infer the existence of an object from the grasping action alone, as far as their looking time reflected no surprise at the screens being empty, but when an object appeared, they expected the object to be congruent with the grasping action that was no longer being performed (but not with a non-object directed act such as flop). This suggests that the grasping action alone does indeed provoke some object expectation, but this might not consist of a full representation of there having to be an object, but something akin to an 'object slot' that only becomes fully activated if an object appears. Alternatively, as suggested by Gómez (2008, 2009), the processing of reality (in this case, the absence of an object) might immediately and completely override any object inference suggested by the contemplation of the action alone. This would be similar to the finding reported by Umiltà et al. (2001) that the mirror neurons for grasping fail to activate when the monkeys know in advance that there is no object behind the screen the human is reaching for. Gómez speculation was that this might mean that the key step in the evolution of the ability to represent objects counterfactually (e.g. in pretence or in false belief attribution)

might have consisted, rather than in the creation of a new cognitive ability, in modifying the way in which representations of reality interact with previous non-matching representations. For example, in pretend play, the ability to represent a banana as a phone does not necessarily require the development of a new cognitive ability, such as mental representation. But to achieve it, available skills are rearranged. The intentionality from the mind to the banana already exists, thereby creating an object-slot for the mind. By retrieving a phone from the memory to fill in the object-slot, the banana is now pretended to be a phone.

The observation of the object-directed action was not enough to trigger the inference of the object, at least not in our paradigm. Note that inferring the object when it had not been displayed earlier in our paradigm was very demanding: we allowed very restricted information for the participants to observe, e.g., we eliminated the cue from gaze in the grasping/flop study. This may have resulted in too few pieces of information to form the whole “intentional relation” picture. In future studies, it will be interesting to investigate if providing more information, for example, grasping paired to looking, will facilitate the participants’ ability to infer about the object, and how much information is essential to trigger the inference.

Our studies add important evidence to the literature on non-human primate regarding their ability to understand object-directed actions (Kano & Call, 2014). We have demonstrated behaviourally that a nonhuman primate is capable of coding object-agent relations when the action and the agent are not perceived together. Until now this skill had only been suggested by the physiological evidence of mirror neuron studies with macaques, and in conditions where the presence/absence of the object was known in advance. Our study shows that the nature of the action displayed can be remembered and this affects the way in which the subsequently perceived object is processed.

The surprising, totally unexpected, finding was the lack of any evidence of object inference in human adults. As mentioned, the assumption was that they would be capable of demonstrating this in the RT paradigm; however, they not only failed to do so, but even failed to show the more modest ability of ‘retrospective’

object-directedness shown by the Capuchin monkeys. This was an indication that some assumptions made about the implicit ToM skills in human adults might be wrong, and this was further re-enforced by our failure to replicate the well-known results of Kovacs et al in another implicit ToM task.

6.3 About Implicit ToM

The second paradigm was developed from a task testing the ability to implicitly register other's belief. Here we investigated the ability of the two species about their implicitly understand of other's belief (chapter 4 & 5). We were particularly interested in our series of experiment with adult humans in Chapter 4. We suggest that more investigation is needed to further assess whether adults' have an implicit understand of other's belief.

6.3.1 Implicit ToM and the two system proposal

Recently, a core cognitive system was suggested to be responsible for spontaneous and implicit understanding of others' beliefs and other mental states in humans (Apperly & Butterfill, 2009; Baillargeon et al., 2010; Butterfill & Apperly, 2013; Thoermer, Sodian, Vuori, Perst, & Kristen, 2012; van der Wel, Sebanz, & Knoblich, 2014).

Implicit ToM tasks with adult humans showed evidence that adults were affected by others' beliefs even though these were irrelevant to the task they were performing (Kovács et al., 2010; van der Wel et al., 2014). Several recent studies show that infants' as young as 7-months-old and far before developing the capability to succeed in the classic false-belief task, were sensitive to false beliefs as well (e.g., Kovács et al., 2010; Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007).

The ability to implicitly understand ToM seems quite robust in humans across different developmental stages. However, when it comes to people with autism spectrum disorder (ASD), the situation becomes complicated. ASD individuals are generally known to have difficulties in ToM abilities (Baron-Cohen, Leslie, & Frith, 1985; Baron-cohen, 2001; Peterson & Bowler, 2000; Senju, 2011; Surian & Leslie, 1999). Although some high-function children and adults in this spectrum can infer

other's mental states when explicitly prompted to do so (Bowler, 1992; Ozonoff, Pennington, & Rogers, 1991), they fail to show the precursors and implicit understanding of false belief (Schneider, Slaughter, Bayliss, & Dux, 2013; Schuwerk, Vuori, & Sodian, 2014; Senju, Southgate, White, & Frith, 2009; Senju, 2012). For example, Senju and colleagues (Senju et al., 2009) used an eye-tracking device on a group of individuals with Asperger syndrome and tested their anticipatory looks in a FB task. They reported that the participants failed to spontaneously anticipate other's behaviour based on other's belief. On the other hand, "they are able to do so in explicit tasks through compensatory learning" (Senju et al., 2009). In another study, Schneider and colleagues (2013) compared typical adults and high-functioning ASD individuals. They found that both of them could pass the explicit false belief task but ASD group failed in an implicit task even with intense practice.

Based on this evidence, Apperly and colleagues (Apperly & Butterfill, 2009; Apperly, Warren, Andrews, Grant, & Todd, 2011; Butterfill & Apperly, 2013) proposed an interesting two-system model regarding tracking beliefs and belief-like states. They suggested that one system is predominantly in charge of implicit belief-understanding. The implicit abilities demonstrated by infants and adults, for instance, would be powered by this system. The other system, which is not only related to ToM abilities but also depends on the facilitation of inhibitory control and language, would be responsible for explicit theory of mind. When children reason about the protagonist's belief in the Sally-Ann task, it involves this second system. In the case of ASD individuals, their implicit ToM system is impaired and it hinders their ToM abilities. Yet through facilitation from language and other abilities, some of them can accomplish the explicit task by logical reasoning (e.g., Pyers & Senghas, 2009; Tager-Flusberg & Joseph, 2005). Apperly and Butterfill further proposed that these two systems operate in different manners, i.e., the former is fast and of low cost while the latter is relatively slower and of high cost. To better explain about the difference between the two systems, they made an analogy with the domain of number cognition. For numbers, a fast but inflexible cognitive system is used to recognize small amounts of numbers (less than 3 or 4) while another more cognitively demanding and flexible system is employed to process larger amounts. The case is

similar for ToM: “a cognitively efficient but inflexible capacity for tracking belief-like states” (cognitive system 1, CS1 hereafter) enables infants and nonhuman animals to understand ToM before they can succeed in the classic FB task; this capacity persists together with “a later-developing, more flexible but more cognitively demanding theory-of-mind abilities” (cognitive system 2, CS2 hereafter) that allows children and adult to deal with more complex social contexts. These two systems work together in providing humans with sophisticated skills in social contexts (Apperly & Butterfill, 2009).

More and more researchers are in favour of the two system hypothesis and have put forward two-system views of different sorts (e.g., De Bruin & Newen, 2012; Low & Perner, 2012). Nevertheless, some questions remain to be answered before we settle on the two systems hypothesis.

6.3.2 Undecided questions

6.3.2.1 The overlap of the two systems?

Firstly, if there are two systems for theory of mind processing, how do the two systems function together? And to ask the question in another and a more fundamental way, is the implicit ToM system distinct from the other one? Or, are the two just different developmental stages of the same system?

Apperly and Butterfill suggested that CS1 allows infants to understand basic social situations without consciously knowing about others’ belief (Apperly & Butterfill, 2009). Adults’ implicit encoding of others’ belief, for example, depends on this very system (CS1). CS2 allows more complicated reasoning and more sophisticated representation of other’s mental states, especially when some peripheral capabilities, e.g., language, executive function, become mature and provide better facilitation of belief understanding. The latter system leads to the capability to understand false belief when children reach 4-years of age (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013).

It is, however, not certain whether the two systems are truly distinct from each other. Firstly, some evidence suggests that implicit ToM abilities in infancy might be

predictive of children's later performance on explicit tasks (Aschersleben et al., 2008; W. A. Clements, Rustin, & Mccallum, 2000; Low, 2010; Wellman et al., 2004; Wellman, Lopez-Duran, LaBounty, & Hamilton, 2008). For example, infant's ability to perceiving or encoding the relation between goal-directed actions and the goal predicts children's understanding of FB at 4-years old (Aschersleben et al., 2008; Wellman et al., 2004). The correlation of performance in implicit understanding of object-directedness and in explicit false belief tasks indicats possible similarity in the fundamental cognitive process. If, on the other hand, the two systems are totally distinct, they are not likely to interfere with each other.

It is also problematic that the investigation into implicit ToM processing itself does not truly distinguish itself from explicit ToM tasks (Schneider Slaughter Dux, 2014). Some researchers argue that whether ToM behaviour occurs without consciousness remains unclear (Schneider Slaughter Dux, 2014). Indeed, the elimination of additional requirements such as language and inhibitory control ruled out some confounding factors (Hale & Tager-flusberg, 2003; Mckinnon & Moscovitch, 2007; Meer, Groenewold, Nolen, Pijnenborg, & Aleman, 2011); nonetheless it doesn't guarantee an isolated and implicit process. It is possible that less experimental demand allows more cognitive resource, which in turn allows success in ToM processing, but that these capabilities are not truly based on another system rather than part of CS2. In this case, once the cognitive load increases, the participants might show a deficit in "implicit ToM" tasks. There is some evidence in support of this suggestion. For example, Schneider and colleagues (2012) reported that when cognitive load increases, e.g., the participants had to listen to strings of letter and to report the letters while doing an implicit ToM task at the same time, implicit ToM processing was disrupted. They suggested that implicit ToM processing might exist, but executive function act as an important factor in affecting it (see also Mckinnon & Moscovitch, 2007; Rowe, Bullock, Polkey, & Morris, 2001; Schneider, Lam, et al., 2012).

Apart from executive function, language development also plays an important role in false-belief reasoning (Astington & Jenkins, 1999; Villiers, 2005; 2007). For example, language ability significantly correlates with performance on explicit false-

belief tasks (Astington, 2006; Low, 2010; Milligan, Astington, & Dack, 2007). In addition, the evidence from deaf children also provides support that language ability correlates with ToM: the development of ToM is delayed for deaf children if their ability in language is hindered. However, if they have acquired sign language from birth, they have no deficit or delay in explicit false belief tasks (Schick, Villiers, Villiers, & Hoffmeister, 2007). Although the correlation doesn't necessarily implicate a causal relationship between the development of language and explicit ToM (Slade & Ruffman, 2005), it is very likely that the development of language assists the explicit expression of belief understanding (Milligan et al., 2007). For example, children who have more conversation using words that involve mental representation (like think, belief, etc.) perform better in ToM tasks (e.g., Lohmann & Tomasello, 2003; Ted Ruffman, Slade, & Crowe, 2002). Some researchers further suggest that language might bridge implicit and explicit understanding of false belief (Juan & Astington, 2011). For example, Juan and Astington (2011) argued that labelling by language reduces the cognitive load for children in explicit ToM tasks and therefore allows explicit demonstration of their implicit understanding of other's belief. In other words, the implicit ability and explicit ability both depend on the same system to track other's belief, but language makes the difference between whether the understanding can be expressed explicitly. Thus they concluded that language development allow the transition from implicit to explicit ToM understanding (Juan & Astington, 2011).

In support of this position, fMRI studies indicate that several common brain areas are activated in both implicit and explicit ToM tasks (Gallagher & Frith, 2003; A. M. Kovács, Kühn, Gergely, Csibra, & Brass, 2014). Though a few differences are detected between the implicit and explicit tasks, they seem to have more commonalities than differences (Thoermer, Sodian, Vuori, Perst, & Kristen, 2012).

To summarize, implicit ToM correlates with explicit ToM abilities. The capability to understand implicit and explicit ToM may depend on the availability of cognitive resources; the capability to explicitly express the implicit understand of ToM might largely depend on language skills; and the activation of brain areas overlap in explicit and implicit Tom tasks. Therefore we suggest that it remains to be

determined whether the two systems are truly distinct from each other. And we tend to believe that they are not totally different, or at least, they share some commonalities.

6.3.2.2 Can human adults attribute belief implicitly?

The second question concerns whether adults can attribute belief implicitly as infants do. It comes from our failure in replicating Kovács et al.'s results, together with many other reports that indicated adults' failure in implicit ToM. Can human adults implicitly and automatically encode others' beliefs?

Kovács et al.'s study (2010) was one of the most highly cited articles among the ones we introduced earlier in support of automatic and implicit ToM processing (Kovács et al., 2010; Leslie, Friedman, & German, 2004; Low & Watts, 2013; Onishi & Baillargeon, 2005; Qureshi, Apperly, & Samson, 2010; Schneider, Nott, & Dux, 2014; Sperber & Wilson, 2002). Our failure to replicate it, however, rings an alarm bell for both the study and its interpretation, as well as other literature that reports adults' ability in implicit ToM understanding. Should the results be attributed to implicit ToM processing, or to other confounding factors?

It has only been very recently that Heyes (2014) has suggested plausible alternative interpretations of the findings about implicit ToM, both in human infants and adults, which might dispute the conclusion that they provide evidence for an implicit ToM system. Heyes suggests that many, if not all of the results, reported in these experiments, could be explained as the result of general-purpose cognitive mechanisms responding to relatively superficial features of the stimuli. Unwittingly, our experiments in chapters 4 and 5, designed and conducted before the publication of Heyes papers, have turned out to be a partial empirical test of some of her claims.

We used an experimental procedure that was slightly different from the original design. Specifically, we felt that it was necessary to modify the order of events displayed in the videos so that they were unified across conditions (Experiment 1 and 2 in chapter 5). With these better controlled stimuli, in which the factor of "retroactive inhibition" identified by Heyes (2014) as a potential explanation of the results, was kept constant in all conditions, we failed to replicate

Kovacs et al's results with human adults. In contrast, once the procedure of video presentations (and data analysis) was exactly the same as Kovács et al's design, the same results emerged, i.e., adults appeared to encode others' beliefs or object-directedness implicitly (Experiment 3 and 4 in chapter 5). However, when the intentional agent was replaced by a pile of boxes which was entirely non-intentional, if the procedure and sequence of events remained exactly the same as in the version of the video with an agent (with the boxes moving in and out of the scene), the participant showed the same results as with an agent, as if they were attributing a belief to the boxes (Experiment 5 in chapter 5). Kovacs et al. had used a "pile of boxes" control in which they found no belief computation; however, they had failed to make the control fully equivalent (e.g., the boxes did not move in and out of the scene, as in our version). As pointed out by Heyes (2014), this might have resulted in retroactive inhibition operating in the agent version, but not the boxes version, of their videos. The results of our controls, therefore, lend empirical support to Heyes' suggestions.

Taking these result together we concur with Heyes that adult participants do not truly display implicit understanding of other's belief or object-directedness, at least not in our paradigm nor in the original paradigm by Kovács et al (2010). The positive results (experiment 3 and 4 in chapter 5; experiment 1 and 2 in Kovács et al., 2010) may have been an artefact provoked by the confounding effects of the experimental design. The empirical support that our results provide for Heyes' critical reinterpretation of Kovacs results from a "sub-mentalizing" perspective opens up the possibility that the notion of implicit systems of theory of mind capable of computing belief-like states may need to be revised, including notions of dual systems of ToM (e.g., Apperly and Butterfield (2009).

In fact, Apperly and Butterfield have themselves reported results casting doubt on the existence of an ability to automatically compute others' beliefs in adults. For example, Apperly and colleagues (2010) reported that adult participants would not automatically track others' belief if not explicitly instructed to do so. In three experiments, they showed participants pictures that involved the location of an object and a character's belief about where the object was and measured the

participants' reaction time in reporting the object's location according to either reality or to the character's belief. Their results showed that unless explicitly instructed to, the participants wouldn't spontaneously keep track of others' belief but would focus their attention on reality. They suggested that belief was not ascribed automatically, but explicitly and intentionally (Back & Apperly, 2010).

Other criticism brings into question the results of the experiments that describe infants' achievements in implicit ToM. For example, Heyes (2014) reviewed more than 20 relevant experiments that use violation of expectancy or anticipatory looking paradigm and suggested that "their results can be explained by the operation of domain-general processes and in terms of low-level novelty". Similarly, Perner and Ruffman (2005) argued that Onishi and Barllargeon's (2005) experiment could be based on behaviour rules and they suggested that neuronal activation could have easily explained their results rather than touching the topic of belief or ToM understanding. Although abundant studies using various methods have tried to investigate infant's ToM abilities, debates and doubts continue as new evidence keeps accumulating.

Here we suggest that more attention should be lent to the negative results. With further investigation, the studies with negative results can provide important evidence either in ruling out confounding factors that prevent the detection of implicit ToM abilities, or on the other hand, prove the implicit ToM proposal wrong. Either way, it would be of great significance in progressing the field of ToM research.

6.3.2.3 Gestalt encoding of object-directedness?

We proposed in the introduction that the ability to automatically infer the goal and understand object-directedness in a gestalt way might be a fundamental process and an important step towards mental representation. When inferring about other's mental states, one has to reason from other's behaviour to imagine and uncover the underlying intention. Similarly, when perceiving object-directedness, one needs to infer from the object-directed actions about the potential goal-object. The processes are alike, i.e., infer the unobservable aspect from the available information, though the latter is much easier. "Mentalistic skills", therefore, might be built depending on

the basic understanding of object-directedness where expectation of potential outcome can be formed (Gómez, 2009).

The partially positive results we obtained with Capuchin monkeys in the grasp versus flop object-directed experiment suggest that the notion of computing intentional relations between agents and objects may play a role in this re-assessment of the notion of implicit mentalism and the significance of the empirical evidence on which it is based. Rather than automatic coding of representations equivalent to the epistemic notions of belief or desire, considered as unobservable mental states separate from their behavioural expressions; implicit ToM encoding could rest on the use of alternative mechanisms, such as computation of intentional relations between agents and objects of the environment which capture the key mental property of “directedness” or “aboutness” (Gómez, 2008, 2009) or computation of relations of “registration” attributed to agents (Apperly & Butterfill, 2009).

However, our results were only partially positive while the majority of the results indicated little supporting evidence. Nonetheless, we suggest that the curtain should not be closed so soon. Firstly, as we have discussed earlier in the introduction (chapter 1), abundant literature had suggested the prospect that primates and human might be able to infer an object from intentional actions. Behavioural studies, such as the object-choice task (Wood & Hauser, 2011a, 2011b) and geometrical gaze following (Tomasello et al., 1999), and physiological studies, such as the mirror neuron study in macaques (Umiltà et al., 2001), all suggested a high probability that an object could be inferred. Given this evidence, it is plausible that intentionality is understood in a gestalt manner. Nevertheless, to support the gestalt proposal, it remains to be determined whether the object can be inferred from the object (Gómez, 2009). Our experiment provides the first direct test of this proposal, although it provides negative evidence. Note that in our paradigm, we allowed very little information for the participant to observe, e.g., we eliminated the cue from gaze in the first paradigm. It may have resulted in too few pieces of parts for the brain to automatically self-organize and form the whole picture. We suggest that

further investigation is essential in confirming our findings and further development of the topic of object-directedness is still warranted.

In addition, although the ability to infer the potential object does not stand firm, there is ample evidence that primates and human share the ability to encode object-directed action together with its goal (Kano & Call, 2014). Given the early emergence of this ability both from the evolutionary perspective and the human-ontogenetic perspective, it is very likely to be a basic and fundamental ability. Moreover, considering the fact that such ability predicts the performance in ToM task when children grows to around 4 years of age, it might be a cornerstone for Theory of Mind development, and in a broader sense, for social cognition (Aschersleben et al., 2008; Wellman et al., 2004; Woodward, 2009). In spite of it, how are the action and the object encoded together and how this ability assists ToM development remains unclear. At least, it seems from the empirical works in this thesis that the gestalt perception of object-directedness fail to answer this question. But with accumulating evidence both about intentionality and about implicit ToM, we suggest future studies move from investigating the behavioural phenomenon to the mechanism of the behaviours. For example, fMRI studies might provide important evidence on whether crucial brains areas involved in ToM tasks are also activated in the object-directed action task, which might provide important support for the above issue. However, to conduct such experiment, participants that can succeed in both tasks are vital. Therefore adult participants should be given more attention as their results might reveal important evidence (Apperly & Butterfill, 2009). However, an oddity has been revealed with our experiment with adult humans: they seem to lack the ability to encode object-directed action and the object together (chapter 3). Given that infant masters this skill, it is not intuitive to see why adults do not encode the relationship between the action and the object any more. We suggest attributing their failure to the experimental paradigm, instead of the subjects' incapability, but it is still essential to re-examine this issue in the future. If adults indeed no longer possess this ability when they have reached a certain developmental stage, it is likely that the ability has transited to a part of ToM ability over development. If so, the study from infants to children to adults, in order to

setup a continuous spectrum of their ability of object-directedness and ToM abilities may help us to understand this process. Another possibility is that adults are dominated by explicit reasoning, instead of the implicit and reflexive behavioural response. In this case, it would also be interesting in investigating how implicit and explicit ToM understanding relate to each other.

Although we didn't find clear evidence from our experiments, we suggest that the gestalt inference of object-directedness and the encoding of the object-directed action together with the object might both provide interesting possibility for intention understanding and ToM development. Further investigation on this topic will be essential in investigating the gestalt inference of intentionality and ToM.

Reference

- Amici, F., Aureli, F., Visalberghi, E., & Call, J. (2009). Spider monkeys (*Ateles geoffroyi*) and capuchin monkeys (*Cebus apella*) follow gaze around barriers: evidence for perspective taking? *Journal of Comparative Psychology*, *123*(4), 368–74.
- Anderson, J. R., Kuroshima, H., Kuwahata, H., & Fujita, K. (2004). Do squirrel monkeys (*Saimiri sciureus*) and capuchin monkeys (*Cebus apella*) predict that looking leads to touching? *Animal Cognition*, *7*(3), 185–92.
- Anderson, J. R., Montant, M., & Schmitt, D. (1996). Rhesus monkeys fail to use gaze direction as an experimenter-given cue in an object-choice task. *Behavioural Processes*, *37*, 47–55.
- Anderson, J. R., Sallaberry, P., & Barbier, H. (1995). Use of experimenter-given cues during object-choice tasks by capuchin monkeys. *Animal Behaviour*, *49*(1), 201–208.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological review*, *116*(4), 953.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological review*, *116*(4), 953.
- Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–4.
- Aschersleben, G., Hofer, T., & Jovanovic, B. (2008). The link between infant attention to goal-directed action and later theory of mind abilities. *Developmental Science*, *11*(6), 862–8.
- Apperly, I. A., Warren, F., Andrews, B. J., Grant, J., & Todd, S. (2011). Developmental continuity in theory of mind: Speed and accuracy of belief–desire reasoning in children and adults. *Child development*, *82*(5), 1691-1703.
- Apperly, Ian A., Dana Samson, and Glyn W. Humphreys. "Studies of adults can inform accounts of theory of mind development." *Developmental psychology* 45.1 (2009): 190.
- Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental psychology*, *35*(5), 1311.
- Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*(1), 54–70.
- Baillargeon, R. (1995). A model of physical reasoning in infancy. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (pp. 305–371). Ablex Publishing Corporation.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–8.
- Baron-Cohen, S. (1991). Precursors to a theory of mind: Understanding attention in others. In A. Whiten (Ed.), *Natural theories of mind Evolution development and simulation of everyday mindreading* (pp. 233–251). Basil Blackwell.

- Baron-Cohen, S. (2001). Theory of mind and autism : a review. *International Review of Research in Mental Retardation*, (23), 196–184.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind” ? *Cognition*, 21(1), 37–46.
- Barth, J., Reaux, J. E., & Povinelli, D. J. (2005). Chimpanzees’ (Pan troglodytes) use of gaze cues in object-choice tasks: different methods yield different results. *Animal Cognition*, 8(2), 84–92.
- Biro, S., & Leslie, A. M. (2007). Infants’ perception of goal-directed actions: development through cue-based bootstrapping. *Developmental Science*, 10(3), 379–98.
- Bowler, D. M. (1992). “Theory of Mind” in Asperger's Syndrome. *Journal of Child Psychology and Psychiatry*, 33(5), 877-893.
- Brannon, E., Call, J., Drea, C. M., Emery, N., Haun, D. B. M., Herrmann, E., ... & Schroepfer, K. K. (2011). How does cognition evolve? Phylogenetic comparative psychology. *Anim Cogn*, 15, 223-38
- Marsh, H. L., Stavropoulos, J., Nienhuis, T., & Legerstee, M. (2010). Six- and 9-Month-Old Infants Discriminate Between Goals Despite Similar Action Patterns. *Infancy*, 15(1), 94-106.
- Marticorena, D. C. W., Ruiz, A., Mukerji, C., Goddu, A., & Santos, L. R. (2011). Monkeys represent others’ knowledge but not their beliefs. *Developmental Science*, 6, 1406–1416.
- Bräuer, J., Call, J., & Tomasello, M. (2005). All great ape species follow gaze to distant locations and around barriers. *Journal of Comparative Psychology (Washington, D.C. : 1983)*, 119(2), 145–54.
- Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, 10(4), 439–48.
- Bräuer, J., Kaminski, J., Riedel, J., Call, J., & Tomasello, M. (2006). Making inferences about the location of hidden food: social dog, causal ape. *Journal of Comparative Psychology*, 120(1), 38–47.
- Bugnyar, T., Stöwe, M., & Heinrich, B. (2004). Ravens, Corvus corax, follow gaze direction of humans around obstacles. *Proceedings. Biological Sciences / The Royal Society*, 271(1546), 1331–6.
- Burkart, J., & Heschl, A. (2006). Geometrical gaze following in common marmosets (*Callithrix jacchus*). *Journal of Comparative Psychology*, 120(2), 120–30.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*.
- Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2007). Enculturated chimpanzees imitate rationally. *Developmental Science*, 10(4), F31–8. doi:10.1111/j.1467-7687.2007.00630.x
- Butterfill, S. A., & Apperly, I. A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, 28(5), 1–36.

- Call, J., & Tomasello, M. (1999). A nonverbal false belief task: the performance of children and great apes. *Child Development, 70*(2), 381–95. Retrieved from
- Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*(5), 187–92.
- Call, J., Hare, B., & Tomasello, M. (1998). Chimpanzee gaze following in an object-choice task. *Animal Cognition, 1*(2), 89–99.
- Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). “Unwilling” versus “unable”: chimpanzees’ understanding of human intentional action. *Developmental Science, 7*(4), 488–98.
- Call, J., Tomasello, M., & Hare, B. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour, 68*(3), 571–581.
- Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science, 15*(2), 292–8.
- Clements, A. W., Clements, W. A., Perner, J., Clements, A. W., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*(4), 377–395.
- Clements, W. A., Rustin, C. L., & McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: A training study of false belief. *Developmental Science, 3*(1), 81–92.
- Cohen, A. S., & German, T. C. (2009). Encoding of others’ beliefs without overt instruction. *Cognition, 111*(3), 356–63.
- Cohen, A. S., & German, T. C. (2010). A reaction time advantage for calculating beliefs over public representations signals domain specificity for “theory of mind”. *Cognition, 115*(3), 417–25.
- Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: mood affects theory-of-mind use. *Emotion, 8*(5), 725.
- Couchman, J. J., Beran, M. J., Coutinho, M. V. C., Boomer, J., Zakrzewski, A., Church, B., & Smith, J. D. (2012). Do actions speak louder than words? A comparative perspective on implicit versus explicit meta-cognition and theory of mind. *The British Journal of Developmental Psychology, 30*(1), 210–21.
- Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 358*(1431), 447–58.
- De Bruin, L. C., & Newen, a. (2012). The developmental paradox of false belief understanding: a dual-system solution. *Synthese, 191*(3), 297–320.
- De Villiers, J. (2007). The interface of language and theory of mind. *Lingua, 117*(11), 1858–1878.

- De Villiers, Jill. "The interface of language and theory of mind." *Lingua* 117.11 (2007): 1858-1878.
- Dunphy-Lelii, S., & Wellman, H. M. (2004). Infants' understanding of occlusion of others' line-of-sight: Implications for an emerging theory of mind. *European Journal of Developmental Psychology*, 1(1), 49–66.
- Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, 24(6), 581–604.
- Flombaum, J. I., Santos, L. R., & Haven, N. (2005). Rhesus Monkeys Attribute Perceptions to Others. *Current Biology*, 15, 447–452.
- Fogassi, L. (2011). The mirror neuron system: How cognitive functions emerge from motor organization. *Journal of Economic Behavior & Organization*, 77(1), 66–75.
- Gallagher, H., & Frith, C. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Sciences*, 7(2), 77–83.
- Gallese, V. (2007). Before and below "theory of mind": embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1480), 659–69.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493–501.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, 7(7), 287–292.
- Gergely, G., Bekkering, H., Király, I., Damme, R. Van, & Wilson, R. S. (2002). Rational imitation in preverbal infants. *Nature*, 415(6873), 755.
- Gómez, J.-C. (2004). *Apes, Monkeys, Children, And the Growth of Mind*. Harvard University Press.
- Gómez, J.-C. (2008). The Evolution Of Pretence: From Intentional Availability To Intentional Non-Existence. *Mind & Language*, 23(5), 586–606.
- Gómez, J.-C. (2009). Embodying meaning: insights from primates, autism, and Brentano. *Neural Networks*, 22(2), 190–6.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59(1), 26–37.
- Griffey, J. A. F. (2011). Human and Non-Human Primate Preferences for Faces and Facial Attractiveness.
- Hale, C. M., & Tager-flusberg, H. (2003). The influence of language on theory of mind: a training study. *Developmental Science*, 3(6), 346–359.

- Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as I do: 7-month-old infants selectively reproduce others' goals. *Developmental Science*, *11*(4), 487–94.
- Hare, B., Tomasello, M., Call, J., & Agnetta, B. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785.
- Hauser, M. D., & Wood, J. N. (2010). Evolving the capacity to understand actions, intentions, and goals. *Annual Review of Psychology*, *61*, 303–24, C1.
- Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *science*, *317*(5843), 1360-1366.
- Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, *9*(2), 131–143.
- Heyes, C. M. (1998). Theory of mind in nonhuman primates. *The Behavioral and Brain Sciences*, *21*(1), 101–14; discussion 115–48.
- Itakura, S., & Tanaka, M. (1998). Use of experimenter-given cues during object-choice tasks by chimpanzees (*Pan troglodytes*), an orangutan (*Pongo pygmaeus*), and human infants (*Homo sapiens*). *Journal of Comparative Psychology*, *112*(2), 119.
- Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–34. doi:10.1016/j.cognition.2008.08.010
- Itakura, S., Agnetta, B., Hare, B., & Tomasello, M. (1999). Chimpanzee Use of Human and Conspecific Social Cues to Locate Hidden Food. *Developmental Science*, *2*(4), 448–456.
- Kano, F., & Call, J. (2014). Great Apes Generate Goal-Based Action Predictions: An Eye-Tracking Study. *Psychological Science*, *25*(9), 1691–8. doi:10.1177/0956797614536402
- Kersken, V., Rollins, R., & Gómez, J.-C. (n.d.).
- Kovács, Á. M., Kühn, S., Gergely, G., Csibra, G., & Brass, M. (2014). Are all beliefs equal? Implicit belief attributions recruiting core brain regions of theory of mind. *PLoS one*, *9*(9), e106558.
- Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830-1834.
- Leslie, A. M. (1987). Pretense and Representation : The Origins of " Theory of Mind ." *Psychological Review*, *94*(4), 412–426.
- Leslie, A. M. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition*, *50*(1-3), 211–238.
- Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind". *Trends in Cognitive Sciences*, *8*(12), 528–33.
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556.

- Lohmann, H., & Tomasello, M. (2003). The Role of Language in the Development of False Belief Understanding : A Training Study, *74*(4), 1130–1144.
- Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: relations with complex syntactical mastery. *Child Development*, *81*(2), 597–615.
- Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: state of the art. *The British Journal of Developmental Psychology*, *30*(Pt 1), 1–13.
- Low, J., & Wang, B. (2011). On the Long Road to Mentalism in Children's Spontaneous False-Belief Understanding: Are We There Yet? *Review of Philosophy and Psychology*, *2*(3), 411–428.
- Low, J., & Watts, J. (2013). Attributing False Beliefs About Object Identity Reveals a Signature Blind Spot in Humans' Efficient Mind-Reading System. *Psychological Science*, *24*(3), 305–311.
- Lurz, R. W., & Krachun, C. (2011). How could we know whether nonhuman primates understand others' internal goals and intentions? Solving Povinelli's problem. *Review of Philosophy and Psychology*, *2*(3), 449–481. Macdonald, C., & Whiten, A. (2011). The "Living Links to Human Evolution" Research Centre in Edinburgh Zoo: a new endeavour in collaboration. *International Zoo Yearbook*, *45*(1), 7–17.
- Martin, A., & Santos, L. R. (2014). The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition*, *130*(3), 300–308.
- McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218.
- Meer, L. Van Der, Groenewold, N. A., Nolen, W. A., Pijnenborg, M., & Aleman, A. (2011). NeuroImage Inhibit yourself and understand the other : Neural basis of distinct processes underlying Theory of Mind. *NeuroImage*, *56*(4), 2364–2374.
- Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850.
- Mikl, A., Soproni, K., & Miklósi, A. A. (2006). A comparative analysis of animals' understanding of the human pointing gesture. *Animal Cognition*, *9*(2), 81–93.
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind : Meta-Analysis of the Relation Between Language Ability and False-belief Understanding, *78*(2), 622–646.
- Moll, H., & Tomasello, M. (2004). 12-and 18-month-old infants follow gaze to spaces behind barriers. *Developmental science*, *7*(1), F1–F9. Mulcahy, N. J., & Call, J. (2009). The performance of bonobos (*Pan paniscus*), chimpanzees (*Pan troglodytes*), and orangutans (*Pongo pygmaeus*) in two versions of an object-choice task. *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *123*(3), 304–9.
- Mulcahy, N. J., & Hedge, V. (2012). Are great apes tested with an object object-choice task? *Animal Behaviour*, *83*(2), 313–321.

- Ocampo, B., & Kritikos, A. (2011). Interpreting actions: the goal behind mirror neuron function. *Brain Research Reviews*, *67*(1-2), 260–7.
- Okamoto-Barth, S., Call, J., & Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychological Science*, *18*(5), 462–8.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science*, *308*(5719), 255-258. Paulus, M. (2011). How infants relate looker and object: evidence for a perceptual learning account of gaze following in infancy. *Developmental Science*, *14*(6), 1301–1310.
- Ozonoff, S., Pennington, B., & Rogers, S. (1991). Executive function deficits in high-functioning autistic Individuals: relationship to Theory of Mind. *Journal of Child Psychology and Psychiatry*, *32*(7), 1081–1105.
- Perner, J., & Ruffman, T. (2005). Psychology. Infants' insight into the mind: how deep? *Science*, *308*(5719), 214–6.
- Perner, J., & Wimmer, H. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.
- Peterson, D., & Bowler, D. (2000). Counterfactual reasoning and false belief understanding in children with autism. *Autism*, *4*(4), 391–405.
- Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition*, *98*(2), 137–55.
- Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*(1), 53–78.
- Phillips, W., Barnes, J. L., Mahajan, N., Yamaguchi, M., & Santos, L. R. (2009). “Unwilling” versus “unable”: capuchin monkeys' (*Cebus apella*) understanding of human intentional action. *Developmental Science*, *12*(6), 938–45.
- Povinelli, D. J., & Eddy, T. J. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, *61*(3).
- Povinelli, D. J., & Eddy, T. J. (1997). Specificity of gaze-following in young chimpanzees. *British Journal of Developmental Psychology*, *15*(2), 213–222.
- Povinelli, D. J., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, *19*(1), 1-28. Premack, D., & Woodruff, G. (1978). Does the Chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *4*, 515–526.
- Povinelli, D. J., Perilloux, H. K., Reaux, J. E., & Bierschwale, D. T. (1998). Young and juvenile chimpanzees' (*Pan troglodytes*) reactions to intentional versus accidental and inadvertent actions. *Behavioural Processes*, *42*(2-3), 205–218.
- Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: evidence from learners of a new sign language. *Psychological Science*, *20*(7), 805–12.

- Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection , not Level-1 visual perspective calculation : Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–92.
- Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental Brain Research*, *71*(3), 491–507.
- Rochat, M. J., Serra, E., Fadiga, L., & Gallese, V. (2008). The evolution of social cognition: goal familiarity shapes monkeys' action understanding. *Current Biology*, *18*(3), 227–232.
- Rosati, A. G., & Hare, B. (2009). Looking past the model species: diversity in gaze-following skills across primates. *Current Opinion in Neurobiology*, *19*(1), 45–51.
- Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). “ Theory of mind ” impairments and their relationship to executive functioning following frontal lobe excisions, 600–616.
- Ruffman, T., Garnham, W., Import, a, & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, *80*(3), 201–24.
- Ruffman, T., Slade, L., & Crowe, E. (2002). The Relation between Children ’ s and Mothers ’ Mental State Language and Theory-of-Mind Understanding, *73*(3), 734–751.
- Santos, L. R., & Hauser, M. D. (1999). How monkeys see the eyes: cotton-top tamarins’ reaction to changes in visual attention and action. *Animal Cognition*, *2*(3), 131–139.
- Savage-Rumbaugh, S. E., Rumbaugh, D. M., & Boysen, S. (1978). Sarah’s problems of comprehension. *The Behavioral and Brain Sciences*, *4*, 555–557.
- Scerif, G. S. C., Gómez, J.-C., Byrne, R. W., & Gomez, J.-C. (2004). What do Diana monkeys know about the focus of attention of a conspecific? *Animal Behaviour*, *68*(6), 1239–
- Schick, B., Villiers, P. De, Villiers, J. De, & Hoffmeister, R. (2007). Language and Theory of Mind : A Study of Deaf Children, *78*(2), 376–396.
- Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science*, *23*(8), 842–7.
- Schneider, D., Slaughter, V. P., & Dux, P. E. (2014). What do we know about implicit false-belief tracking? *Psychonomic Bulletin & Review*, *22*(1), 1–12.
- Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit theory of mind reasoning in autism spectrum disorders: the impact of experience. *Autism*, *19*(4), 459-468.
- Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *The Neuroscientist*, *18*(2), 108-113.

- Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 18(2), 108–13.
- Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science*, 325(14), 883–885.
- Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in integrative neuroscience*, 4, 5. Sodian, B. (2009). Theory of Mind in Infancy. *Infancy*, 5(1), 39–43.
- Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind : A longitudinal study of syntax , semantics , working memory and false belief, 117–141.
- Sodian, B., & Thoermer, C. (2004). Infants' understanding of looking, pointing, and reaching as cues to goal-directed action. *Journal of Cognition and Development*, 5(3), 289-316.
- Sommerville, J. a, & Woodward, A. L. (2005). Pulling out the intentional structure of action: the relation between action processing and action production in infancy. *Cognition*, 95(1), 1–30.
- Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 13(6), 907–12.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18(7), 587-592.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18(7), 580-586.
- Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *The British Journal of Developmental Psychology*, 30(Pt 1), 172–87.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and Brain Sciences*, 28(5), 675–91; discussion 691–735.
- Tomasello, M., Hare, B., & Agnetta, B. (1999). Chimpanzees, Pan troglodytes, follow gaze direction geometrically. *Animal Behaviour*, 58(4), 769–777.
- Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: do infants understand others' beliefs?. *Infancy*, 15(4), 434-444.
- Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing: A neurophysiological study. *Neuron*, 31(1), 155-165.
- Van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, 130(1), 128–133.
- Wass, S. V., & Smith, T. J. (2014). Individual differences in infant oculomotor behavior during the viewing of complex naturalistic scenes. *Infancy*, 19(4), 352–384.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75(2), 523–41.

- Wellman, H. M., & Phillips, A. T. (2001). Developing intentional understandings. *Intentions and Intentionality: Foundations of Social Cognition*, 125–148.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72(3), 655–84.
- Wellman, H. M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology*, 44(2), 618–23.
- Wellman, H. M., Phillips, A. T., Dunphy-Ielii, S., Lalonde, N., Hospital, M. S., & York, N. (2004). Infant social attention predicts preschool social cognition. *Developmental Science*, 3, 283 – 288.
- Wimmer, H., & Weichbold, V. (1994). Children's theory of mind: Fodor's heuristics examined. *Cognition*, 53(1), 45–57.
- Wood, J. N., Glynn, D. D., & Hauser, M. D. (2008). Rhesus monkeys' understanding of actions and goals. *Social Neuroscience*, 3(1), 60–8.
- Wood, J. N., Glynn, D. D., Phillips, B. C., & Hauser, M. D. (2007). The perception of rational, goal-directed action in nonhuman primates. *Science*, 317(5843), 1402-1405. Wood, J. N., & Hauser, M. D. (2008). Action comprehension in non-human primates: motor simulation or inferential reasoning? *Trends in Cognitive Sciences*, 12(12), 461–5.
- Wood, J. N., Glynn, D. D., Phillips, B. C., & Hauser, M. D. (2007). The perception of rational, goal-directed action in nonhuman primates. *Science*, 317(5843), 1402-1405. Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1-34. Woodward, A. L. (1998b). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1–34.
- Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, 22(2), 145–160.
- Woodward, A. L. (2009a). Infants' Grasp of Others' Intentions. *Psychological Science*, 18(1), 53–57.
- Woodward, A. L. (2009b). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, 18(1), 53–57.
- Woodward, A. L., & Guajardo, J. J. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*, 6(3), 297–311.
- Woodward, A. L., Guajardo, J. J., Woodward, & Guajardo, J. J. (2002). Infants' understanding of the point gesture as an object-directed action. *Cognitive Development*, 17(1), 1061–1084.
- Wynn, K. (1998). Psychological foundations of number: Numerical competence in human infants. *Trends in Cognitive Sciences*, 2(8), 296–303.

Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*, 30(2), 111–153.

Appendix

Appendix I. Ethical approval for research with Capuchin Monkeys



University of St Andrews

School of Psychology Ethics Committee

4 May 2012

Project Title:	Understanding of intentional actions and goals in non human primates
Researcher's Name:	Ruoting Fan
Supervisor:	Dr Juan Carlos Gomez

Thank you for submitting your application which was considered at the Psychology School Ethics Committee meeting on the 7th March 2012. The following documents were reviewed:

1. Animal Ethics Form 14/03/2012
2. External Permissions 03/05/2012

The School of Psychology Ethics Committee approves this study from an ethical point of view.

Approval is given for three years. Projects, which have not commenced within two years of original approval, must be re-submitted to the School Ethics Committee.

You must inform the School Ethics Committee when the research has been completed. If you are unable to complete your research within the 3 three year validation period, you will be required to write to the School Ethics Committee to request an extension or you will need to re-apply.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the understanding that the ASAB Guidelines for the Treatment of Animals in Behavioural Research and Teaching (published in *Animal Behaviour*, 2003, 65, 249-255, <http://www.sciencedirect.com/>) are adhered to.

Yours sincerely

Convener of the School Ethics Committee

Cc: Dr J. C. Gomez (Supervisor)
School Ethics Committee
Dr Tamara Lawson (Home Office Liaison Officer)



18 February 2014

Project Title:	Understanding of intentional actions and goals in non-human primates
Researcher's Name:	Ruoting Tao
Supervisor:	Dr Juan Carlos Gomez

Thank you for submitting your application for amendment which was considered at the Psychology & Neuroscience School Ethics Committee meeting on 12th February 2014. The following documents were reviewed:

1. Application for change(s) to a School Ethics Committee Form 12/02/2014
2. Amended Animal Ethics Form 12/02/2014

The School of Psychology & Neuroscience Ethics Committee approves the amendment to this study from an ethical point of view.

Approval is given for three years from the date of approval of the original application. Projects, which have not commenced within two years of original approval, must be re-submitted to the School Ethics Committee.

You must inform the School Ethics Committee when the research has been completed. If you are unable to complete your research within the 3 three year validation period, you will be required to write to the School Ethics Committee to request an extension or you will need to re-apply.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and a further application for amendment submitted where appropriate.

Approval is given on the understanding that the ASAB Guidelines for the Treatment of Animals in Behavioural Research and Teaching (published in *Animal Behaviour*, 2003, 65, 249-255, <http://www.sciencedirect.com/>) are adhered to.

Yours sincerely

Convener of the School Ethics Committee

Cc: Dr J. C. Gomez (Supervisor)
School Ethics Committee
Dr Tamara Lawson (Home Office Liaison Officer)

Appendix II. Test of monkeys' attention across sessions in chapter 2

We assessed whether monkeys' attention to the test phases, i.e., phase two and three together, remained at the same level across sessions (see Figure 2-6), because some monkeys did not seem to remain interested in the experiment during the last sessions. We calculated each monkey's average looking time in each session during the 10-second window of phase 2 and 3 combined, regardless of the action or the outcome. The looking time data across the 9 sessions was examined by a one-way repeated-measure analysis of variance (ANOVA), using session number as the only factor. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(35) = 62.402, p = 0.012$) therefore the Greenhouse-Geisser approach was used to adjust the p value. An ANOVA showed that the difference between sessions was approaching significance ($F(8, 64) = 2.53, p = 0.06$), suggesting that the monkeys average looking time in each session was not equally distributed across the 9 sessions. Paired T-test with Bonferroni adjustment showed no significant difference between the looking time of any pair of the sessions. Nevertheless, the monkey's looking time at the 7th session had dropped below any of the previous session and did not recovered in the last two sessions. This might indicate a trend towards loss of attention and the non significant difference might be due to the sample size ($n=9$) and the large variance.

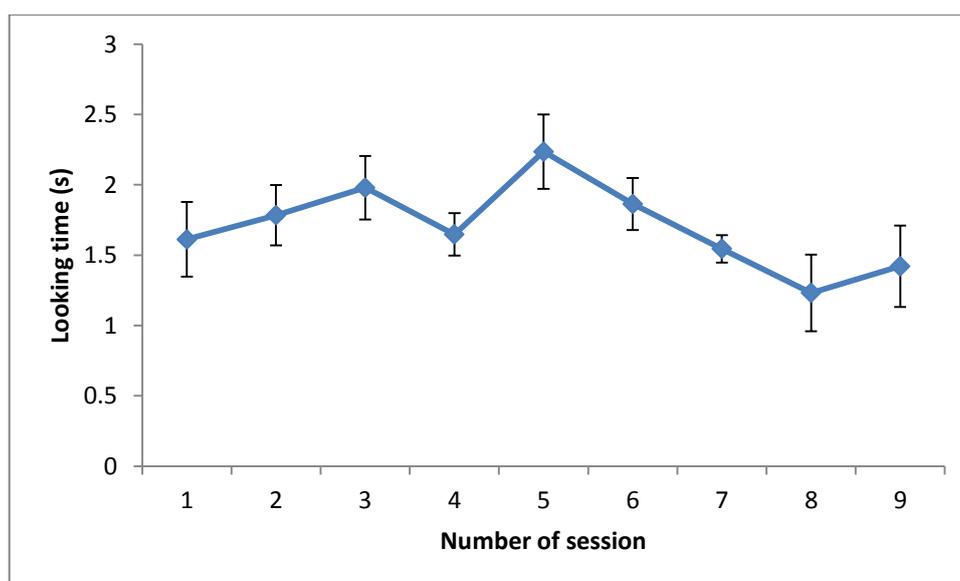


Figure 2 - 10 Monkey's average trial-looking time in each session

Independent of the looking time, the number of attended trials could also indicate the monkeys' attention to the experiments. Table 2-2 showed the monkeys' number of attended trials in each session. About half of the monkeys were only paying attention to one or two trials in the 8th and 9th session. And in the 7th session, one monkey, Diego, did not pay attention to any of the trials at all. Therefore we concluded that the monkey attention dropped in the last few sessions, especially the last three.

Table 2 - 3 Number of attended trials (6 in total) in each session

Monkey	Session number								
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
Anita	2	4	4	4	4	4	5	4	4
Chico	5	4	6	6	6	1	4	2	1
Diablo	5	4	4	4	2	2	4	2	-
Diego	4	2	4	3	3	4	0	1	6
Figo	4	5	4	4	1	6	3	2	3
Inti	5	3	5	6	4	3	3	5	3
Junon	2	1	3	4	3	5	4	2	1
Meako	4	4	5	3	3	4	4	4	2
Ruben	4	6	5	5	3	1	3	3	4
Torris	2	6	2	5	6	4	3	5	3
Ximo	5	1	5	6	4	3	1	3	2
Total	42	40	47	50	39	37	34	33	29

We grouped the first three, second three and the last three sessions together and examined the monkey's average looking time at a trial, regardless of trial-condition to test this proposal. We conducted an one-way repeated-measure analysis of variance (ANOVA) with one factor, i.e., average looking time of trials in session 1-3, session 4-6 and session 7-9 (see Table 2-3 and Figure 2-7). Mauchly's test indicated that the assumption of sphericity had not been violated ($\chi^2(2) = 2.49, p = .29$). The ANOVA showed a significant effect of groups ($F(2, 20) = 4.62, p = .02$) indicating different average looking time of trials in the three session-clusters. The post-hoc test (T-test with Bonferroni's procedure) suggested that the monkeys looked significantly shorter in the last three sessions than the first three sessions ($t(10) = 3.32, p = 0.02$) and the difference between the last three and the second three

sessions was approaching significance ($I(10) = 2.72$, $I = 0.06$), however, no significant difference was found between the first and the second three sessions ($I(10) = 0.22$, $I = 1.000$).

Table 2 - 4 Average looking time of trials in every three sessions.

	Sessions		
	1-3	4-6	7-9
Mean looking time (s)	1.84	1.89	1.38
SD	0.18	0.16	0.16

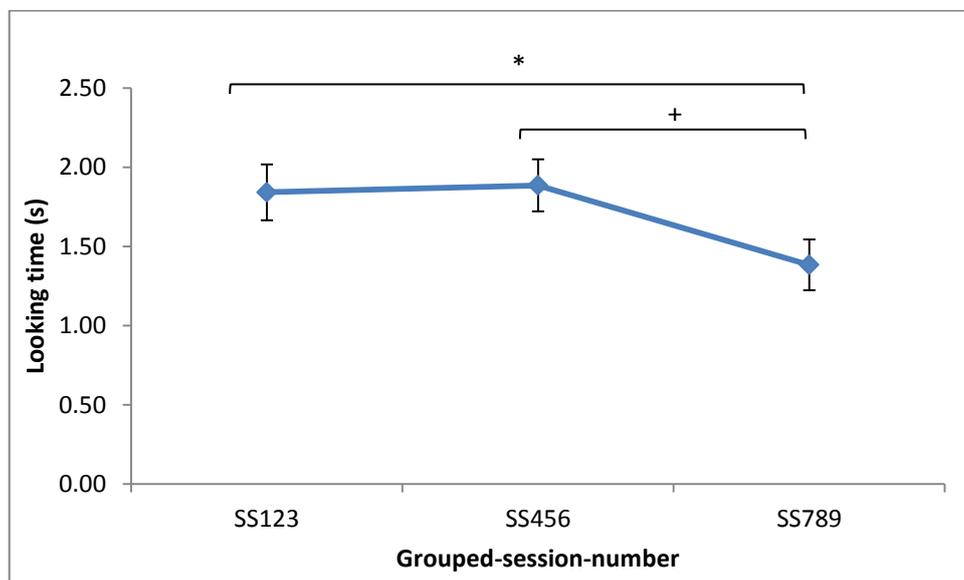


Figure 2 - 11 Monkey's average looking time of trials in every three sessions

Considering the drop of interest in the last three sessions, we decided to only include the first 6 sessions in further data analysis. When not specified, the statistical analysis in the following sections only included the first 6 sessions.

Appendix III. The questionnaire in experiment 1 of Chapter 2

What do you think of the images?

Please tick on the scale which best describes your opinion on the following statements.

1 – 2 – 3 – 4 – 5 – 6 – 7

Strongly disagree – disagree – slightly disagree – neutral – slightly agree – agree – Strongly agree

1. Participant No.

2. Gender *

- Male Female Prefer not answer



3. Please tick on the scale which best describes your opinion on the following statements. *

	Strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	Strongly agree
1. This action indicates the location of the object.	<input type="radio"/>						
2. This action distracts me from the task.	<input type="radio"/>						
3. The actress knows if there's an object.	<input type="radio"/>						
4. This action had nothing to do with the object.	<input type="radio"/>						
5. I ignored the action when making the judgement.	<input type="radio"/>						
6. The actress did the action randomly.	<input type="radio"/>						
7. This action is	<input type="radio"/>						

helpful for me to succeed in the task.							
8. This action is directed to an object.	<input type="radio"/>						
9. I tried to ignore the action.	<input type="radio"/>						
10. The actress did the action for some purpose.	<input type="radio"/>						
11. This action looks random and irrelevant to the task.	<input type="radio"/>						
12. This action does not distract me at all.	<input type="radio"/>						
13. The action was aimed to the object.	<input type="radio"/>						



4. Please tick on the scale which best describes your opinion on the following statements. *

	Strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	Strongly agree
1. This action indicates the location of the object.	<input type="radio"/>						
2. This action distracts me from the task.	<input type="radio"/>						
3. The actress knows if there's an object.	<input type="radio"/>						
4. This action had nothing to do with the	<input type="radio"/>						

object.							
5. I ignored the action when making the judgement.	<input type="radio"/>						
6. The actress did the action randomly.	<input type="radio"/>						
7. This action is helpful for me to succeed in the task.	<input type="radio"/>						
8. This action is directed to an object.	<input type="radio"/>						
9. I tried to ignore the action.	<input type="radio"/>						
10. The actress did the action for some purpose.	<input type="radio"/>						
11. This action looks random and irrelevant to the task.	<input type="radio"/>						

12. This action does not distract me at all.	<input type="radio"/>						
13. The action was aimed to the object.	<input type="radio"/>						

Appendix IV. Example of the form in chapter 3 and 4

Example of forms in chapter 3

Verbal instruction

In the following experiment, we'll present you with several videos. In the videos, there will be two hiding-screens and you'll try to react as quickly and accurate as possible when the hiding-screens open to inform about an object behind the screens. There will also be some questions somewhere, asking what the last action was. So please watch the videos carefully.

During the experiment, please always keep your right hand on the direction keys, with the index finger, middle finger and the ring finger on LEFT, DOWN and RIGHT key, respectively [point out the keys and ask the participants to do so]. Press left if you find the object behind the left screen press right if right, and press down if nothing's there.

You'll finish 4 test blocks in total, before that, you'll have a practice block, where you'll have feedback after each trial, and in the other 4 test blocks, you'll not have any feedback.

The task will last for about 20 minutes. After the reaction task, we'll ask you to finish a questionnaire, which will take less than 5 minutes.

If you have any questions, please come to me.

Participant information sheet

Project Title Where's the apple? A reaction time task

What is the study about?

We invite you to participate in a research project which concerns the reaction time and accuracy of your reaction about the presence of an object at different locations in a scenario.

This study is being conducted as part of my, Ruoting Tao's, PhD Thesis in the School of Psychology and Neuroscience.

Do I have to take Part?

This information sheet has been written to help you decide if you would like to take part. It is up to you and you alone whether or not to take part. If you do decide to take part you will be free to withdraw at any time without providing a reason.

What would I be required to do?

You will be presented with video clips of an actress and two hiding-screens on a computer. The actress will perform an action and after that the two hiding-screens will open. You will then need to press a key on the keyboard to indicate as quickly and accurate as possible if there's an object and behind which hiding-screen it is. Please follow the instructions on the computer during the experiment. This reaction test will last about 20 minutes. After that you'll be asked to fill a brief questionnaire about the task.

Will my participation be Anonymous and Confidential?

Only the researcher(s) and supervisor(s) will have access to the data which will be kept strictly confidential. Your permission will be sought in the Participant Consent form for the data you provide, which will be anonymised, to be used for future scholarly purposes.

Storage and Destruction of Data Collected

The data we collect will be accessible by the researcher(s) and supervisor(s) involved in this study only, unless explicit consent for wider access is given by means of the consent form. Your data will be stored for a period of 5 years since time of publication of the results before

destroyed in an anonymised format in a hard drive which will be locked in a storage cupboard.

What will happen to the results of the research study?

The results will be finalised by 2014 and written up as part of my PhD Thesis and they may be published as part of journal articles or conferences presentations

Reward

£5 for your participation

Are there any potential risks to taking part?

No. If you notd not feel comfortable about any part of the experiment, you could withdraw at any time.

Questions

You will have the opportunity to ask any questions in relation to this project before giving completing a Consent Form.

Consent and Approval

This research proposal has been scrutinised and been granted Ethical Approval through the University ethical approval process.

What should I do if I have concerns about this study?

A full outline of the procedures governed by the University Teaching and Research Ethical Committee is available at <http://www.st-andrews.ac.uk/utrec/Guidelines/complaints/>

Contact Details

Researcher: Ruoting Tao

Contact Details: Email: rt33@st-andrews.ac.uk

Supervisor: Dr. Juan-Carlos Gomez

Contact Details: Email: jg5@st-andrews.ac.uk Tel: 01334 46 2059

Participant Consent Form (Anonymous Data)

Project Title: Where's the apple? A reaction time task

Researcher(s) Name(s)

Ruoting Tao

Supervisors Names

Dr. Juan-Carlos Gomez

Email: rt33@st-andrews.ac.uk

Email: jg5@st-andrews.ac.uk

Tel: 01334 46 2059

The University of St Andrews attaches high priority to the ethical conduct of research. We therefore ask you to consider the following points before signing this form. Your signature confirms that you are happy to participate in the study.

What is Anonymous Data?

The term 'Anonymous Data' refers to data collected by a researcher that has no identifier markers so that even the researcher cannot identify any participant. Consent is still required by the researcher, however no link between the participant's signed consent and the data collected can be made.

Consent

The purpose of this form is to ensure that you are willing to take part in this study and to let you understand what it entails. Signing this form does not commit you to anything you do not wish to do.

Material gathered during this research will be anonymous, so it is impossible to trace back to you. It will be securely stored in a hard drive for 5 years since the time of publication before destroyed. Please answer each statement concerning the collection and use of the research data.

I have read and understood the information sheet. Yes No

I have been given the opportunity to ask questions about the study. Yes No

I have had my questions answered satisfactorily. Yes No

I understand that I can withdraw from the study without having to give an explanation. Yes No

I understand that my data once processed will be anonymous and that only the researcher(s) (and supervisors) will have access to the raw data which will be kept confidentially. Yes No

I understand that my data will be stored for a period of 5 years since the time of publication being destroyed Yes No

I have been made fully aware of the potential risks associated with this research and am satisfied with the information provided. Yes No

I agree to take part in the study Yes No

Participation in this research is completely voluntary and your consent is required before you can participate in this research.

Name in Block Capitals

Signature

Date

Participant Debriefing Form

Project Title: Where's the apple? A reaction time task

Nature of Project

This postgraduate research project was conducted to investigate whether human will automatically relate an intentional action, for example grasping, but not a non-intentional action, a resting movement, to a potential object. If you do infer the object from the action, you'll react quicker when the object position is congruent. In separate studies, we are trying to investigate if non-human primates can do this as well; data with humans will help us to better interpret our results with non-human primates.

Storage of Data

As outlined in the Participant Information Sheet your data will now be retained for a period of 5 years since time of publication of the results before being destroyed. Your data will remain accessible to only the researchers and supervisors if you have given permission on the Consent Form.

What should I do if I have concerns about this study?

A full outline of the procedures governed by the University Teaching and Research Ethical Committee are outline on their website <http://www.st-andrews.ac.uk/utrec/Guidelines/complaints/>

Contact Details

Researcher: Ruoting Tao

Contact Details:Email: rt33@st-andrews.ac.uk

Supervisor: Dr. Juan-Carlos Gomez

Contact Details:Email: jg5@st-andrews.ac.uk Tel: 01334 46 2059

Example of forms in chapter 4

Verbal instruction

PRESENCE

You'll take part in a study investigating visual attention through reaction times. You will watch a series of short movies featuring Mario, a ball and a hiding-screen.

At a certain point in each of the videos, you'll see the hiding-screen scrolls down. Your task is to press the "ENTER" key on the keyboard as soon as the hiding-screen opens and you detect that the ball is present. If the ball is absent, you don't have to do anything. Try to be as fast as possible, press the key as soon as the screen scrolls down. We'll measure how fast you can detect the presence of the ball, while at the same time please avoid making mistakes.

In order to make sure you pay attention to the whole video, we also ask you to press the "SPACE" key when Mario leaves the scene. We will record your response, but not the speed.

Between videos, you'll be asked to press "SPACE" or "ENTER" to continue the experiment. Please do so when you are ready to start.

The study will take about 15 minutes, please try to pay attention to all the videos and crucially remember that you have to press the "Enter" key as fast as you can when the hiding-screen opens and you find the ball was present behind the screen.

Now you will have 4 practice trials. Do you have any questions before we start?

Please keep your hands on the keyboard and follow the instructions on the screen.

ABSENCE

You'll take part in a study investigating visual attention through reaction times. You will watch a series of short movies featuring Mario, a ball and a hiding-screen.

At a certain point in each of the videos, you'll see the hiding-screen scrolls down. Your task is to press the "ENTER" key on the keyboard as soon as the hiding-screen opens and you detect that the ball is absent. If the ball is present, you don't have to do anything. Try to be as fast as possible, press the key as soon as the screen scrolls down. We'll measure how fast you can detect the absence of the ball, while at the same time please avoid making mistakes.

In order to make sure you pay attention to the whole video, we also ask you to press the “SPACE” key when Mario leaves the scene. We will record your response, but not the speed.

Between videos, you’ll be asked to press “SPACE” or “ENTER” to continue the experiment. Please do so when you are ready.

The study will take about 15 minutes, please try to pay attention to all the videos and crucially remember that you have to press the “Enter” key as fast as you can when the hiding-screen opens and you find the ball was absent.

Now you will have 4 practice trials. Do you have any questions before we start?

Please keep your hands on the keyboard and follow the instructions on the screen.

Example of information form

Project Title

Visual detection tasks: how fast can you find the object?

What is the study about?

We invite you to participate in a research project which concerns the reaction time and accuracy of your reaction about the absence of an object in a scenario. This study is being conducted as part of my, Ruoting Tao's, PhD Thesis in the School of Psychology and Neuroscience.

Do I have to take Part?

This information sheet has been written to help you decide if you would like to take part. It is up to you and you alone whether or not to take part. If you do decide to take part you will be free to withdraw at any time without providing a reason.

What would I be required to do?

You will be presented with animation video clips of an agent, a ball, a hiding screen and a bin on a computer. The agent will come into the scene or leave the scene at some points. The ball will roll around in the scene. At the end the hiding screen will vanish and you will need to press a key on the keyboard to indicate as quickly and accurate as possible if the ball was absent behind the screen. Please follow the instructions on the computer during the experiment. This experiment will last about 30 minutes.

Will my participation be Anonymous and Confidential?

Only the researcher(s) and supervisor(s) will have access to the data which will be kept strictly confidential. Your permission will be sought in the Participant Consent form for the data you provide, which will be anonymised, to be used for future scholarly purposes.

Storage and Destruction of Data Collected

The data we collect will be accessible by the researcher(s) and supervisor(s) involved in this study only. Your data will be stored for a period of 5 years since time of publication of the results before destroyed in an anonymised format in a hard drive which will be locked in a storage cupboard.

What will happen to the results of the research study?

The results will be finalised by 2014 and written up as part of my PhD Thesis and they may be published as part of journal articles or conferences presentations.

Reward

£3 for your participation

Are there any potential risks to taking part?

No. If you don't feel comfortable about any part of the experiment, you can withdraw at any time.

Questions

You can ask any questions in relation to this project before completing the Consent Form.

Consent and Approval

This research proposal has been scrutinised and been granted Ethical Approval through the University ethical approval process.

What should I do if I have concerns about this study?

A full outline of the procedures governed by the University Teaching and Research Ethical Committee is available at <http://www.st-andrews.ac.uk/utrec/Guidelines/complaints/>

Contact Details

Researcher: Ruoting Tao

Contact Details: Email: rt33@st-andrews.ac.uk

Supervisor: Dr. Juan-Carlos Gomez

Contact Details: Email: jg5@st-andrews.ac.uk Tel: 01334 46 2059

Example of the consent form

Project Title

Visual detection tasks: how fast can you find the object?

Researcher(s) Name(s)

Ruoting Tao

Email: rt33@st-andrews.ac.uk

Supervisors Names

Dr. Juan-Carlos Gomez

Email: jg5@st-andrews.ac.uk

Tel: 01334 46 2059

The University of St Andrews attaches high priority to the ethical conduct of research. We therefore ask you to consider the following points before signing this form. Your signature confirms that you are happy to participate in the study.

What is Anonymous Data?

The term 'Anonymous Data' refers to data collected by a researcher that has no identifier markers so that even the researcher cannot identify any participant. Consent is still required by the researcher, however no link between the participant's signed consent and the data collected can be made.

Consent

The purpose of this form is to ensure that you are willing to take part in this study and to let you understand what it entails. Signing this form does not commit you to anything you do not wish to do.

Material gathered during this research will be anonymous, so it is impossible to trace back to you. It will be securely stored in a hard drive for 5 years since the time of publication before destroyed. Please answer each statement concerning the collection and use of the research data.

I have read and understood the information sheet. Yes No

I have been given the opportunity to ask questions about the study. Yes No

I have had my questions answered satisfactorily. Yes No

I understand that I can withdraw from the study without having to Yes No

give an explanation.

I understand that my data once processed will be anonymous and that only the researcher(s) (and supervisors) will have access to the raw data which will be kept confidentially. Yes No

I understand that my data will be stored for a period of 5 years since the time of publication being destroyed Yes No

I have been made fully aware of the potential risks associated with this research and am satisfied with the information provided. Yes No

I agree to take part in the study Yes No

Participation in this research is completely voluntary and your consent is required before you can participate in this research.

Name in Block Capitals

Signature

Date

Example of the debriefing form

Project Title

Visual detection tasks: how fast can you find the object?

Nature of Project

This postgraduate research project was conducted to investigate whether human will automatically encode intentionality of an agent to a potential object. The agent in the video does not affect the presence/absence of the ball, however, some evidence suggested that his presence and his belief about the ball will affect your reaction time.

In separate studies, we are trying to examine this phenomenon in non-human primates. The comparison will help us to better understand the evolution of our cognitive abilities in understanding the others.

Storage of Data

As outlined in the Participant Information Sheet your data will now be retained for a period of 5 years since time of publication of the results before being destroyed. Your data will remain accessible to only the researchers and supervisors if you have given permission on the Consent Form.

What should I do if I have concerns about this study?

A full outline of the procedures governed by the University Teaching and Research Ethical Committee are outline on their website <http://www.st-andrews.ac.uk/utrec/Guidelines/complaints/>

Contact Details

Researcher: Ruoting Tao

Contact Details: Email: rt33@st-andrews.ac.uk

Supervisor: Dr. Juan-Carlos Gomez

Contact Details: Email: jg5@st-andrews.ac.uk Tel: 01334 46 2059

Appendix V. The comparison of the experiments in chapter 5

Comparison of experiments three, four and five

The videos of the last three experiments had the same procedure but different scenarios. To examine if the three scenarios had any effect on reaction time, we conducted a two-ways mixed ANOVA using the four conditions as with-in subject factor and the three experiments as between-subject factor. Mauchly's test showed the assumption of Sphericity was not meet ($\chi^2(5) = 25.47, p < .001$), so we reported the ANOVA with Greenhouse-Geisser adjustment. The mixed ANOVA revealed a significant main effect of conditions, ($F(3,219) = 28.10, p < .001$). But the interaction between experiment and conditions was not significant ($F(6,219) = 0.15, p = .99$), indicating that the scenarios did not affect the pattern of reaction time across the conditions.

Therefore we combined the data in the three experiments (see Figure 4 - 15 and Table 4 - 16) and analysed the effect of conditions using a one-way repeated ANOVA. Mauchly's test suggested the assumption of sphericity was not met ($\chi^2(5) = 26.20, p < .001$); therefore the following analysis was adjusted using Greenhouse-Geisser adjustment. The reaction time in the four conditions were significantly different ($F(3,225) = 28.95, p < .001$).

Paired T-test with Bonferroni adjustment showed significant longer reaction time in P-A- than the other three conditions, P+A- ($t(75) = 7.94, p < .001$), P-A+ ($t(75) = 6.48, p < .001$) and P+A+ ($t(75) = 4.01, p = .001$). And the participants responded more slowly in P+A+ condition than P+A- ($t(75) = 4.57, p < .001$) and P-A+ ($t(75) = 3.24, p = .011$). No difference was found between P+A- and P-A+ ($t(75) = 0.88, p = 1.00$). Paired T-test without adjustment showed the same results (see Table 4 - 17).

Table 4 - 16 Participants' average reaction time (ms) to the four conditions in experiment three, four and five combined

	P-A-	P+A-	P-A+	P+A+
Mean	444.0	380.1	385.6	407.0
SD	97.5	85.6	72.4	85.1

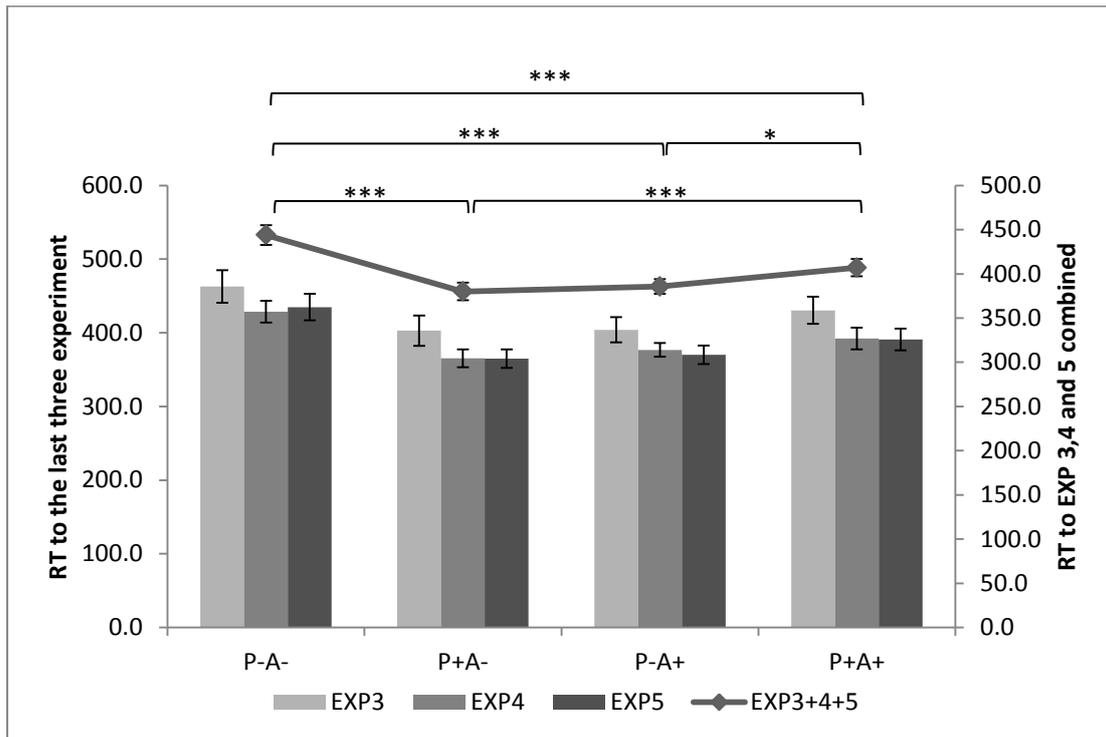


Figure 4 - 15 Participants' average reaction time (ms) to the four conditions in experiment three, four and five combined (error bar: SE)

Table 4 - 17 Results of the paired T-tests for experiment three, four and five combined

Pair of conditions	t(75)	P (2-tailed)	
		Bonferroni adjustment	No adjustment
P-A- vs P+A-	7.94	.000***	.000***
P-A- vs P-A+	6.48	.000***	.000***
P-A- vs P+A+	4.01	.001***	.000***
P+A- vs P+A+	-4.57	.000***	.000***
P-A+ vs P+A+	-3.24	.011*	.002*
P+A- vs P-A+	-.88	1.00	.38

Comparison of the two different procedures

We conducted an additional test to compare whether the two procedures had different effect on the participant's pattern of reaction time to the four conditions. We grouped the data from experiment 1 (the presence group) and 2 together, and the data from the last three experiments together, creating two groups that experienced the two different experimental methods (see Figure 4 - 16 and Table 4 - 18). A two-way mixed ANOVA using the four conditions as the with-in subject factor, and Procedure as the between-subject factor revealed a significant main effect of conditions ($F(3,37) = 24.89, p < .001$), and a significant interaction ($F(3,37) = 5.01, p = .003$). The p-values reported here were adjusted by Greenhouse-Geisser approach because Mauchly's test showed the assumption of Sphericity was not met ($\chi^2(5) = 24.35, p < .001$). The results suggested that the different procedures indeed affected the pattern of reaction time in the four conditions.

For detailed pattern of reaction time for the two procedures, we have reported the results in the last section and in experiment 2. To have a quick recapture, when the procedure of video in each condition were neat (in experiment 1 and 2, the reaction time in P-A- were significantly longer than that in P+A- and P+A+, but not longer than P-A+; when the procedure followed Kovács et al's original design, the reaction time was longer in the P-A- than the other three conditions and the reaction time in P+A+ was longer than the P+A- and P-A+ condition (Figure 4 – 15).

Table 4 - 18 Participants' average reaction time (ms) to the four conditions in experiment for the two experimental procedures

Experiment		P-A-	P+A-	P-A+	P+A+
1,2	Mean	374.0	345.3	353.6	349.1
	SD	81.2	77.0	68.4	97.8
3,4,5	Mean	444.0	380.1	385.6	407.0
	SD	97.5	85.6	72.4	85.1

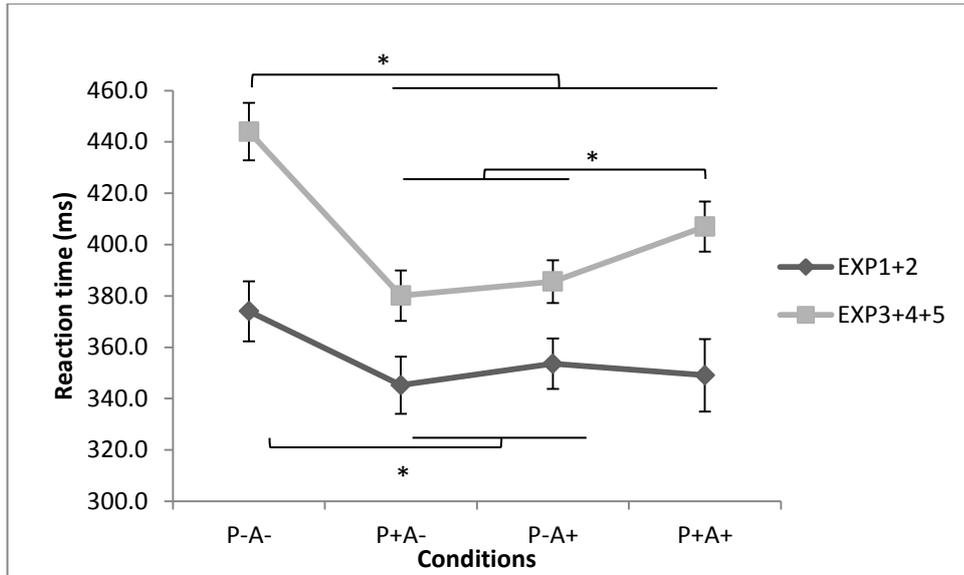


Figure 4 - 16 Participants' average reaction time (ms) to the four conditions for the two experimental procedures (error bar: SE)