# THE FUNCTIONAL AND NEURAL BASIS

# OF BELIEF AND DESIRE REASONING

Sarah Houthuys

A thesis submitted to

The University of Birmingham for the degree

DOCTOR OF PHILOSOPHY

School of Psychology

The University of Birmingham

AUGUST 2010

# ABSTRACT

Beliefs and desires are two core mental states that enable us to understand human behaviour. Research in developmental psychology reveals that children understand desires earlier than beliefs, suggesting that the cognitive processes underlying desire reasoning are less complex or less effortful than the processes underlying belief reasoning. Recent findings from studies with adults have identified a specific neural network associated with belief reasoning and found that belief reasoning is cognitively effortful even for healthy adults. However, there has been very little research into the cognitive and neural processes involved in desire reasoning in adults and the extent to which these processes are distinct to those involved in belief reasoning. In Chapter 2, I directly compare the performance of adults with acquired brain damage in belief and desire reasoning tasks, and I suggest the existence of two separate processes in desire reasoning: one linked to self-perspective inhibition and one linked to the reasoning about avoidance desires. In Chapters 3 and 4 I further assess these processes in healthy adults and show that both lead to a measureable cost in the participants' performance. This provides the first report that desire reasoning (and not only belief reasoning) is cognitively effortful for healthy adults. In Chapter 5, I document a double dissociation in adults with acquired brain damage showing that the processes mediating self-perspective inhibition and those mediating the reasoning about avoidance desires are executive in nature but functionally and neurally distinct. In Chapter 6, I discuss the implications of the findings for the understanding of how beliefs and desires are processed in the adult mind and brain.

## Acknowledgements

I wish to thank both my supervisors Doctor Dana Samson and Professor Glyn W. Humphreys for this amazing experience and opportunity of working in such a dynamic research team.

Further Note to Dana --- j'aimerais profiter de cette occasion pour te remercier de tout ce que tu as fait. Ce fut un honneur d'avoir l'opportunité de travailler sous ta supervision. Il est difficile pour moi d'exprimer ma gratitude par des mots… tout simplement, merci pour tout ce que tu m'as appris, pour les « food for thought/it's all about the learning process » moments, les accusations des bics disparus?, ton support dans le cadre académique et durant mes moments « catastrophiques ».

To all the patients I have had the pleasure to meet - AS, DS, DB, FK, FL, GA, GK, JeB, MH, MP, PH, PW, RH, SP, TM, TT and WBA - without you this would have never been possible - thank you for everything.

To my family and friends, thank you for always believing... you have all contributed towards this adventure, each in your own unique and special way - with memories that I will always cherish...

... Merci du fond du cœur...

**~ IN LOVING MEMORY OF BON PAPA ~**

# Table of Contents

## List of Figures

# List of Tables

# CHAPTER 1

## Introduction

## 1.1. Theory of Mind - belief and desire reasoning

In the movie 'Castaway' by Robert Zemeckis (2000), actor Tom Hanks portrayed a character abandoned on a deserted island, after having survived a horrendous plane crash. The movie carries the viewers through his struggle to survive, and to one day return to civilization. Through his adventures on the island, Tom's character depicts typical characteristics that fundamentally can be linked to humans and our ability to attribute mental states to others, as well as to non living objects. In a hand print made by his own blood, he portrayed a face on a volleyball, which in turn became his notorious friend 'Wilson'.

Over the past 30 years, researchers have been investigating this ability to attribute mental states (e.g., emotions, knowledge, desires and beliefs) to other agents, including ourselves, in order to predict, understand, and explain behaviour (Premack & Woodruff, 1978), an ability often referred to as having a Theory of Mind (abbrev.ToM). Beliefs and desires have been seen as the two core mental states we employ when reasoning about other people's actions (Ziv & Frye, 2003). According to Bartsch and Wellman (1989), both mental states recruit diverse networks of concepts (see Figure 1). Desires stem from basic emotions (e.g., to love), and physiological states (e.g., to hunger), whereas beliefs are built on perceptual experiences (e.g., to see)[1]. It is the combination of both that lead people to act the way they do. For example, in order to explain why my sister purchased gummies at a white night shop, I would need to refer to her desire (e.g., she wants gummies) and to her belief (e.g., she thought that the gummies she likes could be found at that store). The desire was linked to a physiological state, she was craving something with sugar; on the other hand, the belief

[1] It should be noted that beliefs and desires can be transitory mental states and enduring mental states (for review: Overwalle, 2008). Transitory mental states refer to beliefs and desires about specific and changing situations (e.g., the belief about the location of an object; the desire of getting a specific object). Enduring mental states refer to beliefs and desires that are more constant over time (e.g., political or religious beliefs, career aspirations). In this thesis, I am mainly interested in the inference of transitory mental states.

was based on simple perception, she had seen gummies she likes, last time we passed the shop. Her action ended in purchasing the gummies and her reaction, at the fulfilment of her desire, was pure joy as she emptied the pack. Focusing only on the belief that she thought that these particular gummies could be bought in the night shop would not be sufficient to understand why she bought the gummies. Also, knowing that my sister wanted gummies is not sufficient to explain why she specifically went to the white night shop. Similarly, perceptions (my sister seeing the night shop), emotions and physiological states (her craving for sugar) cannot explain or help predict an individual's action when considered in isolation. Bartsch and Wellman's model outlines how both mental states contribute to action prediction and how each originates from different experiences.

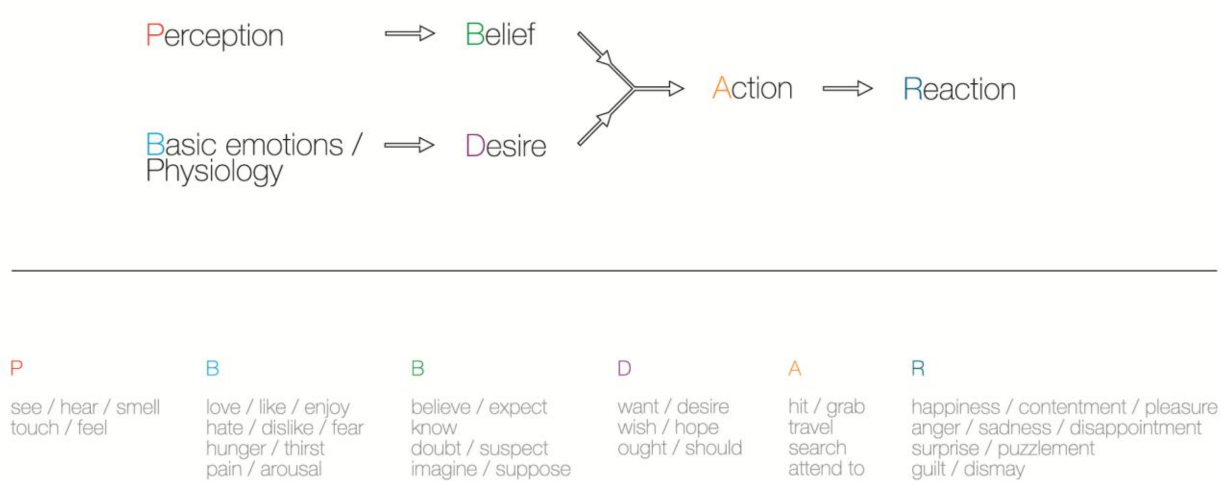



Figure 1. Illustration of Bartsch and Wellman's belief-desire reasoning scheme (1989, p. 947).

## 1.2. Evidence from developmental psychology

As children, we are introduced to a world of imagination and fairy tales. Children's stories often contain information about people's beliefs and desires. For example, the story of Hansel and Gretel (by the Grimm Brothers) explains a tale of a brother and sister who are abandoned in a forest by their father. Their desire was clearly to be able to return home, which was expressed by their behaviour of always leaving a trail behind them. The first time they employed white pebbles and managed to find their way back. However, the second time round, they used bread crumbs, which were unfortunately eaten by birds whilst they were sleeping. If I were to stop the story at this particular point, and to ask children of the age of three about Hansel and Gretel's desires and beliefs, their response would differ to those of children of the age of four for beliefs but not for desires. In the case of desires, if I asked both age groups: "*what did Hansel and Gretel want?*", children would retort, "*well, they wanted to go home*". Conversely, if I asked a belief question: "*where do Hansel and Gretel think the breadcrumbs are?*", children at the age of three would probably respond incorrectly stating that Hansel and Gretel think the birds ate the breadcrumbs. In contrast, children of the age of four would typically respond correctly that Hansel and Gretel think that the breadcrumbs were where they had last left them, as both were still fast asleep and unaware that the crumbs were eaten by birds. Thus, children who have not developed an adult-like ToM would not be able to follow certain subtleties and in many cases misunderstand the plot of the story.

### 1.2.1. Development of desire and belief reasoning

The developmental literature has shown a great interest in this transition children undergo between the ages of three and four. Tasks used to test children's belief reasoning capacities have predominately been based on the change-of-location false-belief paradigm created by Wimmer and

Perner (1983). Typically in such a false-belief task the protagonist is initially aware of an object's current location. The object is then displaced during the protagonist's absence, which leads to a discrepancy between the participant's own knowledge about the current location of the object (reality/self-perspective) and the protagonist's belief about the location of the object (false-belief/other-perspective). Participants are then asked where the protagonist thinks the object is located or where s/he will be looking first to get the object. In order to correctly answer the question, participants must disregard their own knowledge about the location of the object and take into account what the protagonist has and has not seen about the object. On the other hand, tasks that have been used to assess children's ability to reason about desires required children to predict whether a protagonist would approach or avoid a specific object on the basis of prior information about the protagonist's tastes or preferences (e.g., Cassidy, 1998; Repacholi & Gopnik, 1997; Wellman & Wooley, 1990). Using those belief and desire tasks, it is usually observed that children pass the desire reasoning tasks at an earlier age than they pass belief reasoning tasks. In has been debated whether this may be due to desires concepts being acquired earlier than belief concepts, or to the reasoning about beliefs being more dependent on additional cognitive resources, with these additional cognitive resources being acquired later than those required for desire reasoning. These contrasting viewpoints are not necessarily mutually exclusive and fall in three theoretical frameworks that seek to explain the transitional stage between the ages of three and four.

According to the first framework, experience plays a major formative role in the development of a children's TOM. On this view, mental state reasoning relies on domain general mechanisms of theory formation (Gopnik, 1996; Gopnik & Astington, 1988; Perner, 1991; Wellman & Cross, 2001; Wellman, Cross, & Watson, 2001). Authors who adhere to this theoretical position argued that children acquire a qualitatively new skill by the age of four, e.g., the capacity to deal with metarepresentations. Representing someone else's belief, unlike representing someone else's desire, requires metarepresentation skills that children have not yet developed at the age of three (see Figure 2). According to this theoretical framework, children at the age of two predict and explain actions by attributing desires. In this simple desire reasoning stage, they are able to talk about desires in terms of "wants" and "likes", which could be associated with objects and actions, for current or future states of affairs (e.g., *"during the next holidays, I want to go to Disney land"*), and for their own or others' desires (e.g., "*Mommy likes broccoli, but I don't like it*"). Although children of that age can talk about the real state of affairs, e.g., about reality, they do not at this stage possess the capacity to process or analyze reality as a mental representation, e.g., as someone's belief.

An important change occurs at the age of three when children start to use a desire-belief theory. This is in an intermediate stage of reasoning between the nonrepresentational desire theory and a full-blown belief-desire theory (used from the age of four). During the intermediate stage, children talk more about thoughts and beliefs, demonstrating awareness of these mental states. However, they have difficulties dealing with and generally avoid explaining their own and other's actions on the basis of their beliefs. Specifically, these children predominantly use desire concepts when explaining and predicting actions. A study by Cassidy (1998) investigated the ability of three to four year olds to reason about beliefs and desires. The task used the change-of-location false-belief paradigm but, in addition manipulated the desirability of the object displaced (the object was a desirable or a non desirable food). Prior to testing, the child was told of a puppet's food preference and that whichever location the puppet selected, the puppet would subsequently have to eat the content. Cassidy (1998) showed that children who failed on the false-belief question were distracted by the puppet's food preference. In other words, these children reasoned that the puppet would look at a location based solely upon the desirability of the food actually hidden at that location and not based on the puppet's belief about the food location. Thus, children of the age of three simply responded by choosing the locations that contained the desirable object and avoided selecting the location that contained the undesirable one.

The final 'belief desire psychology' stage in the child's development is achieved by the age of four. It is at this stage that children are able to understand that beliefs represent reality and that these may be false. These children start to pass false-belief tasks and start to be able to juggle with both desire and belief concepts to explain human action.

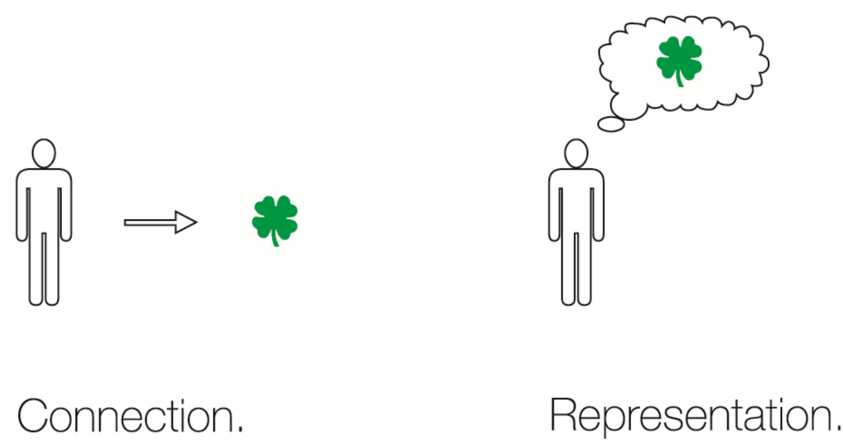



Figure 2. Illustration (taken from Saxe et al., 2004, p. 90) of the two ways in which an agent can be related to an object. The left represents the simple non-representation association between an agent and an object, which can be used to reason about simple desires. The right represents the representation association between the agent and the object, which is necessary for more complex desires and for beliefs.

According to the second theoretical framework (Baron-Cohen, 1995; Fodor, 1992; Leslie, Friedman & German, 2004; Leslie, German, & Pollizi, 2005), mental state reasoning relies on the maturation of domain-specific reasoning processes which are part of an innate cognitive module. A specific formulation of this type of framework is that of Leslie and colleagues (Friedman & Leslie, 2004a, 2004b, 2005; Leslie et al., 2004; Leslie et al., 2005; Leslie & Pollizi, 1998; Leslie & Thais, 1992), who aimed to provide an information processing model for false-belief reasoning. The model incorporates a 'theory of mind module' (TOMM) that matures around the age of two. The main function of this module is to orient children's attention to mental states so that they can then subsequently learn about them. The model also incorporated a 'selection processing' (SP). The SP component was defined as an effortful control process necessary to select among diverse representations proposed by the TOMM. The authors proposed that the TOMM component, which attributes beliefs and desires, is a domain-specific component but that the SP component consists of more domain-general inhibitory processes that would be used in a variety of other cognitive domains. Thus, their model posits an inter-relation between domain-specific and domain-general processes of learning and reasoning.

Leslie and Pollizi (1998) proposed a model to highlight the processing steps involved when people reason about beliefs and desires in a typical change-of-location false-belief task. The model integrates a first stage of default mental state content attribution (via the TOMM component), which makes a particular location salient and then a correction through inhibition (via the SP component), if the initial attribution is not appropriate in the particular context at hand. Figure 3 illustrates the processing steps in four situations, which represent all possible combinations of approach-/avoidance-desire and true-/false-belief conditions that can be found in a false-belief paradigm. The first cell (top left corner) depicts a true-belief and desirable-object condition (the agent wants the object and knows where the object is). Both the true-belief and approach-desire are attributed by default via the TOMM component. The location containing the object becomes thus the most salient information and no inhibition is required to provide the correct response. The next cell (top right corner) depicts a false-belief and desirable object condition (the agent wants the object but has a false-belief about the object's location). Again, the true-belief and approach-desire are attributed by default. The location containing the object becomes again the most salient but this time the selection process is required to inhibit the wrong true-belief attribution and direct attention towards the location where the object is not. The third cell (bottom left corner) shows a true-belief and undesirable object condition (the agent does not want the object and knows where the object is). Again, a true-belief and approach-desire are attributed by default, which makes the location containing the object the most salient. To correctly take into account that the person does not want the object, the location needs to be inhibited and the other location selected instead. Finally, the last cell (bottom right corner) requires what Leslie terms a 'double inhibition'. The experimental condition involves a false-belief and an undesirable object (the agent does not want the object and has a false-belief about the object's location). True-belief and approach-desires are attributed by default which makes the location where the object is the most salient. This time, to correctly take into account the false-belief and the avoidance-desire, two inhibition process would operate in parallel, one to select the location where the object is not because the belief is false and one to select the location where the object is not because the person does not want the object. The result of the double inhibition is that the response selected is the location where the object actually is. Although the response is the same as for a true-belief, desirable object condition, the processes required to get at the correct response are much more effortful so that even children who pass a typical false-belief task with desirable objects would still find it difficult to reason about false beliefs with undesirable objects. Within the context of this model, one reason why children find it easier to reason about desires than beliefs is that researchers have mainly tested children on approach desires which do not require any inhibitory processes. A better comparison to false-belief reasoning would be to reason about avoidance desires, both requiring domain-general inhibitory processes.

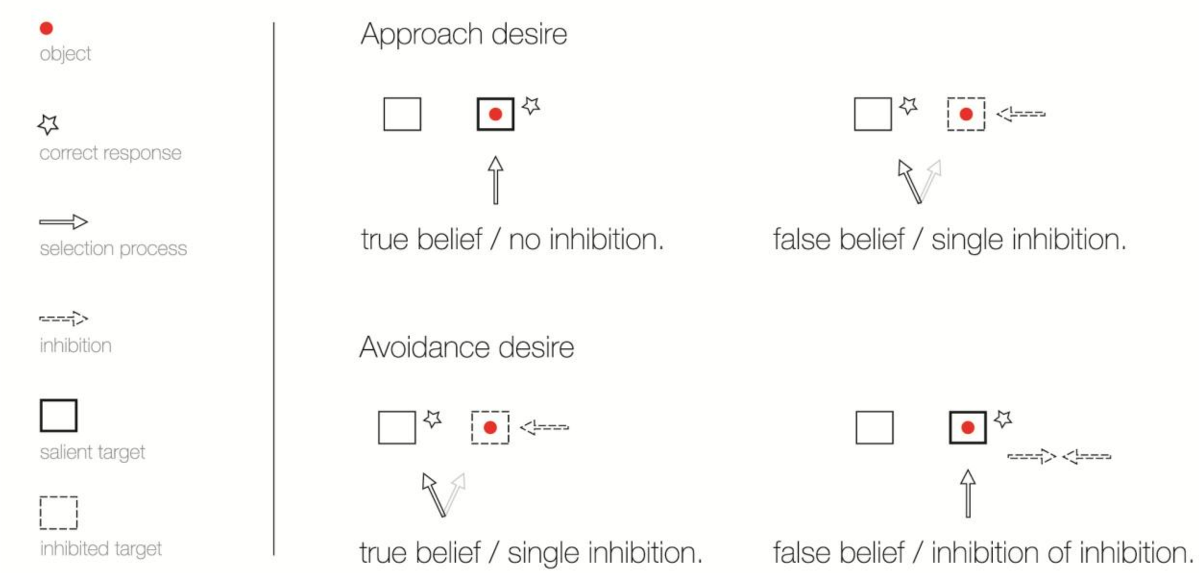



Figure 3. Leslie and colleagues' (Leslie & Pollizi, 1998) model, "Inhibition of Inhibition".

The third theoretical framework puts the idea forth that belief reasoning, especially false-belief reasoning, places higher inhibitory demands than most desire reasoning tasks because the conflict between self- and other-perspectives are higher in the false-belief task (Carlson & Moses, 2001; Carlson, Moses & Breton, 2002; Carlson, Moses & Claxton, 2004; Carlson, Moses & Hix, 1998; Russell, Jarrold & Potel, 1994). In a change-of-location paradigm, the displacement of an object in the protagonist's absence leads to a discrepancy of perspective (the child knows where the object is but the other person doesn't know). It may be this ability to inhibit one's own more salient mental state that shows slower development than the capacity to understand beliefs per se (Russell et al., 1994; Russell, Mauthner, Sharpe & Tidswell, 1991). Typical desire reasoning tasks do not require inhibition of one's own perspective and this may be why children find it easier to solve the tasks. A study by Moore and colleagues (Moore, Jarrold, Russell, Lumb, Sapp & MacCallum, 1995, Experiment 2) directly examined the need for self-perspective inhibition in a desire and a belief inference tasks that were matched for inhibition demands. In their desire task children played a game against a story character called Fat Cat. The game involved a jigsaw puzzle that each player needed to complete in order to win. To obtain a puzzle piece, each player had to take a card from the deck that depicted a colour of a particular box (blue or red). This box contained a designated puzzle piece a player might need in order to complete the set. There were two conditions in the game: children could find themselves in a direct-conflict condition with Fat Cat, where both players wanted a different coloured card, or in a no-conflict condition, where both players wanted the same coloured card. The belief task was based on a traditional change-of-location paradigm using true-belief trials as control (no-conflict/same-perspective). In both tasks children were asked a direct question to infer Fat Cat's perspective (want/think). The results in the desire task showed that in the conflict situation, three year old children falsely identified the colour that they themselves wanted as the card that Fat Cat would want - a similar profile to that found in the false-belief trials where children identified the current location of the object as where Fat Cat would think the object is located. Thus, children found both tasks equally difficult, which was taken as evidence in favour of the view that desire and belief reasoning tap onto general cognitive demands that required inhibition and that it was this capacity to inhibit that developed during the transitional stage when children start to reason about false-beliefs.

In sum, the developmental literature has attempted to explain the maturation of ToM with respect to belief and desire inferences. Studies showed that children usually find it easier to reason about desires than beliefs, with an important transitional shift that occurs between the ages of three and four when children start to pass false-belief tasks. Authors have debated what children acquire during that transitional stage, either new concepts (first framework presented above) or more efficient inhibitory processes required to handle non-salient information (second and third frameworks presented above). Whereas the inhibitory processes associated with false-belief reasoning are described quite similarly in the second and third framework, both frameworks highlight different reasons why inhibition may also be required in desire reasoning when appropriate tasks are chosen: reasoning about an avoidance desire require inhibition of the object (second framework) and reasoning about someone else's conflicting approach desire require inhibition one's own desire (third framework).

### 1.3. Evidence from functional imaging

More recently researchers started to look at theory of mind when it is fully developed in adults. Given that adults do not make errors anymore in typical false-belief tasks, researchers started to look at the brain activation while adults perform various tasks that require them to reason about mental states. In early attempts to identify regions involved with ToM in adults, no clear distinction was made between different types of mental states and researchers only contrasted a ToM versus a non-ToM condition.

Baron-Cohen and colleagues (1994) used single photon emission computerized tomography (SPECT) to examine participants' responses when asked to reason about a list of words referring to mental states (e.g., think) relative to control words (e.g., arm). They found that the right orbito-frontal cortex (OFC) was significantly active when participants were engaged in reading words with ToM content in comparison to control words. This study only focused on this particular region of interest, thus it remained unknown if other areas play a role in theory of mind. Fletcher and colleagues (1995), in a positron emission tomography (PET) imaging study, used story comprehension tasks requiring participants to reason about mental states (ToM condition) versus the physical world (physical control condition) or to simply read unrelated sentences (baseline). Participants were first prompted, which experimental condition they would be given, followed by the text and finally they were given the test question. Unlike Baron-Cohen and colleagues, these authors found no specific activity in the right OFC related to ToM. Instead, their results highlighted areas in the medial dorsal region of the left frontal cortex (Brodmann's Area 8) that were more activated when participants read the ToM stories relative to the physical control condition.

Gallagher and colleagues (2000) presented healthy adults in an fMRI study with the same story comprehension task as the one used by Fletcher et al. (1995), but also with a cartoon comprehension task. In both tasks, the material had some ToM content, no ToM content or consisted of unrelated items (e.g., unlinked sentences/jumbled pictures). For the cartoons, participants were asked to observe the image, and to consider their meaning internally. The ToM condition involved attribution of either a false belief or ignorance, whereas the non ToM content did not (see Figure 4). The presentation of the text and cartoons with ToM content was associated with activation in the medial prefrontal cortex (MPFC) and the temporo-parietal junction (TPJ) bilaterally. However, when directly comparing the ToM and non-ToM content (for both cartoons and stories) only the MPFC (paracingulate cortex) was significantly more activated for the items with ToM content, thereby replicating Fletcher and colleagues' findings.

Using a different type of stimuli in an fMRI study, Castelli and colleagues (2000) presented participants with animated clips representing moving simple geometric shapes (triangles; see Figure 5). There were three different types of animated clips: (1) those with ToM content where the shapes moved in a way that could be interpreted by using mental state terms (e.g., two triangles persuading one another), (2) those with simple goal directed movements (e.g., two triangles dancing with one another), and (3) those with random movements. The results showed four main areas predominately active in the ToM condition: the MPFC, the TPJ, the basal temporal region and the occipital cortex.

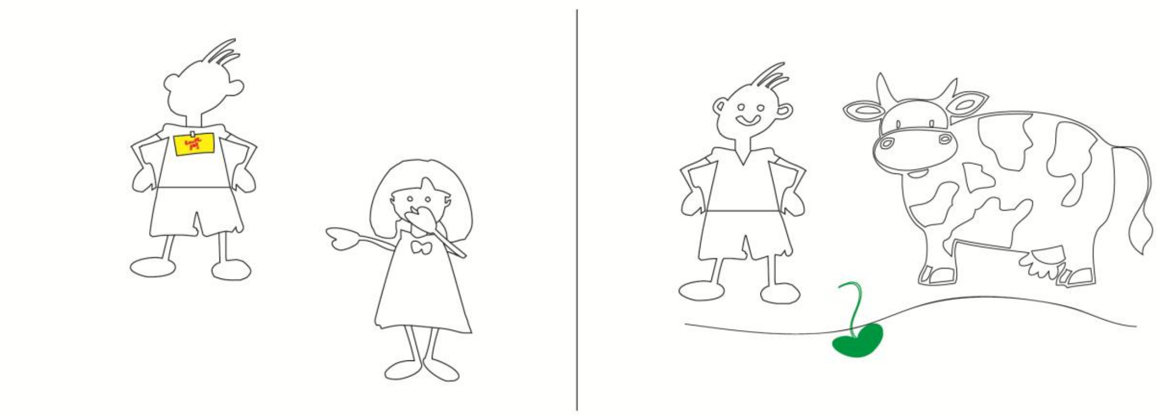


Figure 4. An example of the stimuli employed by Gallagher et al. (2000). On the left, a ToM condition; on the right, a control no-ToM illustration.


Figure 5. Illustration depicting the three different types of animated clips (Castelli et al., 2000).

The picture that emerged from these early attempts to find the neural basis of ToM is that MPFC and the TPJ seem to be the main brain areas that are activated when healthy adults reason about mental states, irrespective of the tasks or the type of material used. The MPFC, in particular, seemed to be more specifically associated with ToM. In the studies mentioned above, it is possible, however, that the extent to which several brain areas are associated with ToM has been underestimated. Baron-Cohen et al. (1994), only focused on one region of interest; in the case of Fletcher et al. (1995), or Gallagher et al. (2000), it cannot be excluded that the control items chosen also invited participants to mentalise. For example, a ToM story would be as followed: "*A burglar who has just robbed a shop is making his getaway. As he is running home, a policeman on his beat sees him drop his glove. He doesn't know the man is a burglar, he just wants to tell him he dropped his glove. But when the policeman shouts out to the burglar, "Hey, you! Stop!", the burglar turns round, see the policeman and gives himself up. He puts his hands up and admits that he did the break-in at the local shop."* The test question was then: "*why did the character do this*?" (Fletcher et al., 1995, p. 125). The matched physical control story was as follows: "*A burglar is about to break into a jewellers' shop. He skilfully picks the lock on the shop door. Carefully he crawls under the electronic detector beam. If he breaks this beam it will set off the alarm. Quietly he opens the door of the store room and sees the gems glittering. As he reaches out, however, he steps on something soft. He hears a screech and something small and furry runs out past him, towards the shop door. Immediately the alarm sounds".* Although their test questions did require to reason about the physical object in the story (e.g., "*why did the alarm go off*?"; Fletcher et al. 1995, p. 125), because a human person was still involved it is possible that even in the control condition, participants were mentalizing. The subtraction between the ToM and physical control condition may have thus masked the involvement of other regions in ToM. The same is true for the cartoons used by Gallagher et al. (2000), as they also involved animate characters (see Figure 4, right). In the case of Castelli et al. (2000), where animated clips involved triangles, even the goal-directed movement control condition could still invite mental state attribution, for example, "*the desire to dance*".

Several subsequent studies using functional imaging have used more fine-grained designs and tasks, more appropriate control condition, and have looked more specifically at belief reasoning and desire and goal attribution.

### 1.3.1. The neural basis of belief attribution

One of the first studies that aimed at finding the brain areas specifically associated with belief reasoning is the study by Saxe and Kanwisher (2003). In one of their experiments (Experiment 2), they compared the brain activation associated with stories in which a character has a false belief with four control condition: false-photograph, desire, physical-people and non-human description. One of the false-belief stories was as follows: "*A boy is making a paper mache project for is art class. He spends hours ripping newspaper into even strips. Then he goes out to buy flour. His mother comes home and throws all the newspaper strips away*" (Saxe & Kanwisher, 2003, p. 1841). The false-photograph condition was formally similar to the false-belief condition in that the photograph would represent an outdated reality but it did not include (explicitly at least) any human character, e.g., "*A photograph was taken of an apple hanging on a tree branch. The film took half an hour to develop. In the meantime, a strong wind blew the apple to the ground"* (Saxe & Kanwisher, 2003, p. 1841). The desire story required participants to ascribe a character's goal/intention, e.g., "*for Susie's birthday, her parents decided to have a picnic in the park. They wanted ponies and games on the lawn. If it rained, the children would have to play inside*" (Saxe & Kanwisher, 2003, p. 1841). The physical people story simply described the physical appearance of people, e.g., "*Emily was always the tallest kid in her class. In kindergarten she was already over 4 feet tall. Now that she is in college she is 6'4". She is a head taller than the others*" (Saxe & Kanwisher, 2003, p. 1841).

Finally, the non-human description simply focused on physical objects, e.g., "*Nine planets, and their moons, plus various lumps of debris called asteroids and comets, make up the sun's solar system. The earth is one of four rocky planets in the inner solar system*." (Saxe & Kanwisher, 2003, p. 1841). The authors found that not only the MPFC but also the left and right TPJ, the posterior cingulate and the anterior temporal sulcus were significantly more activated when participants reasoned about a false belief than a false photograph. A region of interest analysis focusing on the TPJ showed that it was also more significantly activated (especially on the left) in the desire condition than the false photograph, the human and non-human description conditions, suggesting that the TPJ is also involved in desire reasoning and not only in belief reasoning.

In a further fMRI study, Saxe and Powell (2006) contrasted belief stories with stories that described a physical appearance of a protagonist (condition with no internal state), and/or a bodily sensation they were experiencing (condition with an internal state but no mental state). One of the belief stories was as follows: "*Nicky knew that his sister's flight from San Francisco was delayed ten hours. Only one flight was delayed so much that night, so when he got to the airport, he knew that flight was hers*" (Saxe & Powell, 2006, p. 693). One of the appearance stories was for example: "*Joe was a heavy-set man, with a gut that fell over his belt. He was balding and combed his blonde hair over the top of his head. His face was pleasant, with large brown eyes*" (Saxe & Powell, 2006, p. 693). Finally, an example used for the bodily sensation condition: "*Sheila skipped breakfast because she was late for the train to her mother's. By the time she got off the train, she was starving. Her stomach was rumbling, and she could smell food everywhere"* (Saxe & Powell, 2006, p. 693). The results showed that the TPJ (bilaterally) and the posterior cingulate responded selectively to beliefs but not to other internal experiences or physical appearance. The right supramarginal gyrus (SMG) was predominately active when participants reasoned about bodily sensations of the protagonist e.g., hunger. The MPFC showed activity across all conditions. The authors concluded that the MPFC is recruited when participants represent any form of information about another individual.

Perner and Aichhorn (2006) presented healthy adults in a fMRI study with the same false-belief and false-photo stories as those used by Saxe and Kanwisher (2003). In addition, the authors introduced a false-direction sign (which, like false belief, misrepresents reality but which does not involve mental states) and a baseline condition describing a temporal change (with no misrepresentation of reality involved). For the false-direction sign, stories were presented such as: "*The sign to the monastery points to the path through the woods. While playing the children make the sign point to the golf course*". The test question was: "*According to the sign the monastery is now in the direction of the … golf course/woods?"* (Perner & Aichhorn, 2006, p. 248). For the temporal change condition that was used as a control, participants were presented with stories such as: "*The monastery is surrounded by a pond. The pond is being drained and turned into a golf course*". The test question stated: "*Earlier on the monastery was surrounded by a …pond/golf course?*" (Perner & Aichhorn, 2006, p. 248). The authors reported differences between the left and right TPJ, with the left TPJ area being activated not only in false-belief condition but also the false-photograph and the false-sign conditions. In contrast, the right TPJ was mainly recruited when participants reasoned about belief-specific content only. The authors also found activations in the middle temporal gyrus, precuneus and MPFC (similar regions of activation as those reported by Saxe and Kanwisher, 2003), which were, however, not specific to the false-belief condition.

The studies described here which looked more specifically at belief reasoning confirmed the involvement of a ToM neural network, including among other regions the MPFC and TPJ, but point to the TPJ (especially right TPJ) as a region specifically involved with false-belief processing.

### 1.3.2. The neural basis of desire attribution

Not many researchers have investigated desire attribution in adults and few studies, aside from the study described earlier from Saxe and Kanwisher (2003; who found TPJ activation for desire reasoning), have included such conditions in their task. Research that has incorporated desires at an implicit level, engage participants in games that directly measured their online performance when playing against a human or computer player/opponent. In such games, we can expect that participants will process their opponent's desires to predict the opponent's next response as well as their own desires to decide how to respond themselves. McCabe and colleagues (2001) examined adults' performance in an fMRI study when playing a "trust and reciprocity" game with a human and/or a computer. In Figure 6, a diagram illustrates the logic and rules of the game. The players are represented as X1 and X2. X1 in this case begins at the top of the pyramid and has two possible options: (1) either X1 moves to the left and both players leave with 45 dollars each, or (2) he moves to the right and gives the next move to X2. In this case, X2 has two possible options: (1) to choose the left side option of 180/225, where both parties benefit, or (2) to choose the right side option where he takes the money for him/herself. These researchers split their participants into two profiles: cooperative verses non-cooperative, based on a behavioural test done prior to testing. For participants with cooperative behaviour, regions within the prefrontal cortex (anterior paracingulate cortex) were more active when playing with a human in comparison to a computer. With participants who demonstrated non-cooperative behaviour, no difference between the human and computer condition could be observed.

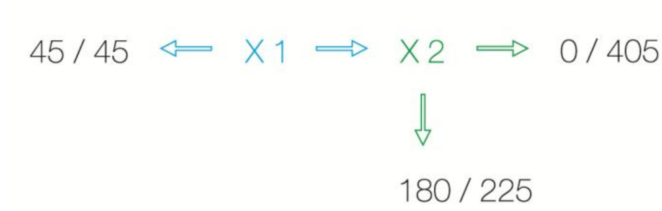



Figure 6. Visual demonstration of the "trust and reciprocity" game (taken from McCabe et al., 2001, p. 11832)

Similarly, Gallagher and colleagues (2003) employed PET to investigate the online processes when participants were engaged in paper-scissor-stones game, played against a human or a computer. There were three conditions, which were identical in structure but differed in instructions prior testing. In one condition, participants were told they would be playing against a human opponent; in the other two conditions that they were told that they would be playing against a computer. In the first condition, participants were asked to outsmart their human opponent. In one of the computer conditions (rule solving condition) participants were told that the computer generated responses, based on previous responses made by the participant. This was outlined as three basic rules: (1) the selection made by the participant, (2) the option that beat the participant response or (3) the option that lost to the participant's selection. The pattern of response generated by the computer varied and participants were instructed to adapt to the rule changes. In the second computer condition (random selection), participants were asked to respond at random and to never select the option the computer previously played. The authors showed activations bilaterally in the anterior paracingulate cortex when participants believed they were playing against another human opponent.

Interestingly in both studies (Gallagher et al., 2003; McCabe et al., 2001), the difference between control and test content relates only with respect to the instructions given to the participants (playing against a real opponent who has mental states or playing against a computer that has no mental states). This manipulation was sufficient to modulate activation in the anterior paracingulate cortex, an area close to the MPFC activation associated with belief reasoning (e.g., Fletcher et al., 1995; Gallagher et al., 2000). For example, Frith and Frith (2003) propose that this region may be involved in dealing with two simultaneous representations, i.e., someone else's belief and reality for false beliefs or competing desires in a game.

In sum, neuroimaging studies show a relatively consistent brain network activated when healthy adults reason about mental states. Researchers have mainly focused on the MPFC and the TPJ, although other regions seem to be also involved. Studies that have looked specifically at false-belief reasoning have highlighted a specific role of the TPJ (especially right TPJ). The specific network associated with desire reasoning has not been investigated in a systematic way but the evidence available so far suggest that it recruits part of the network also involved in false-belief reasoning (e.g., TPJ, see Saxe & Kanwisher, 2003; anterior cingulate as shown in studies that used interactive games, see McCabe et al., 2001; Gallagher et al., 2003).

## 1.4. Evidence from neuropsychology

The neuropsychological approach offers two types of valuable information to understand the neural and cognitive basis of ToM. Firstly, by determining which brain lesions impair a patient's ToM ability, it can show which brain areas are necessary for ToM. Secondly, by analyzing the pattern of association and dissociation of deficits across different ToM tasks, the neuropsychological approach can highlight ToM components that are independent at the cognitive and neural level. The success of the neuropsychological approach depends, however, on the use of appropriate measures which control for other cognitive impairments that the patients may have. Indeed, patients often have general cognitive deficits (e.g., language or working memory impairments) that can hinder their understanding of the task but that do not necessarily prevent from reasoning about mental states per se. To address this issue, researchers have introduced control items in the task and have simplified the material by creating nonverbal tasks or by providing pictorial supports.

### 1.4.1. Brain regions necessary for ToM

Several studies have investigated whether the frontal lobes are necessary for ToM. For example, Stone and colleagues (1998) examined the contribution of the frontal lobes on a series of tasks presented to patients with bilateral damage to the orbito-frontal cortex and unilateral damage to the left dorsolateral prefrontal cortex (with patients with anterior temporal lesions acting as control patients). Three different tasks were administered to these patients: (1) first-order false-belief task, (2) second-order false-belief task, and (3) a social faux-pas task. In the first-order false-belief task, patients were presented with a traditional change-of-location paradigm that measured the ability to infer the belief of a story character. There were four test questions in this task: (1) a belief question "*where does the character think the object is?"*; (2) a reality question: *"where is the object?"*; (3) a memory question: "*where was the object at the beginning?"*; and (4) a physical (inference) question: *"where would there be 'the object in question'?"* (Stone et al., 1998, p. 652). The second-order false-belief task required an extra processing step and the capacity of representing two individual's mental states. Here, participants were presented with a similar change-of-location paradigm, but this time a first protagonist who changed the location of the object was unaware that a second protagonist saw the displacement. Test questions probed the false belief of the first protagonist about the second protagonist's belief (second-order false belief), the true belief of the second protagonist about the object location (first-order true belief), and the participant's knowledge about the current location of the object (reality). In addition, the authors manipulated the memory demands by presenting the tasks without pictorial support (high memory load condition) or with pictorial support which helped to recall key aspects of the story at test (low memory load condition). In the social faux-pas comprehension task, patients were presented with stories that contained information of someone committing/saying something in a maladroit fashion. In other words, an individual would be unaware of certain information and as a result expressed a thought that would offend another individual. In this case, two individuals' mental states needed to be represented in order to understand the faux pas that occurred (the ignorance of one of the characters and the emotion in the second character produced by what the first character said). There were five test questions probing (1) the faux pas detection: "*did someone say something they shouldn't have?"*; (2) the faux pas understanding: *"who said something they shouldn't have said?"*; (3) the understanding of the listener's mental state: *why shouldn't they have said it?"*; (4) the understanding of the speaker's mental state: *why did they say it?"*; and (5) elements of the story, i.e., reality question (Stone et al. 1998, p. 654). Unlike the first- and second-order belief task, no specific measure to control the memory demands was used in the faux-pas task.

Overall, the researchers showed that patients with anterior temporal lesions showed intact ToM reasoning abilities across all three tasks. Patients with bilateral orbito-frontal damage showed patterns of impairment in the faux pas task only, with intact first- and second-order belief reasoning. Furthermore, the authors highlighted that patients with dorsolateral frontal lesions performed well in the faux pas stories but made errors on the first- and second-order false-belief task mainly in the high memory load condition. In that high memory load condition, errors were not only made on the false belief questions but also on the control questions. The authors concluded hereforth that the dorsolateral frontal areas were not essential for ToM per se. This evidence supports early neuroimaging findings which highlighted the role of the OFC in mentalizing (Baron-Cohen et al., 1994).

Gregory and colleagues (2002) adopted a similar type of methodology and examined ToM performance in frontotemporal dementia. Participants were presented with a first- and second-order false-belief task, a faux pas test and a reading the mind in the eyes test (for review, Baron-Cohen, 1997; Stone et al., 1998). The reading the mind in the eyes test presented eyes in isolation that expressed affective and cognitive mental states. Participants were asked to select one out of two suggested possibilities that best described the mental state expressed (e.g., "*ignoring you versus noticing you*"; Gregory et al., 2002, p. 756). Overall, the results showed that the patients were impaired in all tasks but they were able to answer the control trials. Proportionally, more patients were impaired in the faux-pas task compared to the first- and second-order false-belief task. Thus, areas affected in frontotemporal dementia seem to be necessary for ToM.

Happé and colleagues (2001) reported a single case study of a man treated for stereotactic anterior capsulotomy with lesions found in the medial frontal cortex. This patient was presented with three types of tasks: (1) a story comprehension task, (2) a single cartoon task and (3) a paired cartoon task. The story comprehension task included stories that incorporated double bluff, mistakes, persuasion and white lies in a variety of different day-to-day contexts (e.g., x-raying an individual after an accident). Test questions probed the protagonist's thoughts and feelings to the given situation. Control questions involved a particular detail of the story that did not involve any mental state. The single cartoon task took sketches from newspaper magazines that either depicted something funny with a ToM content or violations of the social norms. Test questions here probed: *"why is this particular cartoon funny?"* (Happé et al., 2001, p. 86). For the paired cartoon task, participants were presented with nonverbal cartoons that depicted either something funny or thee mirror image that replaced the comical point of that particular story with a neutral attribute (control). Here, the ToM cartoons were contrasted with cartoons that required inferences of a physical property. Pictures were presented in sets and participants were asked to discriminate between both, by determining which of the two was funny. The results showed that the patient was impaired in all tasks that required a mental state inference for both stories and cartoons. It should be noted that this patient had been reported to have deficits with inhibition and working memory. The authors argued, however, that this patient's ToM impairment could not be explained by these general deficits as the patient showed intact reasoning abilities on the control trials. It is still possible, however, that the ToM trials placed higher or different effortful demands than the control trials.

Another study directly tested whether the medial prefrontal cortex usually associated with ToM in neuroimaging studies was indeed necessary to perform ToM tasks. Bird and colleagues (2004) report a single case study of a patient GT, who following a stroke had bilateral lesions to the medial frontal lobe areas of her brain. The authors used five different tasks, each tapping onto ToM processes and each having been shown to involve the MPFC when healthy adults mentalize. The first task involved picture sequences and was taken from Baron-Cohen and colleagues (1986). Here, the patient was required to form a story out of different puzzle pieces. The second task taken from Happé (1995) was an advanced ToM test that employed similar stories as those reported in Fletcher et al., (1995) and Gallagher et al., (2000). The third task tested the patient's sensitivity to violations of social norms and was devised from Berthoz and colleagues (2002). This presented the patient with stories containing, for example, embarrassing content and questioned whether this was indeed embarrassing, and if such behaviour was appropriate. The fourth task taken from Baron-Cohen and colleagues (1999a) presented patient GT with a social faux-pas paradigm. The fifth and final task presented animated clips similar to those reported earlier by Castelli et al., 2000. Patient GT showed impairment in planning and prospective memory; however, despite these deficits, she showed intact ToM reasoning abilities across all tasks. The authors concluded that the MPFC may not be necessary for ToM.

More recent studies have shown that it is not only the frontal areas that are associated with ToM but that the TPJ also plays an important role. Samson and colleagues (2004) showed that patients with lesions to the left temporal parietal junction were impaired in belief reasoning. The authors created a video-based nonverbal false-belief task in which several control conditions were included (memory control, inhibition control and an additional control to ensure that participants do not use a superficial strategy to solve the task). The TPJ patients performed very well on all the control trials but they did not perform above chance level on the false-belief trials.

### 1.4.2. Dissociation between ToM components

Some authors have tried to decompose the processes involved in belief reasoning using the neuropsychological approach. For example, Samson and colleagues (2005) presented a single case study of a patient WBA (also reported in Apperly et al., 2004), who sustained lesions to the right frontal lobe. The authors presented two types of video-based tasks that required reasoning about someone else's belief. One task was based on a typical change-of-location false-belief paradigm where the participant is aware of an object's change of location whereas a protagonist in the video is not aware of this change. The second task was slightly changed to decrease the salience of the participant's own knowledge of the object location. In that task, the participant witnessed the change of location, without knowing the object's new location. The study showed that WBA was impaired only in the traditional paradigm where his own knowledge of the location of the object was salient. In contrast, when the patient had no precise knowledge about the location of the object, he was able to infer that the other person would have a false belief. The authors concluded that WBA suffered a specific deficit in self-perspective inhibition.

A later study showed that a lesion to the TPJ is not associated with a deficit in self-perspective inhibition but rather linked to difficulties in monitoring the environment for relevant cues on the basis of which someone else's belief content can be inferred (Samson et al., 2007). Here, patients were presented with non-verbal false-belief task that provided three types of response choices. The videos depicted a character placing an object (e.g., a passport) in a box (e.g., a pizza box) and then changing the content with a different object (e.g., scissors). The change was made in the presence (true-belief) or absence (false-belief) of another character. The test question was related to the second character's belief of the content of the box. The three choices given to the patients were (1) the object that would usually found in the box (i.e., a pizza is this is the usual content of a pizza box, appearance based response, incorrect for both true- and false-belief trials), (2) the first object placed in the box (i.e., a passport, correct response for false-belief trials), or (3) the last placed object in the box (i.e., a pair of scissors, reality-based response correct for true-belief trials but incorrect for false-belief trials). This three-choice response allowed these researchers to dissociate two profiles of patients. The first profile corresponded to that of a patient who had lesions in the left TPJ and who mainly made appearance-based errors. The absence of reality-based responses showed that the patient was able to inhibit her own knowledge of the content of the box. Instead, her difficulties seemed to be in keeping track of the relevant information in the environment. The second profile was that of patient WBA with right frontal lesions and who only gave reality-based responses, showing deficits in inhibiting his own salient perspective.

Other authors have looked at other dimensions along which ToM can be decomposed. As highlighted already by the dissociations between the faux-pas task and the false-belief tasks observed in the study by Gregory et al., (2002) and the study by Stone et al., (1998) described above, there could be different cognitive processes each specific task taps upon, with the faux-pas task involving an affective component that is absent in the classic false-belief tasks. Some authors have directly examined whether affective and cognitive perspective taking is dissociable and whether these tap onto different neural networks. For example, Shamay-Tsoory and Aharon-Peretz (2007) examined the performance of a group of patients with limited focal lesions in distinct unilateral PFC and posterior regions on two types of tasks. The first task involved judgments based on images depicting a character's eye gaze and verbal cues (Baron-Cohen, 1995; i.e., the Charlie task). There were three types of conditions (see Figure 7): (1) one probing the object for which the character had a positive or negative emotional attachment, (2) one probing the object a character was thinking of (cognitive perspective condition), and (3) one probing the object that was in physical proximity to the character (physical control condition). The second task involved short stories that either presented an affective or cognitive false belief and also an affective and cognitive ignorance statement. For the affective false-belief condition sentences were structured as such*:* "*Joe and Anna are setting the table for a festive dinner at the dining room. Anna pours Joe a glass of water, but some water spills on his new shirt. Joe says: 'It's nothing, I will change the shirt later'. Anna puts the glass on the table and goes to look for a paper towel to dry Joe's shirt. When she leaves the dining room, Joe gets furious about the wet shirt and kicks the table. Anna peeks into the dining room, sees what Joe is doing and feels guilty. Anna comes back to the dining room*." (Shamay-Tsoory & Aharon-Peretz, 2007, p. 3066). The test questions were: "*1. What does Joe think that Anna feels about the wet shirt, when she returns? 2. What Anna thinks Joe feels about the wet shirt? 3. How does Joe feel? 4. What were Joe and Anna preparing for?"* (Shamay-Tsoory & Aharon-Peretz, 2007, p. 3066). For the cognitive false-belief condition, sentences were structured as such: "*Joe and Anna are setting the table for a festive dinner at the dining room. Anna pours Joe a glass of water, but some water spills on his new shirt. Joe says: 'It's nothing, I will change the shirt later'. Anna puts the glass on the table and goes to look for a paper towel to dry Joe's shirt. When she leaves the dining room, Joe takes his handkerchief and dries the shirt and the table. Anna peeks into the dining room, sees what Joe is doing, and so she doesn't bring a paper towel. Anna returns to the dining room."* (Shamay-Tsoory & Aharon-Peretz, 2007; p. 3066). The test question involved: "1. *What does Joe think that Anna thinks about the shirt's condition, when she returns to the dining room? 2. What does Anna think of the shirt's condition? 3.*

*What is the shirt's condition? 4. What were Joe and Anna doing?"* (Shamay-Tsoory & Aharon-Peretz, 2007; p. 3066). In the case of cognitive irony, the sentences stated i.e., "*Joe went into the bank's manager office and couldn't find anywhere to sit down because all the chairs were occupied with documents and folders. An unorganized pile of letters and documents were randomly set on the table. Joe said to the bank manager 'your office is so tidy!'"* (Shamay-Tsoory & Aharon-Peretz, 2007; p. 3066). Test questions here were: *1. Why did Joe say that? 2. Did Joe think the office was tidy? 3. Was the office tidy? 4. Which office did Joe go to?"* (Shamay-Tsoory & Aharon-Peretz, 2007; p. 3066). Finally, the affective irony condition sentences were constructed as i.e., "*Joe's dad was supposed to pick him up after chess club at 6 p.m. By the time he recalled he had to pick up his son, it was 7 p.m. Dad found Joe standing tired and frightened out in the rain. When they got home, Joe was crying and told his mom what had happened. Mom said to dad 'You are such a good father!'*" (Shamay-Tsoory & Aharon-Peretz, 2007; p. 3066). Test question were: "*1. Why did mom say that? 2. Did mom think dad was a good father? 3. Was dad a good father on this occasion? 4. What kind of club did Joe attend?*" (Shamay-Tsoory & Aharon-Peretz, 2007; p. 3066). The authors showed that across both types of tasks (picture and story tasks), patients with ventromedial (VM) frontal lobe damage tended to make more errors on the affective than cognitive ToM trials. Damage to more extensive prefrontal areas produced affective and cognitive ToM deficits.

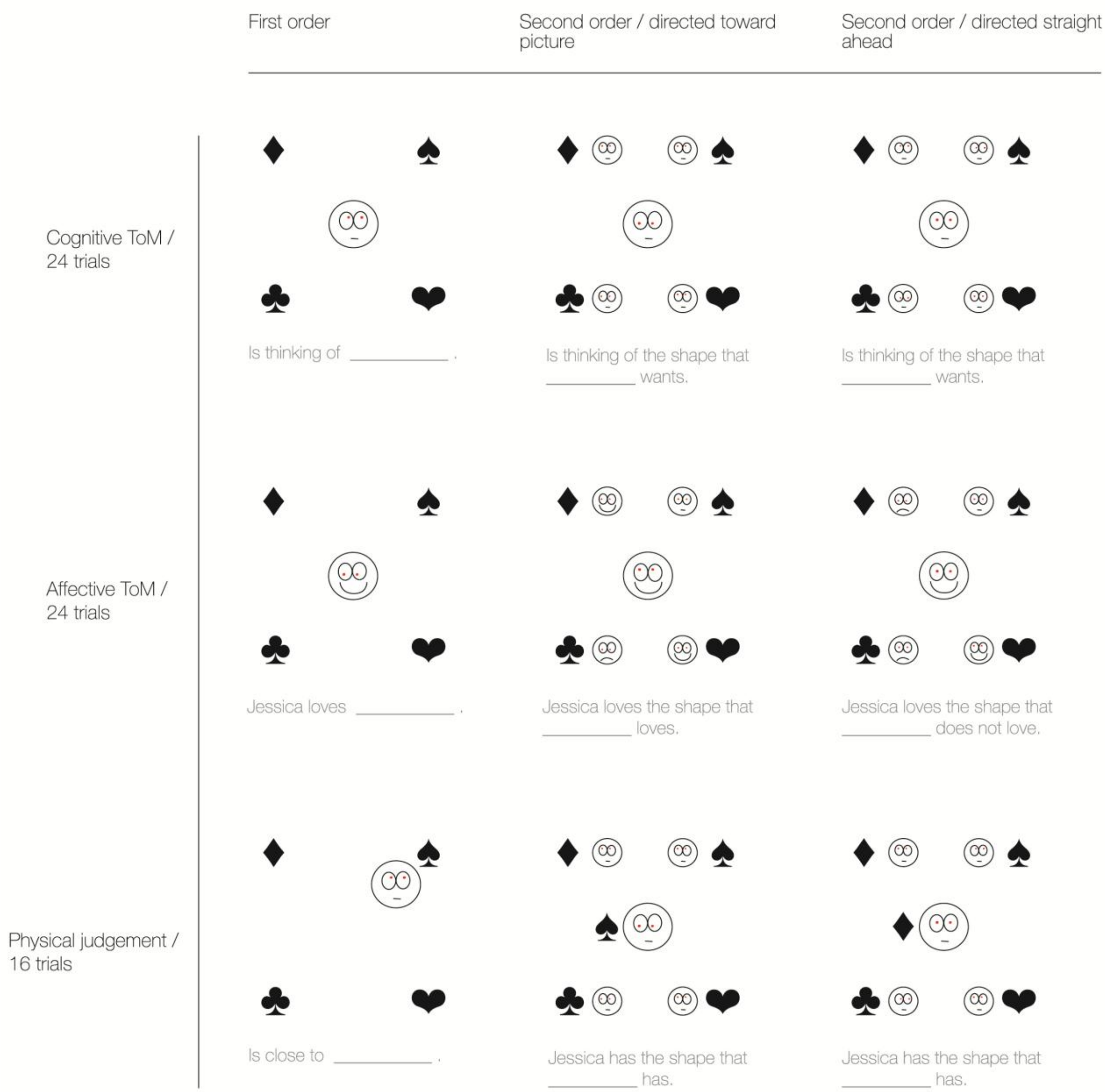



Figure 7. Example of stimuli used by Shamay-Tsoory and Aharon-Peretz (2007, p. 3059).

So far no studies have directly compared desire and belief reasoning in a systematic way. There is some evidence that the patient WBA described earlier with a self-perspective inhibition deficit was impaired not only in belief reasoning but also when reasoning about someone else's desire or emotion, but the patient was only tested on a limited number of desire and emotion reasoning trials (Samson et al., 2005).

In sum, neuropsychological studies have attempted to identify which areas are necessary among those that show increased BOLD activity when healthy adults mentalize. Several studies have shown that the frontal areas (Goel et al., 2004; Rowe et al., 2001; Stone et al., 1998; cf. Bird et al., 2004) and the TPJ (Samson et al., 2004; Saxe & Kanwisher, 2003) play a critical role. There is also evidence for a functional distinction between dorsolateral frontal areas involved in self-perspective inhibition (Samson et al., 2005, 2007), the TPJ involved in monitoring the environment for relevant cues (Samson et al., 2007) as well as the ventromedial prefrontal areas involved in affective perspective taking. To date, there has been no systematic investigation of desire reasoning in patients but there could be a link between desire reasoning and affective perspective taking, since, as I explained at the beginning of this chapter, desires can originate from emotional states. Based on this, a problem in processing someone else's emotion could lead in problems in desire reasoning without affecting belief reasoning.

## 1.5. Evidence from behavioural studies with healthy adults

As adults, we have reached the endpoint of ToM development and are able to correctly respond to tasks presented to children. However, even though one would think that adults would perform flawlessly, in reality, when using fine-grained measures, it has been shown that even adults commit errors or can take some time to respond. Understanding where the cost in errors or reaction time originates can provide valuable information about the cognitive processes underlying the use of one's ToM.

In a study by Birch and Bloom (2007), adults were presented with a false-belief task and were randomly assigned to one of three conditions: (1) ignorance, (2) knowledge-plausible, and (3) knowledge-implausible. Across all conditions, participants were presented with pictures that depicted a girl who places her violin in one of four possible locations (see Figure 8). The story varied depending on the experimental condition. In the ignorance version, participants were told that the violin was moved to another location. For the knowledge-plausible version, they were told the violin moved to the red container. Finally, in the knowledge-implausible version, the story was that the violin was moved to the purple container. All displacements were made by a second character presented in the story. Test questions involved the first character and participants were asked to rate the percentage this character would look in each of the four given locations provided. The authors found that adults underestimated the percentage of time/probability that the protagonist will search for a particular object in the location corresponding to his false belief, thus suffering a "curse of knowledge".

Keysar and colleagues (2003) also demonstrated that adults suffer significant interference of their own perspective in a simple perspective taking task. Here, participants were presented with a grid and were required to move objects in that grid according to the instructions of a speaker. Some of the cells of the grid were occluded so that the speaker could not see which object was behind the occlusion. On the critical trials, the object that the speaker asked the participant to move was ambiguous. For example, the grid would contain a small candle, a medium size candle or a large size candle and the small candle was positioned in an occluded cell. When the speaker asked to move the "small candle", this referred actually to the medium size candle that both the speaker and the participant could see.

Keysar and colleagues showed that on such trials, participants were often interpreting the speaker's instruction from their own point of view not taking into account the speaker's visual perspective (and hence moving the small candle that was not visible to the speaker). The results of these two studies show that even for healthy adults (and not only children or patients with frontal lesions like WBA), it is hard to inhibit their self-perspective.

It also appears that adults do not in all situations spontaneously infer other people's belief. For example, Apperly and colleagues (2006) examined the processing costs required in belief inferences. Participants were presented with videos and asked to monitor the location of an object which was moved from one location to the other. A woman in the video sometimes witnessed the object's change of location and sometimes did not witness the changes. During the viewing of the video, participants would sometimes be probed about events happening in the video (e.g., about the colour of the object, its location or, critically, the woman's belief about the object location). If at the onset, participants were only explicitly required to monitor the object, it took them longer to respond to the woman's belief probe than to the other probes. However, if at the onset, participants were required to monitor both the object and the woman's thoughts, the difference in reaction time disappeared. These results suggest that belief inferences can, but are not necessarily, made online, possibly because a certain amount of effortful control is required to do so.

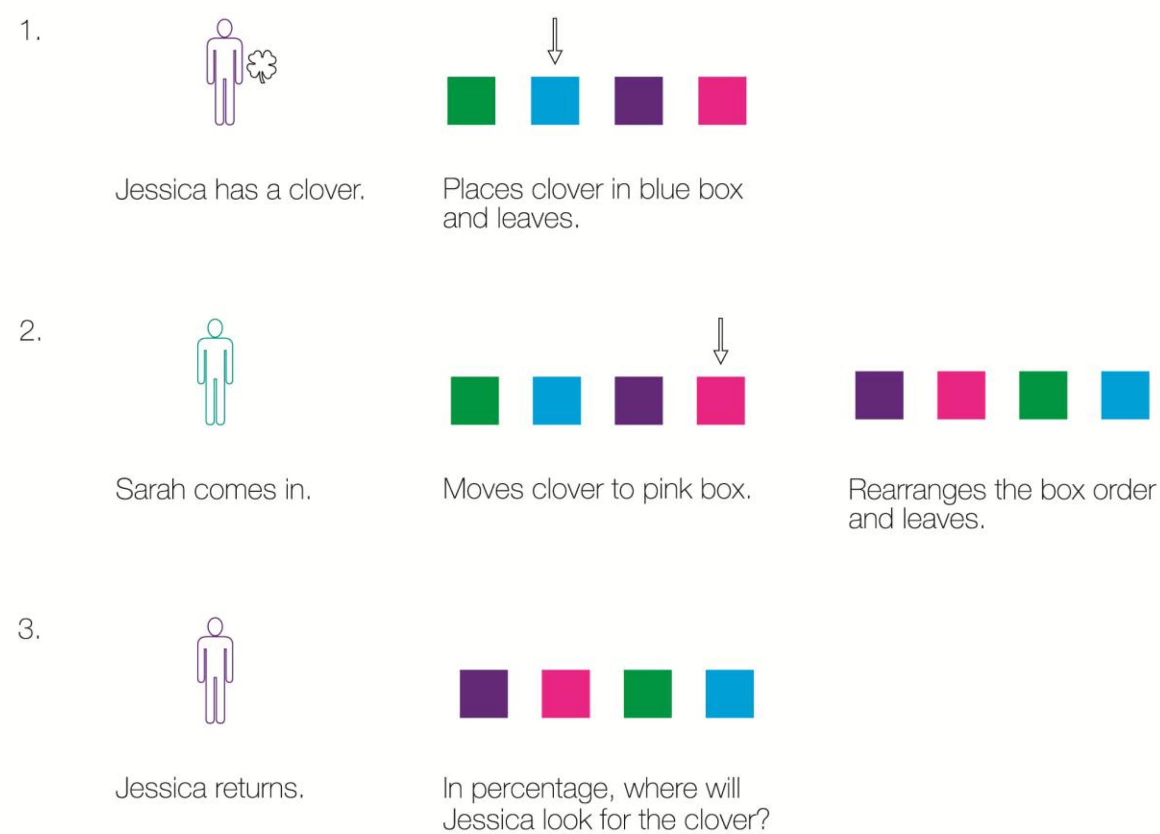



Figure 8. Demonstration of a knowledge-plausible condition in Birch and Bloom's study (2007, p. 384).

Even when no belief inference is required, simply holding a false belief in mind can be cognitively costly. In another study by Apperly and colleagues (2008), participants were asked to read two sentences describing a physical attribute of an object and a person's belief of the object's appearance. Thus, they received two sets of information to hold in memory. In one condition, the real appearance of the object (reality) was in conflict with the person's belief about the appearance of the object (false belief; e.g., "*She thinks the ball is red but really the ball is yellow*"). In a second condition, there was no conflict between belief and reality (the attributes provided for the woman's belief and the reality being unrelated; e.g., "*She thinks the ball is red and the ball is on the chair*"). After the sentence presentation, participants were asked to judge whether a picture probe matched the reality or the belief content described in the sentence. The results showed a cognitive cost when simply holding a false belief in mind even when no inference was required.

To date, few studies have examined desire reasoning in adults. German and Hehman (2006) asked adults to read stories and predict a character's action on the basis of his belief about the situation (either true or false) and his desire (either approach or avoidance desire). The authors presented written scenarios that described the four combinations of beliefs and desires. Their results showed that participants were slower and more error prone on certain combinations (i.e., false belief, avoidance desire) relative to others (true belief, approach desire). The aim of the study was not to compare belief and desire reasoning but to see whether false-belief reasoning could be made more difficult for adults if the desirability of the object was manipulated. Supporting Leslie's model (Leslie & Pollizi, 1998) described earlier, the false-belief + avoidance-desire condition was found to be the most difficult. The true-belief + avoidance-desire condition did not appear to be much more difficult than the true-belief + approach-desire condition. In other words, avoidance-desire reasoning did not seem to produce extra processing costs unless it was combined with false-belief reasoning.

In sum, results from studies conducted with healthy adults show that mental state inference requires effortful processes at different stage of information processing. Evidence so far mainly comes from studies that have looked at belief reasoning and show that effortful controls is required to inhibit one's own perspective, track someone else's belief, and hold a false belief in memory. There is some suggestion that processing an avoidance desire is also cognitively costly when combined with false-belief reasoning, but overall the cognitive processes underlying desire reasoning remains to date largely unexplored.

## 1.6. Theory of mind and executive function

This chapter highlighted that reasoning about mental states relies on effortful and controlled processes, which are commonly referred to as executive function (Burgess, Alderman, Evans, Emslie & Wilson, 1998). As seen in previous sections, there are several ways in which executive function and ToM are related. First, executive function is necessary to deal with the incidental task demands such as monitoring and integrating the sequence of events in ToM tasks. This executive function role is not specific to ToM but it is seen in any reasoning problem in which various elements have to be integrated and held in mind (Apperly, Samson & Humphreys, 2005).

Secondly, executive function also plays a more essential role in mental state reasoning. As shown by studies with children and adults (Carlson & Moses, 2001; Carlson et al., 2002; Keysar et al., 2003), executive function is necessary for inhibiting one's own salient self-perspective, not only when inferring mental states but also when simply holding in mind the other person's perspective (Apperly, Back, Samson & France, 2008). This has been empirically demonstrated in belief reasoning for both children and adults. However, for desire reasoning, this has been mainly shown in the developmental literature. Other executive function processes have been identified as essential in desire reasoning, particularly in reasoning about avoidance. These executive function processes have been identified as those required to shift attention away from the object to be avoided. It has not been investigated yet if inhibiting one's own perspective relies on the same inhibitory processes as those required to shift attention away from the object to be avoided. However, there are suggestions that the inhibitory processes involved in reasoning about mental states may be the same as those involved in any reasoning task that requires resisting interference from salient but irrelevant objects. For example, developmental studies have shown significant correlations between children's performance on false belief and conflicting desire tasks and their performance on standard tests measuring inhibitory control (Carlson & Moses, 2001; Carlson et al., 2002; Rakoczy, 2009). Furthermore, WBA, the patient previously mentioned with a self-perspective inhibition deficit, was also impaired on standard tests measuring inhibitory control (Samson et al., 2005) and his lesion location encompassed a brain region (right lateral prefrontal cortex) found to be important for response inhibition (Aron, Robbins & Poldrack, 2004).

## 1.7. Current Thesis

The current thesis aimed to directly assess the functional and neural mechanisms underlying belief and desire reasoning in adult ToM. A series of novel tasks were created in order to understand the extent executive function plays in mental state reasoning, and whether these were associable/dissociable processes, constituting the architecture of the social brain. This was investigated in both neurological brain injured patients and healthy adults. Four empirical chapters are presented.

In Chapter 2, patients with frontal, temporal and parietal lesions took part in a study that directly compared belief and desire reasoning in two separate ToM tasks. The aim was to investigate processes required towards success, when the self-perspective inhibition demands were controlled for across both tasks. The evidence (Experiment 1) showed three patterns of performance in the different-perspective consistency trials: (1) patients with intact belief and desire reasoning abilities (PH, PF, RH & SP); (2) patients with deficits in beliefs only (DS, GA, JeB, MP & PW); and (3) one patient who showed impairment across both mental states (WBA). Based on these results, a second experiment (Experiment 2), focusing solely on desires, was administered to the same group of patients. This experiment was designed to avoid simple matching strategies that patients may have adapted in order to achieve success in Experiment 1. A second level of desires was introduced and patients were required now to equally approach and avoid a desired target for their opponent's perspective. The data (Experiment 2) showed two patterns of performance: (1) patients who were able to reason about both approach and avoidance desires for other-perspective (DS, JeB, MP, PF, PH, PW, RH and WBA) and (2) patients who showed a specific deficit in avoidance only, across same- and different-perspective consistency trials (GA and SP). Overall, the results of this chapter highlighted the necessary role of self-perspective inhibition and other potential functional processes required towards mental state reasoning. Such processes, unrelated to the self-perspective inhibition dimension, were found in patients who showed intact belief reasoning abilities but failed the desire task, when the cognitive load was increased by introducing avoidance (SP). Also, the reverse dissociation was observed in patients who performed well on the desire reasoning task yet were impaired in the belief reasoning task (DS, JeB, MP, & PW).

Given that we know very little about desires, in Chapter 3, I provide evidence of cognitive costs required when healthy adults formulated desire judgments. In Experiment 3, participants were initially passively involved and asked to monitor two players' perspectives. Within the second half of the experiment, they were told that they were actively involved and asked to formulate self and other judgments. Both parts of the task (active, passive) were identical in structure and players were represented by symbols (green and purple pawns) on screen. The only difference between parts was the instructions given to the participants prior commencing each. The results showed a difference between passive and active involvement trials, whereby merely telling participants they were now engaged in the task made participants faster and less error prone for their own perspective relative to their opponent's. This did not occur when participants were merely observing two players within the first half of the study (baseline condition), where no difference in costs could be observed. Also, across both active and passive trials, costs were greater when asked to avoid a target relative to approach. In Experiment 4, the order of involvement was swapped, where participants were first active and then asked to passively attend to two players. The results showed that participants were responding faster and were less error prone when judging self trials in the active involvement and this association was carried over to the passive trials, where a preference for the previous stated symbol marked as self was apparent. Also, participants were quicker when asked to judge approach trials relative to avoidance across active and passive sections. Interestingly, costs observed between self and other disappeared in the avoidance trials, when participants were initially active in the task. This did not occur in Experiment 3, when the participants were directly engaged after taking part in the passive trials. In Experiment 5, participants were presented with the active trials only, and the symbols representing each player were replaced with photographs taken of the participant prior testing. This was thought to raise the saliency of self within the task and aimed to further investigate the avoidance costs between self- and other-perspective. The results found in the active trials of Experiment 4 were replicated in this study (Experiment 5), highlighting no difference between self and other for avoidance. In conclusion, the evidence reported in this chapter suggested that there were two types of processing costs when healthy adults reason about desires: (1) self-perspective inhibition cost (difference between self and other judgments on approach trials only) and (2) avoidance-desire cost (difference between approach- and avoidance-desire judgments for self-perspective trials only).

In Chapter 4, I aimed to further investigate this evidence and examined the level of processing at which inferences occurred (Experiment 6a and 6b) and whether such results would be replicated under strategic control (Experiment 7). In Experiment 6 (a and b), I manipulated the display time and examined the levels of constructing and maintaining such mental representations. In Experiment 7, participants knew in advance which perspective needed to be attended to, prior receiving the coloured cards. This was done to reduce the perspective demands of the task. Overall, the data supported previous results of two different types of costs: (1) participants were quicker and less error prone when judging self relative to their opponent's perspective, in the approach trials; and (2) they responded faster and were less error prone for the approach trials in comparison to the avoidance trials for self-perspective. Specifically, the effect of perspective in the approach trials occurred during mental representation, whilst maintaining such information in mind and during strategic control. The significant difference found between both two desire types for self-perspective trials showed larger effects during construction of the mental representation (Experiment 6a). In sum, the evidence provided additional proof to the existence of two functional dimensions when healthy adults reasoned about desires.

In Chapter 5, I examined these effects in neurological brain injured patients (GA, PW, SP, and WBA) to determine whether they reflected the existence of functionally and neurally distinct processes. This proved indeed to be the case, where I demonstrated a double dissociation in my patients' pattern of performance. There were two distinct sources of difficulties: one linked to self-perspective inhibition (PW and WBA), the other linked to the inference of an avoidance desire (GA and SP).

In the final Chapter (the general discussion), findings of the empirical chapter are summarized. Overall, I provide the first empirical evidence of two processes (self-perspective inhibition cost and avoidance-desire cost) that were dissociable at a functional and neural level in adult desire reasoning.

# CHAPTER 2

## Belief and desire reasoning in patients with acquired brain injury

## 2.1 Introduction

This chapter is the first stepping stone of my thesis, where I examined brain injured patient's performance on belief and desire reasoning tasks. Inspired by the tasks developed by Moore and colleagues' study (1995), I created two novel, non-verbal ToM tasks: one in which participants were asked to infer the protagonist's belief and the other in which they were asked to infer the protagonist's desire. In both tasks, on half of the trials (different-perspective trials), both the protagonist and participant had different-perspectives (think or want something different). On the other half (same-perspective trials), both parties had the same-perspective (think or want the same). Moore and colleagues (1995) found that when self-perspective inhibition demands were matched across a belief and desire inference task, children found both tasks equally difficult. To properly compare belief and desire reasoning and to investigate any difference in processing between belief and desire reasoning other- than self-perspective inhibition it is thus important to match the belief and desire tasks in terms of self-perspective inhibition demands.

As shown in Chapter 1, neuropsychological evidence (i.e., Apperly et al., 2004; Samson et al., 2004; Samson et al., 2005; Samson et al., 2007) shows that self-perspective inhibition still plays a critical role in adult belief reasoning. One example is the case of patient WBA who, following acquired lesions to the right inferior frontal gyrus, showed a selective deficit in inhibiting his own self-perspective when reasoning about other people's beliefs. Self-perspective inhibition is not the only type of deficit that brain-damaged patients can experience. For example, patients with lesions to the left temporal parietal junction (TPJ) have been shown to have problems in belief reasoning even when there were few demands in self-perspective inhibition (Samson et al., 2004). It has been argued that these patients may have difficulties selecting the relevant cues to infer someone else's belief from the stimuli in the environment (Samson et al., 2007). In sum, these neuropsychological results provide evidence for separate perspective taking components required for successful belief reasoning in adults.

Following Moore et al.'s findings (1995), and based on the observation that different patterns of deficits on belief reasoning tasks emerge in brain damaged patients depending on their lesion site, I selected patients with frontal, temporal and parietal lesions and directly compared their belief and desire reasoning performance on newly designed tasks in which the self-perspective inhibition demands were matched. Crucially, I aimed to see whether any deficits in belief reasoning would be associated with a deficit in desire reasoning. From Moore et al. (1995), I would expect that patients like WBA, who show self-perspective inhibition deficits in belief reasoning tasks, would also have difficulties in desire tasks when the self-perspective inhibition demands are well matched across both types of tasks. For the TPJ patients, the problem seems more with looking for relevant cues in the environment. It is possible that this would be more difficult in belief tasks than desire tasks and so perhaps I may expect to find dissociation between beliefs and desires. Third, I could expect to find the reverse dissociations amongst patients, i.e., deficits in desires with individuals who have been previously reported to pass false-belief tasks. This has never been shown. Instead the developmental literature shows that belief reasoning usually taps onto more cognitive demanding processes, not necessarily required for desire reasoning. However, as seen in Chapter 1, reasoning about desires relies on different concepts (e.g., emotions, physiological states) than reasoning about beliefs (e.g., knowledge access information). Patients may have difficulties reasoning about the concepts associated with desires but not the ones associated with beliefs.

I presented participants with two experiments in this chapter. In Experiment 1, I used two separate tasks investigating participants' ability to reason about someone else's beliefs and desires. The tasks were matched for self-perspective inhibition demands. However, it could be argued that the reasoning demands per se were higher in the belief than the desire task. Therefore in Experiment 2, I further examined desire reasoning by keeping the same self-perspective inhibition requirements but increasing the reasoning demands to create a better fit to the belief task.

## 2.2. Experiment 1: Comparing belief and desire inference

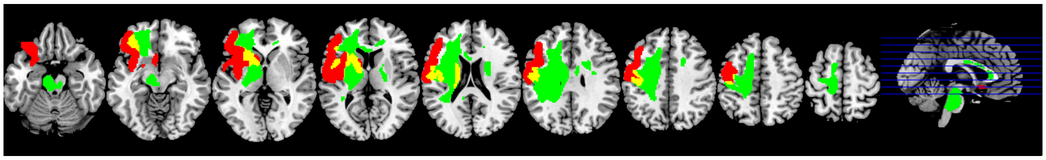
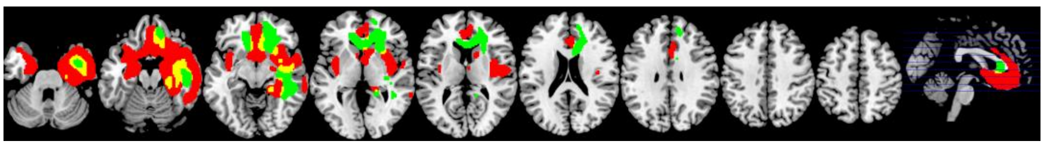
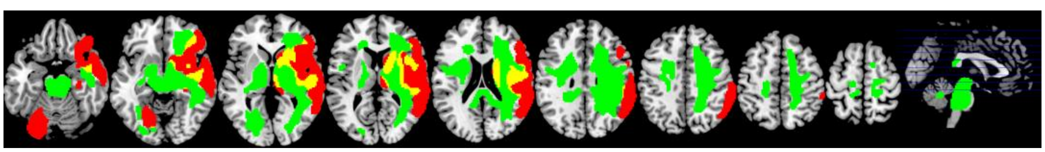
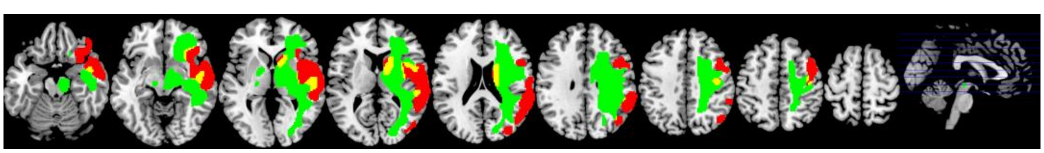
### 2.2.1. Method

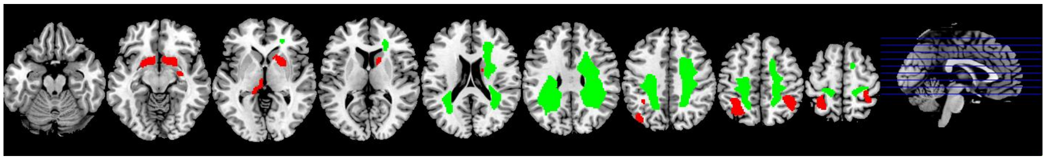
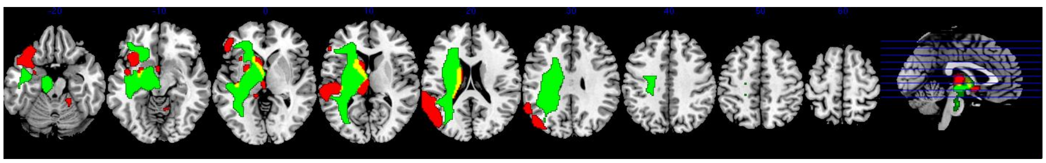
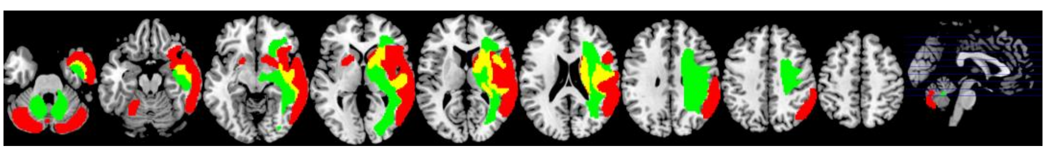
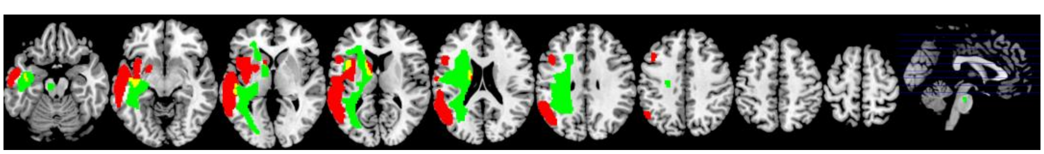
#### *Participants*

Ten patients (DS, GA, JeB, MP, PF, PH, PW, RH, SP, and WBA) with lesions to the frontal, parietal and/or temporal regions of the brain were selected for this study. From previous published (e.g., Apperly et al., 2004) as well as unpublished data, it also appeared that all 10 patients should be in principle able to deal with the incidental task demands. The patients characteristics are described in Table 1 and lesion descriptions can be found in Figure 9. Additionally, five healthy age-matched adult controls (all male, $M_{age}$ = 62 years; SD = 7.3) took part.

Table 1. Patients' characteristics

| Patient | Gender | Age | Hand | Major Clinical Symptoms | Aetiology |
|---|---|---|---|---|---|
| DS | M | 70 | R | aspects of dysexecutive syndrome, aphasia | stroke |
| GA | M | 52 | R | amnesia, dysexecutive syndrome, category specific agnosia | herpes simplex |
| JeB | F | 71 | R | extinction, left neglect dyslexia | stroke |
| MH | M | 53 | R | right extinction, optic ataxia | anoxia |
| PF | F | 58 | R | left extinction, dysgraphia | stroke |
| PH | M | 34 | R | right hemiplegia, aphasia | stroke |
| PW | M | 75 | R | left hemiplegia, dysexecutive syndrome | stroke |
| RH | M | 73 | L | right neglect, aphasia | stroke |
| SP | M | 53 | R | amnesia, dysexecutive syndrome, category specific agnosia | herpes simplex |
| WBA | M | 61 | R | some dysexecutive symptoms | stroke |

| **Patient** | Main Lesion Site |
|---|---|
| **DS** | <br>Left inferior frontal gyrus: pars opercularis, pars triangularis and pars orbitalis, left rolandic operculum, left insula, left middle frontal gyrus, left precentral gyrus, left postcentral gyrus, left caudate & putamen. |
| **GA** | <br>Left and right anterior cingulate gyri. Right parahippocampal gyrus & hippocampus. Left hipocampus. Right and left amygdala. Right fusiform. Right & left insula. Right caudate. Right inferior & middle temporal gyri. |
| **JeB** | <br>Right sided lesion of superior and middle temporal, supramarginal, inferior parietal, angular gyri, extending into postcentral and precentral gyrus, inferior frontal gyrus: pars opercularis, triangularis and orbitalis, thalamus, caudate & insula.<br><br>Small lesion in left lingual gyrus, extending into left cerebellum. |
| **MP** | <br>Right sided lesion affecting the superior and middle temporal gyri, superior & middle temporal poles, precentral gyrus, middle frontal gyrus, inferior frontal gyrus: pars opercularis & orbitalis, rolandic operculum, supramarginal gyrus, angular gyrus, insula & caudate |

| | |
|---|---|
| **PF** |  bilateral lesions in the superior and inferior parietal gyri, left sided lesion extends into the left angular gyrus and the right sided lesion extends slightly into the right postcentral gyrus. There is a small lesion in the left putamen which extends into the right caudate & another small lesion in the right putamen extending into the right caudate. There is also a small lesion in the left thalamus. |
| **PH** |  Left middle occipital gyrus and angular gyrus, which extends into the left superior temporal gyrus, left inferior frontal gyrus: in the pars orbitalis and the pars triangularis. There is a small lesion in the left caudate which extends into the left putamen and left thalamus. |
| **PW** |  Right superior, middle and inferior temporal gyri and angular gyrus, extending into inferior parietal and supramarginal gyri. Right caudate, insula, putamen & inferior frontal (operculum). Small lesion in left putamen. Bilateral cerebellar lesions. |
| **RH** |  Left sided lesion affecting the superior and middle temporal gyri, supramarginal gyrus, angular gyrus, inferior frontal gyrus: pars opercularis, rolandic operculum, insula, pallidum & caudate |

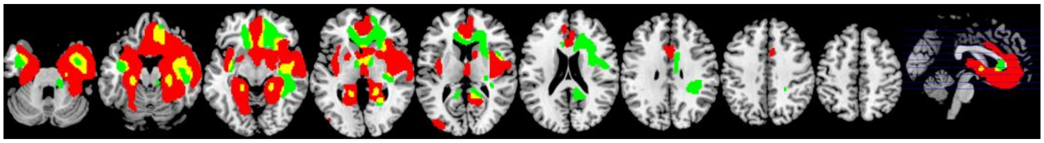
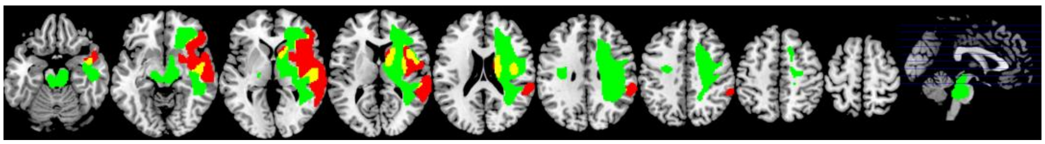
| | |
|---|---|
| **SP** |  Right superior, middle and inferior temporal. Bilateral fusiform and parahippocampalextending into hippocampal gyri. Left and right amygdala, insula and putamen. Left and right anterior ciungulum. Left mid occipital. Right inferior & superior orbital frontal. |
| **WBA** |  Right supramarginal, superior and middle temporal gyri , extending into rolandic operculum, inferior frontal gyrus: pars opercularis, triangularis and orbitalis, insula, thalamus and caudate. |

Figure 9. Lesions created through voxel-based morphological (VBM) analysis of T1 scans using SPM and added as an overlay onto a standard multi-slice template in MRIcron. SPM stats: one sample t- test with 3 covariates: healthy (140 brains aged 40+) v patient, age & sex FWE p=0.1. Extent threshold including only significant blobs containing ≥ 100 voxels Grey matter lesion in red. White matter lesion in green. Yellow represents areas which may be grey or white matter lesion.

### *Stimuli and procedure*

Two types of tasks were presented: one in which the participant was asked to infer someone else's belief and the other in which the participant was asked to infer someone else's desire. Both tasks were presented within the context of a story character named detective O'Reilly, displayed in short animated clips on Microsoft PowerPoint. In both tasks, on half of the trials, detective O'Reilly and the participant had a different perspective (thinking or wanting something different). On the other half of the trials both parties had the same perspective (thinking or wanting the same). The belief and desire tasks were carefully constructed so as to have a similar structure and to be equally engaging (see Table 2). Testing was spread over two sessions of approximately 1 hour each. In the first session, the participant first started with a card game (desire condition) followed by the robber chase task (belief condition). In the second session, the participant was first presented with the robber chase task (belief condition) followed by a card game (desire condition). In each testing session, each task was preceded by practice trials to ensure that participants understood the task.

Table 2. Illustration of the experimental structure of both tasks (belief and desire)

| **Task** | **Belief** | | **Desire** |
|---|---|---|---|
| **Agent** | thinks robber is in A | **Opponent** | wants a red card |
| **Participant** | thinks robber is in A | **Participant** | wants a red card |
| **Perspective Consistency** | *Same* | **Perspective Consistency** | *Same* |
| | | | |
| **Agent** | thinks robber is in A | **Opponent** | wants a red card |
| **Participant** | thinks robber is in B | **Participant** | wants a blue card |
| **Perspective Consistency** | *Different* | **Perspective Consistency** | *Different* |

### *Belief*

The belief task (see Figure 10) was based on a classic false-belief change-of-location paradigm. The animation showed a robber's chase where the participant passively watched detective O'Reilly hunting down the evil villain. Each trial consisted of the robber moving in full view of the participant to one hiding location (e.g., a sandbox in a playground) and then to another hiding location (e.g., into a tree; see Figure 10). The detective was always on the screen and could always see the first hiding place. Just before the robber moved to the second location, a distracting event happened (e.g., a passerby came to ask the detective a question or chickens obstructed a passing car). On half of the trials, the detective's position during the distracting event meant that he could still see where the robber moved to for his second hiding place (same-perspective/true-belief trials where both the participant and the detective knew were the robber was hiding). On the other half of the trials, the detective's position meant that he did not see where the robber moved to (e.g., the detective was facing away from the robber; different-perspective/false-belief where the participant knew where the robber was hidden but where the detective held a false belief about the robber's location). Each time the robber had moved to his second hiding location, the participant was asked where the detective thought the robber was. This question directly probed the detective's belief (i.e., "*Where does detective O'Reilly think the robber is?*"). Response options were depicted on paper by showing the detective with three thought bubbles, each one showing the robber in a different location: the current hiding place (correct response on the true-belief trials), the previous hiding place (correct response on the false-belief trials) and a place where the robber passed by but behind which he never hid (neutral distractor). Sound effects were added to keep the participants engaged in the task. There were in total 24 trials split across 2 sessions, with each session including 6 true-belief and 6 false-belief trials.

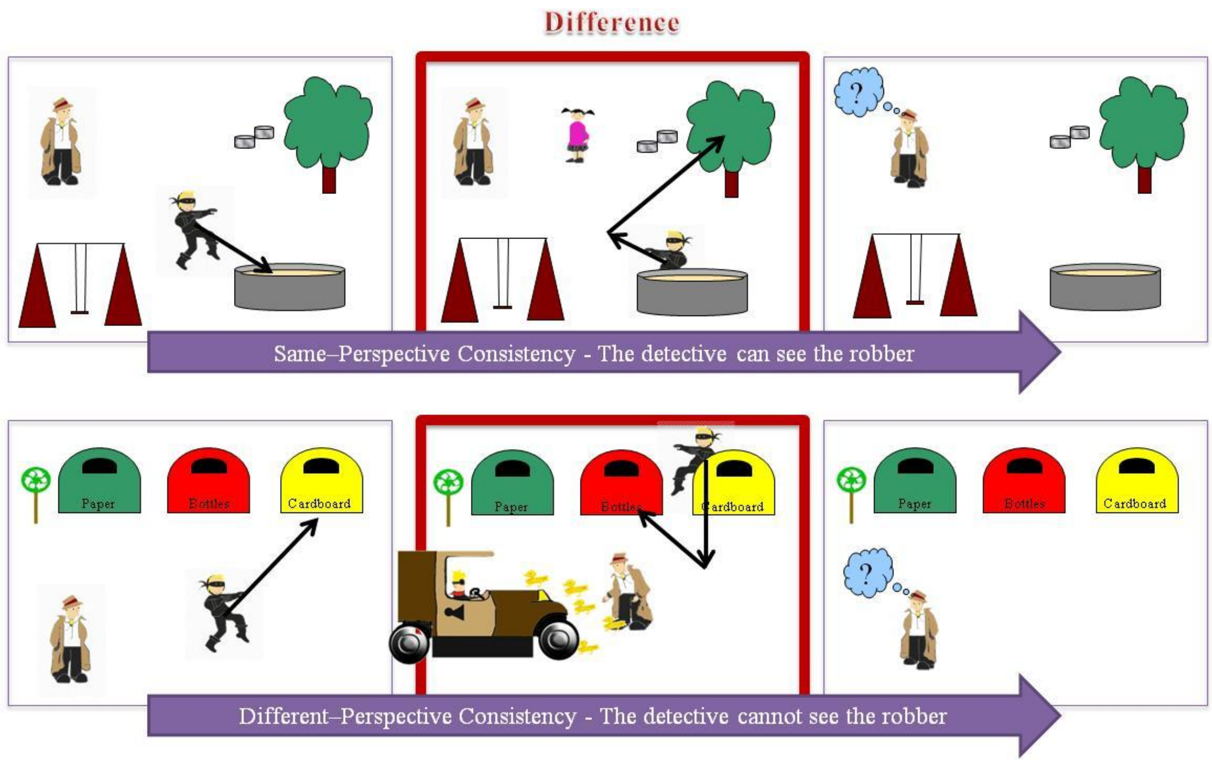



Figure 10. Schematic representation of the belief task for both same- and different-perspective trials.

### ***Desire***

The desire task (see Figure 11) comprised a card game ("show me the money") in which detective O'Reilly directly invited the participant to play a game against him. The rules of the game were then explained to participants. Participants were told that each player would receive a coloured card (blue, red and yellow). A third card would then be turned over and whoever has the same colour as the colour of the third card would win that round. The game was implemented by animations in PowerPoint. After two trials of practice and once the participant had understood the rules, the test trials began.

Each round or trial started with the display of a 'money counter' that indicated how much money was at stake in that particular round (amounts employed ranged from £10 to £100, in £10 increments and presented in random order). Participants and detective O'Reilly then each received their coloured card. The participant's card was always distributed first. The cards remained visible on screen for approximately 500 ms and were then turned over. This was done to avoid as much as possible that the task would be solved by a superficial matching strategy. The colour of the third card was then revealed and the money was distributed to the winning player (in case both players won, the sum was shared between them). On a subset of trials, the game was paused just before the third card was turned over and participants were asked what colour the detective wanted the next card to be. The question directly probed the detective's desire (i.e., "*which colour does detective O'Reilly want the next card to be?"*). On half of the trials, the patient and the detective had the same colour card (same-perspective trials) at the time the question was asked; on the other half of the trials, the patient and the detective had cards with different colours (different-perspective trials). Response options were depicted on paper by showing the detective with three thought bubbles, each one showing a different coloured card (blue, red or yellow). Specifically, this represented the colour matching the opponent's card (correct on all trials), the colour matching the participants' card (correct on same-perspective trials only) and a neutral coloured card that was not in play. The game was split into two sessions, each including 6 desire questions where the perspectives were the same and 6 desire questions where the perspectives were different. To avoid as much as possible that the task would be solved by a superficial matching strategy, the card that each player was given was only displayed for a short period (controlled by the examiner) and was then turned over. So at the time the desire question was asked, the players' colour card was not visible anymore. Thirty-six filler rounds (18 in each session) in which no question was asked were added to the 24 test-rounds to ensure that the game flowed and to keep the task engaging.

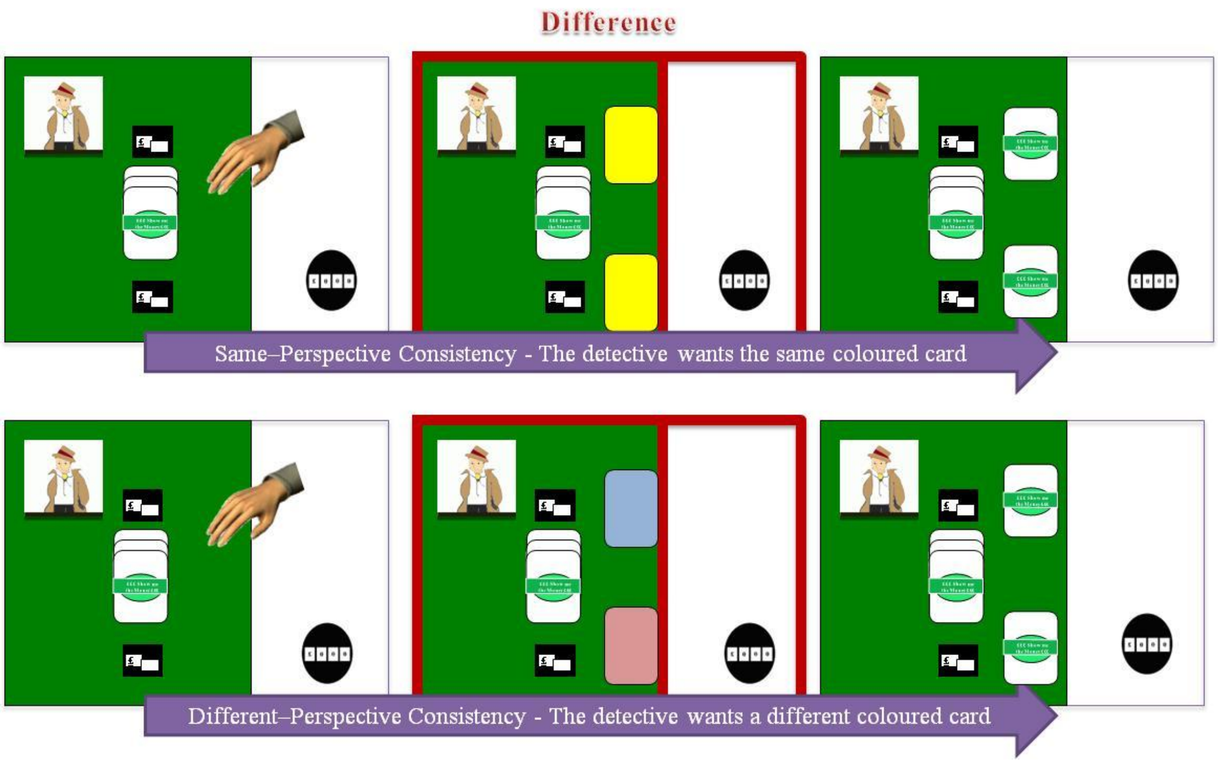



Figure 11. Schematic representation of the desire task for both same- and different-perspective trials.

### 2.2.3. Results and discussion

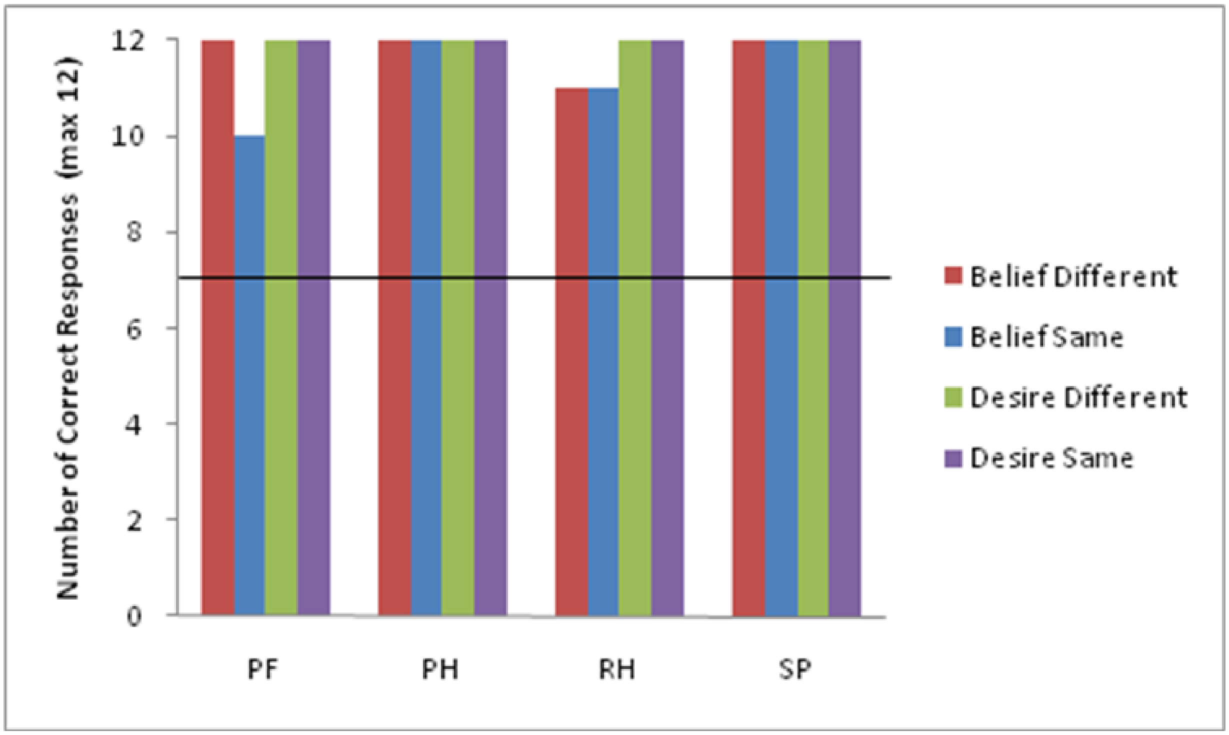
A score of 8 (or more) out of a maximum of 12 responses needed to be achieved for a participant to perform significantly above chance in each experimental condition (8/12 correct with a probability of success of 0.33 on each trial has a one-tailed probability of 0.02 by binomial test). Same-perspective trials were used as a control to ensure that the patients were correctly following the task (see Figure 12, for the overall results). The healthy normal adult controls all performed with 100 % accuracy across both tasks. These data are not included in the overall graph.

There were three distinct profiles of performance across the patients in the different-perspective consistency trials. The first profile, identical to healthy normal controls' performance, was presented by a group of four patients (PH, PF, RH and SP) who passed both experimental tasks, and both same- and different-perspective trials. It should be noted that, as in previous studies (e.g., Apperly et al., 2004, 2005), patient PH was mainly employed as a control patient. Patient PH is an aphasic patient (Broca aphasia, see Apperly et al., 2006, for details) who failed to show any belief reasoning deficit in previous studies despite severe language impairments. The fact that PH did not show any problems in the newly designed tasks indicates that the tasks did not place so many incidental task demands that a patient with a substantial lesion and severe aphasia was unable to deal with it. Noticeably, the TPJ patients performed very well on these tasks and their success in passing the false-belief task here contrasts with their difficulties documented in previous studies. I will discuss these results in the discussion at the end of this chapter.

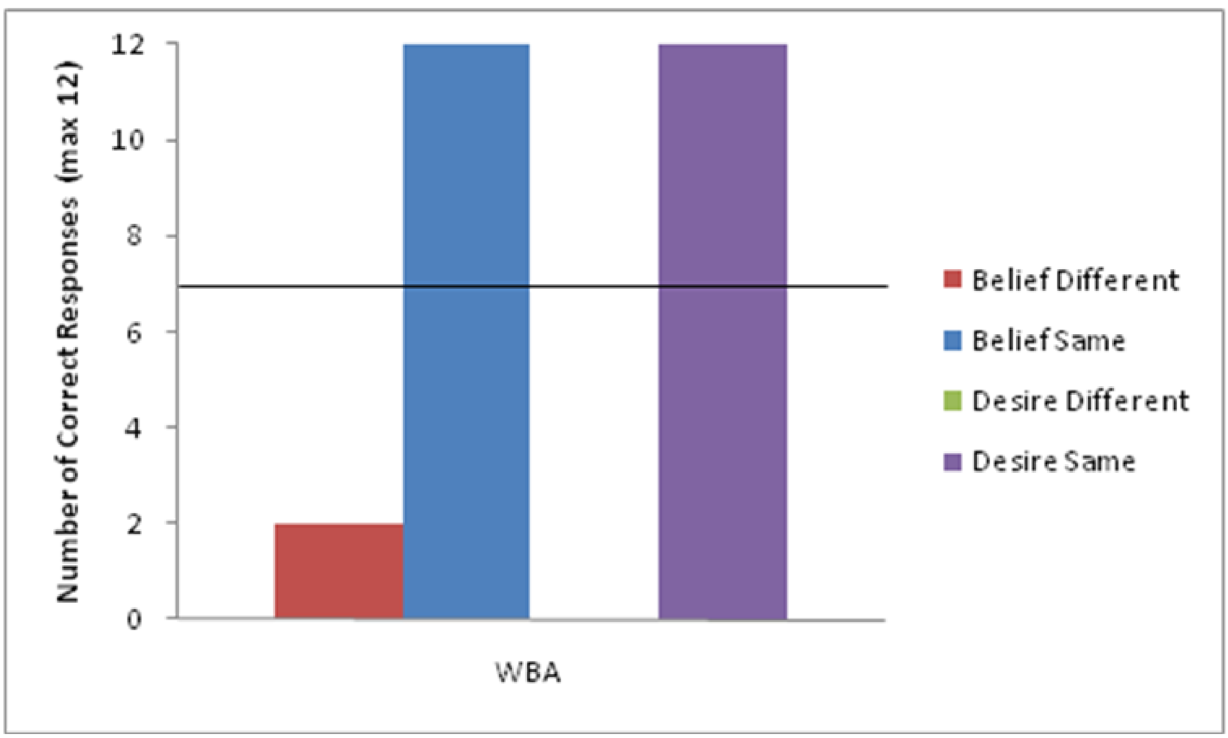
The second profile was presented by patient WBA, who was impaired in both the belief and desire task. WBA performed below chance on different-perspective trials and above chance on same-perspective trials. Importantly, this held across both mental states. The profile of WBA is consistent with previous findings and shows that his own self-perspective interferes with his reasoning for both mental states (i.e., Samson et al., 2005, 2007).

The third profile was shown by five patients (DS, GA, JeB, MP, and PW) who were unable to infer someone else's belief when it conflicted with their own knowledge of the reality but they were able to infer someone else's desire even when it conflicted with their own desire. This was observed despite the belief and desire tasks being matched for self-perspective inhibition demands. There could be several explanations to this finding. Firstly, it could be that the incidental demands were higher in the belief condition (e.g., there may have been more information to monitor). Secondly, it could be that what needs to be monitored was more obvious in the desire than the belief task (e.g., in the desire task, the context of the game could prompt attention to monitor the cards). Thirdly, it could be that the belief tasks placed higher inhibition demands (e.g., knowledge of reality being particularly salient). Fourthly, it could be that patients did not perform the desire condition as a perspective taking task but rather as a simple matching task. To solve the task, patients had simply to give as response the colour that matched the colour of the card that the detective had received.

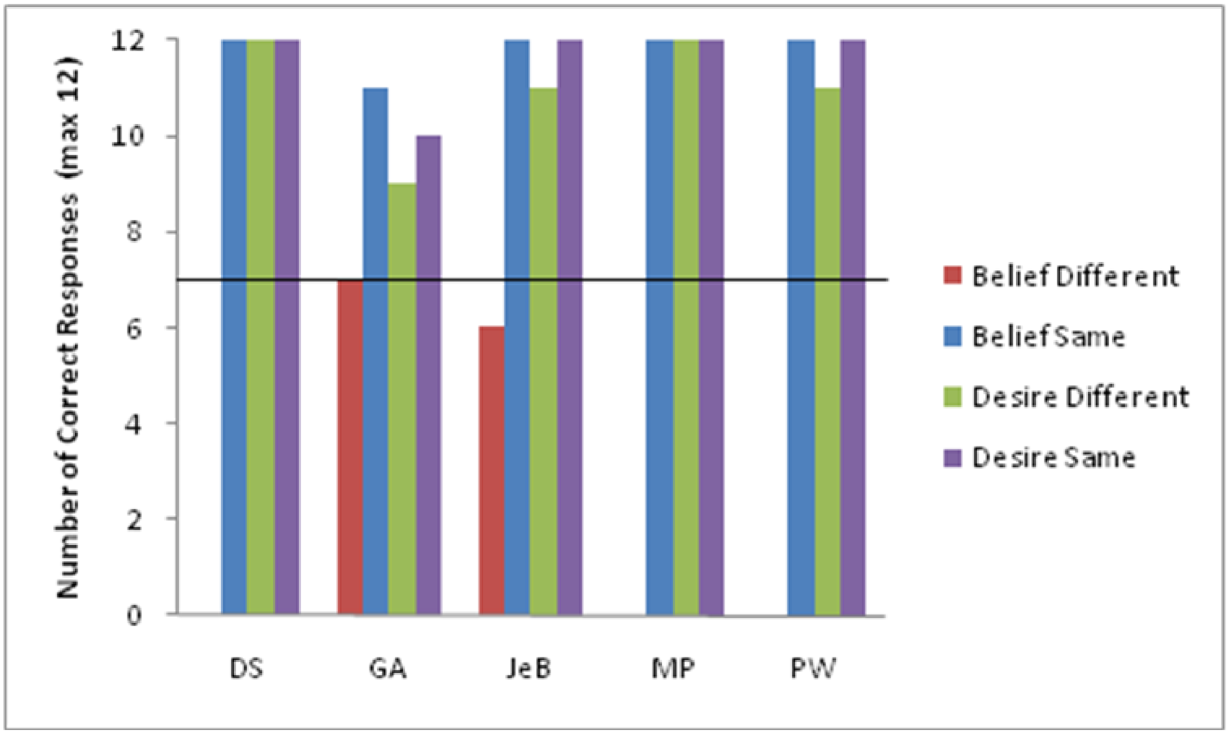
Experiment 2 was designed to test this last possibility. Experiment 2 was another desire reasoning task, but in this case a simple matching strategy was avoided by including avoidance in addition to approach-desire trials. For these avoidance trials, participants were told that, if the third colour card matched the coloured card they had in hand, they would lose a sum of money. Thus, on avoidance trials, when asked to infer which colour the other person would like the next card to be, participants needed to select any colour other than the colour card their opponent had in play. This should prevent a simple matching strategy from being adopted.




Profile 1- Intact belief and desire reasoning abilities




Profile 2 - Deficit in both mental states




Profile 3 - Deficit in belief reasoning only

Figure 12. Performance of ten patients (DS, GA, JeB, MP, PF, PH, PW, RH, SP, and WBA) on the different- and same-perspective trials, for each metal state respectively. The patients are grouped into three different profiles based on their global performance across Experiment 1. The horizontal line indicates the level above which a patient is scoring significantly above chance.

## 2.3 Experiment 2: Approach and avoidance desire

### 2.3.1 Method

#### *Participants*

All patients and controls who took part in Experiment 1 also took part here.

#### *Stimuli and procedure*

Participants were presented with two tasks in which they were asked to infer someone else's desire. Each task was presented within the context of a card game called "blame it on the weather" and participants played against a story character named Simon, displayed in short animated clips in Multimedia Flash 8 (see Figure 13). Similar to Experiment 1, in both tasks, on half of the trials both Simon and the participant had different perspectives (wanting something different). On the other half of the trials, both parties had the same perspective (wanting the same). The test question directly probed the character's desire (i.e., "W*hich colour does Simon want the next card to be?"*)

The first task merely replicated the rules and procedure used in Experiment 1, where participants merely had to match a coloured card in order to win a sum of money (win only game). This was to ensure that visual changes of the experiment would not affect a patient's performance before the new rule was introduced in the second task. The new rule incorporated the concept of loss, where on some trials the players should now want to avoid losing a sum of money and therefore would not want the third card to be of the same colour as the card they have (win and loss game). Thus, there were two desire types within the task, depicted as weather symbols on screen: either a sun symbol (approach trial cue) or a cloud symbol (avoidance trial cue). If a sun symbol was displayed, and the third card matched the colour the player had, then the participant would win the sum of money indicated on the screen. On these trials, a player should want this to occur in order to progress in the game.  In contrast, if a cloud symbol was shown and the third card matched the participant's colour card, then the participant would lose the sum of money displayed on the screen. In this case a player would not want this coloured card to match their own colour. Thus, a trial would start with the display of the weather counter, indicating whether the stake is to win or lose money. Each participant would then receive their coloured card. These would be visible for a short period of time. After that, the colour of the third deciding card was revealed. The card dealer would then hand in/remove money to/from the players depending on the outcome of that round. The scoreboard, displaying both Simon and the participant's name was positioned in the top right hand corner to help participants keep track of their overall gains compared to Simon's.

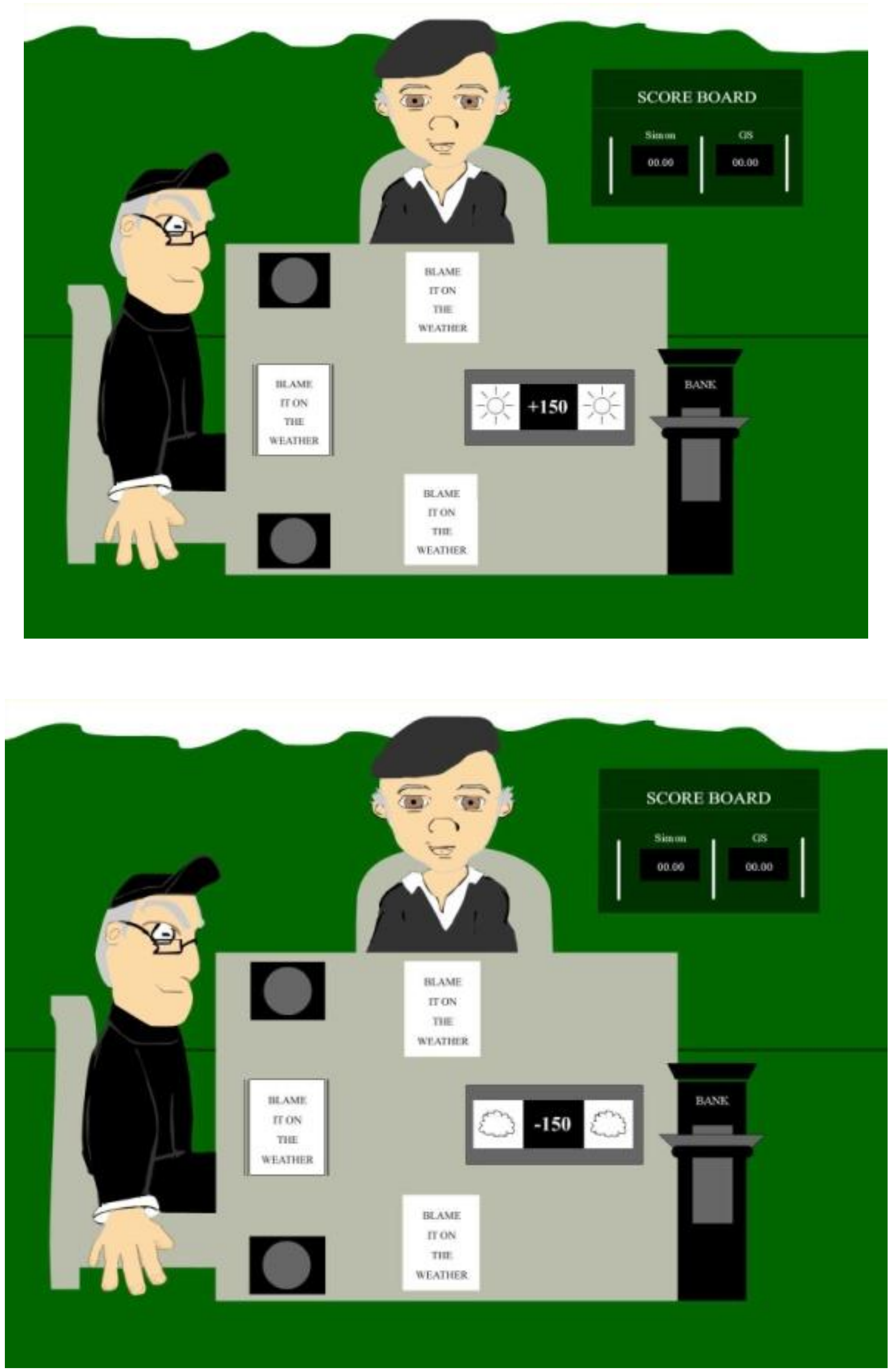



Figure 13. Visual setup of Experiment 2 with approach-desire trials (top half) depicted with a sun and avoidance-desire trials (bottom half) depicted with a cloud symbol.

Certain differences with Experiment 1 should be noted. First, in Experiment 2 the visual setting of the character made the character's presence more salient than in Experiment 1. The task demonstrated an animated character who was sitting at a table facing the participant. Although this character did not talk during the test phase, he did display eye motion toward the selected information that needed to be monitored at the time. This manipulation made the protagonist more salient than in Experiment 1, and was mainly done to enhance attention to the other person's perspective in the task. Secondly, no filler trials were included as previous testing revealed that patients answered after every trial without being prompted. Thus, on each trial, participants were asked the desire question. Finally, a third person was introduced to the task as a card dealer to raise the ecological validity of the task (as opposed to the floating hand used in Experiment 1). This person was always depicted on the left hand side of the screen and distributed the cards first to the patient then to Simon. The video was paused before the card dealer turned the third deciding card of that particular trial and patients were asked the test question before they were shown the identity of the third card.

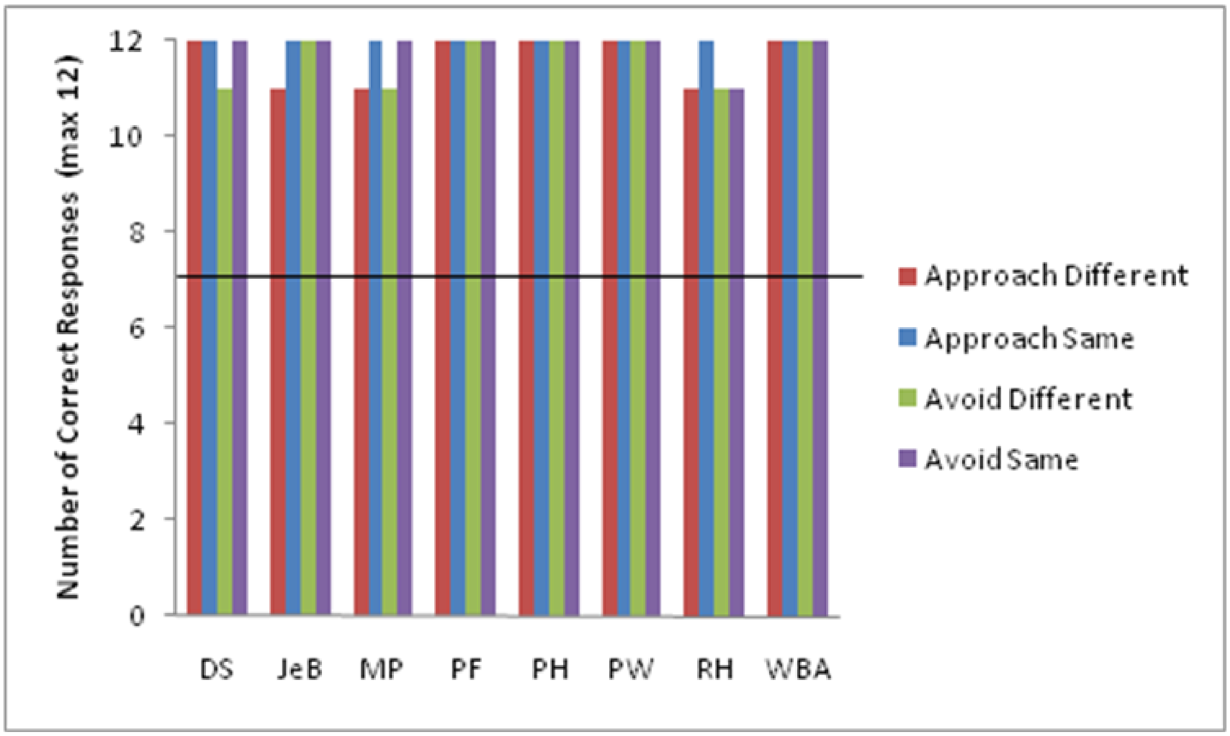
### *2.3.2 Results and discussion*

#### *Win only game*

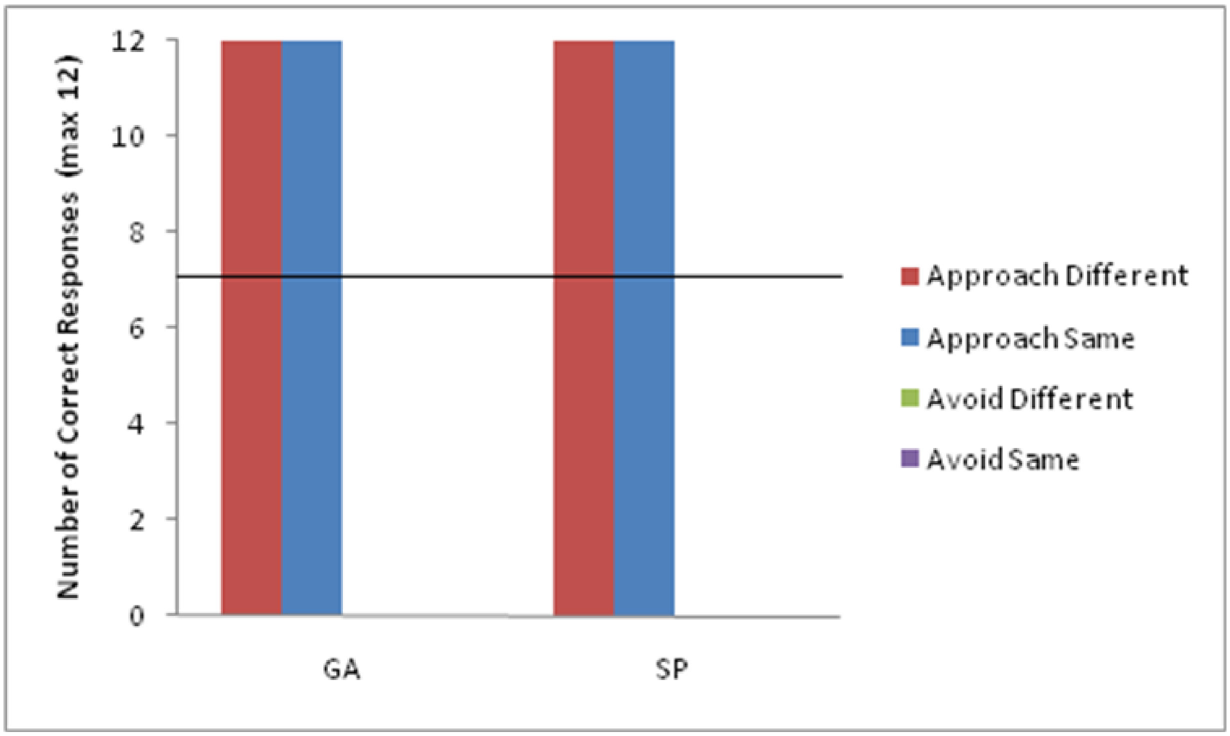
Healthy normal adult controls performed at ceiling (12/12) in the first task, and all the patients performed above chance (all performing at least 9/12 correct in each condition).

#### *Win and loss game*

The results are displayed in Figure 14. Eight of 10 patients (PF, PH, RH, WBA, DS, JeB, MP, and PW) performed above chance on both the approach- and avoidance-desire trials. All but one of these patients (WBA) also performed above chance in the desire reasoning of Experiment 1. The results of Experiment 2 show that they have no problems when the difficulty of the desire reasoning task is increased. PF, RH and PH also performed above chance on the false-belief and different-approach trials in Experiment 1. Thus, these patients seem to have no problem in belief and desire reasoning in the paradigms used here. DS, JeB, MP and PW had problems with false-belief trials and their good performance in all desire reasoning tasks suggests that their deficit may be specific to belief reasoning. This will be considered further in the discussion. More surprising is the pattern of performance of patient WBA, who failed the simpler desire reasoning task in Experiment 1, but passed the more complex desire reasoning task in Experiment 2. Thus, WBA seemed now to be able to inhibit his own perspective in this more complex desire reasoning task. One possible interpretation for this pattern of result is that by only asking him to judge the opponent's desire, WBA may have started to identify himself with the opponent and/or he may have started to appropriate himself the cards given to the opponent. I will review this in the discussion below. Two of the 12 patients (GA and SP) performed above chance level for approach-desire trials but did not perform above chance level for avoidance trials. For both patients, the problem was not confined to different-perspective trials. Thus, the deficit seemed to be unrelated to the self-perspective inhibition demands of the task and could reflect some difficulties with complex forms of desire reasoning per se. One of these 2 patients (GA) had also problems in the belief reasoning task in Experiment 1. His deficit may be thus common to both tasks and related to complex reasoning. Interestingly, SP performed quite well in the false-belief task of Experiment 1. His deficits may be more specific to reasoning about avoidance.




Profile 1 - No deficit




Profile 2 - Deficits in avoidance desires only

Figure 14. Performance of ten patients (DS, GA, JeB, MP, PF, PH, PW, RH, SP, and WBA) on the different- and same-perspective trials, for each desire type respectively. The patients are grouped into two different profiles based on their global performance across Experiment 2. The horizontal line indicates the level above which a patient is scoring significantly above chance.

## 2.4. General Discussion

The current study directly compared belief and desire reasoning abilities in brain injured patients controlling for the self-perspective inhibition demands across both reasoning tasks. Developmental studies suggest that the unbalanced demands of self-perspective inhibition across classic belief and desire reasoning tasks could explain why children often find it easier to reason about desires than beliefs (Moore et al., 1995; Rakoczy et al., 2007). It was thus important to match the belief and desire tasks in terms of self-perspective inhibition demands to reliably assess brain-damaged patients' ability to reason about both types of mental state.

Some of the patients tested had no difficulties with either belief or desire reasoning. One of these patients (PH) showed spared belief reasoning abilities in many other tasks despite an extensive lesion and severe language problems (Apperly et al., 2004, 2006). His success in the newly designed tasks thus shows that the tasks did not place unnecessary high incidental demands. The other two patients who performed well in the belief *and* desire reasoning tasks, PF and RH, were patients with lesions to the left TPJ who have previously been reported to have difficulties in belief reasoning, probably linked to impairments in selecting the relevant cues to infer someone else's belief (Samson et al., 2004, 2005). Here, however, the patients were able to judge the protagonist's perspective in both belief and desire reasoning tasks. The good performance on the belief reasoning tasks here might reflect the practice that the current patients had received on similar tasks (Apperly et al., 2004, 2008; Samson et al., 2005, 2007). Due to this practice, the patients could have become sensitized to the cues required to reason about the beliefs of people in the scenarios, and this could have overcome their problems in detecting relevant cues. Both patients performed well in the desire reasoning tasks despite the fact that they had not been presented with desire reasoning tasks before. In the desire reasoning tasks, the relevant cues to infer the other person's desire simply consisted in monitoring the cards on the table (whether it was a sun or cloud card as well as the coloured card the opponent received). Such cues may have been salient enough for the patients to process them at the onset, and it is therefore perhaps not surprising that the patients performed well in these desire reasoning tasks.

One patient, GA, showed difficulties in both the belief and the desire reasoning tasks particularly when the desire reasoning was better matched to the belief reasoning task in terms of reasoning complexity. Indeed, this patient could correctly infer what someone would want but they could not infer what someone else would not want. Interestingly, the problems in reasoning about avoidance occurred not only on different-perspective trials, but also on same-perspective trials. A deficit in self-perspective inhibition cannot explain such a pattern of performance, as it would have resulted in a good performance on same-perspective trials. Indeed, on same-perspective trials, reasoning about one's own avoidance desire would have led to the correct response. Thus, it seems that GA's problems observed for both belief and desire reasoning may have been unrelated to a self-perspective inhibition deficit and instead could be related to a deficit in complex reasoning.

Patient WBA was previously reported to have a deficit in self-perspective inhibition (Samson et al., 2005, 2007). His performance in Experiment 1 is consistent with this interpretation, as he made many errors in both the belief and the desire reasoning task, when inferring the other person's mental states, but only when that other person's mental state was different to his own mental state. Unexpectedly, WBA performed very well in the desire tasks in Experiment 2, where the self-perspective inhibition demands remained unchanged, but the general reasoning demands were increased. One possible explanation for this improvement is that through repeatedly asking to take the other person's perspective, without ever asking to make a self-perspective judgment, the patient might have taken on the role of Simon, ignoring his own cards and attributing Simon's cards to himself. In that case, using one's own new perspective would lead to the correct answer without any need of self-perspective inhibition.

The profile of the five remaining patients shows at first view a double dissociation. Patient SP performed well in the belief reasoning task (Experiment 1), but was impaired in the desire reasoning task (Experiment 2) that was matched to the belief reasoning task, in terms of both the self-perspective inhibition demands and the reasoning demands. In contrast, patients DS, JeB, MP and PW were impaired in the belief reasoning task (Experiment 1), but they were not impaired in any of the desire reasoning tasks (Experiments 1 and 2), even when the task required to reason about a complex desire such as an avoidance desire so that it could not be passed on the basis of a simple matching strategy, and even when the other person's desire was different to their own desire.

The latter dissociation (impaired belief reasoning but spared desire reasoning) was observed despite the belief and desire tasks being matched for self-perspective inhibition demands, and the effect remained even when the desire reasoning task was designed so that it could not be passed on the basis of a simple matching strategy (Experiment 2). There are several possible explanations for such a deficit and it is possible that the patients in this group did not all have the same underlying problem. Firstly, however, it seems that the interpretation that the patients could not deal with the incidental task demands of the belief reasoning task can be discounted. Indeed, all patients could infer the detective's belief on same-perspective trials (i.e., true-belief trials). Moreover, when making errors on the different-perspective trials (i.e., false-belief trials), they never chose the neutral location as response. This suggests that they were not responding at random and that they processed and understood sufficiently the sequence of events to know where the robber was currently located. At first hand, the dissociation suggests that the patients' deficit was unrelated to a self-perspective inhibition problem. However, it could be that, relative to the desire tasks, the belief task placed higher demands on self-perspective inhibition. I mentioned already in the case of WBA, that, in the desire reasoning task, patients may have taken on the role of Simon, ignoring their own cards. In the desire reasoning task, the cards were clearly spatially segregated and it would be easy to focus only on the opponent's card and ignore one's own card for the entire duration of the game (particularly because the patients were never probed about their own perspective). In the belief reasoning task, on the other hand, the actual location of the robber had always to be monitored in relation to whether the detective had seen it or not. In this case, it was thus mandatory to have knowledge about the location of the robber to process the detective's belief. Thus, whereas in the desire reasoning task, patients could get to the correct response by bypassing the need to inhibit their own perspective, it may have been much harder to do so in the belief reasoning task. Alternatively, it is possible that the patients failed the belief reasoning task because they have lost some conceptual knowledge about beliefs (without loss of conceptual knowledge about desires), or because they find it harder to monitor the relevant cues to infer someone else's belief than someone else's desire. For example, in the belief reasoning task, patients needed to monitor the gaze direction of the detective to work out whether the detective had or had not seen the robber moving. No such gaze processing was required in the desire reasoning task. The results reported here do not yet allow me to disambiguate these different interpretations, but they suggest that there are a multitude of reasons why belief reasoning could be harder than desire reasoning, and this is not necessarily linked to the self-perspective inhibition demands of the task.

The reverse dissociation (spared belief and impaired desire reasoning) displayed by SP suggests that there may also be processing demands, over and above inhibition of one's self-perspective, that may be uniquely linked to desire reasoning, particularly avoidance-desire reasoning. As reported in Chapter 1, in Leslie and collaborators' (1998) model of belief and desire reasoning, avoidance-desire reasoning requires inhibitory processes to shift attention away from the object to be avoided. Such inhibitory processes may be distinct (functionally and perhaps neurally) from the processes required to inhibit one's perspective. Consistent with this, it is important to note that SP made errors on same-perspective and different-perspective trials when reasoning about avoidance desires. If his problem was only with self-perspective inhibition and not with shifting attention away from the object to be avoided, then he should have shown a good performance on same-perspective trials, where reasoning about one's own avoidance desire would have led to the correct response.

To conclude, I found patients who showed an association of spared or impaired performance across belief and desire reasoning tasks, but I also found a double dissociation in the patients' performance across both types of tasks, with some patients having more difficulties with belief than desire reasoning, and other patients showing the reverse profile. This suggests that above the common demands in self-perspective inhibition in belief and desire reasoning there are also more unique processing associated with desires and beliefs. One particular processing worth investigating is the one required to reasoning about avoidance desires. Given that very little is known about what these processes are, I will first present in the next two chapters the results of a series of experiments conducted with healthy adults before addressing the issue again with brain-damaged patients in Chapter 5.

# CHAPTER 3

## Self-perspective inhibition and avoidance inference costs in desire reasoning:

## Evidence from healthy adults

### 3.1 Introduction

In the previous chapter, I highlighted two different aspects of desire reasoning that patients found difficult. Specifically, I showed that some patients had problems inhibiting their own perspective (e.g., patient WBA), whereas other patients had difficulties in reasoning about an avoidance desire compared to an approach desire (GA and SP). We know from previous research that even healthy adults find it difficult to inhibit their own perspective (Birch & Bloom, 2007; Epley et al., 2004; Keysar et al., 2003). However, to my knowledge, it has not been shown yet whether healthy adults also find it difficult to inhibit their own perspective in the context of desire reasoning specifically. Furthermore, there is no clear evidence to date showing that healthy adults find it difficult to deal with avoidance compared to approach desires.

One study that speaks to the relations between desire reasoning in approach vs. avoidance situations has been reported by German and Hehman (2006). In their task, young and older healthy adult participants read stories in order to predict a character's action on the basis of the character's belief about the situation (either true or false) and the character's desire (either approach or avoidance desire). The authors were only interested in the combination of belief and desire types (comparing four conditions: true-belief + approach-desire, true-belief + avoidance-desire, false-belief + approach-desire and false-belief + avoidance-desire) and did not report detailed statistics about the main effect of desire type. However, from the reported mean performance per condition, it appears that performance was quite similar for approach and avoidance desires, except for older adults who seemed to be slower for avoidance than approach desires when these desires were combined with a false-belief. It is worthwhile noting that in German and Hehman's task, the desires were directly expressed in a story and participants had to infer the character's action for example, "*she wants to avoidance some place noisy*" (German & Hehman, 2006; p. 135). Thus, the authors did not measure desire inference per se. Furthermore, there was no time pressure to give a response and the participants' proficiency might have been overestimated. The study here aimed to investigate whether there were two distinct sources of difficulties when healthy adults reason about desires: one linked to self-perspective inhibition, the other linked to the inference of an avoidance desire, similar to what we observed in brain-damaged patients.

Given that healthy adults are proficient in mentalising, I devised a new desire reasoning paradigm that measures reaction time and error rates when participants are required to respond under time pressure. The aim was to compare the participants' performance when they reasoned about someone else's desire compared to their own desire (a measure a self-perspective inhibition cost) and to compare performance when participants reasoned about an avoidance desire compared to an approach desire (a measure of avoidance cost). The paradigm was framed as a card game called 'blame it on the weather', similar to the task that was used with the patients in Chapter 2. Participants were told that the aim of the game was for players to try to have the highest amount of money at the end. In one version of the task, participants simply watched two players playing against each other (passive-involvement condition). In another version of the task, participants were one of the players and played against an opponent (active-involvement condition). Each round started with the display of a symbol that would indicate whether the outcome of that particular round was to win or lose a specified amount of money. Each player would then get a coloured card from the deck and the colour of the third card would determine who wins or loses the specified amount. For each round, participants were asked to judge the desire of one of the players (their own desire or another player's desire) before the third card was turned over. In the win condition, any player would want the third colour to be of the same colour as their own card to win the money (approach-desire). In the lose condition, any player would want the third colour not to be of the same colour as their own card in order to avoid losing money (avoidance-desire). Furthermore, on any round, the two players could either have the same coloured card (and would thus have the same desires) or they could have different coloured cards (and would thus have different desires). The game was presented in three separate experiments. Experiments 3 and 4 directly manipulated participants' involvement in the task. In Experiment 3, participants started by the passive-involvement condition and were then presented with the active-involvement condition. In Experiment 4, the order of the two conditions was reversed. Experiment 5 only included the active-involvement condition, but the salience of the self-perspective was enhanced by using pictorial representations of the actual participants during the game.

The factors manipulated in this study were the level of the participant's involvement (active versus passive, in Experiments 3 and 4 only), the desire type (avoidance versus approach), the perspective judged (in the active condition: self versus other; in the passive condition: player 1 versus player 2) and the consistency between perspectives (conflicting perspectives versus same perspectives). If reasoning about avoidance desires is more difficult than reasoning about approach desires, as suggested by the data in Chapter 2, then I should expect to find in healthy adults a significant effect of the desire type (avoidance versus approach) in participants' reaction times and/or their accuracy. If this is indeed the case, then this would further suggest that the experimental conditions used by German and Hehman's (2006) were not stringent enough to measure the avoidance processing cost. If reasoning about someone else's desire is more difficult than reasoning about one's own desire, then I expect to find a significant effect of perspective in the active-involvement condition. Such an effect should increase as the salience of the self-perspective is increased (Experiment 5). The passive-involvement condition acts as a baseline (i) to ascertain that participants do indeed project themselves in the game and (ii) to measure the general level of difficulty when participants had to deal with two perspectives even when their own perspective was not relevant to the task.

An effect of consistency between perspectives was also expected. Indeed, in the conflicting-perspective condition both players had different coloured cards whereas in the same-perspective condition they had the same coloured card. Based on the fact that processing two cards of different colour may be more confusing than processing two cards of the same colour, I expected to find worse performance on the conflicting-perspective trials. Furthermore, based on the assumption that ignoring one's own card is harder than ignore someone else's card, the consistency of perspectives may have a stronger effect when participants have to process someone else's perspective. Finally, by manipulating the order of presentation of the active- and passive-involvement conditions across Experiment 3 and 4, I assessed whether there were potential carry over effects across the two conditions.

## 3. 2. Experiment 3

### 3. 2.1 Method

#### *Participants*

Twenty students (3 male, $M_{age}$ = 20 years) from the University of Birmingham School of Psychology took part in this study in exchange for course credit. All participants had normal or corrected vision and none were colour blind.

### ***Stimuli and procedure***

Stimuli were presented on a standard desktop using Eprime (Version 2.0; Psychology Software Tools, 1996). The win or lose symbol was presented as a weather counter subtended 2.8 by 1.2 degrees. It displayed either a sun (for a win trial) or a cloud symbol (for a lose trial) as well as a money sum (between £1 and £200). The colour cards subtended 0.5 by 0.7 degrees and were either red or blue. The cards were displayed on a table (subtending 4.1 by 2.9 degrees) situated at the centre of the screen. Two chess pawn pieces (green, purple) were employed as symbolic representation of the players on the screen (the pawn subtended 1.4 by 1.9 degrees). For half the trials the green pawn was presented on the left hand side of the table and for the other half the pawn was presented on the right.

The passive- and active-involvement condition consisted of 192 trials (i.e., game rounds) each. In the passive-involvement condition (first half of the experiment), participants were told that two male players would be playing the game and that the green pawn represented Henry whereas the purple represented, Paul. In the active-involvement condition (second half of the experiment), they were told they were now represented by the purple pawn and that they would be playing against Henry represented by the green pawn. In both conditions, half of the trials were approach-desire (sun cue) trials and half were avoidance-desire (cloud cue) trials. Furthermore, for both approach- and avoidance-desire trials, on half of the trials, both the protagonist and the participant had different coloured cards, thus each would want a different coloured card as the third card (different-perspective trials). On the other half of the trials, both parties had the same coloured card, and thus both would want the same coloured card as the third card (same-perspective trials). On each trial, a cue indicated which perspective participants had to judge ("PURPLE WANTS" or "GREEN WANTS"). A colour card (either blue or red) was then presented, and participants were asked whether yes or no the colour matched what either they or the other person wanted the third card to be. The number of green versus purple perspective trials and matching versus mismatching trials was counterbalanced across desire types (approach versus avoidance) and perspective consistency (same versus different). Finally, the amount of money at stake on each round was either small (£1, £5 or £10) or large (£100, £150, £200) and was counterbalanced across all experimental conditions. All trials were divided into 4 blocks of 48 trials each, preceded by a practice block of 20 trials. Trials were presented in pseudo-random order to avoid more than 3 consecutive trials of the same type.

Participants were seated at a comfortable viewing distance of approximately 60 cm from the screen. Each trial began with a fixation cross that remained on screen for 1000 ms followed by a 1000 ms desire type cue (approach-desire, sun symbol; avoidance-desire, cloud symbol) indicating whether it was a win or loose trial as well as the amount of money at stake. After this cue, the participant would then see two colour cards simultaneously presented in the centre of the screen on the table. Participants were told during the instructions that the card closest to the pawn figure was that player's coloured card. The two cards remained on the screen for 1000 ms and were followed by a 1000 ms perspective cue that indicated which perspective the participant needed to judge (text stating: "GREEN WANTS" or "PURPLE WANTS"). Within the first block, participants were specifically told in the instructions to respond to the following questions, respectively: "Which colour does Henry (green pawn) want the next card to be" or "Which colour does Paul (purple pawn) want the next card to be?" Within the second block Paul's perspective was replaced by the participant's perspective, "Which colour do you (purple pawn) want the next card to be?" Once the perspective cue disappeared, a single coloured card (red or blue) would appear, and participants were asked to press one of two keys ("1" with the index finger or "4" with the middle finger) of the numeric pad of the keyboard to indicate whether the card shown matched ( "1") or mismatched (

“4”) the person’s desire. The coloured card remained visible for maximum 2000 ms or disappeared upon response, following which participants were given immediate feedback and shown a final frame depicting the actual identity of the third card and hence the outcome of the round (see Figure 15). This third coloured card (blue, red) was accompanied with a text stating the winner/loser of that particular round (e.g., “PURPLE WINS”) before a new trial commenced.

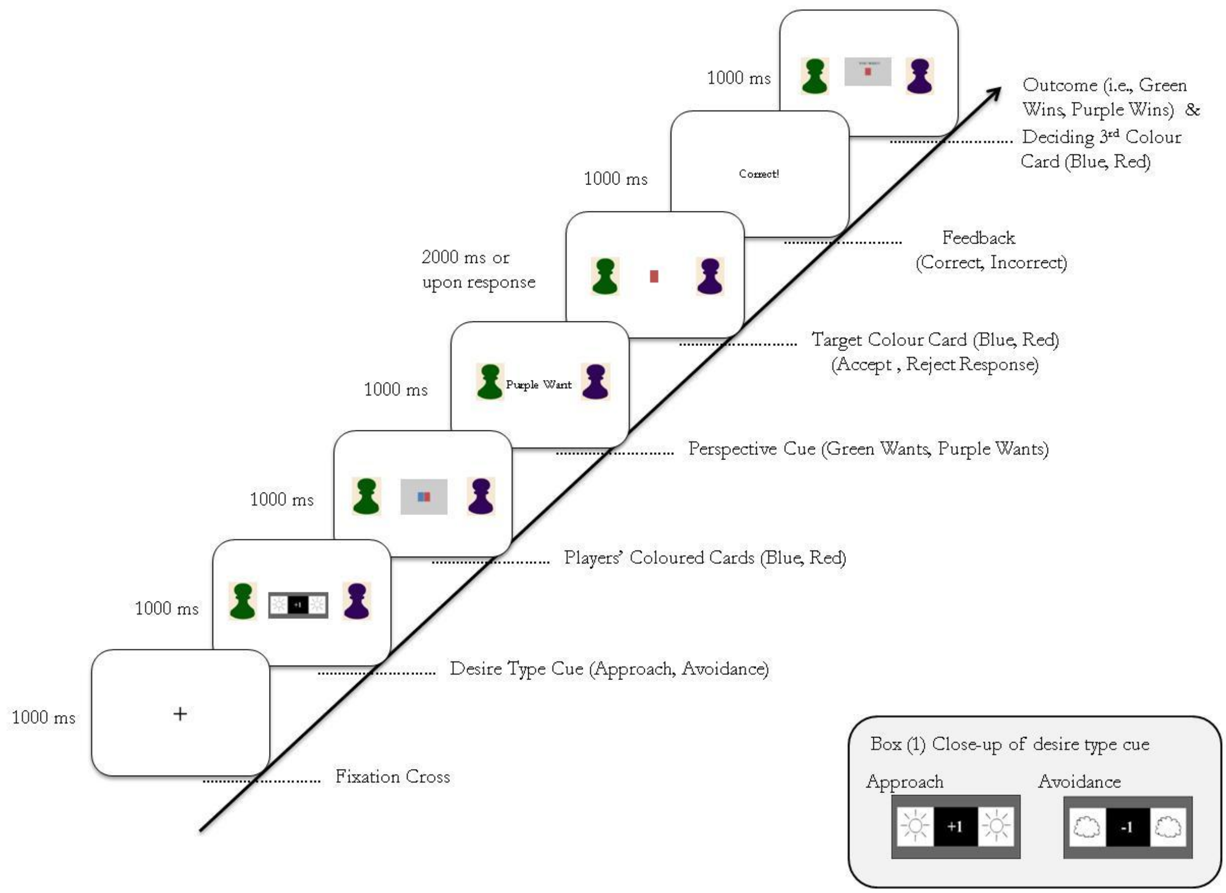



Figure 15. Experimental procedure displaying the sequence of events presented in order as shown on screen in Experiment 3 and 4. Box (1) contains a close-up of both desire type cues (approach, avoidance).

### 3.2.2. Results

Two participants were excluded from the study because their high error rate (accuracy range: 56–59%) meant that they may not have complied with the task instructions and/or that too many data points would be excluded from the RT analyses. These participants were replaced by new participants to keep the total number to 20. The RT data was then explored to see whether it was normally distributed. The skew and kurtosis values were within tolerable ranges to use parametric tests without transforming the data or removing outliers (skew: 1.28; kurtosis: 1.27).

A 2x2x2x2 repeated-measure ANOVA analysis was then carried out, with desire type (approach, avoidance), perspective consistency (different, same), perspective judged (purple, green) and involvement (passive, active) as independent variables. The dependent measures were reaction time (RT) to correct responses and accuracy (the proportion of correct responses). Only the data of matching trials (i.e., "yes" responses) was taken into account. For the RT analysis, responses omitted due to the timeout procedure (1 % of the data) and erroneous responses (9 % of the data) were eliminated from the data set.

#### *Reaction time (RT) analysis*

The four-way ANOVA revealed a significant main effect of desire type ($F(1,19) = 45.07$, $p < .001$, partial $\eta^2 = 0.70$), such that RTs were slower when participants judged an avoidance desire (742 ms) than an approach desire (617 ms). There was a significant main effect of perspective consistency ($F(1,19) = 21.99$, $p < .001$, partial $\eta^2 = 0.54$), with participants being slower at judging trials with different perspectives (729 ms) compared to the trials where the two perspectives were the same (629 ms). The main effect of perspective judged was also significant ($F(1,19) = 5.84$, $p = 0.03$, partial $\eta^2 = 0.24$), and showed that participants were quicker at judging the purple perspective

(669 ms) than the green perspective (690 ms). Finally, there was a significant main effect of involvement ($F(1,19) = 8.51$, $p < .010$, partial $\eta^2 = 0.31$), which showed that participants were slower when watching two people play (708 ms) compared to when they were involved as a player (650 ms).

There were three significant two-way interaction effects. First, there was a significant Desire Type x Involvement interaction ($F(1,19) = 6.63$, $p = 0.02$, partial $\eta^2 = 0.26$; see Figure 16). Paired *t*-tests revealed a significant desire type effect in the passive-involvement condition ($t(19) = 6.59$, $p < .001$; with a 150 ms disadvantage for the avoidance-desire condition) as well as in the active-involvement condition ($t(19) = 5.36$, $p < .001$; with a 102 ms disadvantage for the avoidance-desire condition). Numerically, the avoidance cost was higher in the passive-involvement condition. The effect of involvement was not significant for the approach-desire trials ($t(19) = 1.86$, $p = 0.08$) but significant in the avoidance-desire trials ($t(19) = 3.26$, $p < .005$; with an 82 ms advantage for the active-involvement condition).

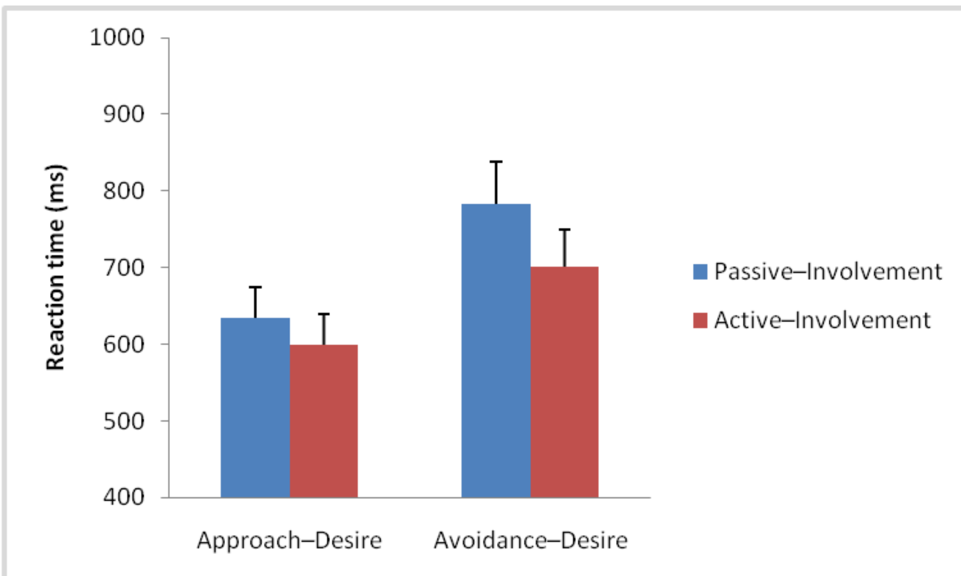



Figure 16. Overall mean (and standard error) for RT of Desire Type x Involvement interaction.

Second, there was a significant Perspective Consistency x Involvement interaction ($F(1,19) = 11.76$, $p < .005$, partial $\eta^2 = 0.38$; see Figure 17). Paired *t*-tests revealed a significant perspective consistency effect in both the passive-involvement condition ($t(19) = 5.56$, $p < .001$; with a 141 ms disadvantage when the perspectives were different) and the active-involvement condition ($t(19) = 2.52$, $p = 0.02$; with a 59 ms disadvantage when the perspectives were different). Numerically, the conflict cost was higher in the passive condition. The effect of involvement was significant when the perspectives were different ($t(19) = 3.81$, $p < .005$; with a 99 ms advantage in the active-involvement block) but not when the perspectives were the same ($t(19) = .86$, $p = 0.40$).

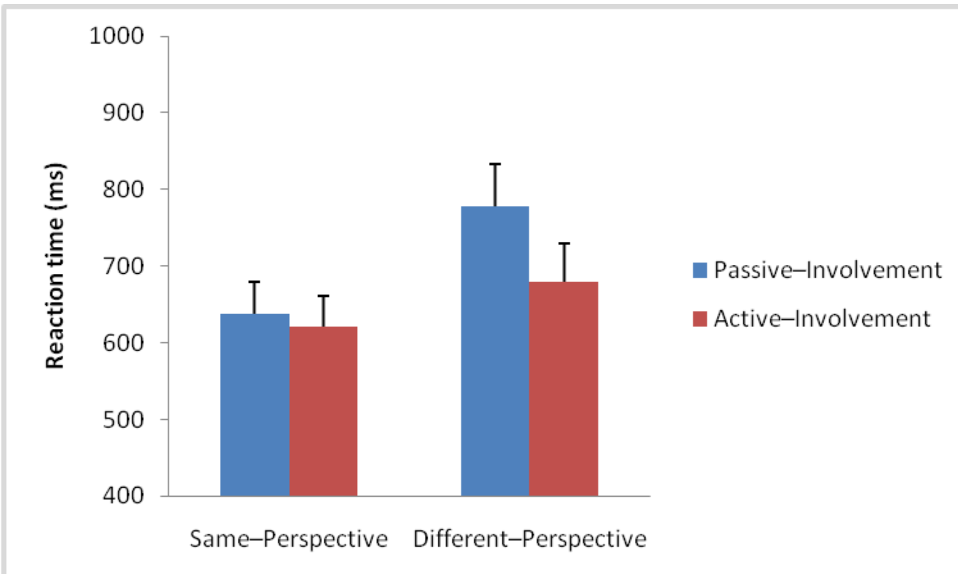



Figure 17. Overall mean (and standard error) for RT of Perspective Consistency x Involvement interaction.

Finally, there was a significant Perspective x Involvement interaction ($F(1,19) = 8.89$, $p < .010$, partial $\eta^2 = 0.32$; see Figure 18.). Paired *t*-tests revealed no significant perspective effect in the passive-involvement condition ($t(19) = 0.44$, $p = 0.67$). However, the effect was significant in the active-involvement condition ($t(19) = 4.26$, $p < .001$; with a 49 ms advantage for self, purple pawn). The effect of involvement was significant for the purple trials only ($t(19) = 3.97$, $p < .005$; with a 85 ms disadvantage for passive condition). No significant effect of involvement was found on the green trials ($t(19) = 1.38$, $p = 0.19$).

No other effects were significant (all *F*s < 3.66, all *p*s > 0.07; see Table 3).

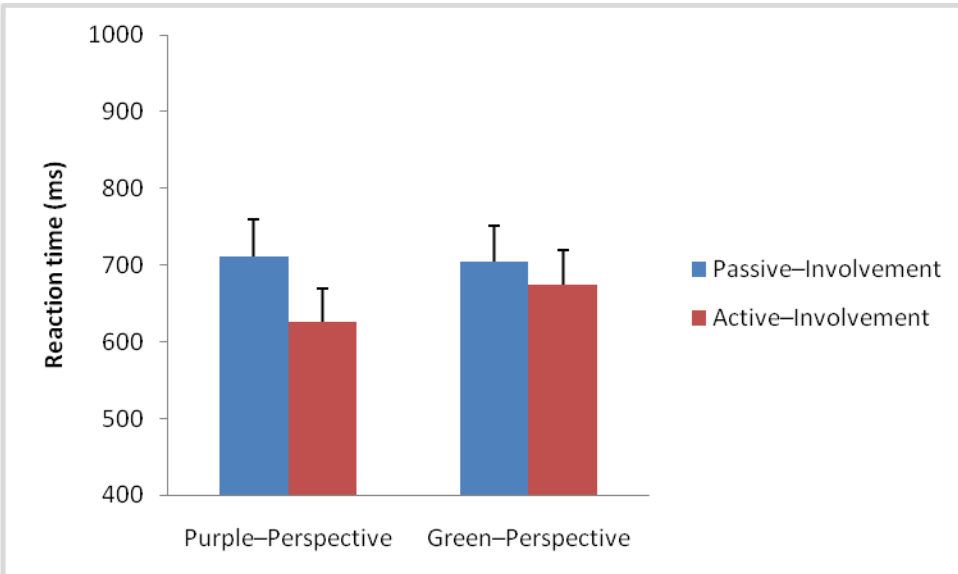



Figure 18. Overall mean (and standard error) for RT of Perspective x Involvement interaction.

Table 3. Overall mean for RT of Experiment 3 across all experimental conditions.

| | **Passive Involvement** | | | | **Active Involvement** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Same** | | **Different** | | **Same** | | **Different** | |
| | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** |
| **Approach** | 566 ms | 584 ms | 700 ms | 684 ms | 564 ms | 612 ms | 601 ms | 621 ms |
| **Avoidance** | 719 ms | 682 ms | 860 ms | 871 ms | 635 ms | 673 ms | 704 ms | 793 ms |

### *Accuracy analysis*

A similar ANOVA on the accuracy data revealed a significant main effect of desire type ($F(1,19) = 12.95$, $p < .005$, partial $\eta^2 = 0.41$); participants were less error prone on the approach-desire trials (92 % correct) relative to the avoidance-desire trials (87 % correct). There was a significant main effect of perspective consistency ($F(1,19) = 16.97$, $p < .005$, partial $\eta^2 = 0.47$), with an advantage for trials were the perspectives were the same (93 % correct) compared to the trials where the perspectives were different (87 % correct). No other significant main effects were found (perspective effect: $F(1,19) = 2.34$, $p = 0.14$, partial $\eta^2 = 0.11$; involvement effect: $F(1,19) = 0.48$, $p = 0.50$, partial $\eta^2 = 0.03$).

No other effects were significant (all *F*s < 3.65, all *p*s > 0.07; see Table 4 for overall mean).

Table 4. Overall mean for accuracy of Experiment 3 across all experimental conditions.

| | **Passive Involvement** | | | | **Active Involvement** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Same** | | **Different** | | **Same** | | **Different** | |
| | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** |
| **Approach** | 0.98 | 0.96 | 0.89 | 0.85 | 0.97 | 0.96 | 0.91 | 0.87 |
| **Avoidance** | 0.91 | 0.86 | 0.84 | 0.85 | 0.89 | 0.89 | 0.86 | 0.85 |

### 3.2.3 Discussion

Firstly, and as predicted, the analysis showed that participants were significantly slower at judging avoidance-desire trials then approach-desire trials, with no speed-accuracy trade-off. This avoidance processing cost was observed in the RTs. The effect was smaller in the active-involvement condition which was presented in the second half of the task. The reduction of the avoidance cost could thus be a result of (i) a familiarization with the task in the second half of the experiment and/or (ii) the involvement of the self per se that may have reduced the avoidance processing cost. Experiment 4 will help disambiguate these two interpretations: if the same effects are observed when the order of the passive- and active-involvement conditions is reversed, then the familiarization explanation can be discounted.

Secondly, the RT results showed no significant perspective effect in the passive-involvement condition. Thus, participants were as fast to judge the purple pawn's perspective as the green pawn's perspective when they themselves were not involved in the game. However, telling participants that they were involved in the task was sufficient to create a significant effect of perspective even though participants saw the exact same pawns as in the passive-involvement condition. Note that participants were as accurate irrespective of which perspective they had to judge. There was thus no evidence for a speed/accuracy trade-off.

Thirdly, the main effect of perspective consistency showed that participants took longer and were more error-prone on to trials where the perspectives were different (i.e., two colours for the cards) than the same (one colour for both card); again no speed-accuracy trade-off was found. For the RTs, the cost of conflicting perspectives decreased in the active-involvement condition (and this was not due to a speed-accuracy trade-off). This is the opposite of what we expected. Perhaps through the familiarization with the task (since the active-involvement condition was presented in the second half of the task) participants may have found a quicker way (or a strategy) to process the two different colours. Experiment 4 will directly test this possibility: if the same effects are observed when the order of the passive- and active-involvement conditions is reversed, then I can again rule out the familiarization explanation.

## 3.3 Experiment 4

### 3.3.1 Method

#### *Participants*

Twenty students (2 male, $M_{age}$ = 21 years) from the University of Birmingham School of Psychology took part in this study in exchange for course credit. All participants had normal to corrected vision and none were colour blind.

#### *Stimuli and procedure*

The method was identical to Experiment 3 in all respects (see Figure 15), except that the order of involvement was reversed. Thus, in the first half of the experiment participants were actively involved in the game and were identified by the purple pawn and played against a male character Harry (green pawn). In the second half of the experiment, participants were told that they would now simply watch two players play the game (passive-involvement condition; Harry being represented by the green pawn and Paul by the purple pawn.).

### 3.3.2. Results

Five participants were excluded from the study because of their high error rate (accuracy range: 55–74%) meant that they may not have complied with the task instructions and/or that too many data points would be excluded from the RT analyses. These participants were replaced by new ones. The skew and kurtosis values for the RT data were within tolerable range for assuming that the data was normally distributed (skew: 1.18; kurtosis: 1.24).

I conducted the same ANOVA analyses as in Experiment 3. Again, only the data of matching trials (i.e., "yes" responses) was taken into account. For RT analysis, responses omitted due to the timeout procedure (1% of the data) and erroneous responses (10% of the data) were not taken into account.

#### Reaction time (RT) analysis

The four-way ANOVA revealed a significant main effect of desire type ($F(1,19) = 21.92$, $p < .001$, partial $\eta^2 = 0.54$); RTs were slower when participants judged an avoidance desire (748 ms) compared with an approach (631 ms) desire. There was a significant main effect of perspective consistency ($F(1,19) = 46.58$, $p < .001$, partial $\eta^2 = 0.71$) with participants being slower at judging trials where the perspectives were different (744 ms) compared to trials where the perspectives were the same (636 ms). The main effect of perspective was significant ($F(1,19) = 10.64$, $p < .005$, partial $\eta^2 = 0.36$) and showed that participants were slower at judging the green pawn's perspective (712 ms) than the purple pawn's perspective (667 ms). Finally, there was a significant main effect of involvement ($F(1,19) = 16.14$, $p < .005$, partial $\eta^2 = 0.46$); participants were slower in the active-involvement condition (738 ms) compared to the passive-involvement condition (641 ms).

Overall, there was a significant Desire Type x Perspective x Involvement interaction ($F(1,19) = 8.84$, $p < .010$, partial $\eta^2 = 0.32$). The interaction was assessed by evaluating the effects of desire type and perspective judged within each involvement condition.

In the active-involvement condition (first block, see Figure 19), paired sample *t*-tests revealed a significant 161 ms avoidance cost when participants judged the purple pawn's (i.e., their own) perspective ($t(19) = 5.29$, $p < .001$) and a significant 92 ms avoidance cost when participants judged the green pawn's perspective ($t(19) = 2.92$, $p < .010$). Participants were also significantly slower at judging the green than the purple (self) pawn when judging approach desires ($t(19) = 3.51$, $p < .005$; with a 79 ms advantage for purple trials). However, for avoidance desires, participants were equally quick at judging the green and purple (self) pawn perspectives ($t(19) = 0.44$, $p = 0.66$).

In the passive-involvement condition (second block, see Figure 20), there was a 102 ms avoidance cost when participants judged the purple pawn's (i.e., their own) perspective ($t(19) = 4.05$, $p < .005$), and there was a significant 114 ms avoidance cost when they judged the green pawn's perspective ($t(19) = 4.03$, $p < .005$). Also, there was a significant 39 ms advantage when judging the purple pawn (Paul)'s perspective on approach-desire trials ($t(19) = 2.54$, $p = 0.02$), and a significant 52 ms advantage when judging the purple pawn (Paul)'s perspective on avoidance-desire trials ($t(19) = 2.21$, $p = 0.04$).

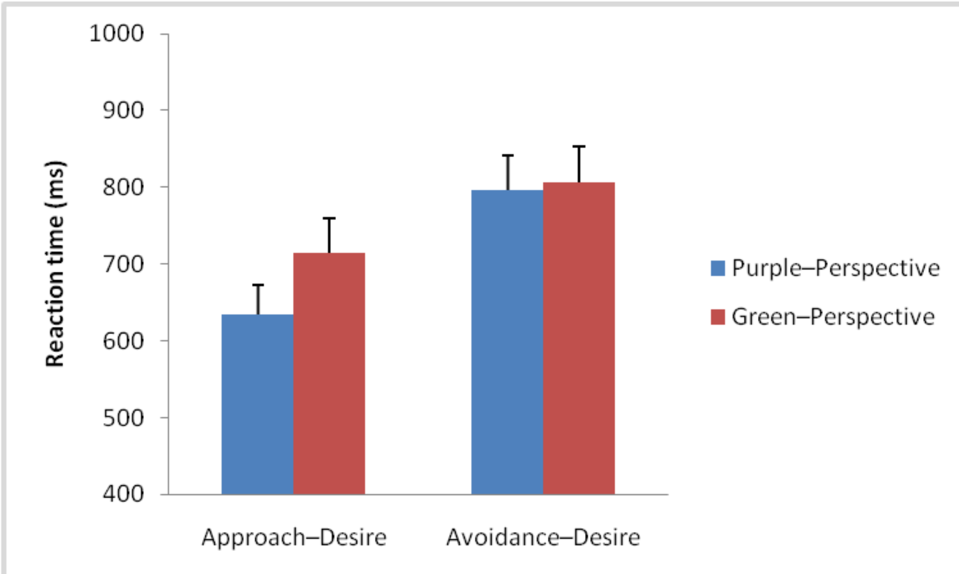



Figure 19. Overall mean (and standard error) RT of the Desire Type x Perspective interaction for the active-involvement condition.

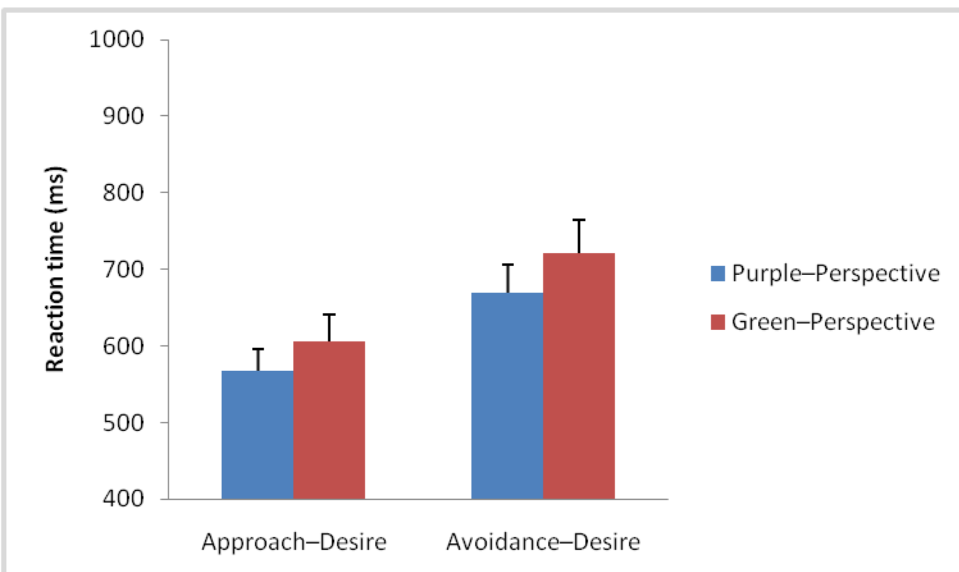



Figure 20. Overall mean (and standard error) RT of the Desire Type x Perspective interaction for the passive-involvement condition.

Separate paired sample *t*-tests were carried out to assess the effects of involvement at each level of the other factors. When judging approach-desire trials, participants were slower in the active- than the passive-involvement condition, both when judging the purple pawn's perspective (second block; $t(19) = 2.40$, $p = 0.03$; with a 68 ms advantage for the passive trials) and when judging the green pawn's perspective ($t(19) = 3.50$, $p < .005$; with a 108 ms advantage for passive trials). When judging avoidance-desire trials, participants were also significantly slower in the active- than passive-involvement trials both, when judging the purple pawn's perspective (first block; $t(19) = 4.73$, $p < .001$; with a 127 ms advantage for passive-involvement trials) and the green pawn's perspective ($t(19) = 2.67$, $p = 0.02$; with a 85 ms advantage for passive-involvement trials). These results show a global enhancement of participants' performance, from the active-involvement trials (first block) to the passive-involvement trials (second block).

No other effects were significant (all *F*s < 3.46, all *p*s > 0.09; see Table 5 for overall means).

Table 5. Overall mean for RT of Experiment 4 across all experimental conditions.

| | **Active Involvement** | | | | **Passive Involvement** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Same** | | **Different** | | **Same** | | **Different** | |
| | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** |
| **Approach** | 586 ms | 653 ms | 685 ms | 776 ms | 531 ms | 576 ms | 604 ms | 638 ms |
| **Avoidance** | 734 ms | 732 ms | 859 ms | 879 ms | 618 ms | 656 ms | 721 ms | 786 ms |

### ***Accuracy analysis***

Here, the ANOVA analysis revealed a significant main effect of desire type ($F(1,19) = 9.32$, $p < .010$, partial $\eta^2 = 0.33$), where participants were more error prone on the avoidance-desire trials (86 % correct) relative to the approach-desire trials (93 % correct). There was a significant main effect of perspective consistency ($F(1,19) = 23.92$, $p < .001$, partial $\eta^2 = 0.56$), with an advantage for trials where the perspective was the same (93 % correct) compared to trials where the perspectives were different (87 % correct). No other significant main effects were found (perspective judged: $F(1,19) = 1.05$, $p = 0.32$, partial $\eta^2 = 0.05$; involvement: $F(1,19) = 3.69$, $p = 0.07$, partial $\eta^2 = 0.16$).

There were two significant two-way interactions, Desire Type x Involvement ($F(1,19) = 6.23$, $p = 0.02$, partial $\eta^2 = 0.25$) and Perspective Consistency x Involvement ($F(1,19) = 5.62$, $p = 0.03$, partial $\eta^2 = 0.23$). These further qualified to a significant Desire Type x Perspective Consistency x Involvement interaction ($F(1,19) = 4.76$, $p = 0.04$, partial $\eta^2 = 0.20$). First, the three-way interaction was assessed by evaluating the effects of Desire Type x Perspective Consistency within each involvement condition.

In the active-involvement condition (first block, see Figure 21), paired sample *t*-tests revealed a significant desire type effect for both, trials where the perspective were different ($t(19) = 3.34$, $p < .005$; with a 11 % advantage for approach trials) and trials where the perspectives were the same ($t(19) = 3.24$, $p < .001$; with a 8 % advantage for approach trials). Numerically, the advantage for approach trials was larger when perspectives were the same. The effect of perspective consistency was significant for both approach-desire trials ($t(19) = 2.63$, $p = 0.02$; with a 7 % advantage for trials with same perspective), and avoidance-desire trials ($t(19) = 6.96$, $p < .001$; with a

10 % advantage for trials with same perspective). Numerically, the advantage for same-perspective trials was larger on avoidance trials.

In the passive-involvement condition (second block, see Figure 22), paired *t*-tests revealed a significant desire type effect for trials where the perspectives were the same ($t(19) = 3.32$, $p < .005$; with a 7 % advantage for approach trials) but not for trials where the perspective were different ($t(19) = 0.50$, $p = 0.62$). The effect of perspective consistency was significant for approach-desire trials ($t(19) = 2.63$, $p < .010$; with a 7 % advantage for trials with same perspective) but not for avoidance-desire trials ($t(19) = 0.36$, $p = 0.72$).

Separate paired sample *t*-tests were carried out to assess the effects of involvement at each level of the other factors. When judging approach-desire trials, no significant difference between active- and passive-involvement could be reported both, when participants viewed different ($t(19) = 0.26$, $p = 0.80$) and same ($t(19) = 0.18$, $p = 0.86$) perspective consistency trials. When judging avoidance-desire trials, participants were significantly less error-prone in the passive- than active-involvement trials, when judging different-perspective consistency trials ($t(19) = 4.02$, $p < .005$; with a 10 % advantage for passive-involvement trials) but not for same-perspective consistency trials ($t(19) = 0.00$, $p = 1.00$). These results show a global enhancement of participants' performance, from the active-involvement trials (first block) to the passive-involvement trials (second block) in the avoidance, different-perspective consistency trials only.

No other effects were significant (all *F*s < 0.64, all *p*s > 0.43; see Table 6 for overall mean).

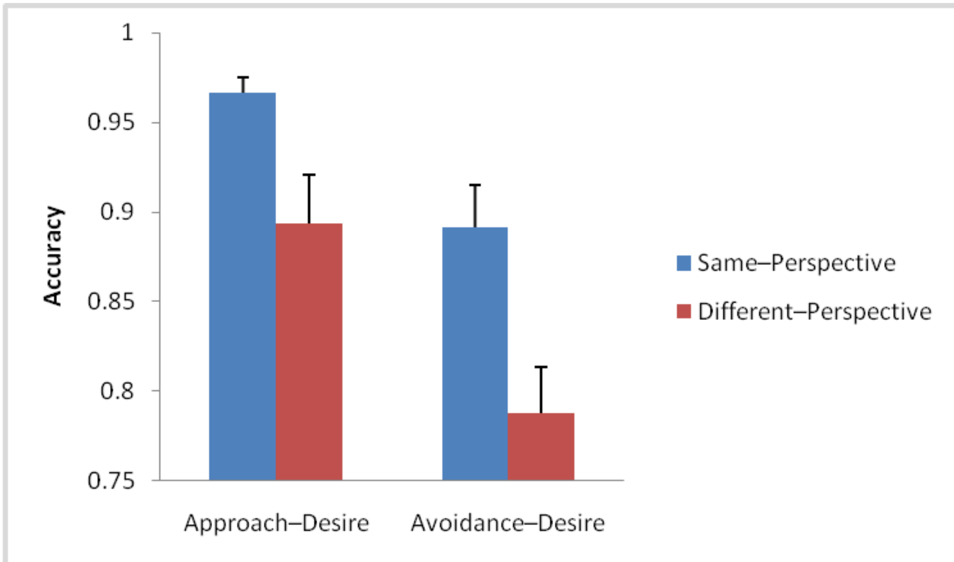



Figure 21. Overall mean (and standard error) accuracy of the Desire Type x Perspective Consistency interaction for the active-involvement condition.

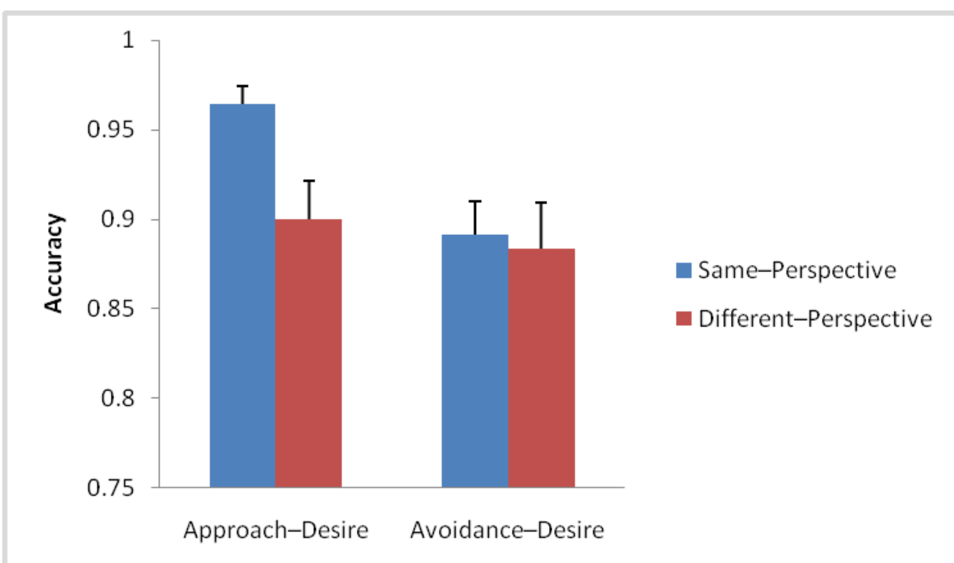



Figure 22. Overall mean (and standard error) RT of the Desire Type x Perspective Consistency interaction for the passive-involvement condition.

Table 6. Overall mean for accuracy of Experiment 4 across all experimental conditions.

| | Active Involvement | | | | Passive Involvement | | | |
|---|---|---|---|---|---|---|---|---|
| | Same | | Different | | Same | | Different | |
| | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** |
| **Approach** | 0.98 | 0.96 | 0.90 | 0.88 | 0.97 | 0.96 | 0.90 | 0.89 |
| **Avoidance** | 0.89 | 0.90 | 0.80 | 0.78 | 0.91 | 0.89 | 0.88 | 0.88 |

### 3.3.3 Discussion

Similarly to Experiment 3, participants were significantly slower and more error-prone at judging avoidance-desire trials than approach-desire trials, with no speed-accuracy trade-off. This avoidance-desire processing cost did not decrease over time when participants judged the green pawn's perspective (Henry's perspective) but it decreased over time when participants judged the purple pawn's perspective and switched from judging their own perspective to judging Paul's perspective. Thus in Experiment 4, familiarisation with the task per se did not decrease the avoidance-desire processing cost while self (active) involvement did in fact increase the avoidance-desire processing cost.

Participants were again affected by the consistency of the perspectives both in their RTs and accuracy. This was particularly the case for the active involvement condition (accuracy analysis). This finding is the opposite of the pattern observed in Experiment 3, where the conflicting perspective cost decreased in the active-involvement condition. It thus seems that it is not the active/passive involvement as such that matters but rather the order of presentation of these trials, with the cost simply decreasing over time. Participants may thus start to adopt a strategy to better deal with situations where the two cards have a different colour. Furthermore, the error analysis in Experiment 4 showed that the conflicting perspective cost was higher for avoidance-desire than approach-desire trials in the active-involvement condition. This is also the opposite of the pattern observed in Experiment 3 (in the error analysis). Again, the order of presentation may have interfered with participant's performance. On the passive-involvement trials presented in Experiment 3, participants observed two people playing the card game and may have used this information to aid their performance in the active-involvement trials. In Experiment 4, this was not the case, and as observed when the self is actively involved in the task, greater costs can be found in avoidance trials.

In Experiment 3, the effect of perspective (advantage for the self-perspective trials) was only found in the active-involvement condition; in Experiment 4, there was an advantage for self-perspective judgments in both the active- and passive-involvement condition. The effect of perspective was, however, larger in the active- than passive-involvement condition. Given that the passive-involvement condition was presented in the second half of the experiment (in Experiment 4), it is possible that participants continued to identify with the purple pawn even though it represented Paul. Interestingly, as evidenced by the significant three-way interaction effect in the RT analysis, in the active-involvement condition, the self (purple pawn)-perspective advantage was only significant on approach-desire trials but not avoidance-desire trials. In contrast, in the passive-involvement condition, the purple pawn advantage was found on both approach and avoidance trials. There was no speed-accuracy trade-off that would suggest that this was a spurious effect. Thus, it appeared that the more salient the self-perspective (as in the active-involvement condition) and the harder it became to judge avoidance-desire self-perspective trials (to the point that it was as hard to make an avoidance judgment from a self-perspective as from the perspective of another person). In order to confirm this surprising effect, I ran an additional experiment which only included the active-involvement condition, but this time, players were represented by a real photograph instead of a pawn. Previous studies have shown that the use of images of oneself is a good method to investigate processes involved with the self (e.g., Kircher et al., 2000). If the strong avoidance cost on self-perspective trials is a reliable effect, I should be able to replicate it in Experiment 5, where participants' own perspective was made even more salient by the use of their own photograph.

## 3.4. Experiment 5

### 3.4.1 Method

#### *Participants*

Twenty students (1 male, $M_{age}$ = 24 years) from the University of Birmingham School of Psychology took part in this study in exchange for course credit. All participants had normal to corrected vision and none were colour blind.

#### *Stimuli and procedure*

The method was identical to the active-involvement condition of Experiments 3 and 4 in all respects, except that the players in the game were represented by real photographs instead of pawns (see Figure 23). The participant's opponent was depicted by a profile shot of a male Caucasian (the photograph subtended 1.4 by 1.9 degrees) and appeared to the left hand side of the table for half of the participants and to the right hand side for the other half of the participants. The picture of the participant (taken prior to the start of the experiment and of similar size to the photograph of the opponent) appeared on the opposite side of the table.

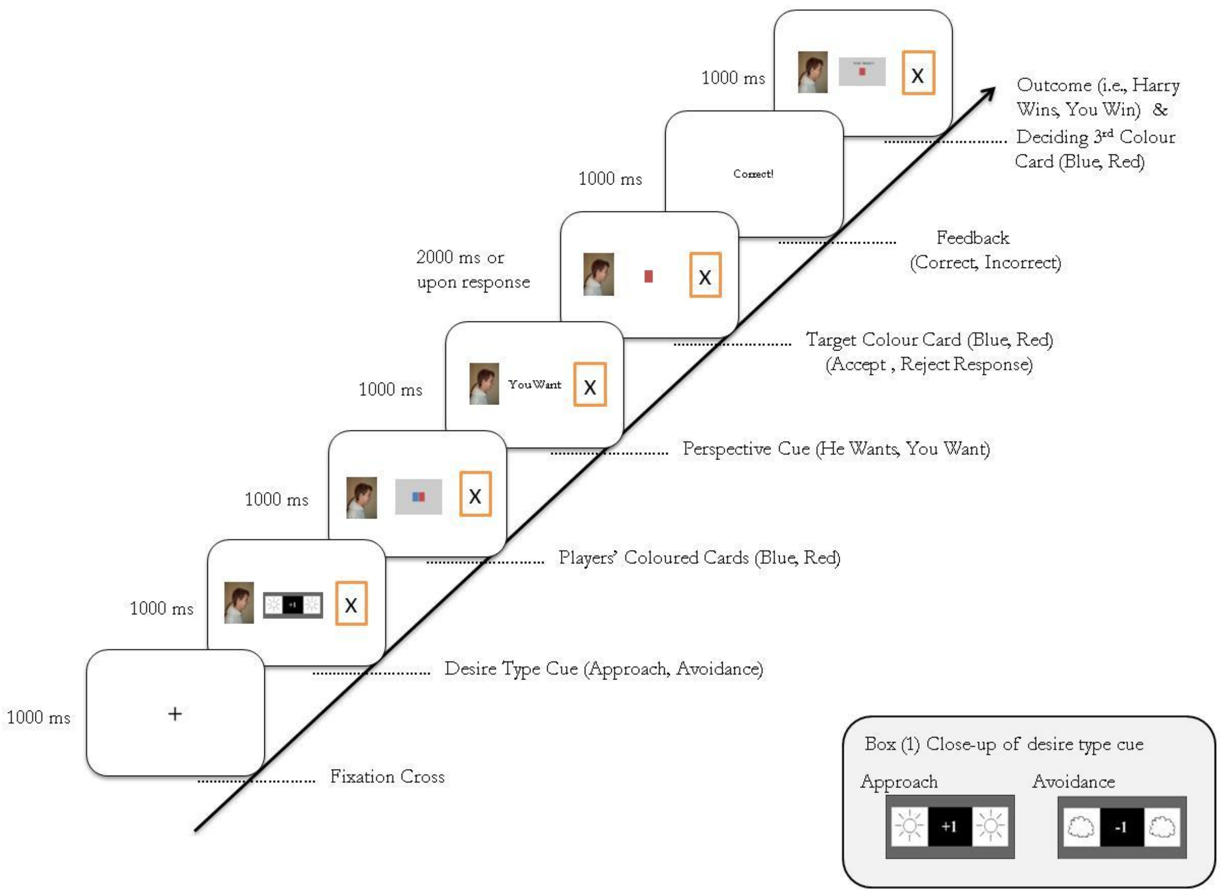



Figure 23. Experimental procedure displaying the sequence of events presented in order as shown on screen in Experiment 5. Box (1) contains a close-up of both desire type cues (approach, avoidance).

### 3.4.2. Results

Here, three participants were excluded (overall accuracy range: 71–73 %) because their high error rate meant that too many RTs had to be discarded from some experimental conditions. They were replaced by new participants. The skew and kurtosis values for the RT data were within tolerable range for assuming a normal distribution (skew: 1.25; kurtosis: 1.19).

The ANOVA analyses were similar to those conducted in Experiments 3 and 4 but excluding the passive-/active-involvement factor. Only the data of matching trials (i.e., "yes" responses) was taken into account. For the RT analysis, responses omitted due to the timeout procedure were (1 % of the data) and erroneous responses (8 % of the data) were not taken into account.

#### *Reaction time (RT) analysis*

The three-way ANOVA revealed a significant main effect of desire type ($F(1,19) = 19.17$, $p < .001$, partial $\eta^2 = 0.50$); RTs were slower when participants judged an avoidance- (706 ms) than approach- (623 ms) desire. There was a significant main effect of perspective consistency ($F(1,19) = 17.98$, $p < .001$, partial $\eta^2 = 0.49$); participants were slower at judgements on trials where the perspectives were different (697 ms) compared to trials where the perspectives were the same (632 ms). The main effect of perspective was significant as well ($F(1,19) = 18.28$, $p < .001$, partial $\eta^2 = 0.49$). Participants were slower at judging the opponent's perspective (701 ms) than their own perspective (628 ms).

There was a significant Desire Type x Perspective interaction ($F(1,19) = 8.99$, $p < .010$, partial $\eta^2 = 0.32$; see Figure 24). Paired *t*-tests revealed a significant desire type effect when participants judged their own perspective ($t(19) = 5.00$, $p < .001$; with a 111 ms advantage in the approach condition). However, this desire type effect failed to reach significance when participants judged their opponent's perspective ($t(19) = 1.71$, $p = 0.10$; with a 35 ms advantage in the approach condition). The effect of perspective was significant on approach trials only ($t(19) = 5.70$, $p < .001$), with participants being slower at judging their opponent's perspective (689 ms) than their own perspective (568 ms). No significant effect of perspective occurred on avoidance trials ($t(19) = 1.90$, $p = 0.08$): participants were as quick to judge their opponent's perspective (724 ms) as at judging their own perspective (679 ms).

No other effects were significant (all $F$s < 2.55, all $p$s > 0.13; see Table 7 for overall mean).

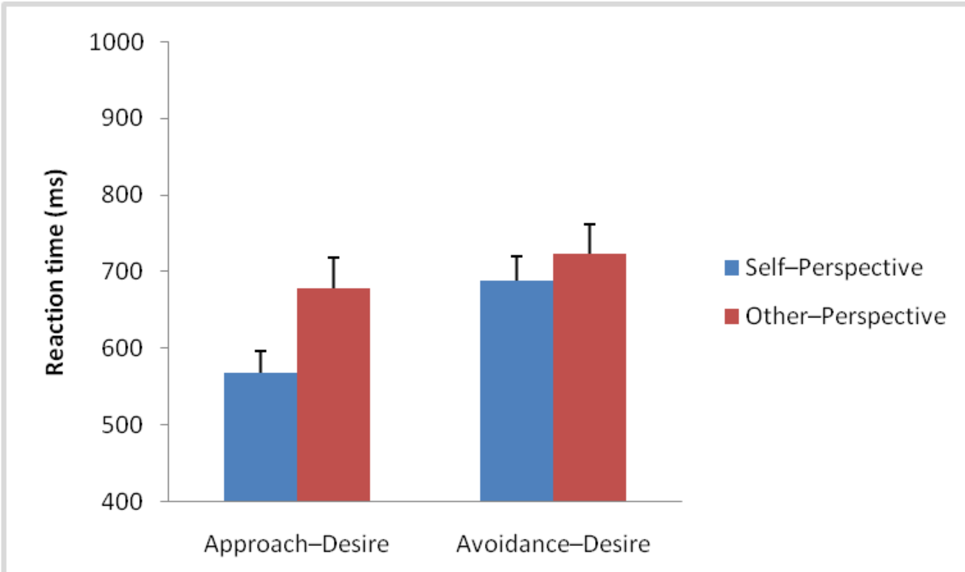



Figure 24. Overall mean (and standard error) RT of the Desire Type x Perspective interaction.

Table 7. Overall mean for RT of Experiment 5 across all experimental conditions.

| | **Same** | | **Different** | |
|---|---|---|---|---|
| | **Self** | **Other** | **Self** | **Other** |
| **Approach** | 554 ms | 656 ms | 581 ms | 702 ms |
| **Avoidance** | 641 ms | 677 ms | 736 ms | 770 ms |

### *Accuracy analysis*

The three-way ANOVA revealed a significant main effect of desire type ($F(1,19) = 6.07$, $p = 0.02$, partial $\eta^2 = 0.24$); participants were more error-prone when they judged an avoidance (89 %) than approach desire (93 %). There was a significant main effect of perspective consistency ($F(1,19) = 4.48$, $p = 0.05$, partial $\eta^2 = 0.19$); participants were more error-prone at judgements on trials where the perspectives were different (90 %) compared to trials where the perspectives were the same (93 %). The main effect of perspective was significant as well ($F(1,19) = 8.14$, $p = 0.01$, partial $\eta^2 = 0.30$). Participants were less accurate at judging the opponent's perspective (93 %) than their own perspective (89 %).

No other effects were significant (all $F$s $< 0.52$, all $p$s $> 0.48$; Table 8 for overall mean).

Table 8. Overall mean for accuracy of Experiment 5 across all experimental conditions.

| | Same | | Different | |
|---|---|---|---|---|
| | **Self** | **Other** | **Self** | **Other** |
| **Approach** | 0.96 | 0.94 | 0.94 | 0.90 |
| **Avoidance** | 0.93 | 0.88 | 0.90 | 0.85 |

### 3.4.3. Discussion

Experiment 5 replicated the overall advantage for self-perspective judgments (main perspective effect), for approach-desire trials (main effect of desire type) and for same-perspective trials (main effect of perspective consistency). More interestingly, Experiment 5 also replicated the perspective x desire type interaction effect from Experiment 4: the self-perspective advantage disappeared when participants judged avoidance trials. This confirms the proposal that reasoning about avoidance desires is particularly hard from a self-perspective point of view.

## 3.5. General Discussion

The aim of this study was to investigate whether there is (i) a cost of self-perspective inhibition in desire reasoning and (ii) a cost of processing avoidance desires for healthy adult participants. This was investigated in three experiments where participants played against an opponent in a card game compared to a baseline where they simply watched two people playing a card game against each other.

Participants were quicker and less error prone when judging the perspective of one of the players when this player was themselves in the active-involvement condition or when this player was identified in the passive-involvement by a symbol that was previously associated with the participant. Thus, if one of the players was linked with the self (directly or indirectly), this made the perspective judgment more efficient. The flip side was that other-perspective judgments became comparatively less efficient. This perspective effect can be interpreted in two ways which are not mutually exclusive. On the one hand, the result can be explained in terms of facilitation for self-perspective judgments. This would fit with the argument that processing is generally facilitated when material (e.g., items to memorize) is related to the self (e.g., Cunningham et al., 2008). On the other hand, the perspective effect may reflect a self-perspective inhibition cost when reasoning about someone else's desire. The question is thus whether there was any cost of self involvement on other-perspective judgments. If there is no such cost then the interpretation of self-perspective facilitation suffices to explain the pattern of results. If there is a cost, then it is necessary to hypothesize the existence of a self-perspective inhibition cost (in addition to any facilitation effect). In principle, Experiments 3 and 4 should allow this to be investigated by comparing the efficiency with which participants judge the green pawn's perspective (i.e., the other person's perspective) in the active- versus passive-involvement condition. Unfortunately, the general familiarization effects in the second part of the experiment make the pattern of results difficult to interpret. For example, in Experiment 3, if there was a self-perspective inhibition cost, it should be reflected by less efficient judgments of the green pawn's perspective in the active- than passive-involvement condition. However, given that participants were first presented with the passive-involvement condition, they may also be expected to judge the green pawn's perspective more efficiently in the active-involvement condition (presented in the second half of the experiment) than the passive-involvement condition. These two effects go in the opposite direction and could cancel each other out—the net result being no difference between the active- and passive-involvement condition when judgements are made from the perspective of the green pawn. Note that this is indeed what I observed. In Experiment 4, if there was a self-perspective inhibition cost, we should again expect less efficient judgments of the green pawn's perspective in the active- than passive-involvement conditions. The active-involvement condition was presented first, so this is another independent reason (and sufficient on its own) why judgments about the green pawn's perspective would be less efficient in the active- than passive-involvement condition. Note that I found the expected difference. The safest conclusion is that, although there is no strong direct evidence in favour of, there is no evidence against the existence of a self-perspective inhibition cost.

Across the three experiments and irrespective of which perspective had to be judged, there was strong evidence that participants were slower and more error prone when judging an avoidance desire than an approach desire. Such a strong avoidance processing cost was not observed in a previous study that investigated avoidance- and approach-desire reasoning in healthy adults (German & Hehman, 2006). One reason for such a discrepancy is that, in German and Hehman's study, participants did not need to infer the avoidance-desire but they needed to infer the resulting action. In the present study, it is the inference of the desire itself that is measured. The other difference between the two studies is that, in German and Hehman (2006), there was no time pressure for participants to process the stimuli and give their response (response times ranged between 1 and 3 sec). In contrast, in the present study, events were rapidly presented within a trial and by the nature of the verification task, response times were given within 1 sec. The paradigm used here is probably more sensitive to measure avoidance-desire processing costs.

What does the avoidance-desire processing cost reflect? An approach trial required participants to mentally represent a colour card, hold this information in mind, and then verify the desire content proposed. In contrast, avoidance trials involved an extra processing step that required participants to mentally negate the colour card (e.g., not red hence blue) before verifying the proposed desire content. That such an extra step would require more processing time and may be more error prone is in line with evidence from linguistic studies, which show that the processing of negation is cognitively taxing (Kaup et al., 2006; Kaup et al., 2007; Macdonald & Just, 1989). Studies that examine negation have often looked at its influence on the accessibility of text information. For example, Macdonald and Just (1989) presented participants with sentences as: "*Elizabeth bakes no* *bread and only cookies*", "*Elizabeth bakes some bread and no cookies*", and "*Elizabeth bakes some bread and some cookies*" (Macdonald & Just, 1989, p. 635). Subsequently, participants were presented with a probe recognition task. In both negated conditions, response times were slower compared to the neutral condition that had no negation involved (third sentence). Thus, the negated marker reduced accessibility to the probed word. Moreover, Kaup and colleagues (2007) argued that people might compute both the affirmative and the negated context, when given a negated context. Thus, for an example such as, "*a bird is in the sky*" versus "*a bird is not in the sky*", Kaup et al. argued that the first affirmative sentence taps onto one representation but the second negated sentence involves two representations (the representation of the negated context i.e., "*bird not in the sky*", and the affirmative statement i.e., "*bird in the sky*"). Consequently, it can be argued that participants on avoidance trials processed two mental representations in that, the negated colour card and the resulting actual colour card in hand. The cost I observed could be due to holding the two representations in mind.

Alternatively, Leslie and colleagues (Leslie & Polizzi, 1998; Leslie et al., 2005) discuss the role of 'target shifting' that occurs at the level of processing avoidance. According to this idea, participants have to inhibit the default approach attribution in order to correctly select the alternative. In line with this theory, the current cost on avoidance trials may reflect the extra inhibitory demands involved and the shifting of attention from one target to the other alternative (Leslie & Polizzi, 1998; Leslie et al. 2005).

Interestingly, the avoidance-desire processing cost seemed to be higher when participants judged their own perspective, at least in Experiment 4 when they started with the active-involvement condition, and also in Experiment 5, where they were only presented with the active-involvement condition and where, in addition, the self was made more salient. If self-involvement enhanced the processing of self-related material, it could be argued that the card that the participants received at the beginning of the round became particularly salient under conditions of active involvement and high self-saliency. Negating that coloured card or shifting focus away from the target may then become even more difficult.

In sum, the study showed a self-perspective processing advantage when healthy adults' process desires. However, it seems that the advantage for processing from a self-perspective had two knock-on detrimental effects. Firstly, it seems to have made other-perspective judgments less efficient (due to a self-perspective inhibition cost). Secondly, and more strikingly, the self-processing enhancement made reasoning about one's own avoidance particularly difficult. There is thus evidence in healthy adults compatible with the idea that there exist two processing costs in desire reasoning: the cost of inhibiting one's own perspective and the cost of inferring an avoidance desire. Chapter 2 suggested that these two processing costs could reflect the existence of functional and neural distinct processes. Chapter 5 will provide even stronger evidence for this claim. But before, I will present in the next chapter another series of experiments conducted with healthy participants to further understand the functional basis of these two components.

# CHAPTER 4

## On the origin of desire processing costs in healthy adults

## 4.1 Introduction

In Chapter 3, I presented evidence for two forms of interference when normal adults reason about desires. First, there was evidence of a self-perspective processing advantage. Second there was evidence of costs when inferring an avoidance desire. Also, the self processing enhancement made reasoning about one's own avoidance desire particularly difficult. In this chapter, I aim to investigate further these two situations where one's own desire interferes with reasoning by exploring (i) the level of processing at which the interference occurs and (ii) whether the interference can disappear or be reduced under strategic control.

In the case of belief reasoning, researchers have started to look at the level of processing at which effortful control is needed to process the information. When inferring a mental state, two broad stages can be identified: one stage in which the mental state content is constructed (i.e., the inference process as such), and another stage in which the information is maintained in working memory to inform subsequent behaviour or judgements. The main processing cost highlighted in belief reasoning relates to the need to resist interference from one's own perspective. Apperly and colleagues (2008) examined whether this interference occurred when participants were processing the incoming information (i.e., when they were integrating all the elements of the scenario to build a mental representation of the other person's belief), or whether interference occurred when participants held the belief in mind while preparing their response (i.e., when participants simply held the belief content in working memory). In order to do so, they manipulated the time available to encode the information. The rationale was the following: by shortening the time available to encode the information, the response time and accuracy measures would still capture the encoding processes as these were not yet completed. In contrast, by lengthening the encoding time, by the time participants would be asked to give a response, the encoding time would be more likely to be completed so that the response time and accuracy measured would now predominantly reflect the processes involved in maintaining the information in working memory. Concretely in Apperly and colleagues' study (2008), participants were first presented with sentences describing someone's belief about an object and the reality about an object. Participants were then shown picture probes, depicting either the person's belief or the reality, and they were asked whether the probe matched the content of the sentences previously presented. In one condition, the person's belief was false and hence conflicted with the state of reality described (e.g., "*He thinks the ball on the chair is red*; *really, the ball on the chair is yellow*"). In another condition, the person's belief was different but not in conflict with the state of reality described. In that latter baseline condition, the belief and the state of reality described simply referred to different objects (e.g., "*He thinks the ball on the chair is red; really, the ball on the table is yellow*"). In Experiment 1, participants were allowed to self-pace their reading, and the results showed that participants were significantly slower and more error prone in the false-belief condition than the control condition, showing interference between the belief content and the state of reality when both were in conflict. In Experiment 2, the time participants were given to read the sentences was manipulated and consisted of a slow or a fast pace. The increased time given to read the sentences in the slow pace condition was meant to give more time to participants to represent the person's belief. The authors reasoned that if the interference effect in the conflict/false-belief condition was still observed in that slow pace condition, this would suggest that the conflict still exists when participants simply hold in mind the belief and the reality; however, if the interference effect disappeared, this would suggest that the conflict mainly arises when participants process the incoming information to build a mental representation of the belief and reality. Similarly, decreasing the time given to read the sentences should give less time to participants to build a representation of the belief and reality. If the interference effect is mainly happening at the stage of building a mental representation, the interference effect should be maximized in this fast pace reading condition. The results suggested that the interference occurred at both the stage of building of a mental representation and at the stage of simply maintaining this information in memory to guide responses.

Using a similar procedure to the one used by Apperly and colleagues, I manipulated the time given to participants to encode the information in the desire reasoning paradigm used in Chapter 3. By either shortening or lengthening the time given to participants to encode the information, I assessed whether the interference of one's own desire when inferring the other person's desire, or when processing one's own avoidance desire, arises either when participants build a representation or when they hold the information in mind. Based on the findings from Apperly and colleagues (2008), I expected that the cost related to the need to resist interference from one's own perspective in the case of desire reasoning should be present both in the encoding and maintenance stage of processing (i.e., the effect of perspective should not be modulated by the manipulation of the encoding time). However, for the cost related to the avoidance-desire reasoning, I expected that the cost may be particularly high during the encoding stage, i.e., at a stage where participants may try to transform the avoidance desire (e.g., I do not want red) into an approach desire (e.g., I want blue). Once the avoidance desire was transformed, there may be no reason anymore to expect effortful control. Thus here, I expected that the effect of desire type could be modulated by the duration of encoding.

I only used the active-involvement condition, where participants directly played against an opponent and players were represented by photographs, as the studies reported in Chapter 3 showed that these conditions elicited the strongest self-perspective inhibition cost as well as the strongest and self/avoidance-desire cost. Of particular interest is the interaction between desire type (approach, avoidance) and perspective (self, other) from which two pure measures can be extracted: (1) the difference between self and other judgments for approach-desire trials only (measuring the self-perspective inhibition cost uncontaminated by the avoidance-desire cost) and (2) the difference between approach- and avoidance-desire judgments for self-perspective trials only (measuring the avoidance-desire cost uncontaminated by the self-perspective inhibition cost).

The paradigm used was framed as the card game called ‘Blame it on the Weather’ with identical rules to the games described in the previous chapter. In one study (Experiment 6a), participants were given less time, compared with previously (Experiment 5, Chapter 3) to process the incoming information (short display version; short DV). In the other study (Experiment 6b), participants were given more time to process the incoming information (longer display version; long DV). Across two experiments (Experiments 6a and 6b), the available response time was varied (shorter in Experiment 6a than Experiment 6b). If the perspective effects on approach-desire trials and the desire type effects on self-perspective judgments disappear or are reduced in the long DV condition, this would suggest that the perspective and desire type effects mainly occur at the stage where participants build a mental representation. In contrast, if the effects disappear or are reduced in the short DV condition, this would suggest that they mainly occur at the stage where participants maintain the information in working memory. Finally, if the effects are maintained in both conditions, this would suggest that they occur throughout processing, from the construction of the mental representation to holding it in mind to formulate a response.

In an additional experiment (Experiment 7), I examined whether the interference of one’s own desire when inferring someone else’s desire could be reduced by strategic control. In the experiments reported previously, participants only knew which perspective to judge after each player’s card was presented. Participants were thus forced to process their own and their opponent’s card. The idea here was to tell participants which perspective they will be asked to judge before presenting the players' coloured cards. The question was whether, under these conditions, participants would be able to focus solely on the relevant perspective and ignore the irrelevant perspective or whether participants would still process both perspectives.

## 4. 2. Experiment 6a and 6b: Short versus long display time

### 4.2.1 Method

#### *Participants*

Forty students from the University of Birmingham, School of Psychology took part in exchange for course credit. All participants had normal to corrected vision and none were colour blind. Half of the participants took part in Experiment 6a (3 male, $M_{age}$ = 21 years) and the other half took part in Experiment 6b (2 male, $M_{age}$ = 21 years).

#### *Stimuli and procedure*

The stimuli and procedure were similar to those used in Experiment 5 (Chapter 3) with the only difference being the time the player's coloured card remained visible on the screen (see Figure 25). In Experiment 6a, the short DV experiment, stimuli remained on screen for either (1) 500 ms (i.e., for half the time they saw it in Experiment 5, Chapter 3) or (2) 1000 ms (same display time as in Experiment 5, Chapter 3). In Experiment 6b, the long DV experiment, stimuli was displayed for either (1) 2000 ms (twice as long as the display time in Experiment 5, Chapter 3) or (2) 1000 ms (same display time as in Experiment 5, Chapter 3).

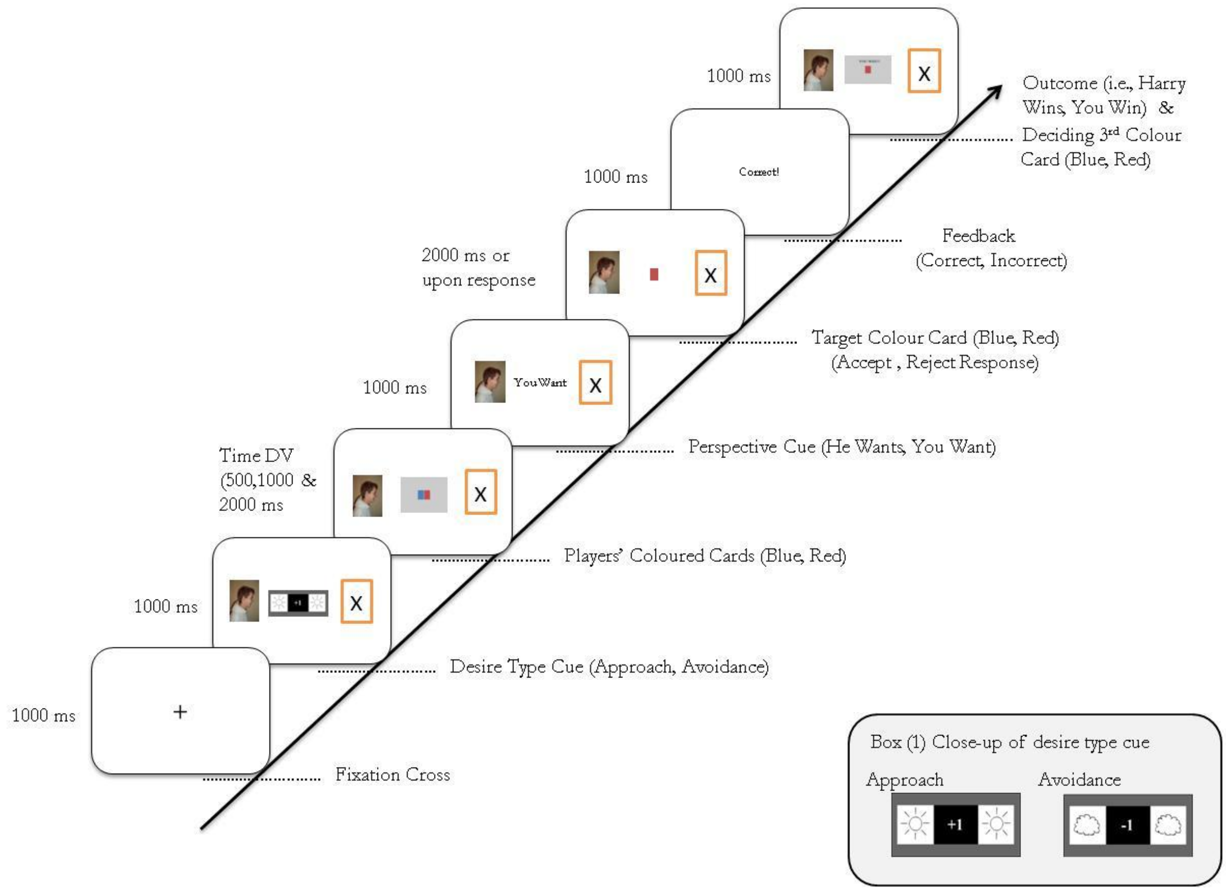



Figure 25. Experimental procedure displaying the sequence of events presented in order as shown on screen. Box (1) contains a close-up of both desire type cues (approach, avoidance).

## 4. 3.2. Results

The method of analysis is similar to Experiment 5 (Chapter 3) in all respects. First, 4 participants were excluded from the study as they performed only 50 % or less correct in at least one experimental condition (overall accuracy range: 49–69 %). These participants may not have complied with the task instructions and/or too many data points would have to be excluded from the RT analyses. They were replaced to keep the total number of 40. Second, the RT data was explored to see whether it was normally distributed. The skew and kurtosis were within tolerable ranges to use parametric tests without transforming the data or removing outliers (overall, Experiment 6a: skew: 1.20; kurtosis: 0.98; Experiment 6b: skew: 1.30; kurtosis: 1.62).

A 2x2x2x2 repeated-measure ANOVA analyses was computed with desire type (approach vs. avoidance), perspective consistency (different, vs. same), perspective (self vs. other) and time (Experiment 6a: 500 ms vs. 1000 ms; Experiment 6b: 2000 ms vs. 1000 ms), as independent variables. The dependent measures, similar to Experiment 5 (Chapter 3), were either RT or accuracy (proportion of correct response). Only the data of matching trials (i.e., "yes" responses) was taken into account. For RT analysis, responses omitted due to the timeout procedure (1% of the data) were eliminated from the data set and erroneous responses (9% of the data) were not taken into account in the analyses.

#### 4.3.2.1. Short display version (500 ms – 1000 ms)

##### *RT Analysis*

This analysis revealed a significant main effect of desire type ($F(1,19) = 31.17$, $p < .001$, partial $\eta^2 = 0.62$) with participants responding quicker on approach-desire trials (635 ms) compared to avoidance-desire trials (769 ms). There was a significant main effect of perspective consistency ($F(1,19) = 16.84$, $p < .005$, partial $\eta^2 = 0.47$), with quicker RTs on the same-perspective trials (664 ms) compared to the different-perspective trials (740 ms). The main effect of perspective was significant ($F(1,19) = 15.23$, $p < .005$, partial $\eta^2 = 0.45$) with participants responding faster for themselves (681 ms) than when asked to take their opponent's perspective (723 ms). There was no significant main effect of time ($F(1,19) = 2.22$, $p = 0.15$, partial $\eta^2 = 0.10$).

Overall, three significant two-way interactions were found, one further quantifying to a significant three-way interaction. First, similar to Experiment 3 (Chapter 3), there was a significant Desire Type x Perspective interaction ($F(1,19) = 10.78$, $p < .005$, partial $\eta^2 = 0.36$; see Figure 26). Paired *t*-tests revealed a significant desire type effect both when participants judged their opponent's perspective ($t(19) = 3.32$, $p < .005$; with a 91 ms advantage for approach-desire trials) and when they judged their own perspective ($t(19) = 6.50$, $p < .001$; with a 177 ms advantage for approach-desire trials). Numerically, the advantage for approach-desire trials was larger when participants judged their own perspective. The effect of perspective was significant on approach-desire trials ($t(19) = 5.19$, $p < .001$) with participants being slower at judging the opponent's perspective (678 ms) than their own perspective (593 ms). However, the effect of perspective was not significant for the avoidance-desire trials ($t(19) = 0.07$, $p = 0.95$).

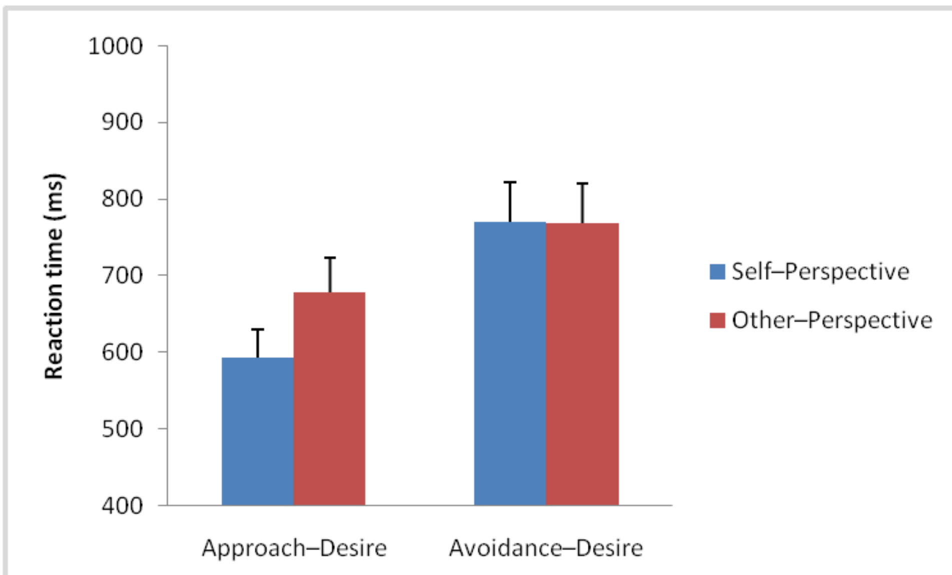



Figure 26. Overall mean (and standard error) for RT of Desire Type x Perspective interaction.

Secondly, there was a significant Perspective Consistency x Perspective interaction ($F(1,19) = 4.81$, $p = 0.04$, partial $\eta^2 = .20$; see Figure 27). Here, paired *t*-tests revealed significant perspective consistency effects when participants judged both, their opponent's perspective ($t(19) = 4.18$, $p < .005$; with a 97 ms advantage when the perspectives were the same) and their own perspective ($t(19) = 3.07$, $p < .010$; with 56 ms advantage when the perspectives were the same). Numerically, the advantage for same-perspective trials was larger when participants judged their opponent's perspective. The effect of perspective was significant both, when the perspectives were different ($t(19) = 3.51$, $p < .005$; with a 62 ms advantage for self) and when they were the same ($t(19) = 2.27$, $p = 0.04$; with a 22 ms advantage for self). Numerically, the advantage for the self was larger when the perspectives were different.

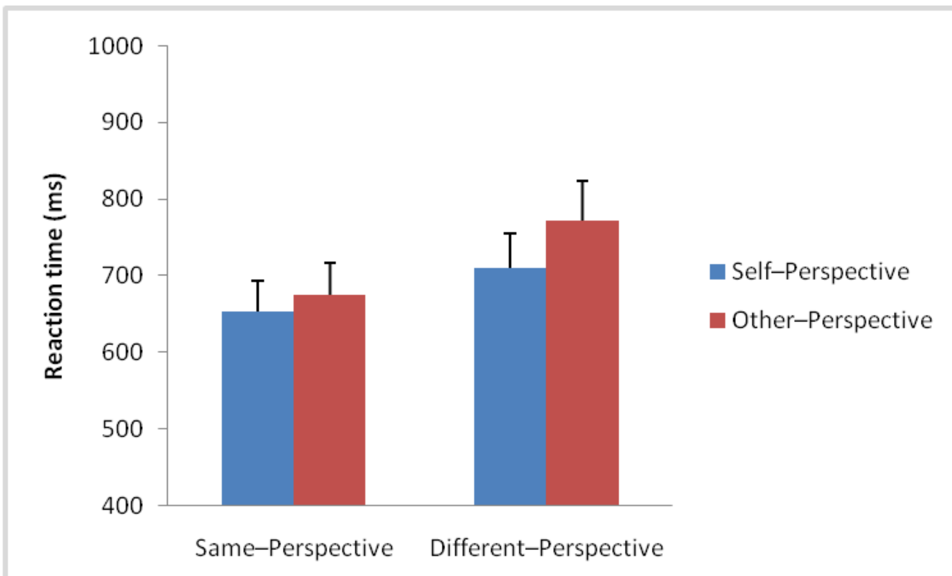



Figure 27. Overall mean (and standard error) for RT of Perspective Consistency x Perspective interaction.

Finally, there was a significant Desire Type x Perspective Consistency ($F(1,19) = 7.52$, $p = 0.01$, partial $\eta^2 = 0.28$) that further qualified to a Desire Type x Perspective Consistency x Time interaction ($F(1,19) = 6.70$, $p = 0.02$, partial $\eta^2 = 0.26$). The data were collapsed across perspective and separate paired sample *t*-tests were carried out to analyse the desire type and perspective consistency effects for each duration condition.

In the 500 ms condition (see Figure 28), there was a significant desire type effect when participants judged trials where the perspectives were different ($t(19) = 2.92$, p < .010; with a 111 ms advantage on approach-desire trials), and when they judged trials where the perspectives were the same ($t(19) = 5.72$, $p < .001$; with a 163 ms advantage on approach-desire trials). Numerically, the advantage for approach-desire trials was larger when the perspectives were the same. The effect of perspective consistency was significant on approach-desire trials ($t(19) = 3.21$, $p < .010$; with a 81 ms advantage for same-perspective trials). However, no significant effect of perspective consistency could be found on avoidance-desire trials ($t(19) = 1.08$, $p = 0.29$).

In the 1000 ms condition (see Figure 29), there was a significant desire type effect when participants judged the trials where the perspectives were different ($t(19) = 6.72$, $p < .001$; with a 153 ms advantage for approach-desire trials), and when they judged trials where the perspectives were the same ($t(19) = 3.41$, $p < .005$; with a 109 ms advantage for approach-desire trials). Numerically, the advantage for approach-desire trials was larger when the perspectives were different. There was a significant perspective consistency effect on approach-desire trials ($t(19) = 3.14$, $p < .010$; with a 76 ms advantage for same-perspective trials), and on avoidance-desire trials ($t(19) = 5.32$, $p < .001$; with a 120 ms advantage for same-perspective trials). Numerically, the advantage for same-perspective trials was larger on avoidance-desire trials.

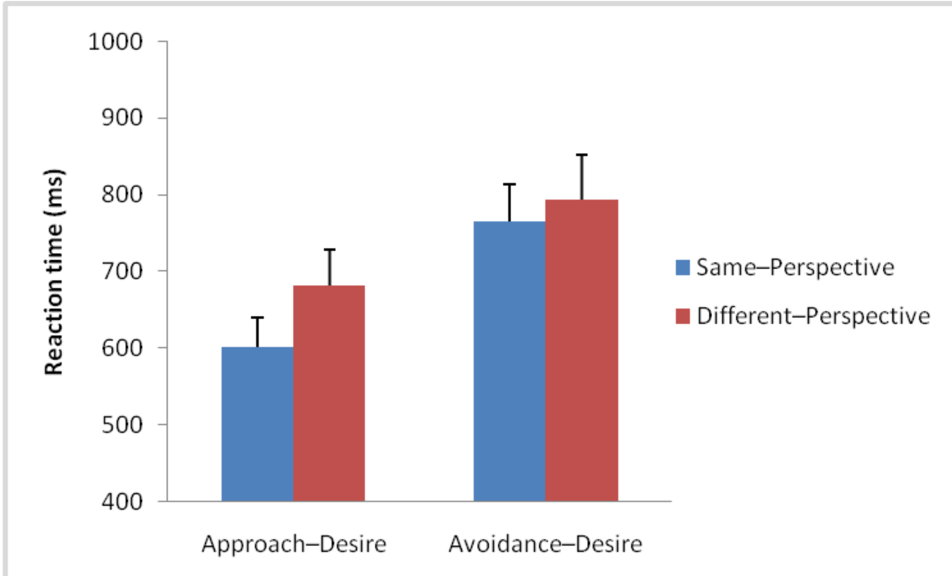



Figure 28. Overall mean (and standard error) for RT of desire type x perspective consistency interaction for the 500 ms trials.

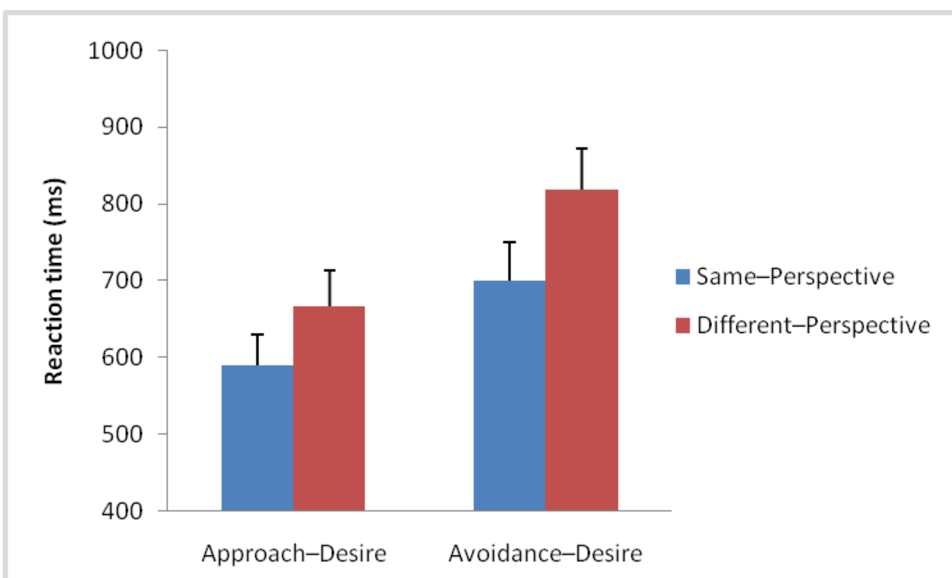



Figure 29. Overall mean (and standard error) for RT of desire type x perspective consistency interaction for the 1000 ms trials.

In order to further investigate the three-way interaction, separate paired sample *t*-tests were carried out to look at the effects of shortening the display time. For same-perspective avoidance-desire trials, there was a significant effect of time ($t(19) = 2.77$, $p = 0.02$; with a 65 ms advantage on 1000 ms trials). However, there was no significant effect of time on the different-perspective avoidance-desire trials, ($t(19) = 1.15$, $p = 0.26$), nor in the same-perspective approach-desire trials ($t(19) = 0.78$, $p = 0.44$) or different-perspective approach-desire trials ($t(19) = 0.70$, $p = 0.49$).

No other effects were significant (all *F*s $< 1.71$, all *p*s $> 0.21$; see Table 9 for overall mean).

Table 9. Overall mean for RT of Experiment 6a across all experimental conditions.

| | **500 ms** | | | | **1000 ms** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Same** | | **Different** | | **Same** | | **Different** | |
| | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** | **Self** | **Other** |
| **Approach** | 581 ms | 622 ms | 623 ms | 741 ms | 555 ms | 626 ms | 611 ms | 721 ms |
| **Avoidance** | 774 ms | 756 ms | 806 ms | 781 ms | 704 ms | 696 ms | 796 ms | 843 ms |

### *Accuracy Analysis*

This analysis revealed a significant main effect of desire type ($F(1,19) = 18.50$, $p < .001$, partial $\eta^2 = 0.49$) with participants being less error prone on approach-desire trials (93% correct) compared to the avoidance-desire trials (85% correct). There was a significant main effect of perspective consistency ($F(1,19) = 7.27$, $p = 0.01$, partial $\eta^2 = 0.28$), with more accurate responses made for same-perspective trials (91% correct) relative to different-perspective trials (87% correct). Neither the main effect of perspective ($F(1,19) = 3.71$, $p = 0.07$, partial $\eta^2 = 0.16$) nor that of time ($F(1,19) = 3.82$, $p = 0.07$, partial $\eta^2 = 0.17$), were significant.

There were two significant two-way interactions, one that further qualified to a significant three-way interaction. First, there was a significant Desire Type x Perspective interaction ($F(1,19) = 12.26$, $p < .005$, partial $\eta^2 = 0.39$; see Figure 30). Paired *t*-tests revealed a significant desire type effect when participants judged their own perspective ($t(19) = 6.15$, $p < .001$; with a 12% advantage on approach-desire trials). However, this desire type effect failed to reach significance when participants judged their opponent's perspective ($t(19) = 1.90$, $p = 0.07$). Furthermore, the effect of perspective was significant on approach-desire trials ($t(19) = 3.70$, $p < .005$; with a 6 % advantage for self-perspective) but not on avoidance-desire trials ($t(19) = 0.46$, $p = 0.65$).

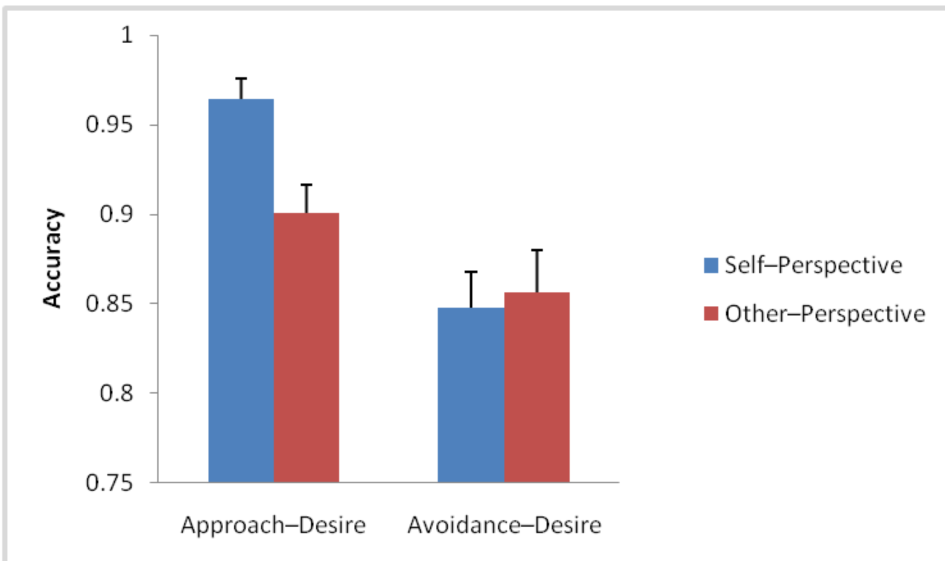



Figure 30. Overall mean (and standard error) for accuracy of Desire Type x Perspective interaction.

Secondly, there was a significant Perspective Consistency x Perspective interaction ($F(1,19) = 6.83$, $p = 0.02$, partial $\eta^2 = 0.26$) that further subsumed to a significant three-way Perspective Consistency x Perspective x Time interaction ($F(1,19) = 7.61$, $p = 0.01$, partial $\eta^2 = 0.29$). The data were collapsed across time and separate paired sample *t*-tests were carried out to analyse the perspective consistency and perspective effects for each duration condition.

In the 500 ms condition (see Figure 31), there was a significant perspective consistency effect when participants judged their opponent's perspective ($t(19) = 2.64$, $p = 0.02$, with a 6% advantage in the same-perspective trials), but no significant perspective consistency effect when they judged their own perspective ($t(19) = 0.43$, $p = 0.67$). The effect of perspective was significant on trials where the perspectives were different ($t(19) = 2.82$, $p = 0.11$; with a 7% advantage for self-perspective trials), but not with same-perspective trials ($t(19) = 0.58$, $p = 0.57$).

In the 1000 ms condition (see Figure 32), there was a significant perspective consistency effect when participants judged their opponent's perspective ($t(19) = 2.73$, $p = 0.01$; with a 5% advantage in the same-perspective trials) and when they judged their own perspective ($t(19) = 2.31$, $p = 0.03$, with a 4% advantage in the same-perspective trials). Numerically, the advantage for same-perspective trials was larger when participants judged their own perspective. The effect of perspective was marginally significant on trials where the perspectives were different ($t(19) = 2.04$, $p = 0.05$, with a 4% advantage for self), but not with same-perspective trials ($t(19) = 1.42$, $p = 0.17$).

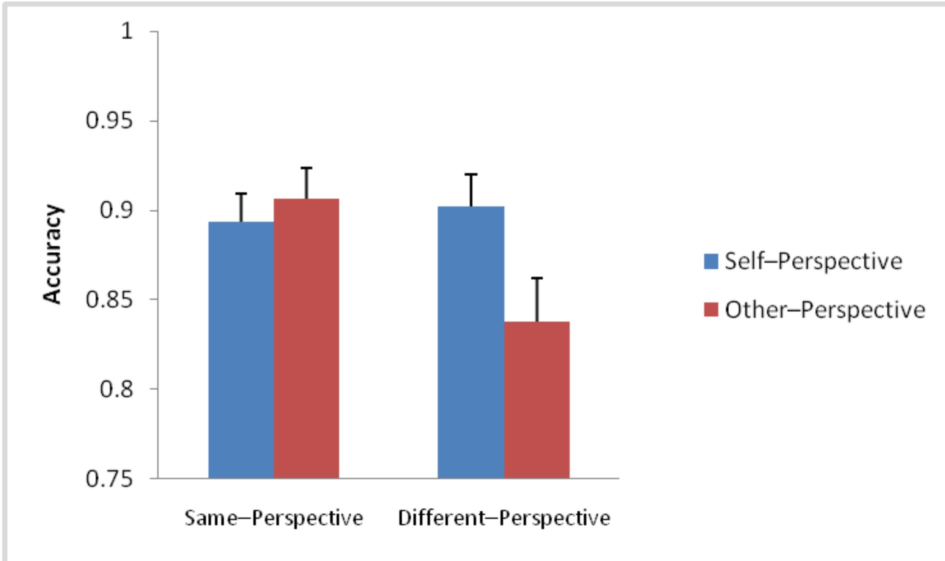



Figure 31. Overall mean (and standard error) for accuracy of Desire Type x Perspective Consistency interaction for the 500 ms trials.

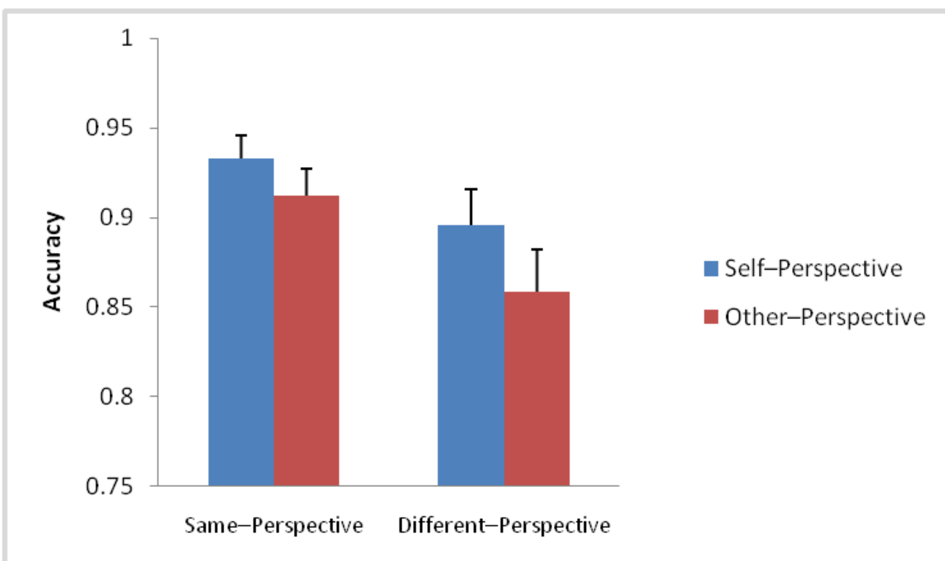



Figure 32. Overall mean (and standard error) for accuracy of Desire Type x Perspective Consistency interaction for the 1000 ms trials.

In order to further investigate the three-way interaction, separate paired sample *t*-tests were carried out to look at the effects of shortening the display time. There was a significant effect of time on the self same-perspective consistency trials, ($t(19) = 2.33$, $p = 0.03$, with a 4% advantage for 1000 ms trials). However, there was no significant effect of time on the same-perspective consistency for other, ($t(19) = 0.36$, $p = 0.72$) nor in different-perspective self trials ($t(19) = 0.42$, $p = 0.68$) or the different-perspective other trials ($t(19) = 1.23$, $p = 0.23$).

No other effects were significant (all *F*s < 1.71, all *p*s > 0.21; see Table 10 for overall mean).

Table 10. Overall mean for accuracy of Experiment 6a across all experimental conditions

| | 500 ms | | | | 1000ms | | | |
|---|---|---|---|---|---|---|---|---|
| | Same | | Different | | Same | | Different | |
| | Self | Other | Self | Other | Self | Other | Self | Other |
| **Approach** | 0.96 | 0.95 | 0.97 | 0.84 | 0.98 | 0.94 | 0.95 | 0.88 |
| **Avoidance** | 0.83 | 0.87 | 0.84 | 0.84 | 0.88 | 0.88 | 0.84 | 0.84 |

## Discussion

The desire type x perspective interaction effect from Experiment 3 was replicated here. There was a perspective effect on approach-desire trials (signature of the self-perspective inhibition cost) and an effect of desire type on self-perspective trials (signature of the self/avoidance-desire cost). The interaction was not modulated by the shortening of the display time. In the RT analysis, the sole effect of shortening the display time was to modulate the ease with which participants were dealing with same-perspective avoidance-desire trials; on these trials there were slower responses in the 500 ms compared to the 1000 ms display time condition. Same-perspective avoidance-desire trials are trials where both players received the same coloured card (e.g., red) and both would want to avoid the matching third coloured card (e.g., red). Each player's resulting approach-desire consists in wanting the third card to be of a colour not presented in that round (e.g., blue). Reasoning in this case may be more difficult if participants not only compute the negated position (blue card) but also the approach-desire condition (red card), and this effect may impact particularly strongly during encoding.

### 4.3.2.2. Long Display Version (1000 ms – 2000 ms)

#### *RT Analysis*

Here, the ANOVA analysis revealed a significant main effect of desire type ($F(1,19) = 19.48$, $p < .001$, partial $\eta^2 = 0.51$), with participants responding quicker on approach-desire trials (627 ms) than avoidance-desire trials (703 ms). There was a significant main effect of perspective consistency ($F(1,19) = 18.15$, $p < .001$, partial $\eta^2 = 0.49$), with quicker RTs on same-perspective trials (634 ms) compared to different-perspective trials (695 ms). The main effect of perspective was significant ($F(1,19) = 4.29$, $p = 0.05$, partial $\eta^2 = 0.18$) and showed a trend for participants responding faster when judging their own perspective (649 ms) than when asked to judge their opponent's perspective (681 ms). The main effect of time was significant ($F$(1,19) = 7.17, $p$ = .015, partial $\eta^2$ = 0.27), with quicker response times in the 2000 ms condition (652 ms) relative to the 1000 ms condition (678 ms) .

Overall, there were two significant two-way interactions that were further qualified by two significant three-way interactions. First, there was a significant Desire Type x Perspective interaction ($F$(1,19) = 6.28, $p$ = 0.02, partial $\eta^2$ = 0.25) which was further qualified by a significant three-way Desire Type x Perspective x Time ($F$(1,19) = 7.04, $p$ = 0.02, partial $\eta^2$ = 0.27). Data were collapsed across perspective consistency and separate paired sample $t$-tests were carried out to look at the effects of desire type and perspective for each duration of display.

In the 2000 ms trials (see Figure 33), the analysis revealed a significant desire type effect when participants judged their own perspective ($t$(19) = 3.21, $p$ < .010; with a 76 ms advantage for approach-desire trials), and when they judged their opponent's perspective ($t$(19) = 2.68, $p$ = 0.01; with a 54 ms advantage for approach-desire trials). Numerically, the advantage for approach-desire trials was larger on self-perspective judgments. The effect of perspective was significant on approach-desire trials ($t$(19) = 2.53, $p$ = 0.02; with a 52 ms advantage for self trials). However, no significant difference of perspective was found on the avoidance-desire trials ($t$(19) = 1.45, $p$ = 0.16).

In the 1000 ms condition (see Figure 34), paired sample $t$-tests showed a significant desire type effect when participants made self-perspective judgments ($t$(19) = 4.78, $p$ < .001; with a 142 ms advantage for approach-desire trials) but not when they made judgments about the opponent's perspective ($t$(19) = 1.55, $p$ = 0.14). There was a significant perspective effect on approach-desire trials ($t(19) = 3.03$, $p < .010$; with a 78 ms advantage for self-perspective trials) but no significant effect of perspective on avoidance-desire trials ($t(19) = 1.06$, $p = 0.30$).

In order to further explore the three-way interaction, separate paired sample *t*-tests were carried out to assess the effects of lengthening the display time. There was no significant effect of time for self-approach trials ($t(19) = 0.15$, $p = 0.88$) and no significant effect of time for self-avoidance trials ($t(19) = 1.42$, $p = 0.17$). There was a significant effect of time for other-approach trials ($t(19) = 3.40$, $p < .005$; with a 68 ms advantage on 2000 ms trials) but no significant effect of time for other-avoidance trials ($t(19) = 0.48$, $p = 0.63$).

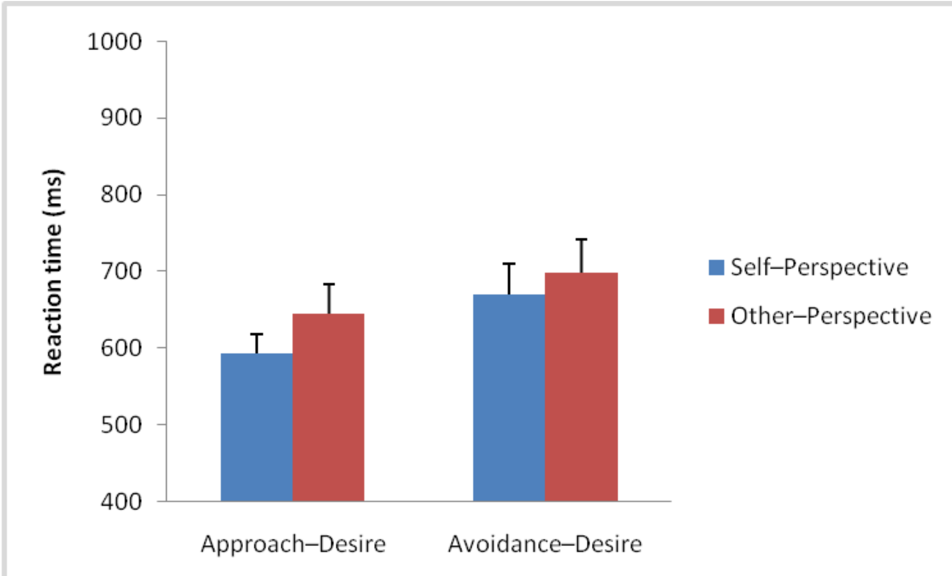



Figure 33. Overall mean (and standard error) for RT of Desire Type x Perspective interaction for the 2000 ms trials.

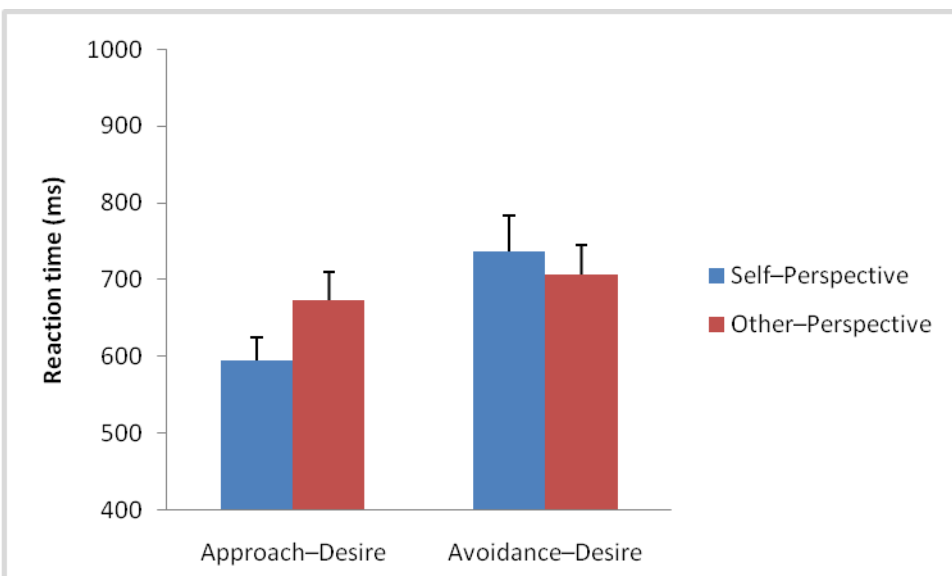



Figure 34. Overall mean (and standard error) for RT of Desire Type x Perspective interaction for the 1000 ms trials.

Secondly, there was a significant Perspective Consistency x Time interaction ($F(1,19) = 26.13$, $p < .001$, partial $\eta^2 = 0.58$) that was further qualified by a significant three-way Desire Type x Perspective Consistency x Time interaction ($F(1,19) = 6.11$, $p = 0.02$, partial $\eta^2 = 0.24$). Data were collapsed across perspectives and separate paired sample *t*-tests were carried out to look at the effects of desire type and perspective consistency for each display duration condition.

In the 2000 ms trials (see Figure 35), the analysis revealed a significant desire type effect when participants judged same-perspective trials ($t(19) = 2.77$, $p = 0.01$; with a 83 ms advantage for approach-desire trials) and when the perspectives were different ($t(19) = 2.19$, $p = 0.04$; with a 48 ms advantage for approach-desire trials). Numerically, the advantage for approach-desire trials was larger when the perspectives were different. The effect of perspective consistency was not significant on approach-desire trials ($t(19) = 1.47$, $p = 0.16$) and no significant effect of perspective consistency was found on avoidance-desire trials ($t(19) = 0.03$, $p = 0.96$).

On the trials with 1000 ms displays (see Figure 36), there was a significant desire type effect when participants judged different-perspective trials ($t(19) = 5.25$, $p < .001$; with a 120 ms advantage for approach-desire trials), and when they judged same-perspective trials ($t(19) = 2.69$, $p = .02$; with a 56 ms advantage for approach-desire trials). Numerically, the advantage for approach-desire trials was larger when the perspectives were different. There was a significant perspective consistency effect on approach-desire trials ($t(19) = 4.64$, $p < .001$; with a 74 ms advantage for same-perspective trials), and on avoidance-desire trials ($t(19) = 7.06$, $p < .001$; with a 138 ms advantage for same-perspective trials). Numerically, the advantage for same-perspective trials was larger when participants reasoned about avoidance desire.

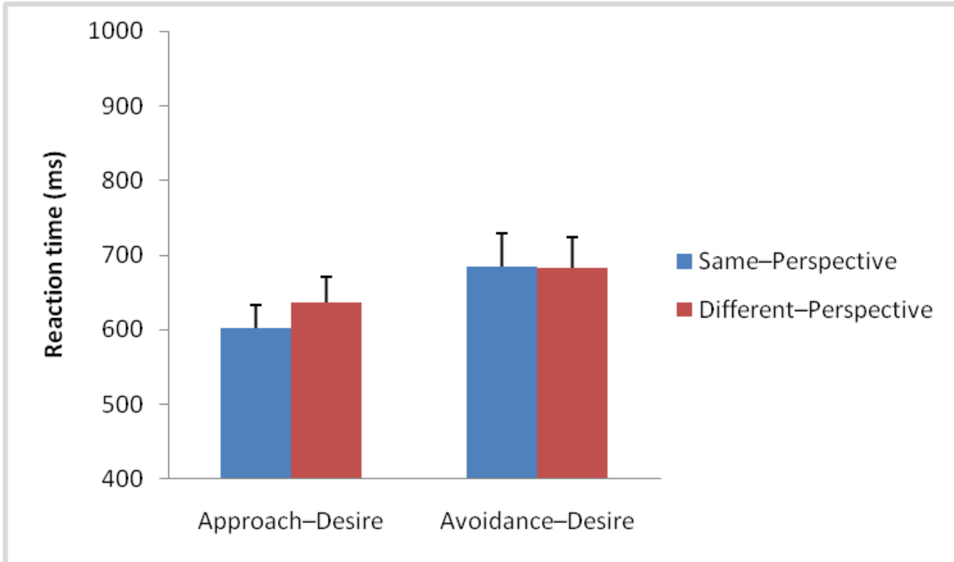



Figure 35. Overall mean (and standard error) for RT of Desire Type x Perspective Consistency interaction for the 2000 ms trials.

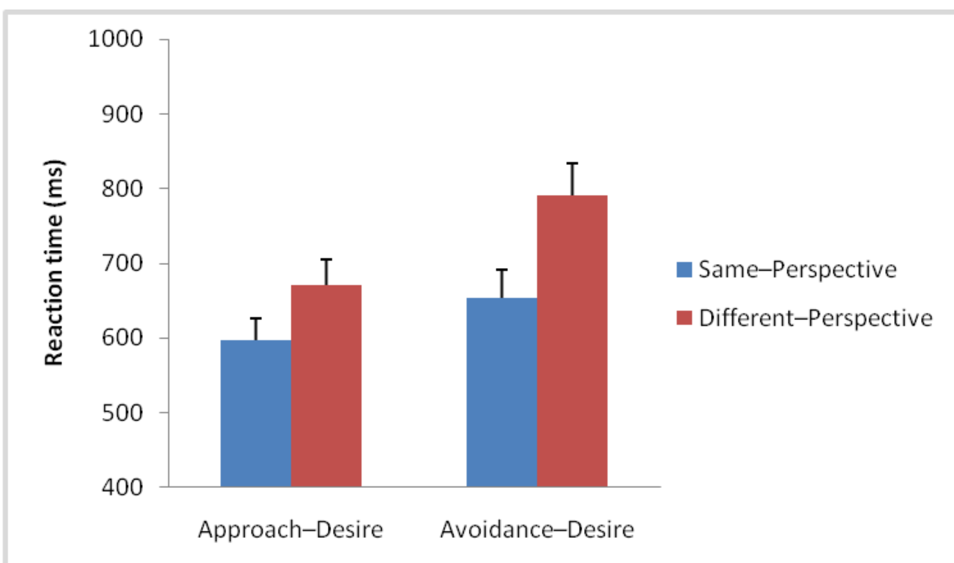



Figure 36. Overall mean (and standard error) for RT of Desire Type x Perspective Consistency interaction for the 1000 ms trials.

In order to further explore the three-way interaction, separate paired sample *t*-tests were carried out to look at the effects of lengthening the display time. There was a significant effect of time in the different-perspective approach-desire trials ($t(19) = 2.14$, $p = 0.05$; with a 35 ms advantage on 2000 ms trials) but not for the same-perspective approach-desire trials ($t(19) = 0.37$, $p = 0.72$). For different-perspective avoidance-desire trials, there was a significant effect of time ($t(19) = 5.12$, $p < .001$; with a 107 ms advantage on 2000 ms trials). However, no significant effect of time could be found in the same-perspective avoidance-desire trials ($t(19) = 1.53$, $p = 0.14$).

No other effects were significant (all *F*s < 2.67, all *p*s > 0.12; see Table 11 for overall mean).

Table 11. Overall mean for RT of Experiment 6b across all experimental conditions.

| | 2000 ms | | | | 1000 ms | | | |
|---|---|---|---|---|---|---|---|---|
| | Same | | Different | | Same | | Different | |
| | Self | Other | Self | Other | Self | Other | Self | Other |
| Approach | 584 ms | 620 ms | 603 ms | 669 ms | 566 ms | 628 ms | 624 ms | 718 ms |
| Avoidance | 676 ms | 694 ms | 663 ms | 704 ms | 669 ms | 638 ms | 805 ms | 777 ms |

### ***Accuracy Analysis***

Here, the analysis of accuracy revealed a significant main effect of desire type ($F(1,19) = 23.92$, $p < .001$, partial $\eta^2 = 0.56$), with participants being less error-prone for approach-desire trials (95% correct) compared to the avoidance-desire trials (89% correct). There was a significant main effect of perspective consistency ($F(1,19) = 13.12$, $p < .005$, partial $\eta^2 = 0.41$), with more accurate responses made for same-perspective trials (94% correct) relative to different-perspective trials (89% correct).The main effect of perspective was significant ($F(1,19) = 7.58$, $p = 0.01$, partial $\eta^2 = 0.29$), with participants responding more accurately for self- (93% correct) than other- (91%) perspective judgments. The main effect of time was also significant ($F(1,19) = 11.84$, $p < .005$, partial $\eta^2 = 0.38$); participants were more error-prone in the 1000 ms condition (91% correct) compared to the 2000 ms condition (93% correct).

Overall there was a significant there was a significant Desire Type x Perspective Consistency x Time interaction ($F(1,19) = 9.27$, $p < .010$, partial $\eta^2 = 0.33$). Data were collapsed across perspective and separate paired sample *t*-tests were carried out to look at the effects of desire type and perspective consistency for each display duration.

In the 2000 ms condition (see Figure 37), the analysis revealed no significant desire type effect on different-perspective trials ($t(19) = 1.43$, $p = 0.17$) but a significant effect for same-perspective trials was found ($t(19) = 3.77$, $p < .005$; with a 5% advantage for approach-desire trials). The effect of perspective consistency was significant for approach-desire trials ($t(19) = 2.81$, $p = 0.01$; with a 4% advantage for approach-desire trials) but not for the avoidance-desire trials ($t(19) = 1.54$, $p = 0.14$).

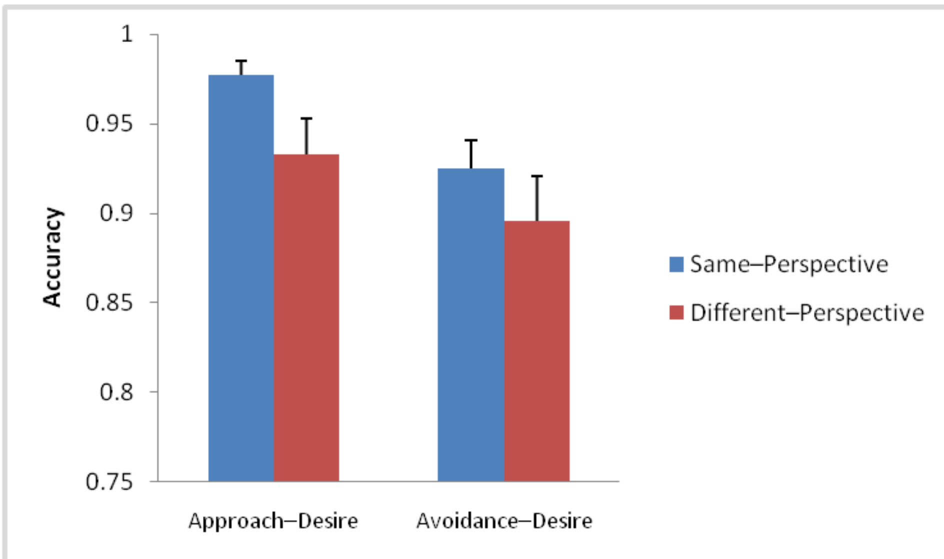



Figure 37. Overall mean (and standard error) for accuracy of Desire Type x Perspective Consistency interaction for the 2000 ms trials.

For the 1000 ms trials (see Figure 38), there was a significant desire type effect when participants judged different-perspective trials ($t(19) = 6.16$, $p < .001$; with an 11 % advantage for approach-desire trials) and when they judged same-perspective trials ($t(19) = 2.37$, $p = 0.03$, with an 9 % advantage for approach-desire trials). Numerically, participants were more error-prone in the avoidance-desire condition. There was no significant perspective consistency effect for approach-desire trials ($t(19) = 1.39$, $p = 0.18$) but there was for avoidance-desire trials ($t(19) = 3.67$, $p < .005$; with an 9% advantage for same-perspective trials).

To further explore the three-way interaction, separate paired sample *t*-tests were carried out to look at the effects of lengthening the display time. There was a significant effect of time in the same-perspective approach-desire trials ($t(19) = 2.13$, p = 0.05; with a 2% advantage for 2000 ms trials) but not for the different-perspective approach-desire trials ($t(19) = 0.26$, $p = 0.80$). For different-perspective avoidance-desire trials, there was a significant effect of time ($t(19) = 3.92$, $p < .005$; with a 7% advantage for 2000 ms trials). However, no significant effect of time could be found in the same-perspective avoidance-desire trials ($t(19) = 0.64$, $p = 0.53$).

No other effects were significant (all $F$s < 3.03, all $p$s > 0.10; see Table 12 for overall mean).

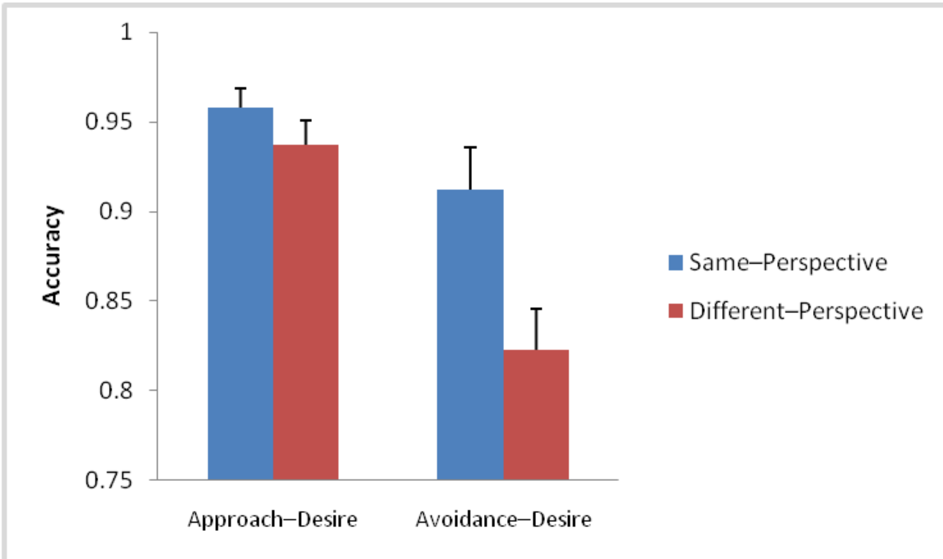



Figure 38. Overall mean (and standard error) for RT of Desire Type x Perspective Consistency interaction for the 1000 ms trials.

Table 12. Overall mean for RT of Experiment 6b across all experimental conditions.

| | 2000 ms | | | | 1000 ms | | | |
|---|---|---|---|---|---|---|---|---|
| | Same | | Different | | Same | | Different | |
| | Self | Other | Self | Other | Self | Other | Self | Other |
| **Approach** | 0.99 | 0.97 | 0.95 | 0.91 | 0.98 | 0.93 | 0.96 | 0.91 |
| **Avoidance** | 0.93 | 0.92 | 0.90 | 0.90 | 0.91 | 0.91 | 0.84 | 0.80 |

### *Discussion*

In Experiment 6b, the Desire Type x Perspective interaction was affected by the duration of the display. For both display durations, the effect of perspective was significant on approach-desire trials and the effect of desire type was significant on self-perspective trials, suggesting that these effects related to both the initial construction of the representations and the later maintenance of these representations.

Furthermore, there were perspective consistency effects in Experiment 6b, but these appeared to be equal for self- and other-perspective judgments (unlike Experiment 6a where the perspective consistency effect was larger when participants judged the other person's perspective). Increasing the display time may have made both players' perspectives more salient and hence produced equal levels of interference. In this case, not only did the participant's own perspective interfere with judgments from the other's perspective, but the other person's perspective interfered with self-perspective judgments.

Finally, the duration of the display particularly affected different-perspective for approach- and avoidance-desire trials. These trials were processed more efficiently (quicker and more accurately) when there was a long display duration (2000 ms). In the different-perspective trials both players received a different coloured card and each one wants to approach and/or avoid their own third matching card. This means that there are two sets of information to be processed (self and other). In this case, the more time available to encode the information, the more efficient the participants' responses becomes.

### 4.2.3. Discussion of Experiments 6a and 6b

Overall, the results of Experiments 6a and 6b showed that, relative to the durations used in Experiment 5, shortening and lengthening the time that participants have to encode the information did not affect the self-perspective inhibition cost or the self/avoidance-desire cost. Together, these results suggest that the costs occurred both when participants constructed a mental representation and when they were holding it in mind.

## 4.3. Experiment 7: Perspective cue first

In all previous experiments, participants had to build and maintain two mental representations, one for their own desire and one for their opponent's desire. This was necessary because they did not know in advance which perspective they needed to judge. It is possible that the advantage for the self-perspective over the other-perspective, observed particularly on approach-desire trials, resulted from the fact that building and holding two representations is cognitively demanding and participants gave more weight to their own perspective. In Experiment 7, participants first saw the perspective cue, before viewing both players' coloured cards. Hence, they knew in advance which perspective they needed to attend to. Since only one mental representation needed to be built and maintained in memory, in this case, then it was expected that the perspective consistency effect would reduce. Furthermore, when judging the other person's perspective, participants would have now a clear opportunity to strategically ignore their own perspective and put equal importance to the other person's perspective. The question is whether participants can easily do this and whether hence the perspective effect (the self-perspective advantage) would reduce in Experiment 7. Finally, it was not expected that presenting the perspective cue first would make the self-avoidance cost disappear. If anything it could increase the cost. This would occur if presenting the self cue first makes the self-perspective even more salient on that trial and hence increases the self-avoidance cost.

## 4.3.1 Method

### *Participants*

Twenty psychology students from the University of Birmingham (2 males, $M_{age}$ = 19 years) took part in exchange for course credit via the research participation scheme. All participants had normal to corrected vision and none were colour blind.

### *Stimuli and procedure*

The method was identical to Experiment 5 (Chapter3) in all respects (see Figure 23) except that the perspective cue was displayed first, followed by the display containing the players' coloured cards. Thus, participants knew in advance which perspective they needed to attend to before seeing the coloured cards.

## 4.3.2. Results

The data were analysed as for Experiment 5 (Chapter 3). Here, 3 participants were excluded from the study because of their low performance rate 50 % or less correct in at least one experimental condition (overall accuracy range: 51–63 %) and then replaced to keep the total number of 20. The RT data was then explored to see whether it was normally distributed. The skew and kurtosis were within tolerable ranges to use parametric tests without transforming the data or removing outliers (overall, skew: 1.18; kurtosis: 1.24).

A 2x2x2x2 repeated-measure ANOVA analyses was computed with desire type (approach vs. avoidance), perspective consistency (different, vs. same) and perspective (self vs. other) as independent variables. Second, the dependent measures were either RT or accuracy (proportion of correct response). Only the data of matching trials (i.e., "yes" responses) was taken into account. For RT analysis, responses omitted due to the timeout procedure (2% of the data) were eliminated from the data set and erroneous responses (11% of the data) were not taken into account in the analyses.

In addition, a between-subject analysis was conducted to directly compare the data with those in Experiment 5.

### RT Analysis

In line with previous findings, there was a significant main effect of desire type ($F(1,19) = 15.69$, $p < .001$, partial $\eta^2 = 0.45$), RTs were slower when participants judged an avoidance (717 ms) compared with an approach (617 ms) desire. There was a significant main effect of perspective consistency ($F(1,19) = 5.84$, $p = .026$, partial $\eta^2 = 0.24$), with participants being slower at judging the different-perspective trials (679 ms) compared to the same-perspective trials (655 ms). A significant main effect of perspective was also found ($F(1,19) = 13.48$, $p < .005$, partial $\eta^2 = 0.42$). Participants were slower at judging the opponent's desire (683 ms) compared with their own desire (651 ms).

No other effects were significant (all $F$s < 1.81, all $p$s > 0.19; see Table 13 for overall mean).

Table 13. Overall mean for RT of Experiment 7 across all experimental conditions.

| | **Same** | | **Different** | |
|---|---|---|---|---|
| | **Self** | **Other** | **Self** | **Other** |
| **Approach** | 576 ms | 645 ms | 609 ms | 637 ms |
| **Avoidance** | 702 ms | 698 ms | 718 ms | 751 ms |

### *Accuracy analysis*

The ANOVA revealed a significant main effect of desire type $F(1,19) = 14.04$, $p = .001$, partial $\eta^2 = .43$) with participants being more accurate when judging an approach (92 % correct) than an avoidance (84 % correct) desire. However, there was no significant main effect of either perspective consistency ($F(1,19) = 3.72$, $p = .069$, partial $\eta^2 = .16$) or perspective ($F(1,19) = 3.3$, $p = .085$, partial $\eta^2 = .15$).

No other effects were significant (all Fs < 2.69, all ps > 0.12; see Table 14 for the overall mean).

Table 14. Overall mean for accuracy of Experiment 7 across all experimental conditions.

| | Same | | Different | |
|---|---|---|---|---|
| | **Self** | **Other** | **Self** | **Other** |
| **Approach** | 0.98 | 0.91 | 0.93 | 0.88 |
| **Avoidance** | 0.85 | 0.84 | 0.84 | 0.84 |

### *Comparison between Experiment 5 & Experiment 7*

In order to further measure the impact of presenting the perspective cue in advance, the results of Experiment 3 and 5 were directly compared, by entering in the ANOVA analysis, the experiment (5 vs. 7) as between-subject variable.

The RT analysis showed no significant effect of experiment ($F(1, 19) = 0.01$, $p = 0.96$, partial $\eta^2 = .00$). The results did show a significant Perspective Consistency x Experiment interaction ($F(1, 19) = 5.24$, $p = 0.03$, partial $\eta^2 = 0.12$; see Figure 39). The perspective consistency effect was numerically larger in experiment 3 (74 ms advantage for same-perspective trials) than experiment 5 (24 ms advantage for same-perspective trials), however there was no effect of experiment on either same-perspective trials ($t(38) = 0.48$, $p = 0.63$) or different-perspective trials ($t(38) = 0.35$, $p= 0.73$).

There was also a significant Perspective x Experiment interaction ($F (1, 19) = 4.65$; $p = 0.04$, partial $\eta^2 = 0.11$; see Figure 40). The perspective effect was numerically larger in Experiment 3 (73 ms advantage for self-perspective trials) than Experiment 5 (32 ms advantage for self-perspective trials), but the effects of experiment was not reliable on either self-perspective trials ($t(38) = 0.50$, $p= 0.62$), or other-perspective trials ($t(38) = 0.4$, $p= 0.74$).

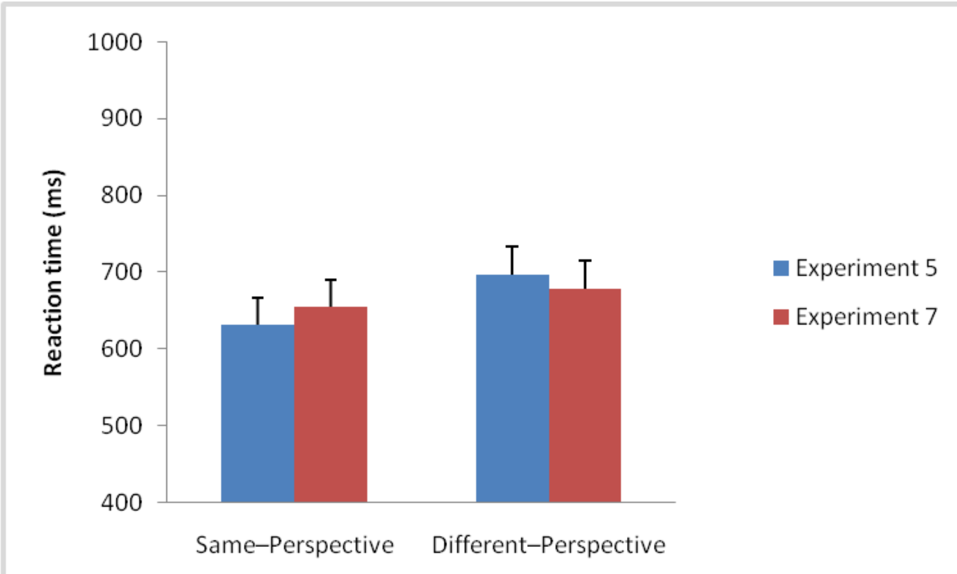



Figure 39. Overall mean (and standard error) for RT of Perspective Consistency x Experiment interaction.

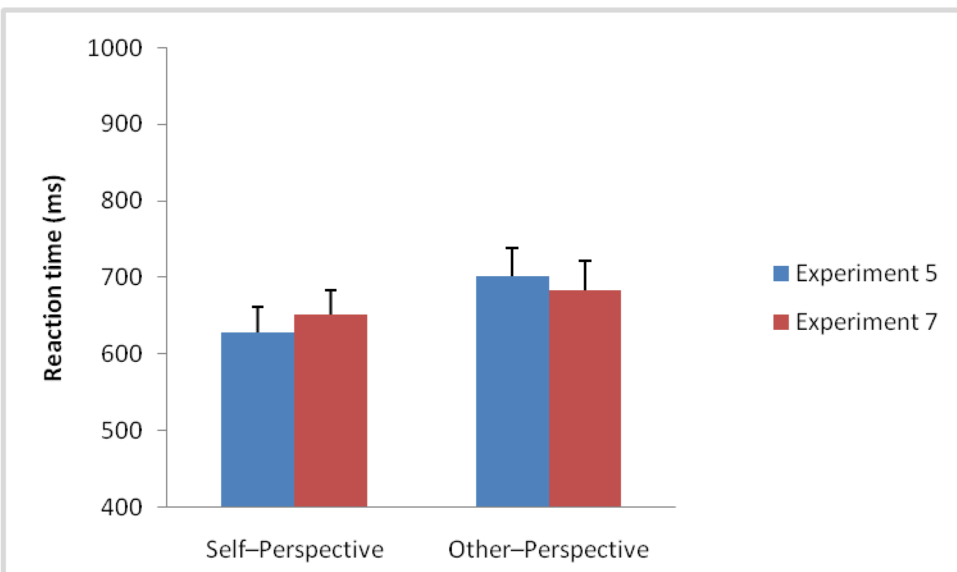



Figure 40. Overall mean (and standard error) for RT of Perspective x Experiment interaction.

No other significant effects were found (all $F$s < 1.56, all $p$s > 0.22). Importantly, it should be noted that no desire type x perspective x experiment interaction ($F(1, 19) = 1.47$, $p = 0.23$, partial $\eta^2 = 0.04$) was found.

Finally, the analysis of accuracy showed no main effect of experiment (between subject factor ($F(1, 19) = 3.30$, $p = 0.08$, partial $\eta^2 = 0.08$) and no significant interactions (all $F$s < 2.39, all $p$s > 0.13).

## Discussion

The effect of perspective on approach-desire trials and the effect of desire type on self-perspective trials were replicated in Experiment 7, showing evidence for a self-perspective inhibition cost and self/avoidance-desire cost. This occurred here even though participants knew in advance which perspective they had to judge (i.e., even when only one mental representation needed to be built and maintain in memory). The direct comparison of Experiment 5 and 7 showed, however, that the perspective effect tended to be reduced in Experiment 7. Thus, to some extent, prior cueing of the perspective helped participants to resist interference from their own perspective and/or put more weight on the other person's perspective. The perspective consistency effect was also reduced, showing that it was easier for participants to ignore the irrelevant perspective under the pre-cueing condition (Experiment 7). In contrast to these results, the self-avoidance cost was not modulated by the presentation of the cue in advance. This might be because the self-approach mode might be the default even when advanced information is presented.

## 4.4. General discussion

In this chapter the experiments further investigated two interference effects in desire reasoning in healthy adults. The focus was particularly on two measures: (1) the difference between self- and other-judgments for approach-desire trials only (measuring the self-perspective inhibition cost uncontaminated by the avoidance-desire cost) and (2) the difference between approach- and avoidance-desire judgments for self-perspective trials only (measuring the avoidance-desire cost uncontaminated by the self-perspective inhibition cost).

A first series of experiments (Experiment 6a and 6b), examined the processing stages at which the costs occurred (i.e., the stage when participants built a mental representation vs. the stage of maintaining that representation in working memory). This was investigated by comparing performance when the display duration was that used previously (Experiment 5) with performance when the time to encode the display was either shortened (Experiment 6a) or lengthened (Experiment 6b). The results showed that the self-perspective inhibition cost and the avoidance-desire cost were present irrespective of the time given to encode the information. Thus, the self-perspective inhibition cost and self-avoidance cost seemed to be present both at the stage of building a mental representation and at the stage of holding this information in mind to subsequently formulate a judgment. The fact that the self-perspective inhibition cost was unaffected by the manipulation of the display durations is consistent with the findings from Apperly and colleagues (2008) and suggests that not only in the case of belief reasoning but also in the case of desire reasoning, effortful control is needed throughout processing to resist interference from one's one perspective. Contrary to my expectation, the avoidance-desire cost was also present irrespective of the display duration. This suggests that participants did not immediately transform avoidance desires into approach desires.

It could be argued that the display durations chosen (500 ms vs. 2000 ms) were not ideal to distinguish between building a mental representation and holding information in mind. The times were chosen arbitrarily and probably do not fit the exact boundary between the two types of processes. However, the fact that there were significant differences between the 1000 ms and 2000 ms display duration conditions (Experiment 6b) suggests that the 2000 ms condition may have captured at least the start of the maintenance of the mental representation in memory.

Whatever the duration of the display presentation, there was a perspective consistency effect, suggesting that interference occurred by the irrelevant card/perspective. At 500 ms and 1000 ms display durations, the consistency effect was stronger when participants judged their opponent's perspective but at 2000 ms the consistency effect was the same irrespective of which perspective was judged. This suggests that, during the building of the mental representation, the self-perspective may have been more salient than the other-perspective but that at longer DV, both perspectives may have become equally salient and interfered with one another. This could be either due to the fact that during the longer DV, participants had more time to pay attention their opponent's card (facilitating the encoding and building of the other-perspective) and/or that once both perspectives are built and held in memory; they are both available to interfere with one another. On this last argument there is not differential interference in memory, but there are differences in the weighting applied as items are encoded, so that the self-perspective is dominant at encoding.

In Experiment 7, unlike all previous experiments, it was not necessary for participants to build and hold in mind two mental representations. Participants knew in advance which perspective they had to judge on a particular trial and thus they could build and maintain just one mental representation. In those conditions, there was still a significant advantage for self- over other-perspective judgments, but the advantage was slightly reduced. Thus, participants were partly able to resist interference from their own perspective or they were able to give more weight to the other person's perspective. Interestingly, however, the effect of perspective did not totally disappear. This is in line with the self-enhancement effect (Cunningham et al., 2008), whereby self-related material is processed more efficiently. The effect of perspective consistency was also reduced but remained still significant. Thus again participants could more easily focus on one perspective (self or other) but not completely ignore the irrelevant perspective. This is probably not surprising, as, in the conflicting-perspective condition, two colour cards were presented in central vision and we would expect some interference from the irrelevant card. The self/avoidance-desire cost was not affected by the experimental manipulation, though. Indeed, the origin of the effect is the transformation of a self/avoidance-desire into a self/approach-desire and these transformations need to be done even when participants know in advance that they have to make a self-perspective judgment.

In conclusion, I highlight two different types of costs that seem to play an important role in adult desire reasoning. When participants had to process an approach desire, they were significantly faster in responding to their own desire than to another's desire. This could be a result of facilitation of self-related material, and/or the need to inhibit one's own perspective (both not mutually exclusive). Such an effect appears when participants build a mental representation and hold it in memory over time and the effect remains (but reduces) even when participants can strategically ignore their own perspective.

Secondly, participants were significantly faster when making self-perspective judgments in approach-desire relative to the avoidance-desire trials. This effect appeared in both constructing and holding such information in mind over time. The effect may reflect the high cognitive control required when a self/avoidance-desire is transformed into the resulting self/approach-desire.

The results of the experiments reported here do not speak to the nature of the processes that are at the origin of the costs observed. In order to directly test whether these costs really reflect inhibitory control, future studies could use the dual task methodology. I would expect for example, that, if healthy controls perform a concurrent task that also requires inhibitory control while they perform the desire reasoning task used in this chapter, the self-perspective inhibition cost and avoidance reasoning cost would increase. Furthermore, this increase in cost should be more significant when the concurrent task taxes specifically inhibitory processes than when it taxes other processes.

In Chapter 5, the self-perspective inhibition and avoidance-desire cost is tested further to examine whether these costs reflect the existence of functionally and neurally distinct processes. This will be investigated in neurological brain injured patients.

CHAPTER 5

# Dissociable inhibitory demands in desire reasoning:

# Evidence from patients with frontal lesions

## 5.1 Introduction

In the previous chapters, I identified two sources of executive control when adults reason about desires. Chapters 3 and 4 showed that healthy adults require more cognitive effort to judge someone else's desire, relative to their own desire, when reasoning about approach-desire trials (the so-called self-perspective inhibition cost), and that they require more cognitive effort to reason about avoidance in comparison to approach desires when making judgements about their own desires (the self/avoidance-desire cost). Chapter 2 suggested that some patients may have more problems with inhibiting their own perspective, whereas, other patients have more problems with reasoning about avoidance desire. These two separate costs may be linked back to distinct control processes required to deal, respectively, with (i) another's desire while ignoring one's own, and (ii) one's own desire to approach rather than avoid a stimulus. In this chapter, converging neuropsychological evidence is sought for the distinction between the two forms of executive control of desire. If these executive control process are distinct both at the functional and the neural level, we would expect to find a double dissociation in patients' performance as well as distinct lesion sites in patients suffering the selective deficit.

Four patients who participated in the experiments reported in Chapter 2 were further investigated here. Two of these patients (WBA and PW) showed large interference of their own perspective when judging someone else's false belief, and interestingly, both patients have lesions in similar brain areas, namely the right lateral prefrontal cortex. Thus, both patients were good candidates to show a selective deficit in self-perspective inhibition. However, in the desire reasoning tasks, WBA's self-perspective inhibition deficit was only apparent in the first task but not in the second task. PW showed self-perspective inhibition problems in the false-belief task but he showed apparently intact desire reasoning abilities across the two desire reasoning tasks. It is thus possible that the patients' deficit was specific to belief reasoning. In principle, a self-perspective inhibition deficit should however be observable on a desire reasoning task. Chapter 3 and 4 showed that even healthy adults show a measurable self-perspective inhibition cost in desire reasoning under particular conditions. In this Chapter, I tested neuropsychological patients using a similar desire reasoning paradigm to the one presented to healthy participants (Chapter 3 and 4). This required patients to judge their own desire and their opponent's, rather than to solely focus on one perspective. Including trials where patients have to make decisions about their own desire should increase the salience of their own perspective and therefore increase the demands in self-perspective inhibition when they are asked to judge someone else's desire. It is possible that a more general deficit in self-perspective inhibition might arise under this circumstance, affecting desire as well as belief reasoning.

The two other patients (GA and SP) showed problems reasoning about avoidance desires but not approach desires (see Chapter 2, Experiment 2). Patient SP only showed deficits in the avoidance-desire task and not in the false-belief task. This was not the case for patient GA, who showed deficits in both. However, the two patients were selected on the basis of their performance in the desire reasoning task only, as well as for sharing common lesioned brain areas that did not overlap with the lesion sites of patients WBA and PW. For this group of patients, the aim was to address whether a problem in avoidance-desire reasoning would emerge when the patients had to judge their own perspective, and whether this was distinct from any problem in inhibiting their own perspective (when reasoning about another's rather than their own beliefs). Chapter 3 and 4 showed that the avoidance-desire cost was especially noticeable in healthy adults on self-perspective trials, and the deficit might be specific to this condition here.

In Experiment 8 all four patients were presented with the exact same task—the same as that used with healthy adults in Chapters 3 and 4. The aim was to extract for each patient the "pure" measures of self-perspective inhibition cost (measured on approach trials only) and self/avoidance-desire cost (measured on self-perspective trials only). If distinct executive control processes are recruited to deal with each demand, then we may observe some patients having selective problems in once case (e.g., due to impaired self-perspective inhibition, in patients WBA and PF) and other patients having the deficit in the other situation (patients GA and SP with problems in avoiding their own positive desire).

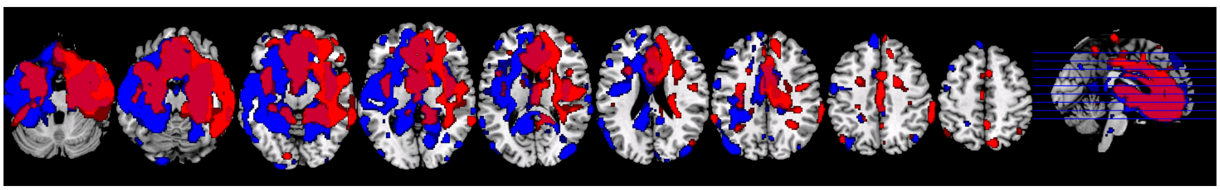
## 5.2. Experiment 8: Switching perspective and types of desires

### 5.2.1 Method

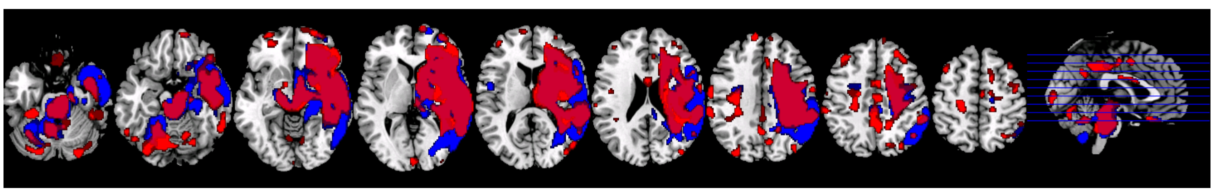
#### *Participants*

Patients GA, SP, PW and WBA participated in this study. Figure 41 shows the extent to which the patients' lesion sites overlapped[2]. GA and SP had common lesioned areas in the medial prefrontal cortex (top panel). Conversely, WBA and PW both sustained damage in areas found in the right prefrontal cortex (bottom panel). The common lesioned area in strict overlap between GA, SP, PW and WBA was the right insula (see Figure 42). The same five healthy aged matched controls (all male, $M_{\text{age}}$ = 62 years) also took part in the study (see Chapter 2).

[2] Scans were pre-processed by applying the advanced segment normalized procedure (Ashburner and Friston, 2005), all scans were transformed to standard MNI space using SPM5 (Statistical Parametric Mapping, Welcome Department of Cognitive Neurology, London UK). Scans were further used in VBM with the following SPM stats: one sample t- test with 3 covariates: healthy grey matter (201 brains aged 40+) vs patient grey matter, age & sex. Coloured areas denote uncorrected significant results (p<.001) with a threshold of 50voxels.



**Red = GA ; Blue = SP**



**Red = WBA; Blue = PW**

Figure 41. Lesion overlaps of the patients with common performance in the desire reasoning task. In the upper half, patients GA (red) and SP (blue), in the bottom half, patients WBA (red) & PW (blue)

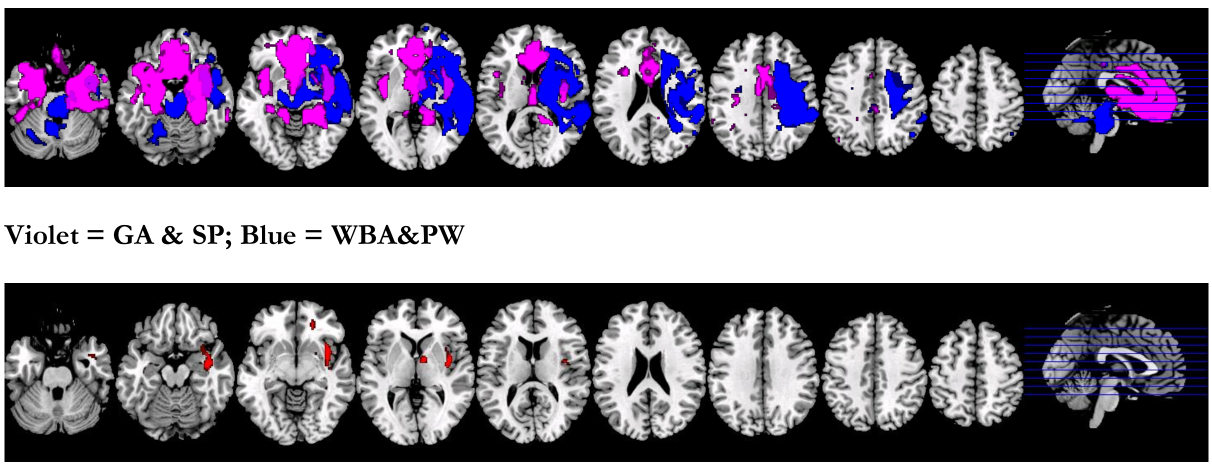



Figure 42. The top half displays the overlapped lesions of both profiles; GA & SP (Violet) and WBA & PW (blue). In the bottom half the image displays the only common area of damage found across all four patients.

### ***Stimuli and procedure***

The same methodology used for the healthy participants (e.g., see Chapter 3, Experiment 5) was used here. However, the duration of the stimuli presentation on screen was increased by 1000 ms in order to help the patients cope with the experimental demands of the task (see Figure 43). All dimensions of stimuli used in this study were identical to those in Chapter 3. For both approach- and avoidance-desire trials, on half of the trials, both the protagonist and the patient had different coloured cards, thus each would want a different coloured card as outcome (different-perspective trials). On the other half of the trials, both parties had the same coloured card, and thus both would want the same coloured card as outcome (same-perspective trials). On each trial, a cue indicated which perspective patients had to judge ("YOU WANTS" or "HE WANTS"). A colour card (either blue or red) was then presented as the perspective content (or desire content) to judge. All trials were divided into 4 blocks of 48 trials each, preceded by a practice block of 20 trials. Trials were presented in pseudo-random order to avoid more than 3 consecutive trials of the same type.

Patients were comfortably seated 60 cm away from the screen. Testing lasted approximately 1 hour. The participants were first explained the rules of the game and given 20 practice trials before commencing the study. At test, each trial began with a fixation cross that remained on screen for 1000 ms followed by a 2000 ms displayed weather counter. After this cue, the patient would then see two colour cards simultaneously presented in the centre of the screen on the table. Patients were told during the instructions that the card closest to each portrait was that player's coloured card (their own and their opponents). The two cards remained on the screen for 2000 ms and were followed by a 2000 ms perspective cue that indicated the perspective the patients needed to judge ("HE WANTS" or "YOU WANTS"). Within the first block, patients were specifically told in the instructions to respond to the following questions, respectively: "*Which colour does Henry want the next* *card to be*" or "*Which colour do you want the next card to be*?". Then, a single coloured card (blue or red) would appear where participants were asked to accept or reject the proposed perspective content. The coloured card remained visible until participants gave their response. Patients responded by pressing one of two keys of the numeric pad of the keyboard: "1" if the card matched the perspective content and "4" if the card did not match the perspective content. Stickers were used on the keyboard to remind which key to press: "yes" (=1) and "no" (=4). After patients made their response, a final frame depicting the actual identity of the third card and outcome of the round, was displayed. This frame was controlled by the examiner who also controlled the onset of the next trial by pressing the spacebar. This was done to ensure that patients were fully attending to the task before the next trial would be presented.

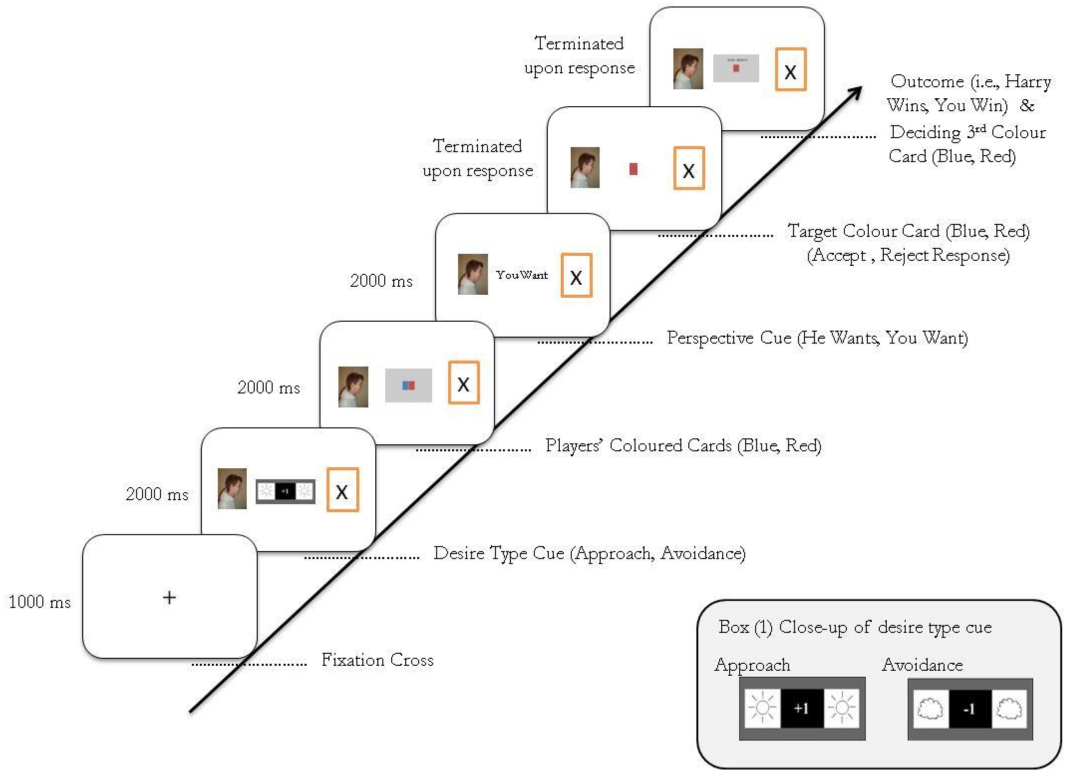



Figure 43. Experimental procedure displaying the sequence of events presented in order as shown on screen. Box (1) contains a close-up of both desire type cues (approach, avoidance).

### 5. 2. 2. Results

Unlike the experiments conducted with healthy participants in Chapters 3 and 4, only accuracy data were taken into account for the patients and their matched control participants. The data were separated to reflect the two indices of executive processing: (1) the self-perspective inhibition cost (the difference between self and other judgments for approach-different trials only) and (2) the self/avoidance-desire cost (the difference between approach- and avoidance-desire judgments for self-perspective same-perspective trials only).

*Z*-scores were computed based on patients' individual performance in each pure measure (self–perspective inhibition cost and avoidance-desire cost; see Figure 44 for overall results). For the first group of patients, WBA showed a significant self-perspective inhibition cost ($Z = 4.91$, $p < .001$, with more errors committed on approach trials when judging the opponent's perspective than his own perspective). Patient PW showed a similar pattern of performance, also showing a significant self-perspective inhibition cost ($Z = 3.76$, $p < .001$). The measure of avoidance desire revealed a significant cost for patient WBA ($Z = 2.86$, $p < .005$), with more errors committed on avoidance-desire trials in comparison to trials when an approach-desire response was required, when judging self-perspective trials. Numerically, the avoidance-desire cost was smaller than the self-perspective inhibition cost. For PW, the avoidance-desire cost was not significant Within the second group of patients, GA showed a significant avoidance-desire cost ($Z = 5.61$, $p < .001$), with more errors committed on self-perspective trials when he had to make an avoidance-desire response in comparison to an approach response. A significant avoidance-desire cost was also observed for patient SP ($Z = 5.36$, $p < .001$). Neither of these patients showed a significant self-perspective inhibition cost; both had similar levels of performance for self and other judgments on approach trials. For healthy aged matched controls, the self-perspective inhibition cost was significant for one out of the five participants: Control 5 ($Z = 5.36$ $p < .001$), with more errors committed on approach trials when judging the opponent's perspective compared with their own perspective. None of the control participants showed a significant avoidance-desire cost.

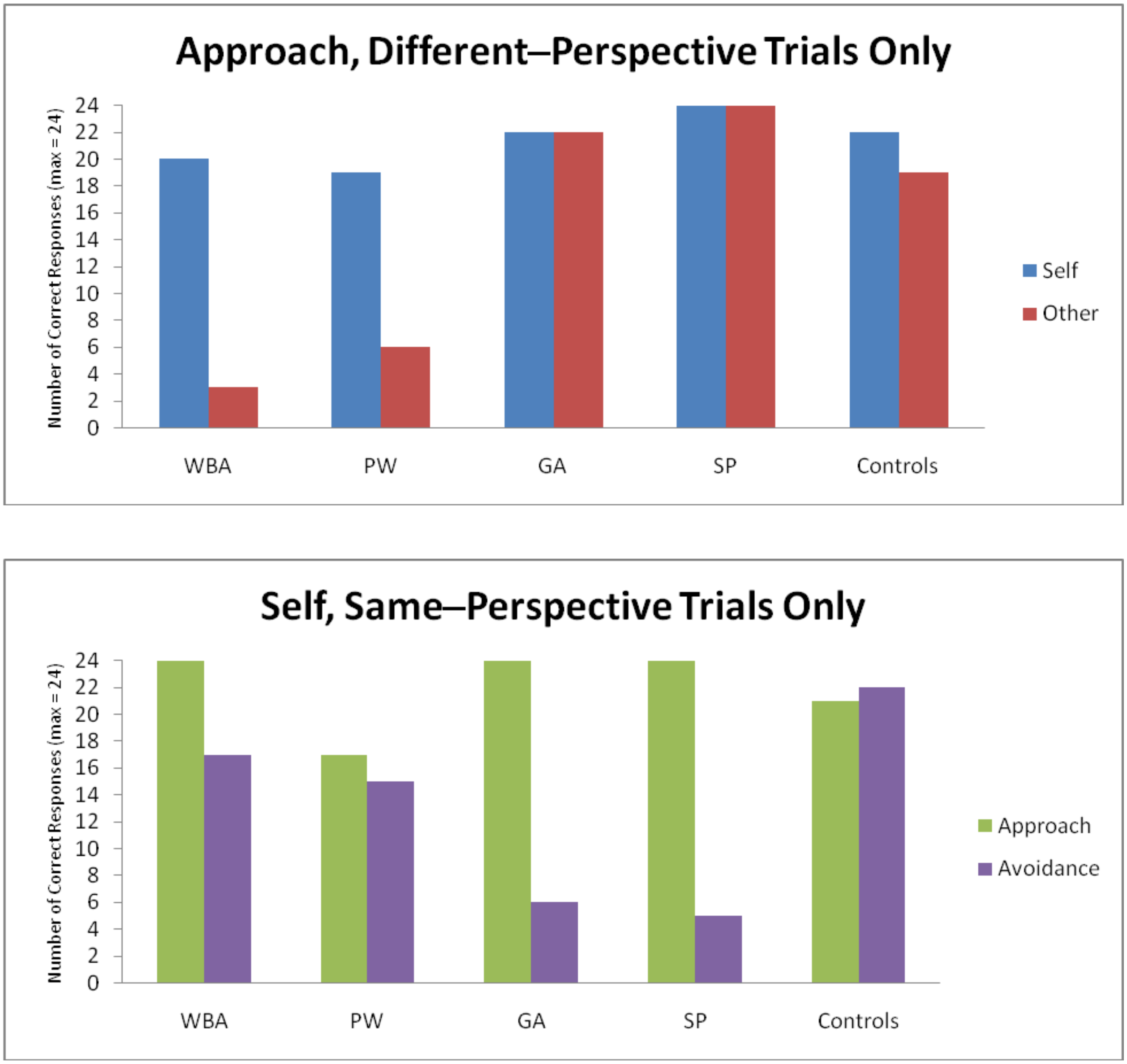



Figure 44. Patients' (GA, PW, SP, and WBA) performance on the self-perspective inhibition level (top panel) and avoidance-desire level (bottom panel). Also, complementary score of the average found across five controls.

### *Discussion*

Patients WBA and PW were hypothesized to have a selective deficit in self-perspective inhibition. In line with this hypothesis, both patients made significantly more errors when judging their opponent's desire than their own desire on approach-desire trials. Compared with the data reported in Chapter 2, increasing the self-perspective inhibition demands in the desire reasoning task made apparent a deficit even in a desire reasoning task. Patient WBA also showed a significant avoidance-desire cost, but this was reliably smaller then for the medial frontal patients (see below). Patient PW showed no significant difference in his ability to reason about his own avoidance and approach desires.  Thus, both WBA and PW showed a disproportionate deficit in inhibiting their own perspective compared to their ability to reason about avoidance desires.

Patients GA and SP were hypothesized to have a specific deficit in reasoning about avoidance desires. Both patients performed significantly worse when required to make avoidance-desire responses compared to their own approach responses. In contrast, neither patient had any difficulty in ignoring their own approach desire to infer someone else's approach desire, showing thus no signs of self-perspective inhibition deficit. In their case, raising the self-perspective inhibition demands of the task did not affect their performance. Here both patients showed the opposite profile to the one shown by WBA and PW.

In Chapter 2, when there was no time pressure for participants to give their response, healthy control participants performed 100 per cent correct across all conditions. However, within this experimental design, participants made several errors and one participant (Control 5) showed a significant self-perspective inhibition cost. These results indicate that even for healthy participants, this particular paradigm was cognitively taxing. In comparison, avoidance-desire trials were much easier to deal with for the healthy controls. The healthy controls in this experiment were aged-matched to the patients and represented an older population than those who participated in Chapters 3 and 4. There is substantial evidence that executive cognitive control declines with age and this could have influenced these participants' performance on the task (e.g., German & Hehman, 2005). Notably, Bailey and Henry (2008) compared young and older participants' performance on tasks with low and high self-perspective inhibition demands. Their results showed that older participants were impaired on trials requiring high self-perspective inhibition processes. The controls did not have problems with making an avoidance-desire rather than an approach-desire response when performing the task from their own perspective, in clear contrast to patients GA and SP.

There could be two explanations for the apparent dissociation in some of the healthy control participants. Firstly, it is possible that age has a larger effect on the executive processes necessary to inhibit one's own perspective. The results reported so far suggest that inhibiting one's own perspective and reasoning about avoidance-desire require distinct types of executive control processes and it is thus possible that age has differential effects on both types of processes. Alternatively, it is possible that with age, the enhancement of processing of self-related items diminish, reducing thereby the executive demands of reasoning about avoidance desires.

### 5.3. Experiment 9: Follow-up study for patients WBA & PW

WBA and PW's self-perspective inhibition deficit in the above desire reasoning task contrasts with their good performance when judging someone else's desire when they have to take the other persons's perspective across the entire task (cf. Chapter 2). One critical difference between the two sets of tasks is the need to switch between two alternating perspectives. This may have facilitaed patients performance, where they could simply focus on the detective's coloured card and play the game based on the information provided, bypassing perspective taking altogether. However, it is also possible that other factors have contributed. In Chapter 2, patients were presented with their coloured card first, followed by their opporonet's coloured card. In Experiment 8 here, both coloured cards were presented simutaneously. Perhaps in Chapter 2, it was easier to remember their opponent's card, due to recency effects as this was presented last. This concern is addressed in Experiment 9 here, where I kept the same experimental procedure and card distribution as in Chapter 2, however, I now introduced a second perspecitve to monitor within the task.

### 5.3.1 Method

#### *Participants*

This was a follow-up study designed for patients WBA and PW who both took part.

#### *Stimuli and procedure*

Similarly to the paradigm used in Chapter 2 (Experiment 2), patients were presented with the card game called "blame it on the weather" and asked to play against a story character named Simon, displayed in short animated clips in Multimedia Flash 8 (see Figure 45). All stimuli used in this study were identical to those in Chapter 2 (Experiment 2). However, unlike the previous experiment, the test questions now directly probed the patients' desire (i.e., "*which colour do you want the next card to be?"*) and their opponent's desire (i.e., "*which colour does Simon want the next card to be?*"). This order of the self and other desire questions was pseudo-randomized throughout the task, so as not to have three consecutive identical responses.

Testing lasted approximately 2 hours and was split in two separate sessions. Participants were seated 60 cm away from the screen. After their consent was given, they proceeded to the instructions of the game followed by 4 test trials. At the beginning for each new trial the amount of money at stake and desire type cue in play (a sun symbol for the winning trials and a cloud symbol for the loosing trials) were displayed. The card dealer then dealt the patient's coloured card to the bottom half of the screen. The coloured card remained visible for 500 ms and was then turned over (front to back). This was followed by the distribution of Simon's coloured card (top half of the screen) that remained visible for 500 ms and was also turned over. The cards were turned over to avoid potential matching strategies later at test. The scene was paused at this instance and the examiner asked the test question ("*which colour does Simon want the next card to be*?" or "*which colour do you want the next card to be?"*). Once a response was given the examiner pressed a button to continue the animation which revealed the colour of the third card. The winner or loser of that round was declared and the scoreboards on screen updated accordingly. This was then followed by a new trial.

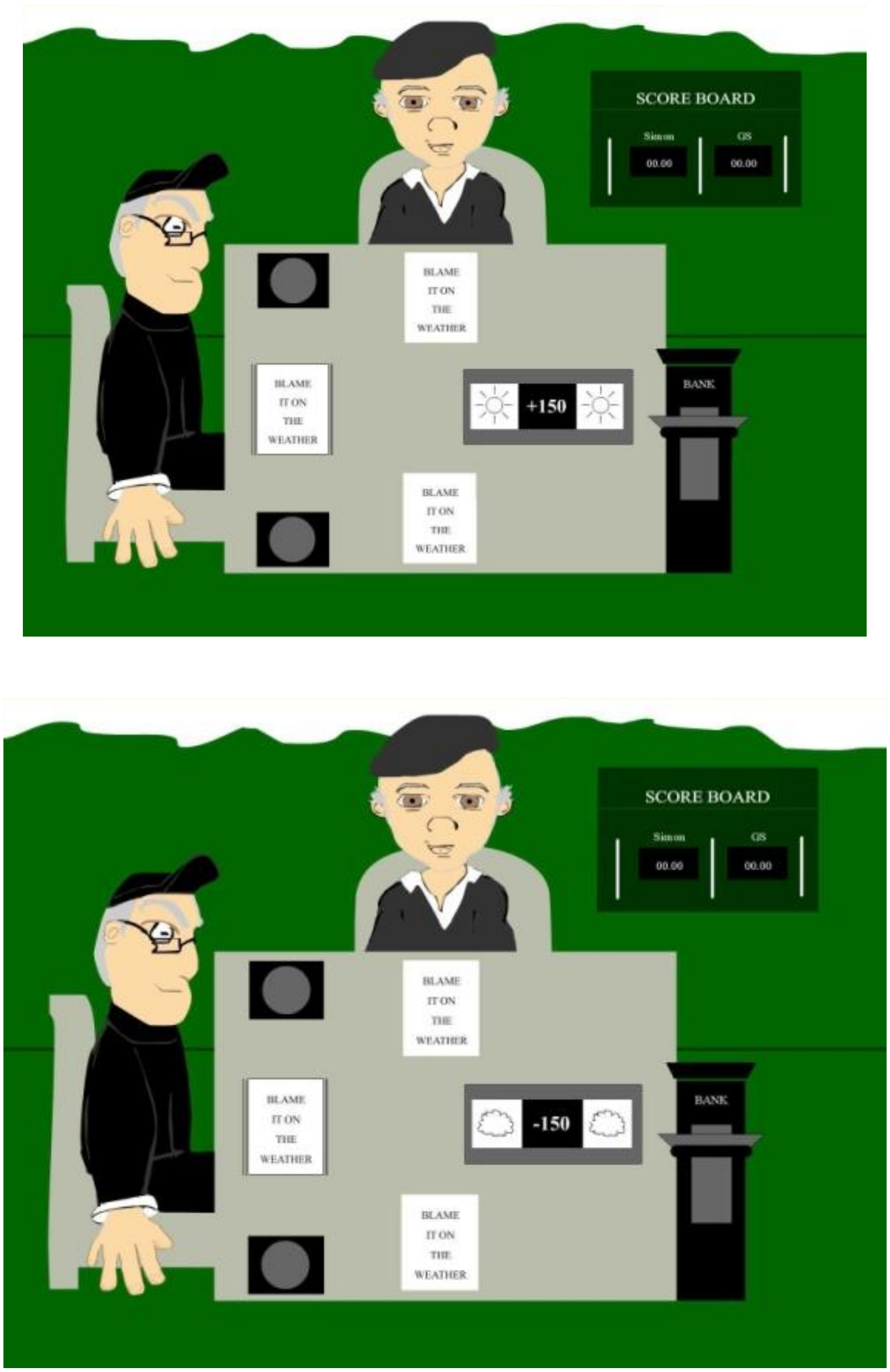



Figure 45. Visual setup with approach-desire trials on the left hand side depicted with a sun and avoidance-desire trials on the right hand side depicted with a cloud symbol.

### 5.3.2 Results

I conducted a similar analysis on accuracy scores as the one conducted in Experiment 8 here above, and specifically report below results obtained for the two executive processes of interest. Unlike Experiment 1, there were 12 correct responses possible for each condition. *Z* Scores were computed based on patient's individual performance in each pure measure (self-perspective inhibition cost and avoidance-desire cost; see Figure 46 for overall graph). There was a significant self-perspective inhibition cost for both WBA ($Z = 3.14$, $p < .001$) and PW ($Z = 4.08$, $p < .001$). Both patients were less error prone when asked to judge their own perspective relative to their opponent's perspective. Both patients performed closed to ceiling for approach and avoidance desires and showed no significant avoidance-desire cost.

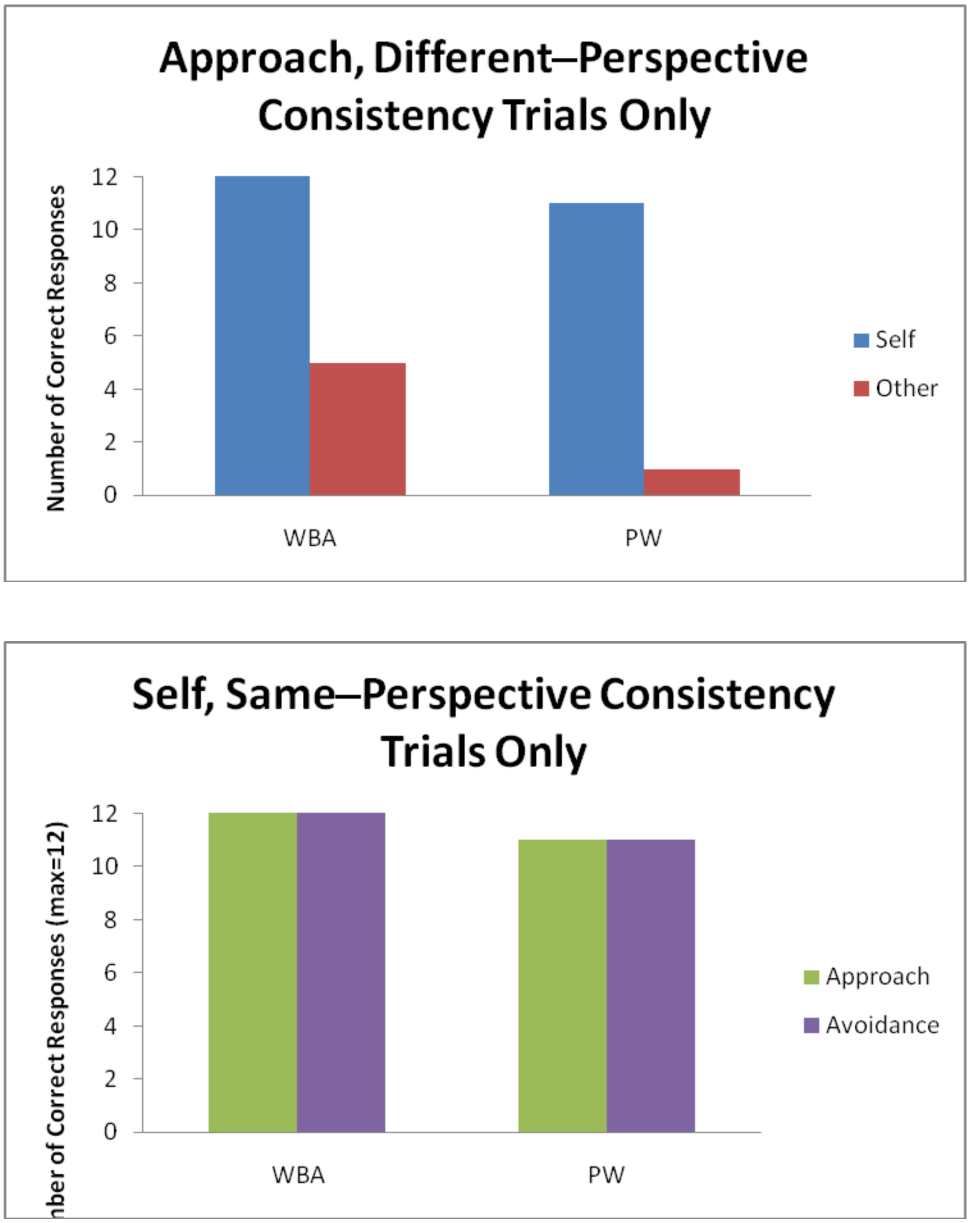



Figure 46. Performance scores of patient PW and WBA for self-perspective inhibition costs (top panel) and avoidance-desire costs (bottom panel).

### *Discussion*

WBA and PW both made far more errors when asked to judge the other's perspective relative to their own perspective on the approach trials but they showed no difficulties in reasoning about their own avoidance desires. Thus, the visual changes (card distribution) from Experiment 1 to Experiment 2, did not affect the patients' global performance. This further supports the interpretation that it is the shifting of perspectives, which was absent in the tasks presented in Chapter 2 but which were now present in the two experiments reported here, that explains the discrepant results with Chapter 2.

## 5.4. Experiment 10: Follow-up study for patients GA and SP

In order to further investigate the origin of GA's and SP's difficulties with avoidance-desire reasoning, I conducted two additional experiments. It is possible that both patients have problems with the executive processes required to negate or shift attention away from the object to be avoided (as it has been hypothesized at the beginning of this chapter). On the other hand, however, it could be argued that both patients have a conceptual deficit and do not understand the concept of avoidance or even loss. Finally, it is also possible that GA and SP are unable to maintain two rules (the loss and win rules) in mind while performing the task and that they therefore always fall back on the easiest rule (the win rule with approach desires). These possibilities were assessed here.

### 5.4.1. Experiment 10a: Avoidance-desire trials only

Here, both patients were presented with just avoidance-desire trials and they were only asked to make judgments from their own perspective (i.e., "*which colour do you want the next colour to be*?"). If GA and SP have problems understanding the concepts of loss and avoidance, they should be unable to perform this task. If they have problems negating or shifting attention away from the object to be avoided, they could still show some deficits in this task as well, but it is also possible that removing the need to shift between approach and avoidance desires reduced the executive demands of negating or shifting away from the object avoided in a similar was as removing the need to shift between perspectives helped WBA and PW to inhibit their perspective (Chapter 2, Experiment 2).

#### 5.4.1.1 Method

##### *Participants*

This was a follow-up study designed for patients GA and SP.

##### *Stimuli and procedure*

The task was similar in structure to that used in Chapter 2, except that it only included avoidance-desire and self-perspective trials. The stimuli involved were identical to those in Chapter 2 (Experiment 2). Unlike the previous experiment, participants started the game with a sum of £1000 and were told that the winner would be the person with the most money left in the bank.

#### 5.4.1.2. Results

Both patients performed either at ceiling (SP) or close to ceiling (GA) when asked to reason about avoidance desires only (see Figure 47). Both patients thus had a good understanding of the concepts of loss and avoidance. Removing the need to shift between approach and avoidance desires facilitated thus their performance either because this reduces the executive demands of negating or shifting attention away from the object to be avoided, or because it reduced the demands of keeping two rules in mind while performing the task. These possibilities were examined further in Experiment 3b.

### 5.4.2. Experiment 10b: Approach- and avoidance-desire trials

This study examined SP and GA's memory capacities when both approach- and avoidance-desires were presented within the same task. When several rules need to be applied in a task, it is often observed that frontal patients fail to implement one or several of the rules, despite the patients maintaining knowledge of the rules throughout the task. This profile has been described as "goal neglect" (Duncan et al. 1996). The problem is not that patients forget some rules but they fail to take them into account or implement them; there is a problem in on-line implementation, even if the rules can be recalled subsequently. This is usually demonstrated by asking patients to recall the rules of a task during and after having performed the task. Patients with goal neglect fail to implement certain rules while performing the task but are able to report in details all the rules of the task during or at the end of the task. A similar procedure was used here. GA and SP were presented with the mixed approach- and avoidance-desire trials and they were prompted to report the rules of the game.

#### 5.4.2.1. Method

***Participants***

Again only GA and SP took part.

### *Stimuli and procedure*

In Experiment 10b, the same paradigm was used as in Chapter 2, with both approach- and avoidance-desire trials. The test question referred to the protagonist of the game (i.e., "*which colour does Simon want the next colour to be*?"). The experimenter initially explained the rules of the game to the patients. In order to proceed to the main study, the patients had to correctly respond to six consecutive practice trials that covered the different experimental conditions. However if a response was judged incorrectly a set of 6 new practice trials was reissued to the patient. During these practice trials, the patients were shown the coloured cards for both players, and the patients had to make a decision about the third coloured card. Also, the patients were specifically asked to report the given outcome of the game, based on the information on screen, once the final deciding card was revealed. This was to ensure that the patients understood the two different rules involved in the game and were able to respond correctly to a series of responses. Once a patient was able to respond correctly, they underwent the same experiment as in Chapter 2. During the game, the examiner asked the patients to report and tell them what the rules of the game were and what each symbol represented. This occurred at the beginning, half way through the study and at the end upon completion of the game.

### 5.4.2.2. Results

Patient GA scored 6 consecutive correct responses after 5 total attempts (each consisting of 6 practice trials). Patient SP required 2 sets of 6 practice trials before commencing the experiment. *Z* Scores were computed based on patient's individual performance in each pure measure (self-perspective inhibition cost and avoidance-desire cost; see Figure 47) both patients started to make errors again for avoidance-desire trials and both showed a significant avoidance-desire cost (GA, (*Z*

= 4.14, $p < .001$); SP, ($Z = 3.80$, $p < .001$). Even though both patients showed an avoidance-desire cost on errors, both were able to report the rules of the game at the beginning, middle and upon termination of the task.

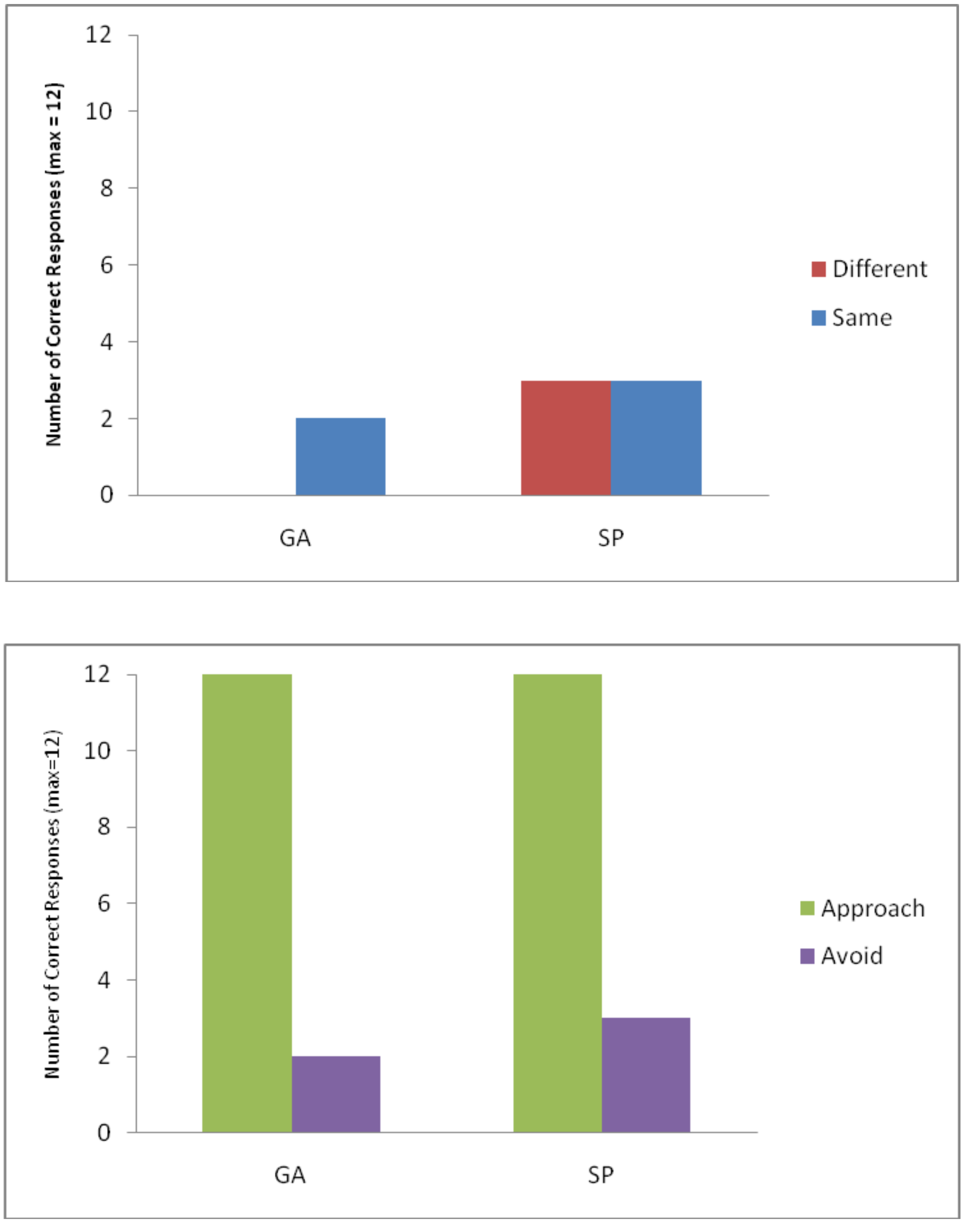



Figure 47. Number of correct responses for both patients GA and SP. The top panel depicts Experiment 10 a and the bottom panel depicts Experiment 10 b.

### 5.4.3. Discussion

GA and SP were able were able to correctly judge an avoidance desire when this was presented in isolation (Experiment 10a). However, by raising the demands of the task and introducing approach as well as avoidance desires there was a drop in performance that was specific to the avoidance-desire trials (Experiment 10b). Thus, both understood the rules of the game, they could apply these rules when solely asked to focus on one, and they could remember the rules through the task. However, when both rules had to be maintained within the task set, the patients responded throughout by applying the approach rule only. This suggests a form of goal neglect that emerges in social decision making when a desire has to be avoided rather than approached. There appears to be a failure to implement the rules on-line, even though the rules remain available for explicit report). It may be that there is an implicit default desire action, which is to make an approach-desire response—and to enact an avoidance-desire response requires extra activation of avoidance-desire rule within the task set. If medial frontal damage disrupts the ability to implement this additional top-down activation, then the de-fault rule may be used on-line even though the avoidance-desire task goal remains available when the patient is directly asked.

## 5.5. General Discussion

The evidence reported in this chapter shows a clear double dissociation between two groups of patients (WBA and PW on the one hand, and GA and SP on the other hand). Patients WBA and PW have substantial difficulties inhibiting their own perspective but fewer problems reasoning about avoidance desires. In contrast, GA and SP have problems reasoning about avoidance desires but not in inhibiting their own perspective. In addition to the dissociation in the patients' behavioural performance, the patients' lesion sites overlap within but not across groups. WBA and PW have overlapping lesions in the right lateral prefrontal cortex whereas GA and SP have more medial prefrontal cortex lesions (in addition to temporal lesions).

In all four patients the deficits seem to be executive in nature and susceptible to manipulations that increase or decrease the cognitive effort required. Patients WBA and PW had particular problems when the tasks required them to shift perspectives. In those conditions, the salience of their own perspective may have been increased and there may have been increased demands too to inhibiting this perspective when it was not relevant. Patients GA and SP had problems particularly when the task required them to shift between approach- and avoidance-desire trials that occurred across both self- and other-perspective. In conditions requiring a response to an approach desire, the salience of the card received may have increased because it was correct to want this card on some of the trials at least. This could have made it more difficult to shift attention away from the received card. Alternatively, by making the win/approach rule (which is also the easiest rule) relevant to some trials, this may become part of the task set. If this is the default set, there may then need to be increased activation of the alternative rule (the avoidance-desire rule) to enable it to compete in implicitly directing action. In the absence of top-down activation to this novel rule (following medial pre-frontal damage), the default rule may dominate; there is neglect of the avoid goal for on-line control of behaviour, even though this goal was still explicitly represented. With the data reported here it is not possible to disambiguate these two alternatives. However, irrespective of which of these two fine-grained interpretations is correct, the results suggest that the origin of GA's and SP's deficit is executive in nature and of a different kind to the deficit shown by WBA and PW.

In conclusion, this Chapter provides supporting evidence for the existence of functionally and neurally distinct executive processes in desire reasoning: those required for inhibiting one's own perspective and those required to reason about avoidance desire. The implications of these findings will be discussed in the next chapter.

# CHAPTER 6

## General Discussion

The current thesis aimed at examining the functional and neural mechanisms underlying belief and desire reasoning in adult TOM. Compared to belief reasoning, very little is known about the processes underlying desire reasoning. Thus, a substantial part of the current thesis examined desire reasoning in both neurological brain injured patients and healthy adults. In the General Discussion here, I will summarise the main findings of the thesis and then discuss what it tells us about the cognitive and neural basis of belief and desire reasoning.

## 6.1. Summary of empirical chapters

The first empirical chapter (Chapter 2) aimed at directly comparing the performance of patients with brain damage across belief and desire reasoning tasks in order to explore the patterns of association and dissociation in performance. The patients had lesions to brain areas that were previously associated with ToM, including medial frontal cortices and the temporo-parietal junction. Novel tasks were created so that the belief and desire reasoning tasks could be matched in terms of self-perspective inhibition demands. Studies from the developmental literature have shown that dissociations across belief and desire reasoning in children can be explained by the lack of matching of the self-perspective inhibition demands across both types of tasks (Moore et al., 1995). It was thus important to overcome this methodological flaw. On the critical trials of the belief reasoning task, participants had to infer someone else’s false belief (and hence they had to inhibit their own knowledge of the real state of the world). On the critical trials of the desire reasoning tasks, they had to infer someone else’s conflicting desire (and thus they had to inhibit their own desire). In one of the desire reasoning tasks (Experiment 1), participants reasoned about the desire to approach an object whereas in the other desire reasoning task (Experiment 2), they had to reason about the desire to either approach or avoid an object. This latter task placed higher reasoning demands and was less easy to solve by means of a superficial strategy.

Interestingly, all combinations of patterns of association and dissociation in performance were observed. There were several cases of association of performance: some patients showed spared performance across belief and desire reasoning tasks (PH, RH and PF) whereas other patients showed impaired performance across both types of reasoning. This was the case for WBA (although his problem with desire reasoning only occurred in the first desire reasoning task) and patient GA (in his case, the impairment in desire reasoning was only observed when the task required reasoning about avoidance desires). There were also several cases of dissociation of performance: some patients were impaired in the false-belief reasoning task but not the desire reasoning task (DS, JeB, MP and PW) whereas another patients was impaired in the desire reasoning task (particularly when the task required to reason about avoidance desires) but not the false-belief reasoning task (SP). No strong conclusion should be drawn from that single study, but the observation of what look like double dissociations, despite the fact that belief and desire reasoning tasks were matched in terms of self-perspective inhibition demands, provides some indication that there may be unique processes or processing demands (beyond self-perspective inhibition) associated with, on the one hand, false-belief reasoning and, on the other, desire reasoning.

The fact that some patients found it harder to reason about false beliefs than conflicting desires is in line with the general idea in the developmental literature that belief reasoning is harder than desire reasoning (Bartsch & Wellman, 1989; Perner, 1991). More surprising is the finding that some patients found it harder to reason about conflicting desires than false beliefs. In those latter cases, it is particularly the reasoning about avoidance desires that was difficult. Reasoning about avoidance desire has been described in the developmental literature as requiring inhibitory processes to direct attention away from the object to be avoided (Friedman & Leslie, 2004a, 2004b, 2005;

Leslie et al., 2004; Leslie et al., 2005; Leslie & Pollizi, 1998; Leslie & Thais, 1992) and this may have been the origin of the patients' difficulties.

There is some evidence that false-belief reasoning is cognitively demanding for healthy adults (Apperly et al., 2008) but there is no evidence that desire reasoning (in particular reasoning about avoidance desires) is difficult for healthy adults as well. Thus, investigating desire reasoning in healthy adults was the logical next step and the focus of the following experiments (Experiments 3 to 7) reported in Chapter 3 and 4.

A new desire reasoning paradigm was created in which healthy adults had to reason about desires under time pressure in the context of a card game. Time pressure was introduced to ensure that the task was sensitive enough to detect processing costs. Furthermore, the paradigm was designed so that two processing costs could be measured: the processing cost associated with the processing of someone else's desire compared to one's own desire (a cost that could be linked to self-perspective inhibition) and the processing cost associated with the processing of an avoidance desire compared to an approach desire (a cost that could be linked to the need to direct attention away from the object to be avoided).

Across all experiments (3 to 7) of Chapters 3 and 4, there was evidence of a processing cost linked to avoidance-desire reasoning. This is in line with the suggestion of Leslie and collaborators (Friedman & Leslie, 2004a, 2004b, 2005; Leslie et al., 2004; Leslie et al., 2005, Leslie & Pollizi, 1998; Leslie & Thais, 1992) that avoiding an object requires to direct attention away from the object to be avoided. It is also in line with research investigating the processing costs of negation (Macdonald & Just, 1989; Kaup et al., 2001; Kaup & Zwaan, 2003; Kaup et al., 2006; Kaup et al., 2007). Interestingly, the avoidance-desire processing cost was greatest when participants reasoned about their own avoidance desire. Indeed, it was particularly high when participants were actively involved in the task (Experiment 3 active involvement condition) and even more so when their involvement was emphasized by the use of their photograph in the task (Experiments 5 to 7). One explanation is that, in the conditions where participants had to reason about their own avoidance, they had to avoid an object that was previously associated with the self. To explain this more concretely, on self-avoidance trials, participants would first receive a coloured card which was presented as "their card". Previous research has shown that the processing of self-related items is enhanced compared to items that are not self-related (Cunningham et al., 2008). It follows that the colour associated with a participant's own card would be very salient in working memory. The avoidance-desire trials, however, required participants to respond with a different colour to that associated with the self. The inhibition of one's own desire may be particularly effortful, disrupting performance in this last condition. Experiment 6a and 6b further showed that the avoidance processing cost was higher when the time given to participants to encode the information was shortened (Experiment 6a) compared to when the time to encode the information was lengthened (Experiment 6b). Thus, the avoidance-desire reasoning cost occurred quite early in processing when participants were building a mental representation of the desire (rather than when they were simply maintaining the representation in working memory). In other words, the avoidance processing cost may be specifically linked to the processing stage where participants transformed the avoidance desire (e.g., not wanting the blue card) into an approach desire (e.g., wanting the red card instead).

Across all experiments (3 to 7) of chapters 3 and 4, there was also evidence for a different processing cost linked to the inhibition of one's own perspective when healthy adults reason about desires. This is in line with studies in developmental psychology which showed that young children have self-perspective inhibition difficulties not only in false-belief reasoning tasks but also when reasoning about conflicting desires (Moore et al., 1995). The self-perspective inhibition cost was observed only when participants were directly involved in the task (Experiment 3, active involvement condition and Experiments 5 to 7) and seemed to occur both when participants were building a representation in mind and when they simply held it in working memory (Experiment 6a and 6b). Finally, the cost was observed even when participants knew in advance which perspective they had to judge (Experiment 7), indicating that the other-perspective was processed even when it was not relevant to the tasks.

The results of Chapter 3 and 4 thus showed there are two sources of processing costs when healthy adults reason about desires: one type of cost is linked to the need to inhibit one's own perspective when we reason about other people's desires; the other type of cost is related to the reasoning about avoidance-desire independently of perspective taking. The observation of separate processing costs in healthy adults could provide a plausible explanation for some of the dissociations observed in Chapter 2 in the patients' performance. In particular, those patients who had problems with reasoning about avoidance desires but not false beliefs may have difficulties directing attention away from the object to be avoided (a process of mental negation) but not necessarily with self-perspective inhibition. It could then be also expected that patients who have problems in inhibiting their own perspective (e.g., patient WBA) should not necessarily have problems when reasoning about avoidance desires (there were hints that this may be the case in Chapter 2).

Chapter 5 (Experiments 8 to 10b) was aimed at directly testing whether the processing of avoidance desire dissociates from processing linked to self-perspective inhibition in brain-damaged patients. The patients selected where those who, based on their performance in Chapter 2, were good candidates for a dissociation between self-perspective inhibition and avoidance-desire reasoning. The patients were presented with the same paradigm as the one used with healthy adults and a clear double dissociation could be observed: patients WBA and PW had many more problems inhibiting their own perspective than reasoning about avoidance desires; patients GA and SP showed the opposite profile of performance. Interestingly the two groups of patients had also lesions in different brain areas. The data supported the argument that self-perspective inhibition (in the particular context where participants have to switch between self- and other-perspective taking) is associated with the right lateral prefrontal cortex, whereas reasoning about avoidance (in the particular context where participants have to switch between approach- and avoidance-desire reasoning) is associated with more medial areas of the prefrontal cortex. The fact that the impairments were most prominent under conditions of switching perspective or desire type further highlights the executive nature of the deficits.

### 6.2. Implications for the cognitive and neural basis of belief and desire reasoning

Researchers have only just started to unravel the cognitive and neural basis of belief and desire reasoning. Neuroimaging studies with healthy adults as participants have highlighted a network of brain areas that are consistently activated when participants reason about other people's mental states. Two regions in particular are consistently activated: the medial prefrontal cortex and the temporo-parietal junction (Saxe et al., 2004). These two areas seem to be involved in both belief and desire reasoning suggesting that there are at least some overlapping areas between the two types of reasoning. Until very recently, there had been no direct comparisons between belief and desire reasoning in adults. However, one recent study has compared event-related brain potentials in adults while they reason about other people's beliefs or desires (Liu, Meltzoff, & Wellman, 2009). Participants had to reason either about two other people's divergent beliefs (where one person thinks that object A is in a box whereas the other person thinks that object B is in the box) or desires (one person likes the object that is in the box whereas the other person does not like the object that is in the box). The authors found overlapping brain activity in mid frontal brain areas for both belief and desire reasoning but a more specific right posterior activity when participants were reasoning about beliefs.

The above finding is in line with claims from the developmental literature that belief reasoning is harder for children than desire reasoning and that belief reasoning requires some extra processing steps. However, as highlighted by some authors (Moore et al., 1995; Russell, 1996), the desire reasoning tasks usually chosen by the authors are often very easy and this alone can explain the difference in performance. In other words, there are some desire reasoning tasks that are harder than others, and these are the ones that should be chosen when comparing belief and desire reasoning. Two harder conditions for desire reasoning have been highlighted by researchers: conflicting-desires tasks (where participants need to inhibit their own desire in order to infer the other person's desire; Moore et al., 1995) and avoidance desires (where participants need to disengage from the object to be avoided; Friedman & Leslie, 2004a, 2004b, 2005; Leslie et al., 2004; Leslie et al., 2005; Leslie & Pollizi, 1998; Leslie & Thais, 1992). It is worthwhile noting that in the study by Liu et al. (2009), participants were asked to reason about avoidance desires but the self-perspective inhibition demands were quite low. In fact, it could be argued that the self-perspective inhibition demands were lower in the desire than belief condition. In the belief condition, although participants were simply reasoning about two other people's beliefs about the content of a box, participants themselves were made aware of the content of the box (and had thus to disengage from their own knowledge of the content of the box). In contrast, in the desire reasoning task, participants were not probed about their own desire and because some of the objects were not really relevant to adults (e.g., half of the objects were toys), participants may not have spontaneously processed whether they themselves liked the object or not. Thus, it cannot be excluded that the additional activation found for belief reasoning was linked to the fact that the belief and desire reasoning tasks were not properly matched in terms of self-perspective inhibition.

As illustrated with the above issues, in order to match or contrast belief and desire reasoning tasks, it is important to understand the collection of processes that are required to solve the tasks. It is only with a fine-grained task analysis that progress will be made in understanding the cognitive and neural basis of belief and desire reasoning.

### 6.2.1. Identifying the processes involved in belief and desire reasoning

There have been several suggestions about the separate processes that may be involved in belief reasoning. At the core is knowledge about beliefs. One needs to understand that people's beliefs do not necessarily match the real state of the world and this is part of one's ToM knowledge. There is also the need to *use* that knowledge when reasoning about someone else's belief. Here a distinction has been made between the need to inhibit one's own knowledge of the real state of the world (Moore et al., 1995; Samson et al., 2005) and the need to monitor the environment or past events for relevant cues on the basis of which a content can be given to someone else's belief (Samson et al., 2007).

Similarly, there is some conceptual knowledge required to reason about desires but there are also processes required to implement that knowledge in reasoning. They may be the need to inhibit one's own desire (Moore et al., 1995), in the case of avoidance reasoning, there is the need to disengage from the object to avoid (Friedman & Leslie, 2004a, 2004b, 2005; Leslie et al., 2004; Leslie et al., 2005, Leslie & Pollizi, 1998; Leslie & Thais, 1992), and it can be argued that one would also need to monitor the environment (including past events) for relevant cues to give a content to someone else's desire.

In the following sections, I will discuss the empirical findings in Chapter 2 to 5 in relation to the different processes highlighted above.

### 6.2.2. Knowledge about beliefs and desires

Knowledge about beliefs and desires may be quite different. Some authors have argued that knowledge about desires is acquired earlier than knowledge about beliefs, and that beliefs, unlike most types of desires, require understanding the representational nature of mental states (Wellman & Cross, 2001; Wellman, Cross, & Watson, 2001). Thus, it is possible, in principle, that both types of knowledge are represented in different areas of the brain. Patients who are impaired in belief and/or desire reasoning may in fact have lost some conceptual knowledge about beliefs and/or desires. Furthermore a double dissociation between belief and desire reasoning could suggest that knowledge about beliefs and desires is represented in different areas of the brain.

Seven of the patients reported in this thesis showed an impairment in belief and/or desire reasoning. For four of these patients (GA, SP, WBA and PW), we can exclude a conceptual deficit. Patients GA and SP had problems in avoidance-desire reasoning but only when they had to switch between approach- and avoidance-desire reasoning. When there was no switching demand, the patients could reason about avoidance desire. They had thus still knowledge about what an avoidance desire is. Similarly, patients WBA and PW could reason about their own desires, even avoidance desires, so they had no conceptual deficit for desires. Previously published data also showed that WBA had spared knowledge about beliefs, because in false-belief tasks where the demands in self-perspective inhibition were reduced, he had no problems anymore (Samson et al., 2005)

For patients DS, JeB and MP, who all had problems in the belief reasoning but not desire reasoning, we cannot exclude that they have lost some conceptual knowledge about beliefs. The patients would need to be tested with tasks that directly assess their conceptual knowledge to address this issue. It is worthwhile noting that these patients did not have lesions in the same brain areas. Patients JeB and MP had overlapping lesions encompassing the right TPJ and extending into the right prefrontal cortex whereas DS had left frontal lesions. Thus perhaps only one of the two subgroups had a conceptual deficit. As I will describe below, there are also other possible explanations for the dissociation observed between belief and desire reasoning.

### 6.2.3. Monitoring relevant cues

The cues that need to be monitored to give content to someone else's belief or desire are quite different. As explained in Chapter 1, beliefs arise from perceptual observations (e.g., what someone has or has not seen) whereas desires in many cases arise from observations/inferences about physiological states or people's preferences. In addition to these qualitative differences, ToM tasks also differ in terms of the salience of the cues provided. In the particular context of the tasks used in the current thesis, there were important differences in the salience of the cues across the belief and the desire reasoning tasks. In case of the belief reasoning task, patients had to track what the detective had or had not seen whereas in the desire reasoning task, patients had simply to take into account of what colour card the other person had received. Furthermore, in most false-belief tasks, detecting what the other person had or had not seen is quite easy because the protagonist leaves the scene and is physically not present when the state of the world changes. However, in the false-belief task used in Chapter 2, determining whether the detective had seen the robber's displacement required much more fine-grained processing. Firstly, in the false-belief task the detective stayed on the screen and computation of his line of sight was required to indicate whether he saw the robber or not. Secondly, the reason why the detective had not seen the robber is because a salient event attracted the detective's attention away from the robber. Presumably, this distracting event also attracted the attention of the participants and therefore participants had less time left to compute the line of sight of the detective. In other words, the belief reasoning task was clearly more demanding than the desire reasoning task in terms of the processing of the relevant cues. This difference could explain why some patients (DS, JeB and MP) were impaired in the belief reasoning task but not the desire reasoning tasks. As I mentioned earlier, it worthwhile keeping in mind that the three patients had not all the same lesions and that therefore they may not all have the same underlying deficit.

In previous published studies, it is the TPJ (particularly on the left), that has been associated with the monitoring of relevant cues in the environment in ToM tasks (Samson et al., 2007). But there is also evidence outside the domain of ToM that the TPJ could play a critical role in monitoring the environment for relevant cues. Firstly, some studies have shown that the TPJ plays a critical role in directing attention towards or away of salient external stimuli. This processing would be important for detecting which cues are relevant and which ones are not relevant for the particular desire or belief reasoning task at hand. For example, Mevorach and colleagues (2005) showed that the left parietal region plays a crucial role in biasing attention away from irrelevant but salient objects in the external world whereas other studies showed the right parietal region guides attention towards salient relevant objects (Mevorach et al., 2006; Hodsoll et al., 2008).

Secondly, the TPJ has been associated with processing gaze direction and computing what someone can or cannot see. This processing is particularly relevant for belief reasoning but can be also relevant for some cases of desire reasoning (e.g., when a person shows his or her preference by looking at an object; Baron-Cohen, 1995). For example, several studies have shown that the superior temporal sulcus (STS) is activated when participants process gaze direction or what someone else can see (Aichhorn et al., 2005; Allison, Puce, & McCarthy, 2000; Pelphrey et al., 2004). Akiyama and colleagues (2006) reported a single case study of a patient who sustained lesions to the right STS and who showed specific deficits in discriminating gaze direction.

Two of the three patients (JeB and MP) who showed impairments in the belief reasoning task but not the desire reasoning task had a lesion encompassing the right TPJ. These patients may have had problems computing the line of sight of the detective and/or paying attention to the relevant events in the detective story. The third patient (DS) had lesions to the left prefrontal cortex and it is possible that he was not automatically cued to the location gazed at by the detective, which may have prevented the computation of what the detective had or had not seen. In healthy participants, it is usually shown that observing someone else's averted gaze automatically directs attention to the location the other person gazed at (Driver et al., 1999). A study by Vecera and Rizzo (2004), however, reported the case of a patient with frontal lobe damage who did not show this automatic gaze cueing effect. This could be further tested by directly assessing the DS's ability to process gaze direction and his knowledge about what someone else can or cannot see.

As I discussed in Chapter 2, the patients with left TPJ lesions (PF and RH) had no problems in the belief and desire reasoning tasks. This contrasts with their difficulties reported in previous published work (Samson et al., 2004, 2007). It is possible that because these patients had been extensively tested on various false-belief paradigms in the past five years, they had no difficulties anymore identifying any of the relevant cues in the false-belief task that I designed.

It would be interesting to test JeB, MP and PW, who showed a dissociation between belief and desire reasoning tasks, along with the patients with left TPJ lesions, on a different type of desire reasoning task that is better matched to belief reasoning tasks in terms of the demands in monitoring the relevant cues in the environment. If the patients' problem is a general problem in directing attention to relevant features in the task, they should be impaired in more demanding desire reasoning tasks. However, if the problem is specific to the cues that are mostly relevant to belief reasoning (like computing the line of sight or computing what someone else can or cannot see), then the patients should not show impairments in desire reasoning tasks that do not require monitoring these types of cues.

### 6.2.4. Self-perspective inhibition and disengaging from a salient object

It is interesting to note that in a false-belief task, the inhibition of one's own knowledge of the state of the world may in fact relate to two distinct inhibition processes. Disengaging from one's own knowledge of the state of the world means that we have to disengage from one's own perspective but also disengage from a salient object in the real state of the world. In typical false-belief reasoning tasks, the two types of disengagement cannot be easily disentangled but they can be disentangled in desire reasoning tasks as seen in Chapters 3 to 5.

The results of Chapter 5 showed that two of the patients had specific problems with the inhibition of their own perspective (patients PW and WBA). Both patients had overlapping lesions in the right lateral prefrontal cortex, a region that has been associated with self-perspective inhibition in an fMRI study as well (Vogeley et al., 2001). The right lateral prefrontal cortex, particularly the right inferior frontal gyrus, has also been previously associated with general response inhibition abilities (Aron et al., 2004). However, the contrast between PW's and WBA's difficulties in inhibiting their own perspective in the face of their relatively spared ability to reason about avoidance desires suggests that their inhibition deficit is not entirely domain-general as the patients are able to disengage their attention from an salient object in the external world.

Two other patients reported in the current thesis, GA and SP, showed the opposite profile to the one displayed by WBA and PW. Their problems seem to be more related to avoidance reasoning, particularly when they had to switch between approach and avoidance trials. One hypothesis is that these patients could not easily disengage from a salient object in the external world. This deficit was observed independently of perspective taking, as it was seen even when the patients reasoned about their own desires. Interestingly both GA and SP had overlapping lesions in distinct areas (more medial and ventral areas of the prefrontal cortex) to the ones affected in the case of WBA and PW.

Although the two types of process (disengagement from one's own perspective and disengagement from a salient object in the environment) can be disentangled in a desire reasoning task, it seems that both are intrinsically linked in false-belief reasoning tasks. This could, at least in part, explain why false-belief reasoning is so hard and often harder than desire reasoning. As a consequence, I would expect that patients who have a deficit in any of the two types of processes would also have difficulties in the false-belief task. This was the case for patients WBA, PW and GA but not for patient SP. Thus, there may have been some features in the false-belief task that I used that helped SP to disengage from the location where the robber was hiding. Or alternatively, the avoidance-desire reasoning used in my studies placed particularly high demands in terms of disengagement from a salient object. SP's deficit may be noticeable only when the task is sensitive enough.

The dissociation observed between self-perspective inhibition and the disengagement of a salient object in the environment suggests that the inhibitory processes involved in mental state reasoning are not unitary. However, these inhibitory processes may not be specific to reasoning about mental states. The dissociation observed could reflect a separation between domain-general processes for resisting interference from irrelevant internally generated thoughts on the one hand, and for resisting interference from irrelevant external objects on the other hand (a similar distinction has been proposed by Gilbert et al. 2005). A failure to resist interference from irrelevant internally generated thoughts would not only impair performance on ToM problems but also performance on other problem solving tasks. In those tasks, patients could persevere with irrelevant solutions they generated. Likewise, a failure to resist interference from irrelevant external information would not only impair the reasoning about avoidance desire but also any other non-social reasoning in which distracting salient object have to be ignored. Future studies could test these hypotheses by directly contrasting the patients' performance on social and non-social reasoning tasks.

## 6.3. Conclusions

When comparing belief and desire reasoning it is important to consider the collection of processes that are involved in both types of reasoning. Whereas researchers have started to look at the cognitive processes involved in belief reasoning, little work has looked at the cognitive processes involved in desire reasoning in adults. In this thesis, I provide the first empirical evidence that desire reasoning can be effortful for healthy adults, particularly in two situations: when they have to inhibit their own desire to take into account someone else's desire and when they have to reason about avoidance desires and disengage their attention from the object to be avoided. I have shown that the processes at the origin of these two processing costs are dissociable at the functional and neural level. This was demonstrated most clearly with the reports of a double dissociation in brain-damaged patients' performance. Both types of process (disengaging one's attention from one's own perspective and disengaging one's attention from a salient object in the environment) seem to be involved in false-belief reasoning. When directly comparing behavioural performance or brain activation across desire and belief reasoning tasks, it is thus important to match both types of tasks in terms of these processing demands. As discussed in this Chapter, there are also other types of process that need to be considered when investigating the cognitive and neural basis of belief and desire reasoning, and these include conceptual processing and processing of the relevant cues in the environment to give a content to someone else's mental state. Investigating the extent to which desire and belief reasoning differ along those other processes will be a worthwhile programme for future research.

## REFERENCES

Aichhorn, M. Perner, J. , & Kronbichler, M. (2006). Do visual perspective tasks need theory of mind? *NeuroImage 30*, 1059–1068.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 4*, 267–278.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic*? Psychological Science, 17(10)*, 841–844.

Apperly, I. A., Samson, D., Carroll, N., Hussain, S., & Humphreys, G. W. (2006). Intact 1st and 2nd order false belief reasoning in a patient with severely impaired grammar. S*ocial Neuroscience, 1*, 334–348.

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and left temporoparietal contributions to theory of mind: Neuropsychological evidence from a false belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience, 16*, 1773–1784.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2005). Domain-specificity and theory of mind: Evaluating neuropsychological evidence. *Trends in Cognitive Sciences, 9*, 572–577.

Aron, A.R., Robbins, T.R., & Poldrack, R.A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*, 170–177.

Ashburner, J. & Friston, K. J. (2005). Unified segmentation. *Neuroimage, 26,* 839-851.

Bailey, P.E. & Henry, J. D. (2008). Growing Less Empathic With Age: Disinhibition of the Self Perspective. *Journal of Gerontology: Psychological Sciences*, *63B (4),* 219–226.

Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, B., Costa, D., & Ell, P. (1994). Recognition of mental state terms: Clinical findings in children with autism and a functional neuroimaging study of normal adults. *British Journal of Psychiatry, 165,* 640–649.

Baron-Cohen, S. (1995). Mindblindness: An essay on autismand theory of mind. Cambridge, MA: MIT Press.

Bartsch, K., & Wellman, H. M. (1989). Young Children's Attribution of Action to Beliefs and Desires. *Child Development*, 60 (4), 946–964.

Bartsch, K., & Wellman, H. M. (1995). Children talk about the mind. New York: Oxford University Press.

Berthoz, S., Armony, J.L., Blair, R.J.R., & Dolan, R.J. (2002) An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain*, *25 (8)*, 1696–1708.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18(5)*, 382–386.

Bird, C. M., Castelli, F., Malik, O., Frith, U., & Husain, M. (2004). The impact of extensive medial frontal lobe damage on 'theory of mind' and cognition. *Brain, 127*, 914–928.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77(1)*, B25–31.

Burgess, P.W., Alderman, N., Evans, J., Emslie, H., Wilson, B.A. (1998). The ecological validity of tests of executive function. *Journal of International Neuropsychological Society*, *4*, 547–558

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72(4)*, 1032–1053.

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

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Cassidy, K. (1998). Three- and four-year-old children's ability to use desire- and belief- based reasoning. *Cognition, 66*, B1–B11.

Cunningham, S.J., Turk D.J., Macdonald ,L.M., Macrae, C.N. (2008). Yours or mine? Ownership and memory. *Consciousness and Cognition, 17*, 312–318.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6(5)*, 509–540.

Duncan, J., Emslie, H. & Williams, P. (1996). Intelligence and the Frontal Lobe: The Organization of Goal-Directed Behavior. *Cognitive Psychology*, *30*, 257–303 .

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327–339.

Ferguson, F.J., Sanford A.J., Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Research.*

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57 (2)*, 109–128.

Friedman, O., & Leslie, A. M. (2004a). Mechanisms of belief-desire reasoning: Inhibition and bias. *Psychological Science, 15,* 547–552.

Friedman, O., & Leslie, A. M. (2004b). A developmental shift in processes underlying successful belief-desire reasoning. *Cognitive Science, 28*, 963–977.

Friedman, O., & Leslie, A. M. (2005). Processing demands in belief-desire reasoning: Inhibition or general difficulty? *Developmental Science, 8*, 218–225.

Frith, C.D., Frith, U., 1999. Interacting minds - A biological basis. *Science, 286*, 1692–1695.

Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalising. Philos. Trans R. Soc., *B 358*, 685–694.

Fodor, J.A., 1992. A theory of the child's theory of mind. *Cognition, 44*, 283–296.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358,* 459–473.

Gallagher, H.L., Frith, C.D., 2003. Functional imaging of theory of mind. *Trends in Cognitive Science*. *7*, 77–83.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, *38(1)*, 11–21.

German, T.P., & Hehman, J.A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition*,*101*, 129–152.

Gilbert, S.J., Frith, C.D., Burgess, P.W. (2005). Involvement of rostral prefrontal cortex in selection between stimulus-oriented and stimulus-independent thought. *European Journal of Neuroscience*, *21*, 1423–1431.

Gopnik, A. (1996). The Scientist as Child. *Philosophy of Science, 63(4)*, 485–514.

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, 57*, 36–37.

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron- Cohen, S., Hodges, J.R., (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. *Brain, 125*, 752–764.

Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., Dolan, R., Frackowiak, R., Frith, C., (1996). Theory of mind in the brain. Evidence from a PET scan study of Asperger syndrome. *NeuroReport, 8*, 197–201.

Hodsoll, J., Mevorach, C. & Humphreys G.W. (2008). Driven to Less Distraction: rTMS of the Right Parietal Cortex Reduces Attentional Capture in Visual Search. *Cerebral Cortex*, 1–9.

Kaup, B., Lüdtke, J., Zwaan, R.A., 2006. Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed. *J. Pragmat. 38*, 1033–1050.

Kaup, B., Yaxley, R.H., Madden, C.J., Zwaan, R.A., Lüdtke, J., 2007. Experiential simulations of negated text information. *Q. J. Exp. Psychol. 60*, 976–990.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89 (1)*, 25–41.

Kircher, T.T.J, Senior,C., Phillips,M.L., Benson, P.J., Bullmore,E.T., Brammer,M., Simmons, A., Williams,S.C.R., Bartels, M., & David, A.S. (2000).Towards a functional neuroanatomy of self processing: effects of faces and words. *Cognitive Brain Research, 10 (1-2)*, 133–144.

Leslie, A. M. (1987). Pretense and representation: The origins of 'theory of mind'. *Psychological Review, 94,* 412–426.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences, 8*, 528–533.

Leslie, A. M., German, T. P., & Pollizi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, 50,* 45–85.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science, 1,* 247–254.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*, 225–251.

Liu, D., Meltzoff, A. N., & Wellman, H. M. (2009). Neural Correlates of Belief- and Desire-Reasoning. *Child Development*, *80(4)*, 1163-1171.

MacDonald, M.C., Just, M.A., 1989. Changes in activation levels with negation. *Journal Experimental Psychology Learning Memory and Cognition, 15*, 633–642.

McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci.* U. S. A. *98*, 11832–11835.

Mevorach, C., Humphreys, G.W., & Shalev L (2006). Opposite biases in salience-based selection for the left and right posterior parietal cortex. *Nature Neuroscience*, *9 (6)*, 740–742.

Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., & MacCallum, F. (1995). Conflicting desire and the child's theory of mind. *Cognitive Development, 10*, 467–482.

Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience*, *23(17)*, 6819–6825.

Perner, J. (1991). Understanding the representational mind. Cambridge, MA: MIT Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515–526.

Premack, D., &Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515–526.

Overwalle, F. (2008). Social Cognition and the Brain: A Meta-Analysis. *Human Brain Mapping.* 30, 829–858.

Rakoczy, H. (2009). Executive function and the development of belief-desire psychology. *Developmental Science*, *in press.*

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*, 12–21.

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. *Brain, 124,* 600–616.

Russell, J. (1996). *Agency: Its role in mental development.* Hove, Sussex, England: Psychology Press.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2008). Seeing it their way: What other people see is calculated by low-level and early acting processes. *Journal of Experimental Psychology: Human Perception and Performance*.

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, *7*, 499–500.

Samson, D., Apperly, I. A., & Humphreys, G.W. (2007). Error analyses reveal contrasting deficits in 'theory of mind': Neuropsychological evidence from a 3-option false belief task. *Neuropsychologia, 45(11)*, 2561–2569.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: A case of a selective deficit in inhibiting self-perspective. *Brain, 128*, 1102–1111.

Samson, D. (2009). Reading other people's mind: Insights from neuropsychology. *Journal of Neuropsychology* , *3*, 3–16.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55*, 87–124.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in theory of mind. *Neuroimage*, *9(4)*, 1835–1842.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17 (8)*, 692-699.

Schneider, Z., Eschman, A., & Zuccolotto, A. (2002). E-Prime (Version 1) [Computer software]. Pittsburgh, PA: Psychology Software Tools.

Shamay-Tsoory, S.G. Shur, S., Barcai-Goodman L. , Medlovich,S., Harari, H. & Levkovitz, Y. (2007). Dissociation of cognitive from affective components of theory of mind in Schizophrenia. *Psychiatry Research*, *149*, 11–23.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10,* 640–656.

Stuss, D.T., Gallup Jr, G.G., Alexander, M.P. (2001). The frontal lobes are necessary for "theory of mind". *Brain, 124*, 279–286.

Vecera, S.P., & Rizzo, M., (2004) What are you looking at?: Impaired 'social attention' following frontal-lobe damage. *Neuropsychologia*, *42 (12)*, 1657–1665.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., et al., 2001. Mind reading: neural mechanisms of theory of mind and self-perspective. *NeuroImage*, *14,* 170–181.

Wimmer, H. M., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Wellman HM, Cross D. 2001. Theory of mind and conceptual change. *Child Development*, *72*, 702–707.

Wellman, H. M,  Cross, D., Watson J. (2001) Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72,* 655–84.

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*, 245–275.

Ziv, M., & Frye, D. (2003). The Relation Between Desire and False Belief in Children's Theory of Mind: No Satisfaction? *Developmental Psychology*, *39* (5), 859–876.