Anticipation, event-plausibility and scene constraints: Evidence from eye movements

Gitte Henssel Joergensen

PhD

Psychology

December 2011

ABSTRACT

We often use language to refer to items within our immediate proximity whereby the constraints of the visual context serves to restrict the number of possible referents, making it easier to anticipate which item will most likely be referred to next. However, we also use language to refer to past, future, or even imagined events. In such cases, anticipation is no longer restricted by the visual context and may now be influenced by real-world knowledge. In a set of eye-tracking experiments we explored the mapping of language onto internal representations of visually available scenes, as well as previously viewed scenes. Firstly, we were interested in how event-plausibility is able to influence our internal representations of described events and secondly, how these representations might be modulated by the nature of the visual context (as present or absent). Our findings showed that when describing events in the context of a concurrent scene the eye movement patterns during the unfolding language indicated that participants anticipated both plausible and implausible items. However, when the visual scene was removed immediately before the onset of spoken language participants anticipated plausible items, but not implausible items – only by providing a more constraining linguistic context did we find anticipatory looks to the implausible items. This suggests that in the absence of a visual context we require a more constraining linguistic context to achieve the same degree of constraint provided by a concurrent visual scene. We conclude that the conceptual representations activated during language processing in a concurrent visual context are quantitatively different from those activated when the visual context to which that language applies is absent.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first and foremost like to thank my supervisor, Gerry Altmann for his guidance, help and support throughout my PhD research. Also, many thanks to Silvia Gennari and the members of my research committee – Gareth Gaskell and Jelena Mirković for helpful discussions and advice.

DECLARATION

This thesis contains original work completed solely by the author under the supervision of Professor Gerry Altmann.
The data reported in chapters 2, 3, 4 and 7 has been presented at the following conferences:

Conference talks

Joergensen, G., & Altmann, G. *Anticipation, event-plausibility & scene constraints.* Talk given at the Experimental Psychology Society Meeting, Nottingham, UK, 2011.

Joergensen, G., & Altmann, G. *Prediction during situated language processing.* Talk given at the 17th Meeting of the European Society for Cognitive Psychology, San Sebastian, Spain, 2011.

Conference posters

Joergensen, G., & Altmann, G. *Anticipating moved objects: Event-plausibility matters, but not always…* Poster presented at the 15th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Barcelona, Spain, 2009.

Joergensen, G., & Altmann, G. *Anticipating moved objects: Event-plausibility matters, but not always…* Poster presented at the 23rd Annual CUNY Conference on Human Sentence Processing, New York, USA, 2010.

Joergensen, G., & Altmann, G. *Anticipating moved objects: Event-plausibility and context-dependence…* Poster presented at the 16th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), York, England, 2010.

Joergensen, G., & Altmann, G. *Anticipation, event-plausibility & scene constraints.* Poster presented at the 24th annual CUNY Conference on Human Sentence Processing, San Francisco, USA, 2011.

CHAPTER 1

INTRODUCTION

We are all aware that some events and actions are more likely to occur than others. For example, we know from experience that people are much more likely to place a glass on a table, than on a lamp. But how do we process and mentally represent such plausible and implausible events when we are not actually performing the actions ourselves, but instead listening to descriptions of other people's actions? When reading a story we are often able to visualise described events as they unfold – imagining what the characters look like, feel like and even the environment in which the story takes place. By relating a story to our knowledge and experience about similar actions and events, we are also able to make inferences, or anticipate upcoming events before they unfold within the narrative.

In everyday life we regularly use language to refer to items and people within our immediate proximity. For example, you may be having dinner with a friend, notice a pot of salt at the other end of the table and ask him if he would mind passing you the salt. In this scenario, both the item you refer to and the person to which the question is directed are right there in front of you. Thus, assuming that you and your friend are the only people in the room he will automatically be inclined to think that the question is directed to him and upon hearing 'would you mind passing me the...', he is further likely to infer that the item you are requesting is; a) something

‘passable’, and b) something within his reach. When inferring the upcoming item in this scenario, language combined with a visual context allows us to restrict the number of possible outcomes, thereby making it easier to anticipate which item will most likely be referred to next (e.g. Altmann & Mirkovic, 2009, Rayner, Warren, Juhasz & Liversedge, 2004).

However, we don’t use language exclusively to refer to items within a visual context. Indeed, one of the great features of language is that it allows us to convey discourse regardless of visual context, time and probability. Going back to the previous example of ‘passing the salt’ we can then ask what would happen if rather than experiencing the situation ourselves, we are merely hearing about for instance, Peter having dinner with a friend. When hearing Peter asking his friend if he would mind ‘passing the...’ our anticipation of the upcoming item is no longer restricted by the visual context and while we would still infer that the upcoming item is something ‘passable’, without a visual context any number of ‘passable’ items could be within Peter’s friend’s reach. In this case our anticipation about the upcoming item is no longer bound by a visual context, but instead restricted only by language.

The aim of my PhD research is to explore the mapping of language onto internal representations of visually available scenes, as well as previously viewed scenes – more specifically, we were interested in how event-plausibility may influence our internal representations of described events and further how this might be modulated by the nature of the visual context (as present or absent). Previous studies have shown that issues such as temporal and spatial distance can affect representational accessibility (e.g. Kelter, Claus & Kaup, 2004; Morrow, Bower & Greenspan, 1989; Morrow, Greenspan & Bower, 1987; Rinck, Williams, Bower & Becker, 1996; Zwaan, Madden & Whitten, 2000). In contrast, the research presented in this thesis investigated the extent to which plausibility of context and location is able to influence the accessibility of objects

within event-representations.

1.1. Mental representations of language

A full understanding of discourse often relies on factors beyond a simple understanding of spoken words. It has been proposed that people construct a mental model (Garnham, 1996; Glenberg & Langston, 1992; Johnson-Laird, 1981, 1983), or a situation model (Kintsch, 1988, 1992; Morrow, Bower & Greenspan, 1990; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998), forming an internal representation that to some extent mirrors the described events. Johnson-Laird (1983) proposed that these models enable us to create a mental representation of the situation described by the language we encounter by relating discourse to the world through perception and conception. When constructing a mental model we integrate linguistic information with our knowledge and experience about similar situations and this allows us to go beyond the linguistic input of discourse in order to provide a rich and more complete representation of described events. This integration of language and knowledge further enables us to form a rich and dynamic mental representation of the situation described – as opposed to a mental representation of language itself (Glenberg, Meyer & Lindem, 1987; Zwaan, 1999). For example, when we hear a sentence such as *"on her way home from work Anna realised that she had forgotten her umbrella, so she used a newspaper instead"* we draw on general knowledge and personal experience, which makes us assume that; a) it must be raining, and b) since Anna did not have her umbrella with her she must instead have used a newspaper to protect her from the rain. Even though the sentence does not explicitly state what Anna used the newspaper for, by relating our knowledge and experience to the language we hear, we are able to form a more complete mental representation of the described situation. This in turn gives us a rich and detailed understanding of the described event. As such, mental representations play an important role in our

understanding of discourse, since our personal knowledge and experience are able to influence how we process and interpret language.

The idea of mental/situational models becomes particular important for the comprehension of several sentences, since longer narratives rarely describe static and unchanging scenarios. Often during descriptive events the location of the characters may change regularly, objects are moved to new locations, different events occur, new characters are introduced and so on (Zwaan, Magliano & Graesser, 1995; Zwaan & Madden, 2004). Such changes require us to continuously update our mental representations of the described situation as it unfolds, in order to gain an up-to-date full representation and comprehension of the described event. One of the questions we consider in this thesis is on what basis, or in other words, what kind of knowledge is recruited when we update representations of events.

1.2. Dimensions of situation models – space

Researchers have identified several dimensions of situation models such as causation, intentionality, protagonist, time and space (e.g. Chafe, 1979; Gernsbacher, 1990; Graesser, Singer & Trabasso, 1994; Sundermeier, van den Broek & Zwaan, 2005; Zwaan, Langston & Graesser, 1995; Zwaan & Radvansky, 1998; Zwaan, Radvansky, Hilliard & Curiel, 1998) that we use in order to successfully construct a mental representation of a described event. A number of studies have investigated how language can evoke elaborate mental representations of the spatial properties of an event (e.g. Denis & Cocude, 1989; Erlich & Johnson-Laird, 1982; Haenggi, Kintsch & Gernsbacher, 1995; Morrow, 1994; Morrow & Clark, 1989; O'Brien & Albrecht, 1992; Rinck, Hähnel, Bower & Glowalla, 1997; Tversky, 1991). Since mental representations rely on personal knowledge and experience, it may be assumed that internal representations of space are likewise influenced by people's actual

experience and knowledge of the spatial world and several studies have shown that representations of space and motion are highly influenced by people's knowledge and previous experiences with similar situations (e.g. Clark, 1973; Freyd, 1983; Kaschak et al., 2005; Miller & Johnson-Laird, 1976; Mishra & Singh, 2010; Shepard & Hurwits, 1984; Tversky, 1991, Zwaan, Madden, Yaxley & Aveyard, 2004) whereby an integration of language and knowledge serves to mentally recreate the spatial elements of a described situation.

A study by Glenberg et al. (1987) showed that spatial information about distance is able to influence accessibility. Participants read short stories that included a target word. The target word was either spatially associated with the character, or in other words, in close proximity to the character of the story (for example, *"after doing a few warm-up exercises John put on his sweatshirt and went jogging"*), or spatially dissociated *("after doing a few warm-up exercises John took off his sweatshirt and went jogging"*). Afterwards participants read a filler sentence (always referring to the character, but never to the target word) and then they made a recognition response to the target word. Glenberg et al. found that participants' response and reading times were longer after reading the spatially dissociated sentences, compared to the spatially associated sentences. This suggests that the spatial properties between the character and the target object were included in participants' representation of the event. As such, the spatial proximity between "John" and "his sweatshirt" was able to influence comprehension, since a small spatial distance made the sweatshirt more prominent in the representation of the event and therefore more accessible.

Later research by Rinck and Bower (2000) extended these findings. They examined the effects of spatial proximity on accessibility by manipulating spatial distances in narratives. For the first part of the experiment participants were required to study and learn the layout of a fictitious building with a number of rooms, each containing certain critical objects. Afterwards, they read a series of

narratives describing the character's actions and (critically) whereabouts in the building (for example, Calvin walked from the repair shop into the experiment room). Spatial distance was determined by the objects' proximity to the location of the character. After the narrative participants read one to six intervening sentences and were then presented with a probe question inquiring whether an object was located in a specific room inside the building (for example, is the bed located in the lounge?). The findings showed that participants were faster to respond to the probe questions when the described location was situated near the current location of the character, as opposed to a previously entered room, or an unmentioned room. As such, spatial distance influenced response times, making the objects and rooms more accessible when these were located close to the whereabouts of the character. This shows that internal representations are able to contain complex spatial layouts, as well as continuously keeping track of a character's whereabouts within the narrative. Zwaan (1999) suggests that such findings may be related to our everyday interactions with the environment, whereby 'close at hand' objects are often more relevant to us, compared to objects that are further away, or out of reach.

Matlock (2004) provided further evidence suggesting that people's experience of both space and motion influences how language is processed. In this study participants read short stories about travel, which included either a fictive motion sentence (a sentence that included a motion verb, but did not describe any explicit movement or change of state, for example, "the road *runs* through the valley"), or a non-fictive motion sentence (in which no movement is implied, for example, "the road *is* in the valley"). Her findings showed that after reading fictive motions sentences, which described slow travel, difficult terrain and long distances, participants took longer to decide if a subsequent test sentence was related to the story, compared to when they read sentences that described fast travel, easy terrain and short distances. Importantly, these differences were not found for

non-fictive motion sentences, which suggest that when reading the fictive motion sentences participants formed an internal representation that mentally simulated the movement described and this influenced language processing. A subsequent eye tracking study by Richardson and Matlock (2007) provided further evidence for this point, demonstrating that differences in described terrain were also reflected in participants' inspection times and eye movements towards visual displays depicting the described scenes. They found that when listening to fictive motion sentences, participants' inspection times and eye movements (scanning) along the road were longer when the terrain was described as difficult, compared to when it was described as easy. As in the previous study by Matlock there were no differences when listening to non-fictive motion sentences. This indicates that mental simulations of fictive motion integrated themselves with visual processing, thereby allowing language to influence visual perception.

In a related experiment Zwaan et al. (2004) examined the extent to which language comprehension is able to activate representations of visual motion. In this experiment participants listened to sentences that described either the motion of a ball in a direction towards the participants (for example, the shortstop hurled the softball at you), or in a direction away from the participants (for example, you hurled the softball at the shortstop). After hearing the sentences two visual displays were presented in quick succession, showing the image of a ball either getting gradually larger (as if coming towards the participants), or gradually smaller (suggesting movement away from the participants). After each trial participants were required to judge if the two pictures displayed were of the same object. Zwaan et al. found that participants responded faster when the pictures matched the direction of the ball, as it had been described by the sentences. For example, if the sentence described the ball as moving towards you, participants responded faster when the pictures showed the ball getting bigger (suggesting movement towards the participant),

compared to pictures showing the ball getting smaller (suggesting movements away from the participant). A number of related studies support these findings (e.g. Glenberg & Kaschak, 2002; Kaschak, Zwaan, Aveyard & Yaxley, 2006; Meteyard, Bahrami & Vigliocco, 2007; Yaxley & Zwaan, 2007; Winter & Bergen, 2009), showing that participants' mental representations involve a perceptual simulation of the event described in the sentence, whereby a match between the perceived motion (generated by language) and actual perception results in faster judgements of similarity. Together, these studies illustrate how real experience and knowledge of distance and spatial relations are not only able to influence our internal representations of space, but also that language can affect how we view visual images.

1.3. Interactions between vision and language

Several studies have demonstrated how previous experience is reflected in our visual representations of objects and events (e.g. Borghi, Glenberg & Kaschak, 2004; Dahan & Tanenhaus, 2002; Stanfield & Zwaan, 2001; Zwaan & Yaxley, 2003). Such representations are generally believed to be experiential in nature, or in other words, grounded in our knowledge and experience about similar situations (e.g. Barsalou, 1999; Marmolejo-Ramos, Elosu´a de Juan, Gygax, Madden, Mosquera, 2009; Zwaan, 2004). A study by Zwaan, Stanfield and Yaxley (2002) explored how perceptual representations are influenced by previous experience of similar events. In this experiment participants read sentences in which the shape of the described objects would differ according to where the objects were located. For example, participants would read either *"the ranger saw the eagle in the sky"* or "*the ranger saw the eagle in its nest"*. While these sentences are very similar, our experiential knowledge of the shape of the eagle differs depending on its location, since real world knowledge informs us that in order to fly an eagle must have its wings outstretched, whereas an eagle in a nest is much more likely to have its wings folded against the body. After

reading each sentence participants were shown a picture that either matched the shape of the eagle with its described location, or mismatched. When asked to decide if the depicted object had been mentioned in the sentence, participants responded faster when the picture (an eagle with outstretched wings) matched the described location (in the sky). This suggests that participants relied on experiential knowledge to infer the implied shape of the described objects, which in turn allowed them to mentally represent the object's most appropriate shape (according to its location).

In a similar experiment, Richter and Zwaan (2009) explored the extent to which words describing different colours are able to activate perceptual representations of colour. Participants were first shown a coloured square, then a word describing a colour and finally the same coloured square. They were slower to decide if the colour word described the same colour as shown in the square when there was a mismatch, compared to when the depicted colour matched the word. This suggests that participants had formed a perceptual representation of the colour of the word whilst reading it. Recent research by Huettig and Altmann (2011) provides further support for this notion, showing that when hearing a word (for example, pea) that is typically related to a certain colour (green) participants looked more to the perceptually related objects (a green blouse), compared to unrelated distractors (a yellow trumpet). Importantly, since blouses are not typically green, but come in a variety of colours, participants' looks to the green blouse appear to be guided by a representation of colour that had been activated moments earlier upon hearing the word "pea". A subsequent study by Richter and Zwaan (2010) expanded these findings by showing that participants also are able to represent several perceptual dimensions simultaneously (shape and colour) when presented with words implying a prototypical shape and colour (for example, a tomato being round and red). Such perceptual representations of colour and shape support the idea of mental representations being experiential in nature.

This notion of experiential knowledge being able to influence perceptual representations of shape may be further extended to see how our knowledge of actions and object-based affordances are able to mediate eye movements. As noted earlier, several dimensions are often used to successfully construct a mental representation. Theories of situation models often assume that people's internal representations of events occur in a shared spatial-temporal framework (e.g. Zwaan et al., 1995a, 1995b), since temporal information alongside spatial information creates an important integrated dimension that is necessary for people to fully comprehend and interpret described situations. Furthermore, language is rarely used to describe static and unchanging events. Just as in real life, objects are moved to different locations, actions are initiated and completed, and time passes while events take place. In order for unfolding language to remain coherent in such circumstances it becomes necessary to either incorporate these changes into our current situation model (Zwaan & Madden, 2004), or alternatively create a new situation model with which to represent these new developments of events (Radvansky, Zwaan, Federico, & Franklin, 1998). For example, when relating a described event to a static visual scene we must interpret whether the scene corresponds to the beginning, the middle, or the end of that event. As such, it may be argued that (in certain circumstances) a complete representation also relies on the temporal properties of the event structure and several studies have shown how language can evoke elaborate mental representations of the temporal properties of events (e.g. Anderson, Garrod, & Sanford, 1983; Carreiras, Carriedo, Alonso, & Fernandez, 1997; Coll-Florit & Gennari, 2011; Zwaan, 1996).

In everyday life we regularly use language to refer to items and people within our immediate proximity and in such circumstances eye movements to objects within a specific visual context allows us to explore the interaction between a concurrent visual environment and language as it unfolds. In the following section we present some of

the early studies, which helped to establish this particular methodology.

1.4. The visual world paradigm

The common goal of research within the 'visual world paradigm' is to explore how language interacts with the external environment. The general structure of these types of studies consists of recording peoples' eye movements while they are looking at visual scenes and listening to either a set of instructions, or descriptions that are related to those scenes (e.g. Allopenna, Magnuson & Tanenhaus 1998; Cooper, 1974; Eberhard, Spivey-Knowlton, Sedivy & Tanenhaus, 1995; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). This methodology allows us to observe the underlying processes that we use during discourse comprehension and precisely when language becomes integrated with the visual world. Studying the time course of people's eye movements towards certain items within a visual display also allows us to make certain inferences about language comprehension – for example, how and when language is able to direct our visual attention towards either named, related, or anticipated objects.

Initial research by Cooper (1974) showed that eye movements tend to be rapidly directed towards items when these are directly referred to by spoken language. In this study participants would view a visual scene consisting of nine different items while they listened to short stories containing words that were either directly or semantically related to the depicted items. The findings showed that when hearing the name of one of the depicted items (for example, a zebra) participants looked more toward that item, compared to the unnamed items. Interestingly, a similar pattern occurred when participants heard a semantically related word (for example, Africa). In this case they would look more towards the zebra than the other unrelated items. This suggests that there is a tight integration between linguistic and visual processing, whereby eye movements

are able to reflect our continuous on-line understanding and semantic interpretation of language.

More recent research by Tanenhaus et al. (1995) further demonstrated that during language comprehension we are able to very rapidly establish reference to objects within our visual environment. In one study participants were placed in front of a visual display and asked to perform a set of instructions. For example, participants would be presented with a visual display containing four items differing in colour, shape and marking and be asked to "*touch the starred yellow square*". Upon hearing these instructions, participants looked to the target square approximately 250 milliseconds after the offset of the word that uniquely determined which of the four items was referred to. When hearing a more complicated set of instructions that could refer to several of the displayed items, participants looked sequentially to each potential item until more specific information was given allowing them to determine which of the four items was the target. As soon as participants had enough information to identify which item was referred to, they would look to the target and then carry out the instruction. These studies show a rapid and sequential integration between language comprehension and eye movements. In other words, as each word unfolds, we look to whichever items in our visual environment that language (at each of these moments in time) might refer to. This sequential interaction between language and the external environment consequently allows us to explore the underlying mental processes that are employed during language comprehension.

1.5. Visual world – experiential knowledge

While language often occurs in conjunction with a visual context, modular theories of language comprehension (e.g. Coltheart, 1999; Fodor, 1983) assume that early stages of language processing are guided solely by syntactic constraints, whereby a single grammatical

interpretation is selected or ranked on the basis of the features of the unfolding linguistic composition (Frazier, 1987; Frazier & Clifton, 1996). According to this approach non-linguistic constraints such as real-world knowledge and visual information are not employed until later stages of processing. In contrast, constraint-based approaches (e.g. MacDonald, Pearlmutter & Seidenberg, 1994; Tanenhaus, Spivey-Knowlton & Hanna, 2000; Tanenhaus & Trueswell, 1995) assume that people continuously evaluate multiple syntactic options, employing a range of constraints, which may be derived from both linguistic and non-linguistic sources of information. According to these accounts, any available and relevant information is continuously assimilated in order to determine the most suitable interpretation of language as it unfolds. In this way constraints derived from both syntactic, as well as visual/non-linguistic information, are able to influence language comprehension from the earliest stages of processing.

More recently, Knoeferle and Crocker (2006) proposed an account, which emphasised the coordinated interplay of information derived from the visual scene, linguistic information and real-world knowledge during language comprehension. According to this account language comprehension occurs in an incremental fashion, whereby we interpret and anticipate upcoming language gradually as each word unfolds. As such, unfolding language guides eye movements to named and anticipated items, or events within a visual scene, which is then able to influence comprehension. According to this account, online comprehension relies on multiple sources of available information, which each serves to constrain our interpretation of upcoming language. Furthermore, in the context of a concurrent visual context this tight temporal interaction between different types of information leads us to visually inspect named items/events at the earliest possible moment (during anticipation before the item is named). This early inspection increases item-salience, thereby leading to a greater reliance on information derived

from a visual scene, compared to experiential knowledge associated with the described event.

In line with constraint-based theories and the coordinated interplay account Altmann and Kamide (2007) explored the representational basis for anticipatory eye movements, specifically the extent to which a combination of the temporal aspects of a verb and the affordances of depicted objects are able to guide anticipatory eye movements towards the most suitable object within a visual scene. They showed participants images depicting for example, a man, a table with a full glass of beer, an empty wine glass, some cheese, and some Christmas crackers on the floor (see figure 1.1.). Participants would then hear either *"the man will drink all of the beer"*, or *"the man has drunk all of the wine"*. Upon hearing "the man will drink all of..." participants looked more towards the full glass of beer. In contrast, when hearing "the man has drunk all of..." there were more anticipatory eye movements to the empty wine glass.

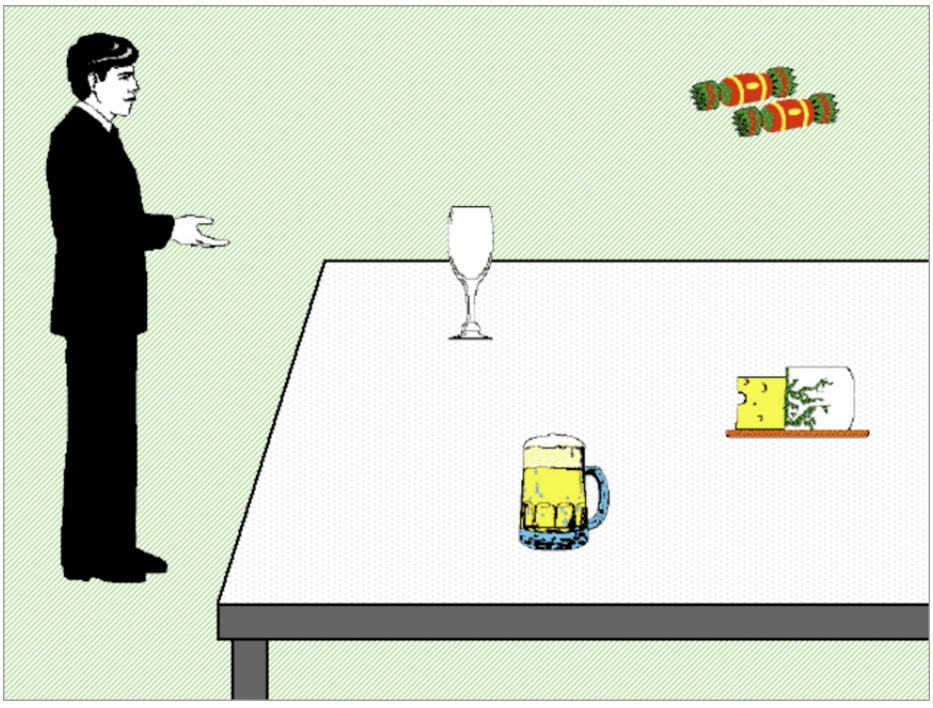


Figure 1.1. Example scene from Altmann and Kamide 2007.

This suggests that participants' knowledge about past and future events activated the objects' affordances (i.e. the full glass afforded a

future drinking event, whereas the empty glass could only afford a past drinking event), which then guided anticipatory eye movements toward the most plausible object in the visual scene. These experiments show that the combination of temporal information with the intended action is able to restrict anticipatory looks to the most affordable item within the visual scene.

But what happens if we are told that the affordable item has suddenly been moved to a new location within the visual scene? Will eye movements reflect the (original) location of the object as depicted in the visual scene, or the new location of the object as described by language? Using a blank screen paradigm, Altmann and Kamide (2009) explored the mapping between language and internal representations of previously viewed visual scenes (the methodology of the 'blank screen paradigm' is similar to that of the 'visual world paradigm', except that in these types of experiments the visual scene is removed before the onset of the spoken language). Participants viewed visual scenes – for example, an image depicting a woman, a wine glass and bottle on the floor, an empty table, and a bookshelf (see figure 1.2.).

Figure 1.2. Example scene from Altmann and Kamide 2009.

After five seconds the scene was removed and replaced by a grey screen. A moment later participants would hear either;

1. The woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Moved condition)

Or

2. The woman is too lazy to put the glass onto the table. Instead, she will pick up the bottle, and pour the wine carefully into the glass.

   (Unmoved condition)

During the sentence-final "the glass" participants who had heard "the woman will put the glass onto the table…" looked more towards the new location of the glass, or in other words, the region corresponding to the previous location of the table. In contrast, participants who had heard "the woman is too lazy to put the glass onto the table…" looked more towards the original location of the glass (previously depicted on the floor). As such, participants' eye movements in the moved condition reflected the location of the glass as conveyed by the spoken language (table), as opposed to the visual image (depicting the glass on the floor). A similar pattern of eye movements was observed for the anticipatory eye movements during "the wine carefully into". In this case, participants' knowledge of the described action led to certain expectations of the upcoming item, which consequently guided eye movements toward the location of the glass as conveyed by the spoken language in both the 'moved' and the 'unmoved' conditions.

These results demonstrate that participants mapped the sentences onto an internal representation of the previously presented scene and further show that language is able to mediate a dynamic updating of this mental representation as the described event

unfolds. This suggests that both anticipatory and concurrent eye movements reflect language as it is processed online, whereby our internal representations are continuously updated as we interpret an unfolding event. Altmann and Kamide speculated that the eye movements toward the language-mediated location of the glass were grounded in event-representations based on experiential knowledge and prior experience about similar situations. According to this theory, participants anticipated and looked towards the previous location of the table when hearing “the wine carefully into” as experiential knowledge informed them that the table not only provided an affordable location on which to place the glass, but further because the sentence described the table as a future location for the glass. In other words, it is the interaction between the affordances of the table and participants’ knowledge of the future location of the glass that lead them to anticipate the table. In the context of the described event it matters that the woman moved the glass onto the table, since according to our experiential knowledge a table constitutes both a plausible and affordable location for the placement of a glass. If eye movements are to some extent guided by experientially based event-representations we could further speculate that the eyes would not move so readily to the location of the glass, if that location was implausible (for example, the woman will put the glass onto the lamp...). The plausibility of an object’s future location would be important, since object-representations that are part of implausible event-representations should be less accessible than those that are part of more plausible events. Consequently, we would expect to see a different pattern of eye movements when an object is described as being moved to an implausible location, as opposed to when that same object is being moved to a plausible location. This hypothesis provided the starting point for my research and further allowed us to explore the extent to which we rely solely on syntactic information (in line with modulate theories of language processing), or multiple sources of visual,

linguistic and experiential information (in line with constraint-based theories and the coordinated interplay account) to constrain our interpretation of upcoming language. More specifically, we were interested in the weighting of constraints provided by a concurrent, or recently encountered visual context and the extent to which visual information is able to influence comprehension.

1.6. Thesis overview

The following chapter will begin with a review of the literature concerning anticipatory eye movements, the blank screen paradigm and plausibility, followed by a description of two eye-tracking experiments that extend the findings of the previous experiment by Altmann and Kamide (2009). In these experiments items were either moved to a plausible or an implausible location within the visual scene, or not moved at all. When the scene was removed before the auditory input we found anticipatory looks to the plausible location, but not the implausible location. In contrast, when the scene remained onscreen there were anticipatory looks to both locations. For both experiments there was no difference in looks during the sentence-final noun. As such, plausibility only influenced anticipatory eye movements after we removed the visual scene.

In chapter 3 we present two more studies (one in which we remove the visual scene before the auditory input and another in which the visual scene remains onscreen), which manipulated the plausibility of objects within a specific context (for example, a cat, or a penguin in a kitchen). We found a similar pattern of eye movements to that observed in experiments 1 and 2, suggesting that the effect is not exclusively related to the plausibility of location.

In chapter 4 we attempted to uncover why we anticipate implausible locations when the visual scene is available, but not when the scene has been removed before the auditory onset. The fifth experiment investigated the extent to which a higher proportion of initial eye movements to the plausible locations (observed in

experiment 1) could have enhanced the accessibility of those locations during later anticipation. The findings showed that the proportion of initial looks made the implausible regions more accessible later in the narrative, but not the plausible regions. This makes it difficult to explain the pattern of eye movements observed in experiment 1 in terms of enhanced accessibility due to the higher proportion of 'initial' eye movements to the region of the table.

In chapter 5 we explored the implications of introducing an item purely through language (for example, the woman will move *a* glass onto the table/lamp). The results showed a similar pattern of eye movements to experiment 1 (where the glass was first presented in the visual scene and then referred to by spoken language), suggesting that a purely linguistic representation of the glass did not increase the proportion of anticipatory eye movements to the implausible locations.

In chapter 6 we propose an alternative account explaining why plausibility affects anticipatory eye movements in the context of a blank screen, but not in the context of a concurrent visual scene. Our explanation is based on the assumption that there is a quantitative difference in how we process and anticipate upcoming discourse when language refers to something outside our visual environment, as opposed to when language refers to entities within our immediate visual proximity. During language comprehension we assume that participants in an event will be drawn from our visual environment. In the visual scene the referred target is the only item that affords the described action and this is why we anticipate its location regardless of plausibility. However, when we have to rely on our memory/internal representation of a visual scene, anticipation is no longer bound exclusively to the visual context, but may now be further influenced by real-world knowledge. When this is the case we might anticipate a variety of affordable items.

In chapter 7 we present two experiments, which aimed to explore the extent to which both visual and linguistic constraints are able to

guide anticipatory eye movements to appropriate (yet implausible) locations. In experiment 8 we manipulated the number of affordable items within the context of a concurrent visual scene. The findings showed that the inclusion of a second (affordable) glass did not weaken affordance-based constraints sufficiently to show an influence of plausibility. In experiment 9 we linguistically restricted the number of affordable items within the context of a blank screen (e.g. the woman really wanted some wine, but she could only find one glass…). In contrast to experiments 1 and 7, there were no more anticipatory eye movements to the plausible locations than there were looks to the implausible locations. This shows that when spoken language unambiguously restricts the number of affordable items the linguistic context is able to guide anticipatory eye movements to the appropriate, yet implausible location of the glass. Thereby, we require a more constraining linguistic context in the absence of a visual context in order to achieve the same degree of constraints provided by a concurrent visual scene.

Chapter 8 concludes the thesis by providing an overview of the experimental findings and discussing the implications of how differently weighted constraints are able to influence our anticipation of upcoming events.

CHAPTER 2

MANIPULATING PLAUSIBILITY OF LOCATION

The following two experiments explored how event-plausibility might affect our internal representations of described events – specifically the extent to which plausibility of location is able to influence the accessibility of objects within these event-representations. Secondly, we investigated how this may be modulated by the nature of the visual context (as present or absent).

2.1. Visual world – anticipatory eye movements

During language comprehension we often rely on experiential knowledge in order to predict which upcoming objects and events will be referred to next (e.g. Kamide, 2008; Matsuki, Chow, Hare, Elman, Scheepers & McRae, 2011; McRae & Matsuki, 2009; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005). For example, when reading a sentence such as "*as it started to rain Anna searched her handbag and hoped that she had remembered to bring her…*" we can combine our knowledge of rainy weather with knowledge about which items might be useful in such situations. This, in turn, provides us with certain constraints and expectations about the nature of the upcoming word – for instance, we know that most people don't like getting too wet when it rains. As such, it is likely that Anna is looking for something that will be able to prevent her from getting wet. Furthermore, since Anna is searching through her handbag, the item

in question must be something that is small enough in size to fit into a handbag. By combining this knowledge most people would probably predict that Anna was looking for her umbrella.

A study by Altmann and Kamide (1999) looked at how language (as it unfolds) is able to rapidly direct our attention towards anticipated objects and more specifically, the extent to which semantic information derived from a verb is able to direct eye movements towards the most suitable item within a specific visual context. In this experiment participants were presented with a scene depicting for example, a boy sitting on the floor surrounded by a toy train set, a toy car, a balloon, and a birthday cake (see figure 2.1.). Whilst viewing the scene participants would hear either *"the boy will move the cake"*, or "*the boy will eat the cake"*. It is important to note that while the verb 'move' could refer to any of the items within the visual scene, the verb 'eat' could only refer to one of the items (the cake).

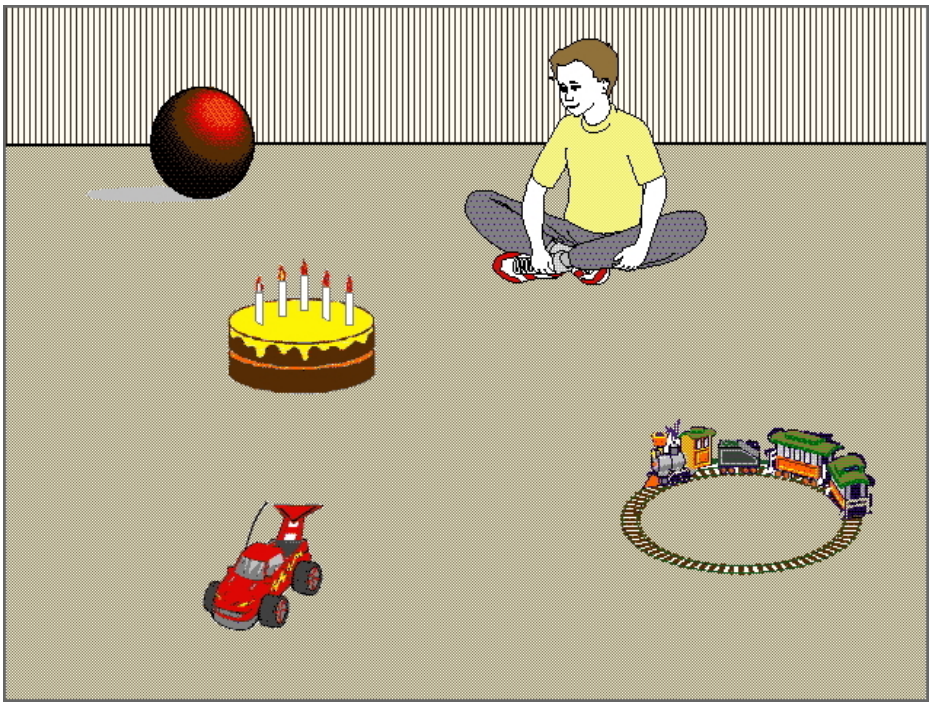


Figure 2.1. Example scene from Altmann and Kamide 1999.

The findings showed that participants looked more towards the cake upon hearing the verb 'eat', compared to when they heard the verb

"'move'. This suggests that participants used verb-related information (i.e. the verb eat implied that the upcoming referent must be something edible) in order to anticipate which item would most likely be referred to next. This demonstrates that verb-related information is able to direct eye movements to the most appropriate item within a visual context, and that these 'anticipatory' eye movements occur even before the onset of the target word.

As such, anticipatory eye movements can be used to explore how we process and interpret language as it unfolds. For example, if we assume that participants construct an internal representation of an event using a combination of the visual context and the linguistic input, we may further assume that this representation is continuously updated as upcoming language is being processed. Thereby, the semantic information derived from the verb is rapidly integrated with the features and affordances of the items presented in the visual context. This in turn, allows participants to instantly restrict the number of possible upcoming referents, resulting in anticipatory eye movements towards the most appropriate item.

Kamide, Altmann and Haywood (2003) extended this research by investigating whether anticipatory eye movements are exclusively related to semantic information derived from the verb, or alternatively if we rely further on information derived from a combination of both the action that is being carried out (the verb), the person performing this action (the agent), and the item onto which the action is being performed (the goal object). Participants viewed a visual scene, for instance, depicting a man, a young girl, a motorbike and a carousel (see figure 2.2.). They then heard either *"the man will ride the motorbike"* or *"the girl will ride the carousel"*. As in the previous experiment the motorbike and the carousel were the only items that fulfilled the selectional restrictions of the verb 'ride', or in other words, they were the only items that could afford to be ridden. Similarly, the man and the girl were the only depicted agents that could perform the described action (ride).

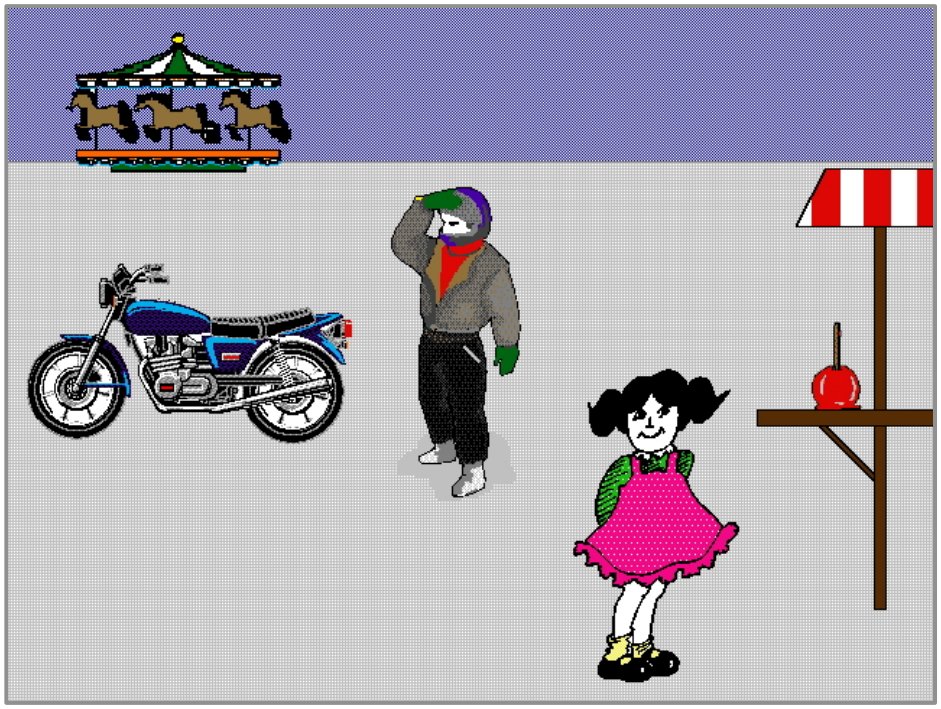


Figure 2.2. Example scene from Kamide, Altmann and Haywood 2003.

As such, both the verb and the agents were restricted in terms of which items could afford 'riding on' and 'who' could perform the action of riding. However, one might also argue that experiential knowledge could imply further restrictions, specifically in terms of plausibility. Real world knowledge and experience informs us that a man would be much more likely to ride a motorbike, compared to a carousel and likewise, that a young girl would be more likely to ride a carousel than a motorbike. If anticipatory eye movements are guided solely by verb-related information we would expect to see a similar proportion of looks to the motorbike and the carousel, regardless of who is performing the action. However, if anticipatory eye movements rely on a combination of linguistic information and experiential knowledge we would expect to see a difference in the proportion of looks to the motorbike and the carousel, depending on who is performing the action. Kamide et al. found that upon hearing "*the man will ride the...*" participants made more anticipatory eye movements toward the motorbike. In contrast, when hearing "*the girl will ride the...*" there were more anticipatory looks toward the

carousel. This shows that when anticipating the most appropriate outcome, participants integrated verb-related information with experiential knowledge (in this case plausibility) about the subject performing the action, as well as the item on which the action was performed. Thereby anticipatory eye movements were not simply mediated by selectional restrictions derived from the verb, or the subject. Rather, it is the combination of linguistic information with experiential knowledge, which guided anticipatory eye movements toward the most appropriate goal. This suggests that while visual and linguistic information allowed participants to construct a mental representation of the situation described, as language unfolded they made further use of relevant background knowledge and stereotypical information. In other words, participants' experiential knowledge provided additional information about the plausibility of possible actions and outcomes, consequently facilitating the update of event-representations as language unfolded.

2.2. Plausibility

In the following two experiments we explored how event-plausibility might influence our internal representations of described events. More specifically, we were interested in whether the plausibility of an object's location is able to affect its accessibility within these event-representations. During comprehension we often rely on real-world knowledge and experience in order to correctly anticipate upcoming events before they unfold within a narrative and while language most often has a plausible outcome, real-world knowledge provides additional information which allows us to further constrain the number of possible outcomes in terms of their graded status of likelihood (e.g. Kamide et al., 2003). In other words, as language unfolds we often have to rely on the likelihood, or plausibility of events in order to anticipate the most probable outcome and this occurs regardless of whether the outcome of the unfolding event turns out to be plausible, or implausible. For example, we know from experience that people

are more likely to eat a tub of ice cream than a tub of butter, even though both options are perfectly edible.

As mentioned earlier, during comprehension people integrate language with real-world knowledge and this integration allows us to construct a rich and dynamic mental model of described events, rather than a mental representation of the linguistic information itself (Glenberg et al., 1987; Zwaan, 1999). As such, mental models play an important role in discourse comprehension, since real-world knowledge and experience are able to influence how we process and interpret language. If it is the case that our mental representations are influenced by experiential knowledge we may speculate that the likelihood of described events occurring in real life would be of further importance, since mental representations of unlikely events will be less anticipated and consequently less accessible, compared to more plausible events. By manipulating the likelihood of events, we can use plausibility as a tool to further explore the processes by which we construct event representations during language comprehension. In the remainder of this section, we briefly review different approaches to plausibility, in order to motivate the operational definition that will be used in the empirical studies described later in this and subsequent chapters. We will, after this section, review the empirical techniques that will be used in these studies.

Several event-related potential (ERP) studies have shown that people use experiential knowledge when anticipating upcoming words, or events (e.g. Camblin, Gordon & Swaab, 2007; Federmeier & Kutas, 1999; Hagoort, Hald, Bastiaansen & Petersson, 2004; Kuperberg et al. 2003; Otten & Van Berkum 2007; Urbach & Kutas, 2010; Van Berkum, et al., 2005), demonstrating that comprehension to some extent is driven by expectations derived from real-world knowledge and experience about similar situations. In a study by DeLong, Urbach and Kutas (2005) participants read sentences such as *“the day was breezy so the boy went outside to fly...”*. Prior to the experiment DeLong et al. normed the cloze probability of a number of

words in terms of their likelihood of completing the target sentence (see also Kutas & Hillyard, 1984). The most likely continuation of the sentence above was judged to be *"a kite"*, while a less likely continuation was *"an airplane"*. A number of previous ERP studies have shown that very likely and therefore expected upcoming words such as "kite" tend to elicit a small N400 effect, whereas relatively unexpected words such as airplane normally elicits a large N400 effect (e.g. Kutas & Hillyard, 1984). Using the example above, DeLong et al. argued that the difference in N400 in response to "kite" and "airplane" could simply be due to the difference in meaning between the two words, whereby the word "kite" might be easier to integrate into the unfolding sentence due to people's schematic associations of the event. In order to remove the difference in word meaning, DeLong et al. decided to focus on the preceding indefinite articles as in, 'a' kite and 'an' airplane, since these only differ in phonological form. Thereby, any N400 difference between these two indefinite articles would provide evidence of people's early and highly specific expectations of upcoming words. Their findings supported this idea, showing a larger N400 during 'an' and a smaller N400 during 'a'. This suggests that there is a rapid and incremental integration of upcoming words and further that upcoming words are constrained by contextual probability. In other words, our experiential knowledge of the event as it unfolds is able to determine the probability of upcoming words and thereby constrain the number of alternative options.

In the following studies we define plausibility in terms of probability, or in other words, event likelihood. Coming back to the earlier example, we propose that eating a tub of ice cream is more plausible than eating a tub of butter, because experiential knowledge informs us that in real life people are much more likely to eat a whole tub of ice cream, compared to a whole tub of butter. Thereby, we consider actions, or events to be plausible in terms of their compatibility with our previous knowledge and experience about such

actions and events (e.g. Collins & Michalski, 1989; Connell & Keane, 2004; Johnson-Laird, 1983).

In contrast, other studies have applied somewhat different operational definitions of implausibility. Below we provide different definitions of implausibility and anomaly. Murray (2006) argued that "there is nothing in the empirical literature to suggest a functional distinction between implausibility and anomaly" (p. 80). Rather, he agrees with McCawley (1971) who suggests that when considering imaginary beliefs, dreams and cartoons, selectional and semantic restrictions differ little from implausibility. However, in order to think along those terms, an appropriate (and imaginary) context should arguably be provided. Studies have shown that people initially evaluate language against real-world knowledge and that this occurs even in a counterfactual context (Ferguson & Sanford, 2008; Ferguson, Scheepers & Sanford, 2010) – if this weren't the case we would happily believe it if someone told us that they had built a car out of cucumbers and driven it to the moon. Only by adding an appropriate context such as "I had a dream last night that I..." does the described event become believable, and more importantly possible in the context that it occurred (e.g. Filik, 2008; Nieuwland & Van Berkum, 2006).

On the other hand, several studies have distinguished between anomaly/impossibility and implausibility (e.g. Joseph et al., 2008; Warren & McConnell, 2007), as well as syntactic and semantic anomaly (e.g. Braze, Shankweiler, Ni, & Palumbo, 2002; Ni, Fodor, Crain & Shankweiler, 1998; Yang, Wang, Chen & Rayner, 2009). One such study, by Rayner et al. (2004) investigated the effects of reading plausible (a), implausible (b), or anomalous (c) sentences such as these below (p.1292).

a. John used a knife to chop the large carrots for dinner.
b. John used an axe to chop the large carrots for dinner.
c. John used a pump to inflate the large carrots for dinner.

The first sentence is plausible since we would normally use a knife to chop carrots. In contrast, using an axe to chop carrots is quite unlikely, but nonetheless affordable. The third sentence however, is impossible (in a real world context) since carrots do not afford inflating. Rayner et al. found an earlier pattern of disruption when reading the anomalous sentences, compared to the implausible sentences. They proposed that this might be because the anomalous sentences provided an instant cue to violation between the verb (inflate) and the noun (carrots)[1]. In comparison, the implausible sentences required a complete semantic evaluation of the event in order to detect any violations. These results show the importance of separating implausible and anomalous/impossible events since it appears that each elicits a different type of disruption to processing narrative events.

We now turn to the empirical methods that will be used in the experiments described below. In these experiments we explored the extent to which plausibility is able to influence the accessibility of objects within our internal representations of described events and further how this accessibility may be modulated by the presence, or absence of a visual context.

2.3. Visual world – the blank screen paradigm

Previous eye tracking studies have shown that a combination of linguistic information and experiential knowledge allows us to anticipate the most likely upcoming item within a concurrent visual scene. While language is often used to refer to items within our visual proximity, we may also use language to refer to items, people, and events when these are absent from our visual context. As such, we can distinguish between 'embedded language use', where language

[1] But see Matsuki et al. (2011) who found immediate processing difficulty in response to implausible/atypical sentences that did not violate selectional restrictions (e.g. Donna used the hose to wash her filthy hair). This suggests that processing difficulties either due to a violation of selectional restrictions, or atypical event-knowledge is not necessarily distinct from each other.

refers to the current communicative situation (Spivey & Richardson, 2009) and 'displaced language use', where language refers to past, distant, or even imagined events (e.g. Zwaan, 2009). For example, you might meet a friend in the street and tell him about the car you just bought, or the restaurant you went to the previous weekend. What happens to our mental representations when spoken language refers to items, or events that are no longer within our visual context? Will language on its own provide enough information to form a coherent representation of the described event? The 'blank screen paradigm' allow us to further look into questions such as these. The method behind the blank screen paradigm is essentially the same as the method used in the 'visual world paradigm', except that in these experiments the visual scene is removed before the onset of the spoken language. This manipulation allows us to explore how we map language onto our internal representations of previously encountered visual scenes and various studies have shown that people tend to look at the previous region of an object when asked to imagine, or recall information related to that item (e.g. Bourlon, Oliviero, Wattiez, Pouget & Bartolomeo, 2011; Brandt & Stark, 1997; Johansson, Holsanova & Holmqvist, 2005; Laeng & Theodorecu, 2002; Spivey & Geng, 2001).

A study by Richardson and Spivey (2000), showed participants a sequence of visual scenes divided into four equally sized quadrants. Each scene depicted a different face in a different quadrant and each face would deliver a fact of general knowledge (*e.g. "the Pyrenees is a mountain range separating France and Spain"*) whereafter it would disappear from the screen. Next, participants heard a second statement, which was related to one of the four previously provided facts (e.g. *"the Pyrenees is a mountain range"*) and were required to say if the statement was either true or false. The findings showed that when participants were formulating their answers, they were more likely to look at the quadrant that had previously contained the

person who delivered the fact in question and that this occurred even though the spatial information was irrelevant to the task.

In a similar study by Hoover and Richardson (2008) participants watched animals burrowing underground and then emerging from a molehill. For example, in one condition a rabbit would first emerge in one location during which the participants would hear a piece of information, for example about Cleopatra. The rabbit would then descend the molehill and a moment later an identical rabbit would emerge from a different location (this time no information was given) and descend shortly afterwards. Participants were then asked a (yes/no) question about the previous information. The second condition was similar to the first, except this time the rabbit would descend the molehill whereafter burrowing would begin in a different off-screen location. This presentation sequence suggested that the rabbit associated with the provided information was different from the second rabbit to emerge. In contrast, the on-screen burrowing in the first condition connected the appearances of the rabbit, suggesting that the second rabbit to emerge was the same as the first. When answering the questions participants looked at both locations when the rabbit appeared to be the same. However, when there appeared to be two separate rabbits, participants only looked to the first location. As in the previous study by Richardson and Spivey (2000) participants associated the spoken information with certain locations on the screen and used this as a spatial cue when later recalling the information. When what appeared to be the same rabbit emerged in two different locations, participants kept track of these locations and relied on (and looked to) both locations when answering questions. In contrast, when what appeared to be two different rabbits emerged in two locations participants only relied on the location that had been associated with the spoken information. In this case, participants were able to keep track of each individual rabbit as it moved around the screen and as such constrain the spatial cues to those exclusively associated with the spoken information. Being able to

associate visual objects with spoken information and further keeping track of these spatial cues is particularly useful in a dynamic visual environment where people and objects regularly move from place to place.

Another study by Altmann (2004), explored when, and to what extent eye movements would be directed to the previous location of an anticipated or named item within a scene, when that scene had been removed before the onset of the spoken language. Participants were shown a visual image (e.g. depicting a man, a woman, a newspaper and a cake) (see figure 2.3.). After five seconds the image was removed and replaced with plain white screen.
A moment later participants would hear either "*the man will eat the cake"*, or *"the woman will read the newspaper"*.

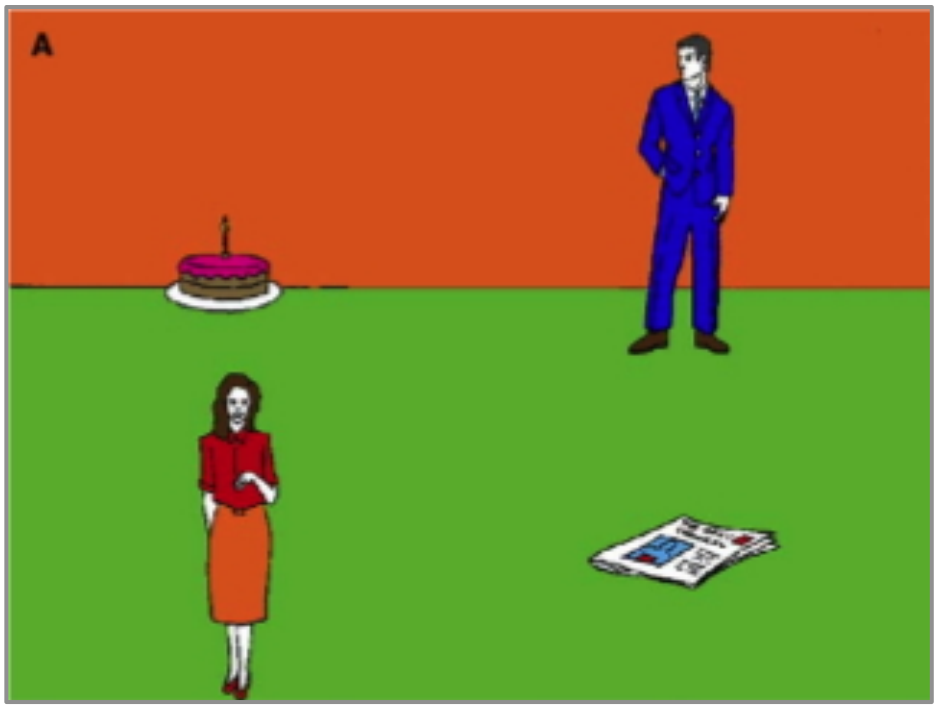


Figure 2.3. Example scene from Altmann 2004.

Upon hearing "the man will eat..." participants looked more towards the previous region of cake, compared to when they heard "the woman will read..." and likewise there were more anticipatory looks toward the previous region of the newspaper when hearing the verb

'read', compared to when hearing the verb 'eat'. By removing the visual image participants were required to hold an internal representation of the scene in order to keep track of any descriptive changes that might occur in the event. Thereby, participants were essentially required to map the sentence onto a mental representation of the previously presented scene – as opposed to the visual scene itself. The increase in looks to the prior location of the cake/newspaper shows that anticipatory eye movements are not necessarily dependent on seeing a concurrent visual scene. In the absence of a visual context, language is simply mapped onto an internal representation of the previously encountered scene. Furthermore, the higher proportion of eye movements to the previous region of the appropriate target demonstrates that the spatial layout of the scene had been incorporated into participants' mental representations of the described event (see also Altmann & Kamide, 2004, for a review).

2.4. Comparing presentation types – blank vs. concurrent

When comparing the effects of plausibility on presentation type it might be useful to look at other studies which have looked at the importance between visual and linguistic information and how the two interact. An interesting experiment by Magliano, Miller and Zwaan (2001) explored the extent to which people monitor and indicate (via button press) changes in time and space when viewing a movie. They compared these findings with results from previous reading studies and found that viewing a movie elicited a similar pattern in participants' perception of changes in time and space to when reading narrative texts (e.g. Zwaan & Radvansky, 1998). This suggests that participants construct similar situational models in terms of temporal and spatial dimensions regardless of whether they are watching a movie, or reading a narrative text.

Another study by Wassenburg and Zwaan (2010) investigated the extent to which prior exposure to visual images is able to influence

language processing at a later point in time. In the first part of the experiment participants were asked to perform a picture-verification task (always depicting the target item in a horizontal or vertical position). The second part consisted of a 15-minute filler task. Finally, participants' eye movements were monitored as they read sentences which described objects in certain locations, implying specific orientations of those objects that either matched or mismatched the position of the previously seen object (e.g. *"he pounded the nail into the wall"* – implies a horizontal position of the nail, whereas *"he pounded the nail into the floor"* implies a vertical position of the nail). They found that participants were faster to read sentences that matched the orientation of the previously viewed object, suggesting that the visual memory of an object is able to influence language processing and that this can occur even after a certain amount of time has passed between the two events.

A recent ERP study by Knoeferle, Urbach and Kutas (2011) showed participants visual images (e.g. depicting a journalist and a gymnast – the gymnast either clapping her hands, or extending one hand wearing a boxing glove). Participants viewed the image for at least three seconds whereafter they read *"the gymnast applauds/punches the journalist"*. The data showed an effect of prior visual context on language comprehension. During the verb there was a larger N400 effect when the sentence mismatched the previously presented image, than when the sentence matched the image. This interplay between prior visual information and language processing might suggest that participants' visual memory of previously viewed scenes would have a considerable influence on how they (shortly after) process spoken sentences related to those scenes. As such, prior visual information can arguably play an important role in comprehension and language processing, suggesting that concurrent visual information is further likely to have a considerable amount of influence on how we anticipate and process language.

Knoeferle and Crocker (2006) investigated the importance and interaction between language-mediated world knowledge and visual information by measuring participants' eye movements as they listened to descriptions of accompanying visual scenes. In this study participants were shown images, depicting for instance a pilot, a wizard and a detective (see figure 2.4. taken from Knoeferle and Crocker, p. 503). While the pilot was always depicted as a patient (not performing an action), the wizard and the detective were depicted in different agent roles, which importantly defied stereotypical knowledge about the actions they would normally perform. For example, the detective was depicted serving a plate of food to the pilot. In contrast, the wizard was depicted spying on the pilot, this action stereotypically being associated with the detective, rather than the wizard. Whilst viewing the scenes participants heard sentences such as "the detective /wizard will soon spy on the pilot" (however in the original German sentences the pilot was mentioned first marked in the accusative case, whereas the detective/wizard was mentioned last marked in the nominative case – "den piloten bespitzelt gleich der detective/zauberer"). This structure allowed Knoeferle and Crocker to explore the extent to which anticipatory eye movements are driven by stereotypical knowledge associated with the verb (detectives typically spy), or in contrast, whether eye movements are predominately guided by knowledge associated with the concurrent visual information (depicting the wizard performing the action of spying). In other words, when anticipating the upcoming agent participants were forced to choose between the spying (non-stereotypical) wizard and the serving (stereotypical) detective.

Figure 2.4. Example scene from Knoeferle and Crocker 2006.

The findings showed that participants looked more towards the wizard after hearing “spies on” than they looked to the detective. As such, participants relied more on the information associated with the depicted event, than knowledge associated with the description of the event. However, a different pattern of eye movements emerged when the wizard’s telescope was replaced by non-related item, such as a roll of toilet paper. When this was the case participants did rely on verb-based knowledge, looking more towards the detective than the wizard. In a subsequent study (using the same stimuli) Knoeferle and Crocker (2007) removed the visual scene before the onset of the spoken language. Their findings replicated the previous results, showing a predominant reliance on prior visual information over verb-based information, but only in cases where the non-stereotypical agent was depicted performing the action usually associated with the stereotypical agent. As in the previous experiment, when neither agents were depicted performing the described action participants relied more on real-world knowledge derived from language.

The experiments mentioned above show that even when visual information has been removed before encountering language, the

visual memory, or internal representation of the image appears to influence how we anticipate and process language. Moreover, this influence seems to differ little depending on whether we are viewing a concurrent visual scene, or the scene is no longer available. This seems to suggest that we should see a similar pattern of eye movements regardless of whether we present spoken language in the context of a concurrent visual scene, or in the context of a blank screen.

In terms of plausibility however, one might argue that the removal of the visual scene before the onset of the spoken sentence might encourage participants to use their imagination more when constructing a mental model of the described events. The element of imagination may be especially relevant when forming representations of implausible events. Barsalou and Prinz (1997) argue that productivity in imagination leads to greater creativity, allowing people to simulate never before encountered objects such as pink bananas, or talking flowers. Our imagination even allows us to simulate all types of implausible and even impossible events by combining existing concepts in novel and infinite ways (e.g. Barsalou, 1999). Taking this into account we might expect people to more readily accept implausible scenarios when they have to rely more on their imagination, as is the case when having removed the visual scene. In contrast, we might find implausible events more difficult to process when having to relate them to a static and concurrent image of the event.

This highlights the importance of potential differences in how we process and anticipate upcoming discourse when language refers to something outside our visual environment, as opposed to when language refers to entities within our immediate visual proximity. Coming back to the earlier example of being asked to ‘pass the salt’ it matters that the referred item is located within the a concurrent visual context – upon hearing the request e.g. “would you mind passing me the...” the linguistic information allows us to infer that the

upcoming item is something ‘passable’ and therefore also likely to be located within the concurrent visual context. Assuming that we don’t normally ask a person to pass us something that isn’t within a reachable distance, we can also use the concurrent visual context to restrict the number of possible outcomes, which in turn makes it easier to anticipate which item will most likely be referred to next. As such, both the properties of language, as well as the concurrent visual context are able to restrict the number of upcoming references, thus making it easier to anticipate the most plausible role-fillers within a described event (e.g. Altmann & Mirkovic, 2009; Rayner et al., 2004). In contrast, when language refers to something outside our visual environment our anticipation of the upcoming item is no longer restricted by the visual context, but now constrained purely by language. While language allows us infer that the upcoming item is something ‘passable’, without a visual context any number of ‘passable’ items might be referred to. According to this account, we would expect a difference in eye movements depending on the absence/presence of a visual context, whereby a concurrent visual scene will provide a stronger degree of constraint, possibly making implausible object-representations more accessible and therefore easier to anticipate.

2.5. EXPERIMENT 1

In the studies described below we define plausibility in terms of event likelihood. The intention was to construct the implausible events as being very unlikely, but not impossible, since this would require participants to evaluate the described events against real-world knowledge, rather than simply detecting semantic or syntactic violations. By doing so we hoped to explore the properties of the mental representations that are formed during language comprehension and the extent to which experiential knowledge is able to make representations of plausible events (and the object-representations they contain) more accessible than those of

implausible events. In other words, is a representation of a fish more accessible when the fish is in an aquarium, as opposed to in a toilet bowl? Secondly, we investigated how the accessibility of plausible and implausible object-representation may be modulated by the nature of the visual context. In experiment 1 we removed the visual scene before the onset of spoken language, whereas in experiment 2 the visual scene remained available throughout the duration of the experimental trials. These two measures allowed us to explore how differently weighted constraints (visual and linguistic) are able to influence how we process plausible and implausible events.

We expanded the stimuli used in Altmann and Kamide (2009) to include a third condition where items were described as being moved to implausible locations (as opposed to items being moved to plausible locations or not moved at all). On the basis of their (2009) data Altmann and Kamide speculated that eye movements towards the language-mediated locations were grounded in event-representations based on experiential knowledge and prior experience about similar situations (see chapter 1 for further details about this study). If this is the case we would expect the plausibility of an object's future location to influence accessibility, consequently eliciting a different pattern of eye movements when an object is described as being moved to an implausible location, as opposed to when that same object is being moved to a plausible location.

2.5.1. Method

Participants

Forty-eight students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Linguistic stimuli

Twenty-four experimental scenes (see figure 2.5. for an example) were matched with three conditions:

1. The woman will move the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Plausible moved)

2. The woman will move the glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Implausible moved)

3. The woman is too lazy to move the glass onto the table. Instead, she will pick up the bottle, and pour the wine carefully into the glass.

   (Unmoved)

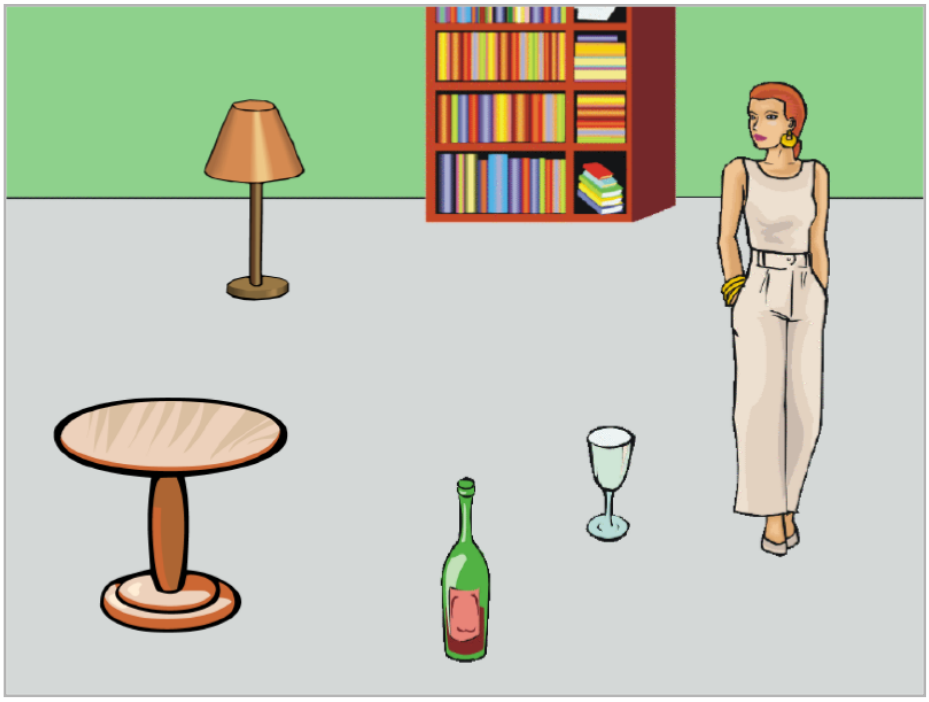


Figure 2.5. Example of one of the visual scenes paired with sentences: 1), 2), or 3).

In conditions (1) and (2), the first sentence always described the agent moving the target object to a new location. In condition (1) the target object would be moved to a plausible location, whereas in

condition (2) the target object would be moved to an implausible location. In contrast to condition (1) and (2), the first sentence in condition (3) described the target object as staying in its original location. The second (target) sentence was always the same for all three conditions (see appendix 1 for a full list of the experimental sentences). In addition to the 24 experimental trials, 24 sentence-picture pairs were included as fillers (see appendix 2 for an example of the filler items).

Norming for plausibility of location

In order to ensure that the plausible and implausible locations were perceived as intended we normed the experimental stimuli for likelihood of location. We presented 91 participants with the 24 experimental sentences and asked them to rate the likelihood of the target items being moved to either plausible or implausible locations. The likelihood was rated on a scale from 1-7 (1 being “very likely” and 7 being “not very likely”). Two lists were created, ensuring that participants rated either the plausible or the implausible version of each event. Thereby list (a) presented 50% of the plausible events and 50% of the implausible events, whereas list (b) presented the opposite versions of the plausible and implausible events. The presentation order of the sentences was randomised across both lists. The mean rating for the plausible locations was 3.06 (SD = 0.96) and 6.47 (SD = 0.28) for the implausible locations. The difference in likelihood between the plausible and implausible locations differed significantly t (23) = -17.735, p < .001, the plausible locations consistently judged to be more likely than the implausible locations.

Visual Stimuli

The visual stimuli were created using commercially available ClipArt packages and presented at a resolution of 800 x 600 pixels (see appendix 3 for a full set of experimental pictures). The positions of

the items representing the plausible locations (e.g. the table), the implausible locations (e.g. the lamp), the original locations (e.g. the glass), and the distractor locations (e.g. the bookcase) were counterbalanced to ensure that the location of each item varied across the full set of experimental scenes. The sentences were recorded by a male native speaker of British English and sampled at 44.1 KHz. The audio files were played via a mono channel split across two speakers that were positioned on each side of the screen. We also noted the onsets and offsets of the critical words in the experimental sentences (using a sound editing program) for carrying out later analysis. The sentences were coupled with their corresponding scenes and allocated to three conditions in a fixed-random order. Three lists of stimuli were made, which included all of the 24 experimental scenes, but only one version from each of the sentence-pairs. As such, the visual scenes were always the same for all three conditions, while the sentences differed. The same 24 filler sentences and scenes were used in all three conditions. Participants were allocated to either condition 1, 2, or 3 and presented with a total of 48 trials presented in the same randomised order for all conditions.

Procedure

The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye. Participants were seated approximately 60 cm. from the display screen. They were told that first they would be shown a picture, the picture would then disappear, and they would hear a sentence. Participants' task was simply to look at the pictures and listen to the sentences. Prior to presenting the trials, a nine-point calibration procedure was performed, followed by a validation of the calibrations. Once this had been performed successfully participants were presented with four practice trials, whereafter the remaining 48 trials were presented. Between each trial a single centrally located dot was presented in the centre of the screen, which corrected any drift in the eye-track

calibration. The images were presented for five seconds and then replaced by a grey screen. The onset of the audio (corresponding to the scenes) occurred one second after the scene had been removed and the trials finished 11 seconds after the audio onset. As such, each trial lasted a total of 17 seconds. After every sixth trial, the eye-tracker was recalibrated. The total duration of experiment was approximately 45 minutes.

Analysis

We defined four identically sized regions of interest within each scene (see figure 2.6.); one, corresponding to the plausible location of the moved glass (the previous region of the table top), another corresponding to the implausible location of the moved glass (the previous region of the lamp), and a third area corresponding to the unmoved location of the glass (the previous region of the glass). Finally, we included a fourth area that corresponded to a distractor item (the previous region of the bookcase). This region was included since it would provide a 'baseline' for the proportion of looks to an unnamed item, thereby allowing us to compare the proportion of looks to previously named locations (plausible and implausible) and unnamed locations. Participants' eye movements were examined during certain time points in the spoken sentences with the percentage of saccades as the dependent measure. These critical time points occurred during 'the wine carefully into' (anticipatory eye movements), and during the final noun phrase 'the glass'.

To begin with, we compared the proportion of saccades to the region of the table and the glass in the moved and the unmoved conditions. This comparison allowed us to see if we were able to replicate the previous findings by Altmann and Kamide (2009). Secondly, we compared the proportion of saccades to the region of the table (plausible locations) and the region of the lamp (implausible locations). This allowed us to see if the plausibility of these locations had any influence on eye movements. Finally, we compared looks to

the region of the table/lamp with looks to the region of the books (distractor). This provided information about whether the proportion of saccades to the plausible and implausible locations was higher than the 'baseline' proportion of looks to an unnamed distractor item. In order to evaluate the extent to which any effects could be generalised across participants and items we analysed the data separately by-subjects *(t1)* and by-items *(t2)*.

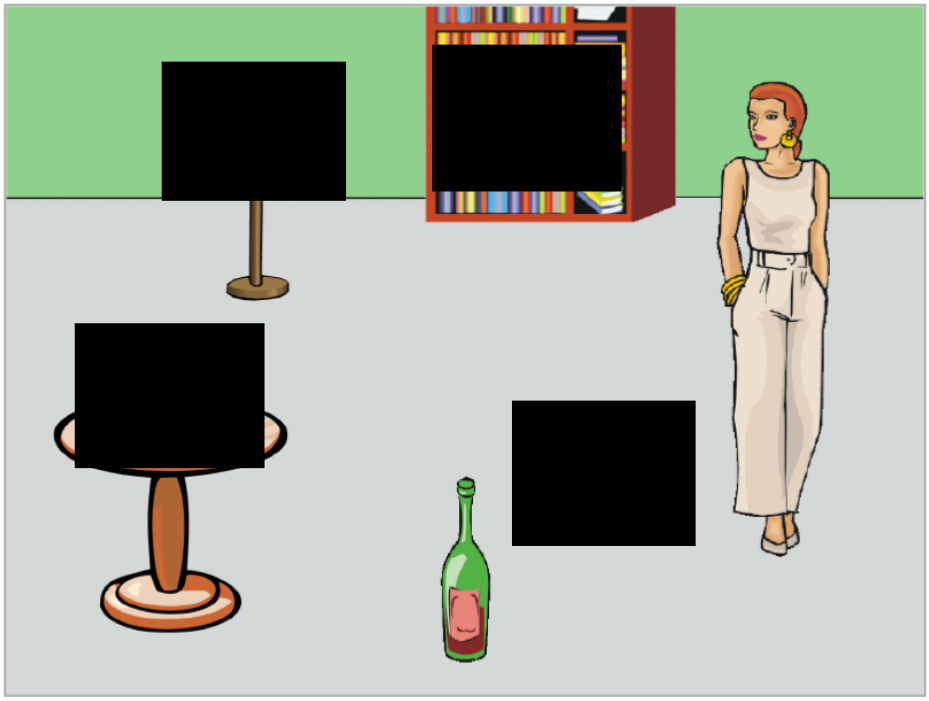


Figure 2.6. Example of the regions of interest, shown in black, superimposed over a visual scene.

2.5.2. Results

Moved vs. unmoved

During 'the wine carefully into' (anticipatory eye movements) there were more looks toward the previous location of the glass in the unmoved condition (the woman is too lazy to move the glass…), compared to the plausible moved condition (the woman will move the glass onto the table…) (*t1* (47) = -2.135, $p < .05$) (*t2* (23) = -2.435, $p < .05$). A similar pattern of eye movements was observed during "the glass" (*t1* (47) = -3.538, $p < .01$) (*t2* (23) = -3.401, $p < .01$).

Likewise, there were more looks toward the previous location of

the table in the moved condition, compared to the unmoved condition, both during “the wine carefully into” (*t1* (47) = 2.387, *p* < .05) (*t2* (23) = 3.273, *p* < .01) and “the glass” (*t1* (47) = 3.808, *p* < .001) (*t2* (23) = 4.985, *p* < .001) [2] (see figures 2.7., 2.8., and table 2.1.). These results replicate previous findings by Altmann and Kamide (2009).

Time	glass (unmoved)	glass (moved)	table (unmoved)	table (moved)
the wine carefully into	7	3	5	10
the glass	7	2	1	8

Table 2.1. Percentage of trials with looks to the previous location of the glass and the table during “the wine carefully into the glass”. Percentages calculated from the total number of trials.

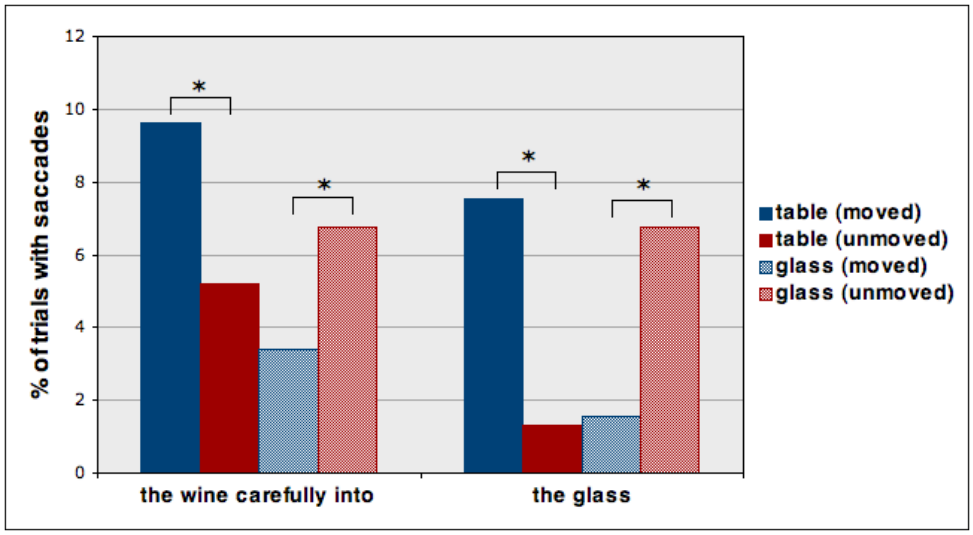


Figure 2.7. Looks to the previous location of the glass (original location) and the table (plausible location) during “the wine carefully into the glass”.

[2] Arcsine transformations were applied to the data prior to t-tests in order to better comply with the assumption of a normally distributed dataset.

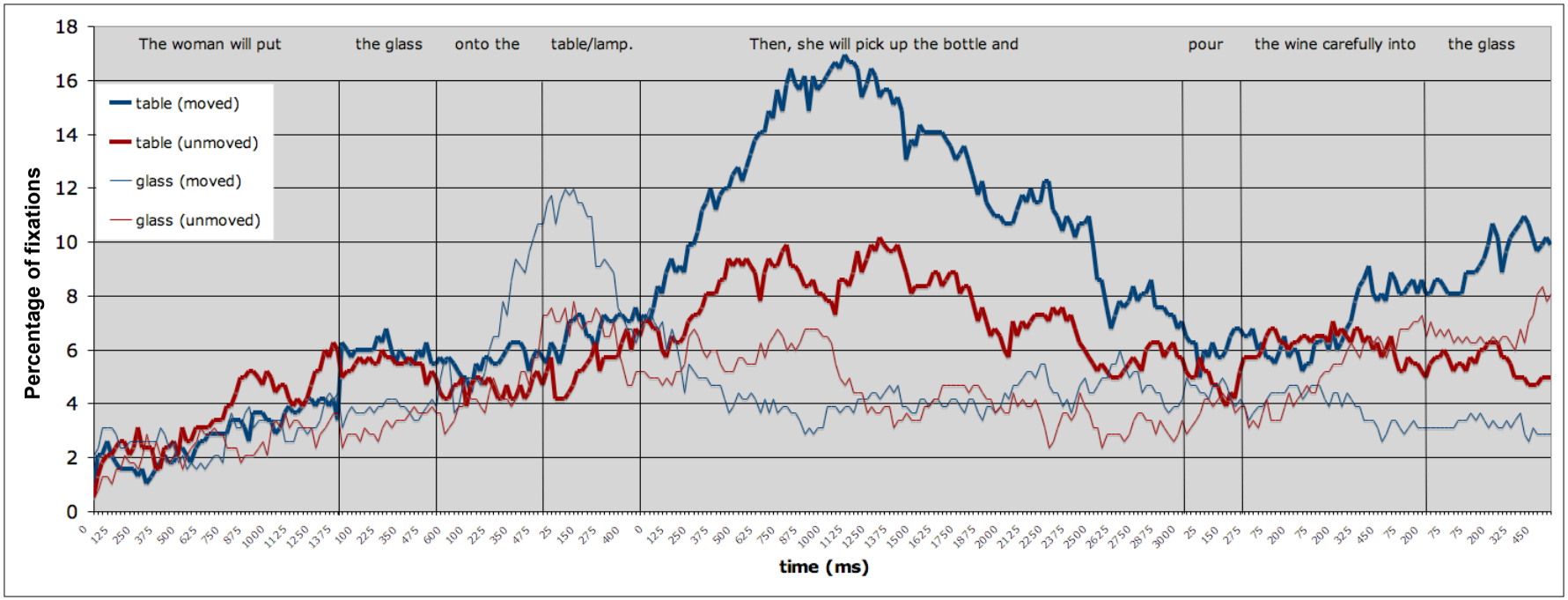


Fig 2.8. Percentage of trials with fixations toward the previous region of the table and the glass in the moved and unmoved conditions. The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

Plausible vs. implausible

During 'the wine carefully into' there were more anticipatory looks toward the previous location of the table in the plausible condition (the woman will move the glass onto the table…) than there were looks to the previous location of the lamp in the implausible condition (the woman will move the glass onto the lamp…) (*t1* (47) = -3.154, $p < .01$) (*t2* (23) = -2.163, $p < .05$). However, during 'the glass' there were no more looks toward the previous location of the table in the plausible condition, than there were looks to the previous location of the lamp in the implausible condition[3] (*t1* (47) < 1) (*t2* (23) <1).

It is further worth noting that during 'the wine carefully into' there were more looks to the previous region of the table than the books (distractor) in the plausible condition (*t1* (47) = 4.076, $p < .001$) (*t2* (23) = 4.461, $p < .001$), but there were no more looks to the previous region of the lamp than the books in the implausible (*t1* (47) = 1.374, $p > .05$) (*t2* (23) = 1.602, $p > .05$). This suggests that participants anticipated the table (plausible locations) more than the unnamed distractor, but they did not anticipate the lamp (implausible locations) any more than they anticipated the distractor. During the final reference to 'the glass', there were more looks to the region of the table than the books in the plausible condition (*t1* (47) = 3.884, $p < .001$) (*t2* (23) = 4.036, $p < .01$) and similarly more looks to the region of the lamp than the books in the implausible condition (*t1* (47) = 3.091, $p < .01$) (*t2* (23) = 2.558, $p < .05$) (see figures 2.9., 2.10., and table 2.2.).

[3] This pattern of eye movements has since been replicated in a later experiment (see appendix 4 for methods and results).

Time	lamp (impl.)	table (pl.)	books (impl.)	books (pl.)
the wine carefully into	4	10	2	2
the glass	7	8	2	1

Table 2.2. Percentage of trials with looks to the previous location of the lamp, table and books (distractor) during “the wine carefully into the glass”. Percentages calculated from the total number of trials.

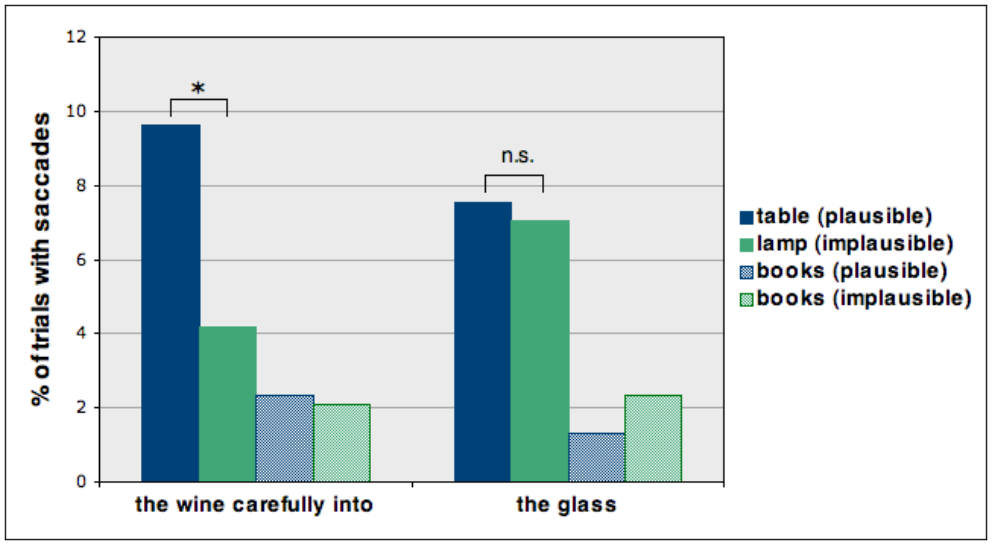


Figure 2.9. Looks to the previous location of the lamp (implausible location), table (plausible location) and the books (distractor) during “the wine carefully into the glass”.

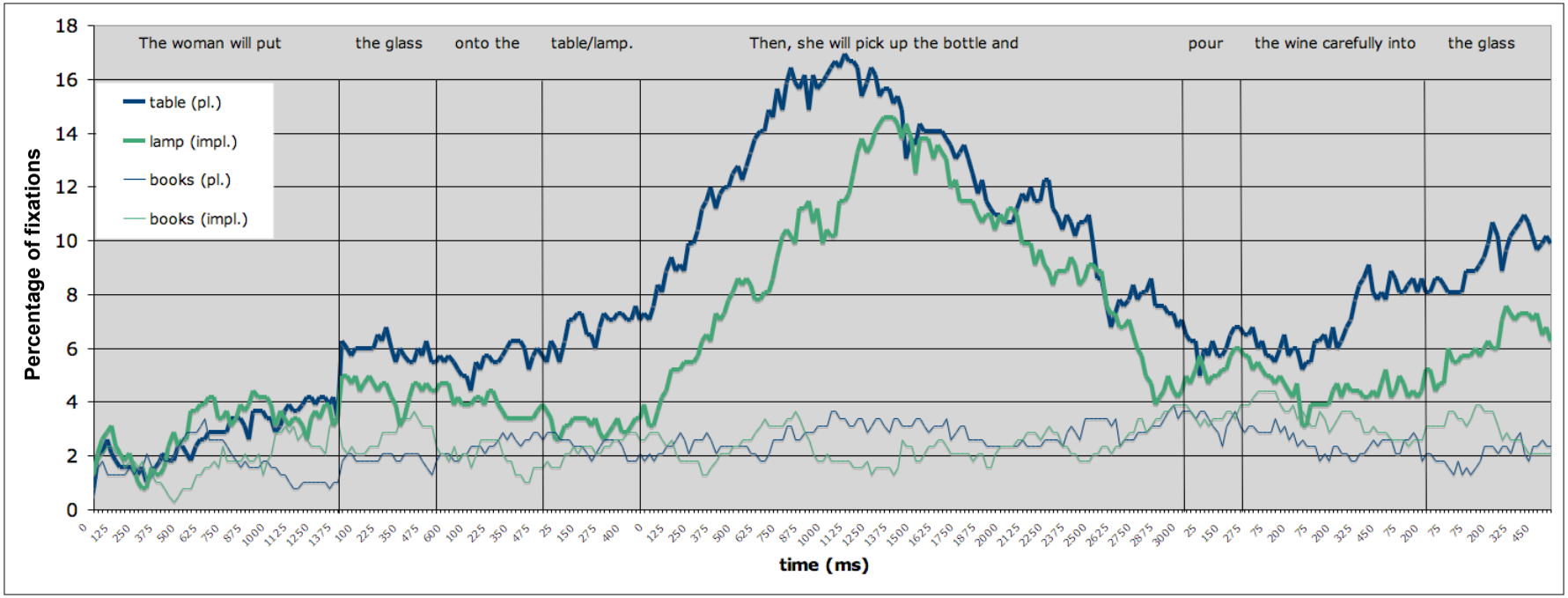


Fig 2.10. Percentage of trials with fixations toward the previous region of the table (plausible condition), the lamp (implausible condition) and the books (distractor). The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table/lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

2.5.3. Discussion

The data showed that plausibility only influenced anticipatory eye movements – at the end of the narrative there was no difference in looks to the plausible and implausible locations. This finding might initially be unexpected – if the effect of plausibility was simply a result of the implausible locations being less accessible we would expect to see a similar pattern of eye movements later in the sentence during "the glass". However, during this latter part of the sentence there were as many looks to the implausible locations, as there were looks to the plausible locations. Furthermore, we see a delay in looks to the lamp in the beginning of the sentence ("the woman will move the glass onto the..."). This delay is presumably a result of anticipatory looks to the table due to the likelihood of this being the upcoming location. Similarly, Altmann and Kamide (1999) showed that verbs such as 'eat' were able to direct anticipatory eye movements to the appropriate object within a visual scene. While the verb 'move' is not as constraining as the verb 'eat', it is possible that participants' expectations about the upcoming location might have been further constrained by experiential knowledge. This notion further relates to a study by Kamide et al., (2003) who showed that a combination of verb-related information and real-world knowledge is able to guide anticipatory eye movements towards the most appropriate and plausible (goal) item, even when this item is not presented immediately after the verb. For example, when hearing *"the woman will spread the butter… "* they found more anticipatory looks (after the verb) to the bread in the visual scene. In contrast, when hearing *"the woman will slide the butter…"* there were more anticipatory looks to the man in the visual scene. Following those principles we may speculate that in the current experiments experiential knowledge similarly guided participants' eye movements towards the most likely (and consequently, most anticipated) location for the glass. As such, participants initially anticipated (and looked at the table), only directing their eye movements towards the lamp upon hearing "…the

lamp". It may then be possible that participants' processing was affected by this delay later in the sentence, resulting in fewer eye movements to the implausible location. However, this account does not explain why we only see an effect of plausibility during the anticipatory region of the sentence. Could it be the case that participants initially processed the implausible sentences slower, yet once this implausible information had been processed it allowed them to 'catch up' during the final region of the sentence when hearing "…the glass"? Another possibility is that the difference in anticipatory eye movements is related to the likelihood of the verb (e.g. pour) referring to one of the previously introduced entities. In other words, it may be the case that when the glass is described as having been moved onto the table it presents a plausible referent for 'pouring the wine carefully into'. However, when the glass is described as having been moved onto the lamp, the unlikelihood of this location makes it a much less plausible referent[4]. If the observed plausibility effect is exclusively related to the linguistic processing of plausible and implausible scenarios we would expect to see the same pattern of eye movements if the visual scene had remained onscreen throughout the entire trial. However, if the effect is somehow modulated by the nature/constraints of the visual context we would expect to see a different pattern of eye movements depending on the presence/absence of the visual context. The second experiment aimed to explore these options further by making the visual scene available throughout the entire duration of each trial.

2.6. EXPERIMENT 2

The second experiment was identical to the previous experiment, except the visual scenes remained onscreen throughout the duration of the trials. This design allowed us to explore the extent to which the effect of plausibility is solely related to linguistic processing (in which

[4] We return to this possible explanation in chapter 5.

case we would expect to see the same pattern of eye movements as in experiment 1), or somehow mediated by the absence/presence of a visual context (in which case we would expect to see a different pattern of eye movements to that observed in experiment 1).

2.6.1. Method

Participants

Forty-eight students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

The auditory and visual stimuli were identical to those used in experiment 1.

Procedure

The procedure for this experiment was the same as in experiment 1, except that the scenes remained on the screen throughout each of the experimental trials and the visual stimulus was presented for only 1000 ms before the onset of the auditory stimuli[5]. The trials ended 11 seconds after the audio onset, lasting a total of 12 seconds. The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

We used the same four regions of interest as in experiment 1. However, since the scenes remained onscreen throughout each of the trials, we defined the regions of interest according to the outline

[5] The different timings for the picture preview (5000ms for blank screen experiments and 1000ms for concurrent screen experiments) are based on previous studies (e.g. Altmann & Kamide, 2009). This ensures that the visual scenes are presented for a similar period of time before the onset of the critical part of the narrative (e.g. '…the wine carefully into the glass').

of the target object. This way participants' eye movements had to be directed to one of the pixels occupied by each object within the scene, as opposed to rectangular regions surrounding the objects. As in experiment 1 we compared the proportion of saccades to the table and the glass in the moved and the unmoved conditions, since this would let us know whether we were able to replicate previous findings by Altmann and Kamide (2009). In order to see if plausibility of location had any influence on eye movements we compared the proportion of saccades to the table (plausible location) with the proportion of saccades to the lamp (implausible location). Finally, we compared looks to the region of the table/lamp with looks to the region of the books (distractor). This provided information about whether the proportion of saccades to the plausible and implausible locations was higher than the 'baseline' proportion of looks to an unnamed distractor item.

2.6.2. Results

Moved vs. unmoved

During 'the wine carefully into' there were no more anticipatory eye movements toward the glass in the unmoved condition (the woman is too lazy to move the glass…), compared to the plausible moved condition (the woman will move the glass onto the table…) (*t1* (47) < 1) (*t2* (23) <1). Similarly, there was no difference in the proportion of looks toward the table in the moved condition, compared to the unmoved condition (see figure 1) (*t1* (47) < 1) (*t2* (23) < 1). While Altmann and Kamide (2009) found no difference in anticipatory looks to the glass, they did see more anticipatory looks to the table in the moved condition, compared to the unmoved condition. We did not see this difference here. One explanation may lie in the inclusion of the new 'implausible' condition. This would have resulted in participants having heard objects being moved to highly unlikely locations in one third of the experimental trials, which could explain why we see fewer anticipatory looks to the table in this experiment. In

Altmann and Kamide's (2009) experiment objects were either moved to a plausible location, or not moved at all. In this case, as soon as participants heard "pick up the bottle, and pour the wine carefully into…" they would automatically anticipate the most likely upcoming location (e.g. the table). However, in this experiment objects were either moved to the table, or the lamp, which could have resulted in fewer anticipatory looks to the table.

During 'the glass' there were marginally more looks to the glass in the unmoved condition, compared to the moved condition, although this difference was not statistically significant (*t1* (47) = -1.959, $p > .05$) (*t2* (23) = -1.645, $p > .05$). In the by-subjects analysis there were more looks to the table in the moved condition compared to the unmoved condition (*t1* (47) = 2.236, $p < .05$), however the effect was not statistically significant in the by-items analysis (*t2* (23) = 1.721, $p = .09$), (see figures 2.11., 2.12., and table 2.3.). Overall, the results show a similar pattern of eye movements to that previously observed by Altmann and Kamide (2009).

Time	glass (unmoved)	glass (moved)	table (unmoved)	table (moved)
the wine carefully into	27	26	19	21
the glass	26	21	14	19

Table 2.3. Percentage of trials with looks to the glass and the table during "the wine carefully into the glass". Percentages calculated from the total number of trials.

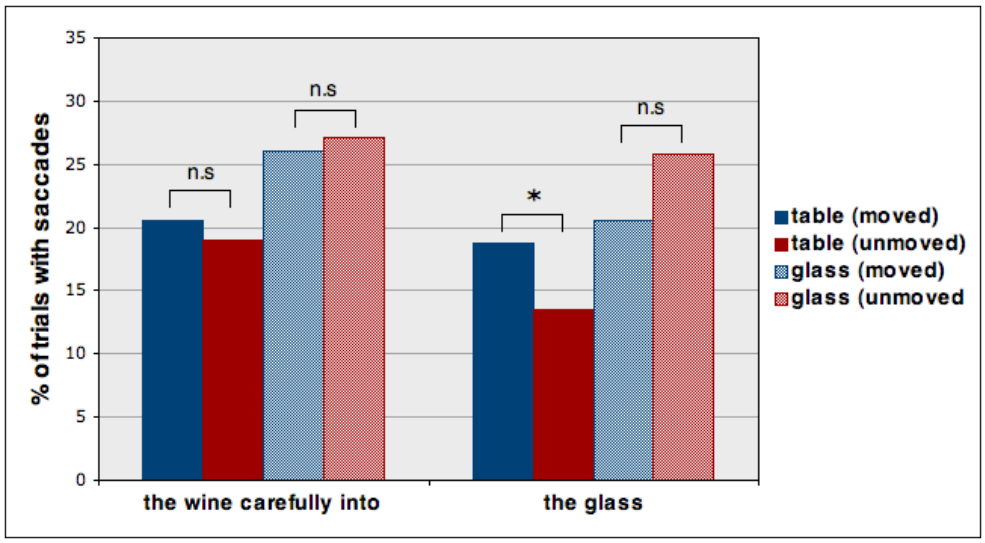


Figure 2.11. Looks to the glass (original location) and the table (plausible location) during “the wine carefully into the glass”.

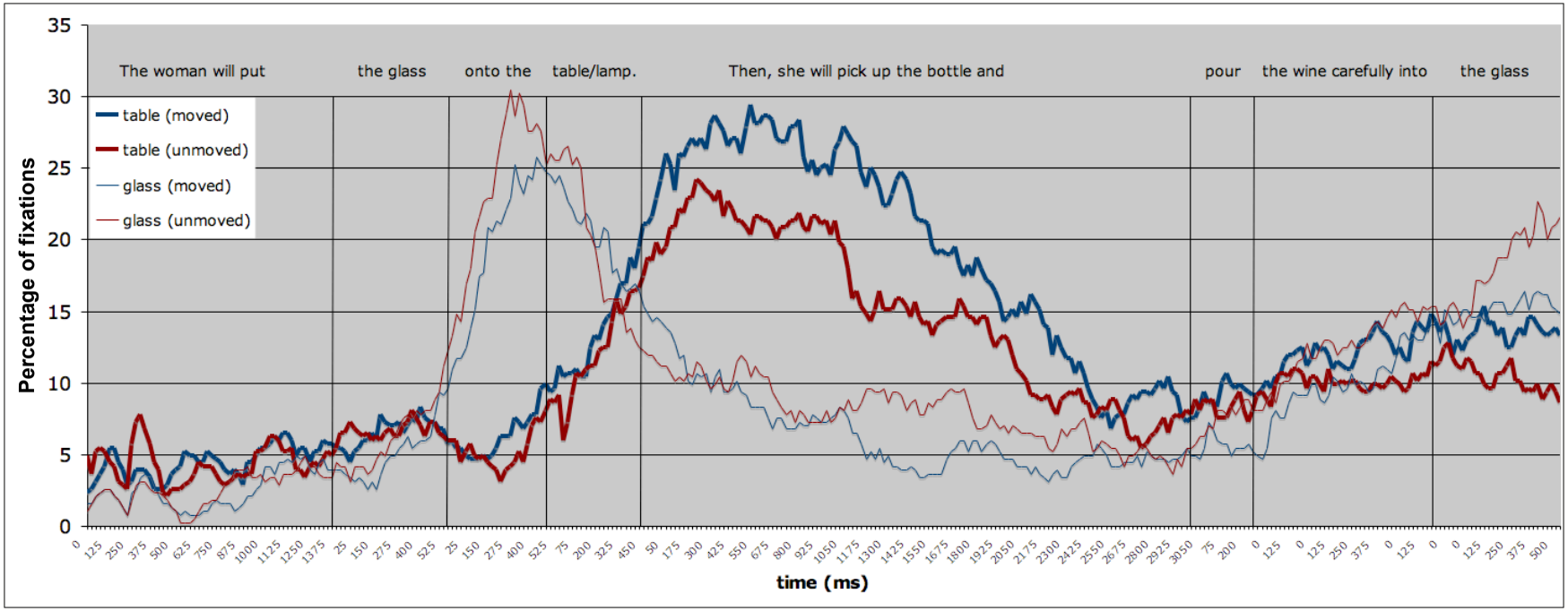


Fig 2.12. Percentage of trials with fixations toward the table and the glass in the moved and unmoved conditions. The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

Plausible vs. implausible

During 'the wine carefully into' (anticipatory eye movements) there were no more looks toward the table in the plausible condition (the woman will move the glass onto the table) than there were looks to the the lamp in the implausible condition (the woman will move the glass onto the lamp) (*t1* (47) < 1) (*t2* (23) < 1). As in experiment 1 there was no difference in looks to the table and the lamp and during 'the glass' (*t1* (47) = -1.265, $p > .05$) (*t2* (23) < 1) (see figures 2.13., 2.14., and table 2.4.).[6]

It is also worth noting that during 'the wine carefully into' there were more looks to the table than the books (distractor) in the plausible condition (*t1* (47) = 3.525, $p < .01$), although the effect narrowly failed to be statistically significant in the by-items analysis (*t2* (23) = 2.015, $p = .056$). Likewise, there were more looks to the lamp than the books in the implausible condition (*t1* (47) = 5.107, $p < .001$) (*t2* (23) = 3.403, $p < .01$). During the final reference to 'the glass', there were similarly more looks to both the region of the table and the books in the plausible condition (*t1* (47) = 7.755, $p < .001$) (*t2* (23) = 4.860, $p < .001$) and the region of the lamp and the books in the implausible condition (*t1* (47) = 4.544, $p < .001$) (*t2* (23) = 4.885, $p < .001$).

[6] In concurrent experiments where the visual scene is represented alongside spoken language it is not general practice to compare looks to one region against looks to another region because of likely confounds due to saliency in size, colour, screen positioning etc. In blank screen experiments, we assume that such confounds doesn't determine eye movements since the visual stimulus is no longer available. However, it is nonetheless possible that saliency confounds are able to make certain items more accessible in terms of participants' internal representations/visual memory of the previously presented visual scene.

Time	lamp (impl.)	table (pl.)	books (impl.)	books (pl.)
the wine carefully into	20	21	9	10
the glass	16	19	5	3

Table 2.4. Percentage of trials with looks to the lamp, table and books (distractor) during “the wine carefully into the glass”. Percentages calculated from the total number of trials.

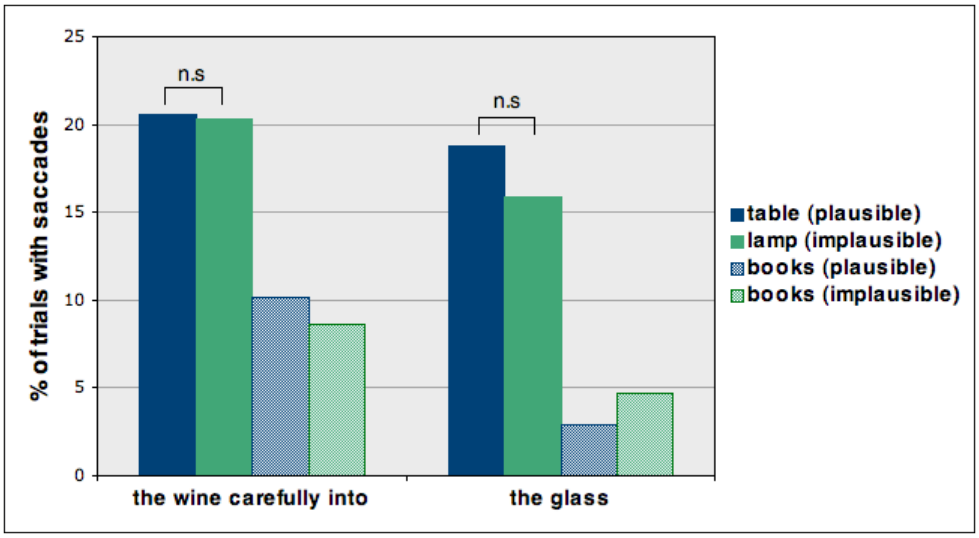


Figure 2.13. Looks to the lamp (implausible location), table (plausible location) and the books (distractor) during “the wine carefully into the glass”.

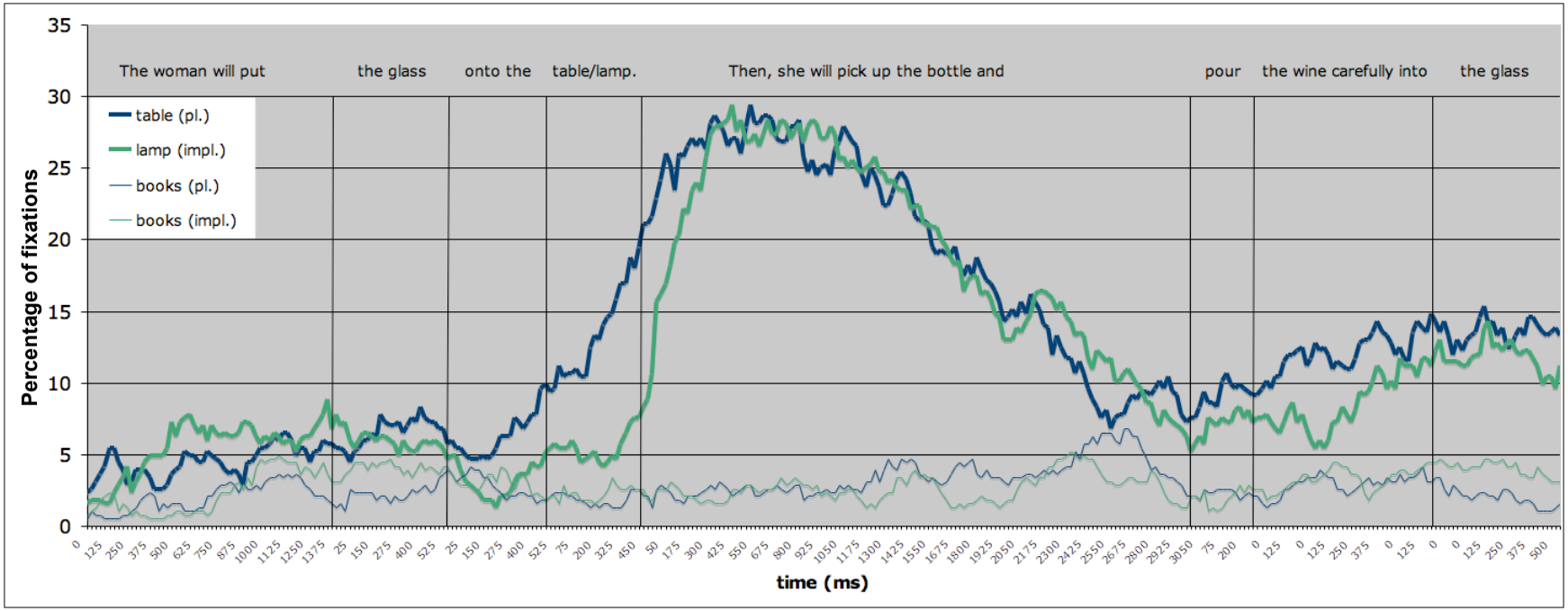


Fig 2.14. Percentage of trials with fixations toward the table (plausible condition), the lamp (implausible condition) and the books (distractor). The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table/lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

2.6.3. Discussion

In contrast to experiment 1 there was no effect of plausibility on anticipatory eye movements to the event-specific location of the glass when the visual scene was presented alongside the spoken sentences. This difference between the two experiments suggests that the effect of plausibility we observed in experiment 1 is not exclusively related to linguistic processing. In experiment 1 we observed a delay in looks to the lamp in the beginning of the sentence ("the woman will move the glass onto the..."), presumably due to the likelihood of the table being the upcoming location. In the second experiment we see a similar delay, yet no difference in anticipatory looks to the table and the lamp. We earlier proposed that this delay might explain the results from experiment 1 – i.e. that the slower processing of the first mention of the implausible location (lamp) resulted in a processing delay which consequently lead to the lack of anticipatory eye movements to the implausible location later in the sentence. However, as we only observed an effect of plausibility during the anticipatory region of the sentence (not at the end of the sentence) we had to further assume that participants' slower processing of the implausible location somehow managed to 'catch up' to the same of the participants' processing of plausible locations. The results from the second experiment rules out this possibility. While we observe the same delay in looks to the lamp pattern of eye when it is first mentioned, the pattern of eye movements observed in the concurrent version of the experiment shows an equal proportion of subsequent anticipatory looks to both the plausible (table) and implausible (lamp) locations.

To sum up, the difference in eye movements occurred in looks directed to the table and the lamp – when the scene was absent participants anticipated the pouring event to take place on the table (after hearing that the glass has been moved to this location), but not on the lamp. However, when the scene was present participants anticipated the pouring event to take place both on the table and the

lamp. Findings by Wassenburg and Zwaan (2010) showed that participants were faster to read sentences that matched the orientation of previously seen objects, suggesting that the visual memory of an object is able to influence language processing. Accordingly, participants' visual memory of previously viewed scenes should likewise be able to influence their processing of spoken sentences related to those scenes much in the same way that current visual scenes would influence sentence processing. However, if eye movements were driven solely by participants' visual memory of the scene we would not expect a difference in anticipatory looks in response to concurrent and absent scenes. Rather, we would expect more looks to the region of the glass after hearing 'pour' as opposed to the table or the lamp, since our visual memory of the scene statically represents the glass in its original and unchanged location on the floor. Similarly, findings by Altmann and Kamide (2009) showed a higher proportion of eye movements to the table when the glass was said to have been moved there, compared to when the glass was said to not have moved at all. As such, eye movements (after 'pour') reflected the moved location of the glass as conveyed by the spoken language (describing the glass as being on the table), as opposed to the depicted location of the glass (floor). These results demonstrate that participants mapped spoken sentences onto internal representations of the previously presented scenes (rather than their visual memory of those scenes), thus allowing language to mediate a dynamic updating of mental representations as the described event unfolded.

The findings from experiments 1 and 2 suggest that the availability/absence of a visual scene is able to influence our anticipation of upcoming events (at least when those events are implausible). On the other hand, Knoeferle and Crocker's (2007) blank screen experiment replicated the results from their previous study (2006) in which the visual scenes were presented alongside spoken sentences – both these studies demonstrated a stronger

reliance on concurrent, or previously seen visual information as opposed to language-mediated real-world knowledge. In contrast, the results from experiments 1 and 2 showed a different pattern of anticipatory eye movements depending on whether the visual scene is present or absent. Interestingly, this difference only occurred for anticipatory eye movements directed to the implausible locations – when comparing eye movements to the plausible locations we found a similar pattern of eye movements irrespective of whether the visual scene remained onscreen, or was removed before the onset of spoken language. This suggests that in terms of plausibility the presence/absence of visual information is able to influence our anticipation of implausible events. In the current studies the presence of a visual scene is able to guide anticipatory eye movements to both plausible and implausible locations, whereas the absence (or memory) of a visual scene is only able to direct anticipatory looks to plausible locations. This supports the notion that a concurrent visual scene provides an additional level of (visual) constraints, which consequently restricts the number of possible referents, making the implausible object-representations more accessible and as such easier to anticipate and keep track of when referred back to later in the sentence.

2.6.4. Summary & questions

The findings from experiment 1 (where we removed the visual scene before the onset of spoken language) showed that participants only anticipated plausible locations. When locations were implausible there was little or no evidence of anticipatory eye movements. In contrast, results from experiment 2 (where the visual scene remained onscreen throughout the duration of spoken language) showed no influence of plausibility – in this experiment we found anticipatory eye movements to both the plausible and implausible locations. Why do we see this difference in anticipatory looks between the two presentations types (blank screen and concurrent screen)? One

possible explanation is that we process and anticipate upcoming discourse differently when language refers to something outside our visual environment, as opposed to when language refers to entities within our immediate visual proximity. During language comprehension we assume that participants in an event will be drawn from the visual context if one is available (e.g. Altmann & Mirkovic, 2009; Cooper, 1974; Tanenhaus et al., 1995). In terms of anticipation the concurrent visual context, alongside language serves to restrict the number of potential participants and this makes it easier to anticipate which item will most likely be referred to next. However, in the context of a blank screen the anticipatory activation of appropriate representations is no longer bound exclusively by the visual context, but may now be influenced by real-world knowledge. While experiential knowledge informs us that a table is a very likely location for a glass, it also informs us that a lamp is a very unlikely location. This unlikelihood may in turn lead us to anticipate a number of alternative options.

A potential confound in these experiments is that we describe items as being moved to distinct (plausible and implausible) locations. As noted earlier, in concurrent experiments it is not general practice to compare looks to different regions since confounds such as colour, size and screen positioning may make one item more salient than another. While these issues do not appear to have influenced anticipatory eye movements in the context of a concurrent visual scene, in the context of a blank screen it is nonetheless possible that differences in saliency could have made certain locations more accessible within participants' internal representations, or memory of the previously depicted scenes. It is possible that this higher accessibility of the plausible locations could have made it easier to anticipate these regions when referred to the second time around.

Experiments 3 and 4 aimed to explore whether the pattern of eye movements we observed in experiments 1 and 2 can be generalised

and extended to contextual plausibility. In these experiments we controlled for saliency of location by describing contextually plausible and implausible items as being moved to the same location. This manipulation further allowed us to explore how an early visual introduction of implausible items (as well as the proportion initial eye movements to the plausible and implausible items) might influence the accessibility of these items when referred to later in the spoken narrative.

CHAPTER 3

MANIPULATING CONTEXTUAL PLAUSIBILITY

In the previous chapter we explored the extent to which plausibility of location is able to influence the accessibility of objects within event-representations and further, how this may be modulated by the nature of the visual context (as present or absent). We found that in the context of a blank screen (experiment 1) participants only anticipated plausible locations. When items had been moved to implausible locations there was little or no evidence of anticipatory eye movements. However, in the context of a concurrent visual scene (experiment 2) participants anticipated both the plausible and implausible locations. In both experiments there was no influence of plausibility during the final reference to the glass. Experiments 3 and 4 aimed to explore whether the pattern of eye movements we observed in experiments 1 and 2 can be generalised and extended to contextual plausibility. Instead of moving objects to different locations we now moved contextually plausible (e.g. a cat in a kitchen scene) and implausible items (e.g. a penguin in a kitchen scene) to the same location.

In experiments 1 and 2 plausibility was manipulated purely through language. In other words, the visual scenes always depicted a plausible setting and only the sentences determined whether the glass was moved to a plausible or an implausible location (e.g. the woman will move the glass onto the table/lamp). As such,

participants only became aware of the implausible component after having viewed the visual scenes. In the current set of experiments we introduced the component of plausibility immediately, presenting it within the visual scenes – in these experiments half of the experimental scenes depicted items that were contextually implausible, or in other words, unlikely considering the scene that they were presented in. Thereby, participants became aware of the implausible component immediately upon seeing the visual scene and this allowed them to process and accommodate the aspect of implausibility before the onset of spoken language. This early introduction of the implausible components of the scene gave participants more time to prepare for (and perhaps anticipate) the implausible nature of upcoming events and this could have made the implausible items more accessible and therefore easier to anticipate and keep track of when referred to later in the narrative. This manipulation of 'visual plausibility' further allowed us to examine participants' initial eye movements to contextually plausible and implausible items in the absence of language and whether initial attention to an item might make it more accessible when referred to at a later point in time. For example, it may be the case that a contextually implausible item such as 'a penguin' will be more salient and therefore more memorable, precisely because it is unusual compared to the other (contextually plausible) items within the scene. Loftus and Mackworth (1978) proposed that implausible items are often seen as being more distinctive and as such more informative for later discrimination compared to plausible items. On a similar note, Friedman (1979) proposed that participants' identification of unexpected items leads to more processing of local visual details, whereas identification of expected items results in a more global level of processing (in the next section we discuss the findings of these studies). According to these theories we might similarly expect the implausible items (as well as the locations associated with these items) to be more distinctive than the plausible items and therefore

more accessible when later referred to by spoken language.

Experiments 3 and 4 also differ from the previous experiments in terms of how the target items within the scene were referred to. In experiments 1 and 2 the reference during the second sentence to the plausible or implausible location was somewhat indirect. In other words, these locations were only directly referred to at the beginning of the narrative (e.g. 'the woman will move the glass onto the table/lamp'). Subsequent eye movements to the table and the lamp were due to the second and final reference to the glass (e.g. 'then, she will pick up the bottle, and pour the wine carefully into the glass'), whereby participants' knowledge of the location of the glass as described by spoken language led to an increase in eye movements to this location. In contrast, the current experiments manipulated the plausibility of the referent itself (e.g. 'the woman will lift the cat/penguin onto the table. Then she will quickly feed the cat/penguin'). As such, the location of the referent was always the same, whereas the referent itself was either contextually plausible, or implausible. This manipulation allowed us to explore how plausibility of the referent (as opposed to location) might influence eye movements when described as being moved to an identical location. As mentioned earlier, if implausible items are more salient and memorable than their plausible counterparts they should similarly be more accessible when later referred to[7]. If this is the case, we should see an increase in anticipatory looks to the location of the implausible items – both when referred to in the context of a concurrent or blank screen. On the other hand, a similar pattern of eye movements to that observed in experiments 1 and 2 would suggest that the influence of plausibility is not specifically related to plausibility of location, but may be extended to the contextual plausibility of the referent itself.

[7] A previous study by van Gompel and Majid (2004) showed that participants processed pronouns were faster when the lexical frequency of the antecedent was low compared to when the lexical frequency of the antecedent was high. They proposed that this was due to a stronger saliency of the infrequent antecedents.

3.1. Object-plausibility in the context of a visual scene

From time to time we may come across objects that seem strangely out of place considering the context that we encounter them in (e.g. a mouse in the bedroom), whereby our initial encounter with this item might take us by surprise and therefore attract our attention in a different manner from items we would expect to encounter within that same context. An early study by Loftus and Mackworth (1978) investigated how participants view pictures containing either contextually plausible objects (e.g. a tractor in a farmyard scene), or contextually implausible objects (e.g. an octopus in a farmyard scene). Participants' eye movements were recorded as they viewed pictures, which they had previously been told required a subsequent recognition task. The data showed that participants fixated more often on contextually implausible objects, compared to contextually plausible items and further that participants' first fixations to the implausible objects were longer than their first fixations to the plausible objects. These findings suggest that the context in which an item is presented can influence how we view that item. Specifically, in terms of processing and encoding visual information for subsequent recognition, implausible items are seen as being more distinctive and as such more informative when later having to discriminate between similar visual scenes. Loftus and Mackworth further explain their findings in relation to schema theory, suggesting that implausible objects require more processing in order to incorporate them into participants' schematic and prototypical representation of the depicted event (but see Henderson, Weeks & Hollingworth, 1999; Underwood & Foulsham, 2006 for an alternative explanation). In the current study we similarly manipulated the contextual plausibility of items since this allowed us to examine eye movements to contextually plausible and implausible items in the absence of language and the extent to which increased initial attention to an item might make it more accessible when later referred to.

Friedman (1979) similarly recorded participants' eye movements while they viewed pictures containing expected (e.g. a sink in the context of a kitchen) and unexpected (e.g. a hockey stick in the context of a kitchen) objects. After seeing a picture for 30 seconds participants had to discriminate the target pictures from distractor pictures in which either the expected or unexpected items had been changed. Friedman proposed that participants' identification of unexpected items would lead to more processing of local visual details, whereas identification of expected items would results in a more global level of processing. Her findings supported this theory, showing that participants' first fixations to the unexpected items were approximately twice as long as first fixations to the expected items. During the discrimination task participants noticed when both the expected and unexpected items had been removed from the visual scene, or replaced with different items. However, when the items were slightly different participants tended to only recognise the changes made to the unexpected items and this occurred regardless of having fixated on both the unexpected and expected target items when viewing the pictures. This supports Friedman's notion of unexpected items being processed locally, because they are implausible and therefore more informative for later discrimination. In contrast, scenes containing expected items were processed more globally, since our stereotypical knowledge about such scenes already entailed the presence of such objects, whereby they were deemed less informative for later discrimination. Therefore, participants' recognition and discrimination of expected items appeared to rely more on their prototypical knowledge of a familiar scene, whereas later recognition and discrimination of unexpected items relied more on the visual memory of the scene itself.

These studies demonstrate the importance of the context in which objects are presented and further, how 'out of place' objects are sometimes able to attract our attention in a different manner from items we expect to encounter within specific contexts. If contextually

implausible items attract more attention and participants rely more on their visual memory for such items we may further speculate that implausible items will be more accessible when referred to in the context of a blank screen. If this is the case, we would expect to see a higher proportion of looks to the regions of the implausible items, than the plausible items (as well the locations of these items) when referred to later in the narrative.

3.2. EXPERIMENT 3

In experiment 3 we used the blank screen paradigm to explore the extent to which the plausibility effect observed in experiment 1 is solely dependent on plausibility of location, or if this effect can be extended to contextual plausibility. In other words, will we see a similar pattern of eye movements when we manipulate the plausibility of the referent itself, rather than its location? We created a new set of stimuli, which presented either contextually plausible items (e.g. a cat in a kitchen), or contextually implausible items (e.g. a penguin in a kitchen). The items were always described as being moved to the same location. This design further allowed us to examine participants' initial eye movements to the contextually plausible and implausible items in the absence of language and whether increased initial attention to the implausible items might make them more accessible and therefore easier to anticipate when referred to later in the narrative.

3.2.1. Method

Participants

Thirty-two students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

Sixteen experimental scenes (see figure 3.1.) were matched with two conditions:

1. The woman will lift the cat onto the table. Then, she will quickly feed the cat.

   (Contextually plausible)

2. The woman will lift the penguin onto the table. Then, she will quickly feed the penguin.

   (Contextually implausible)

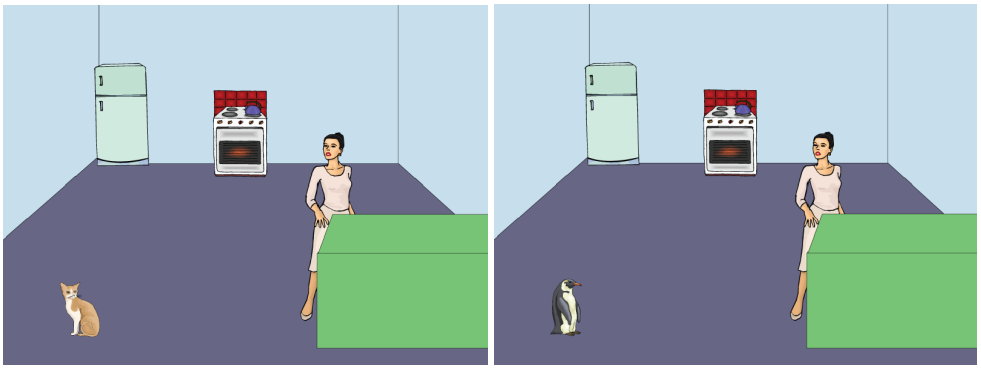


Fig 3.1. Example of the visual scenes paired with sentence 1 and 2 as shown above.

In both conditions the first sentence always described the agent moving the target object to a new location. In condition (1) the target object (cat) was plausible, given the context it was presented in. In condition (2) the target object (penguin) was implausible, given the context of the scene (see appendix 5 for a full list of the experimental sentences). In both conditions the target objects were always moved to the same locations[8]. In addition to the 16 experimental trials, 32 sentence-picture pairs were included as fillers (see appendix 6 for an

[8] In experiments 1 and 2 we compared eye movements across three conditions (plausible moved, implausible moved and unmoved) since this design allowed us to replicate the results from Altmann and Kamide (2009) in addition to exploring the influence of plausibility. As the experiments presented in this thesis were mainly interested in the influence of plausibility we did not include an 'unmoved' condition in the subsequent experiments.

example of the filler items). Twelve of the filler scenes depicted a contextually implausible item, however this item was never referred to. These items were included so that participants would not be able to automatically assume the subsequent relevance of the implausible items.

The visual stimuli were created using commercially available ClipArt packages (see appendix 7 for a full set of the experimental pictures) and presented at a resolution of 800 x 600 pixels. The positions of the objects representing the plausible and implausible items (e.g. the cat and the penguin), the locations that these were described as being moved to (e.g. the table), and the distractor locations (e.g. the oven) were counterbalanced to ensure that the location of each item varied across the full set of experimental scenes. The sentences were recorded by a male native speaker of British English and sampled at 44.1 KHz. The audio files were played via a mono channel split across two speakers that were positioned on each side of the screen. The onsets and offsets of the critical words in the experimental sentences were noted (using a sound editing program) for carrying out later analysis.

The sentences were coupled with their corresponding pictures and allocated to two conditions in a fixed-random order. Two lists of stimuli were made, which included one of the 16 sets of experimental scenes alongside the corresponding version from each of the sentence-pairs. The same 32 fillers were used in both conditions. Participants were allocated to either condition 1 or 2 and presented with a total of 48 trials presented in the same randomised order for all conditions.

Norming for action & contextual plausibility

To ensure that the items were perceived as intended we normed the experimental stimuli for contextual plausibility and likelihood of the actions being performed on the items. In the first norming study we presented 46 participants with the 16 experimental scenes and asked

them to consider the target item (e.g. a penguin) alongside the context that it was presented in (e.g. a kitchen). Participants were then asked to rate the likelihood of the item given its surrounding context. Plausibility was rated on a scale from 1-7 (1 being “not very plausible” and 7 being “very plausible”). Two lists were created, ensuring that participants rated either the plausible or the implausible version of each event. Thereby list (a) presented 50% of the plausible events and 50% of the implausible events, whereas list (b) presented the opposite versions of the plausible and implausible events. The presentation order of the sentences was randomised across both lists. The mean rating for the plausible contexts was 6.02 (SD = 0.76) and 1.89 (SD = 0.69) for the implausible contexts. The difference in likelihood between the plausible and implausible contexts differed significantly, t (15) = 15.262, p < .001; items were consistently judged as more likely when presented in a plausible context.

In the second norming study we presented 55 participants with the 16 experimental sentences and asked them to consider the target item and rate how likely they would be to perform a specific action on this item (e.g. suppose you had a cat/penguin – how likely would you be to feed it?). Likelihood was rated on a scale from 1-7 (1 being “not very likely” and 7 being “very likely”). The mean rating for the actions performed on the plausible items was 4.41 (SD = 1.83) and 3.54 (SD = 1.33) for the implausible location. There was no significant difference in likelihood between actions being performed on the plausible and implausible items, t (15) = 1.892, p > .05.

Procedure

The visual scenes were presented for five seconds and then replaced by a grey screen. The onset of the audio (corresponding to the scenes) occurred one second after the scene had been removed and the trials finished 11 seconds after the audio onset. As such, each trial lasted a total of 17 seconds. The total duration of the

experiment was approximately 40 minutes. The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

We defined four identically sized regions of interest within each scene; one, corresponding to the new location of the cat/penguin (e.g. the previous region of the table top), another corresponding to the original location of the cat/penguin (the previous region of the cat/penguin), and a third area that corresponded to an unnamed distractor item (the previous region of the oven) (see figure 3.2.).

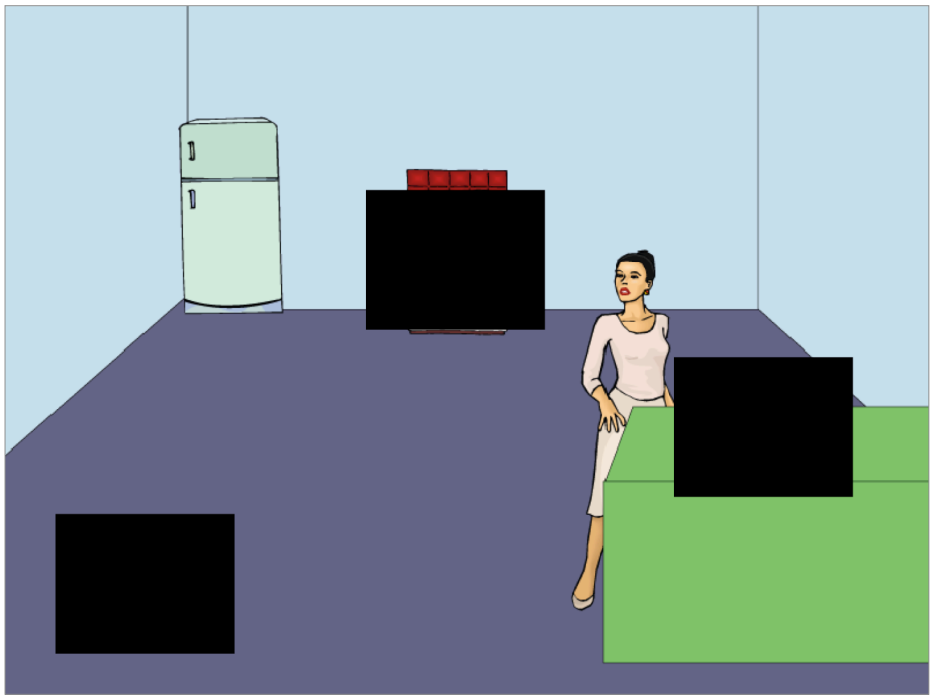


Figure 3.2. Example of the regions of interest, shown in black, superimposed over a visual scene.

Participants' eye movements were examined during certain time points in the spoken sentences with the percentage of saccades and percentage of fixations as the dependent measures. These critical time points occurred during 'quickly feed' and during the final noun phrase 'the cat/penguin'. It is important to note that in contrast to the critical time points analysed in experiments 1 and 2 where the

anticipatory region of the sentence occurred after the final verb (e.g. 'the woman will put the glass onto the table/lamp. Then, she will pick up the bottle and pour *the wine carefully into* the glass'), the anticipatory region in the current experiments occurred earlier, beginning at the onset of the word immediately preceding the final verb, and ending at the offset of the final verb (e.g. 'the woman will lift the cat/penguin onto the table. Then, she will *quickly feed* the cat/penguin'). This difference in the anticipatory region is due to the structure of the experimental sentences and allowed us to measure anticipatory eye movements at the earliest possible moment, both before and during verb-related information[9].

To begin with, we compared the proportion of saccades to the region of the table in the plausible (cat) and implausible (penguin) conditions. This allowed us to see if the contextual plausibility of the items moved to this location had any influence on eye movements. We also compared looks to the table with looks to the oven (distractor). This provided information about whether the proportion of saccades to the table (in both the plausible and implausible conditions) was higher than the 'baseline' proportion of looks to an unnamed distractor item. Finally, we compared the proportion of saccades to the region of the cat (plausible item) and the region of the penguin (implausible item). This comparison allowed us to see if the unexpected nature of the contextually implausible items would lead to a higher proportion of looks than the contextually plausible items.

[9] Kamide, Altmann and Haywood (2003) showed that in Japanese (where the verb is typically presented at the end of a sentence), information derived from pre-verbal arguments was able to facilitate anticipatory eye movements to the most appropriate target. Later we return to the discussion of what can be anticipated during this early region.

3.2.2. Results

Table – plausible vs. implausible

During 'quickly feed' there were more anticipatory looks toward the previous location of the table after hearing 'the woman will lift the cat onto the table' (plausible condition), than after hearing 'the woman will lift the penguin onto the table' (implausible condition) (*t1* (31) = 2.982, $p < .01$) (*t2* (15) = -2.573, $p < .05$). Interestingly, we see the same difference even before the onset of the verb – during 'quickly' there were more looks to the region of the table when the cat was said to have been moved there, than when the penguin was said to have been moved there (*t1* (31) = 3.418, $p < .01$) (*t2* (15) = -2.374, $p < .05$). With respect to the proportion of anticipatory eye movements it is further worth noting that during 'quickly feed' there were more looks to the table than the oven (distractor) in the plausible condition (*t1* (31) = 4.996, $p < .001$) (*t2* (15) = 2.961, $p < .05$). In contrast, there were no more looks to the table than the oven in the implausible condition (*t1* (31) = 1.158, $p > .05$) (*t2* (15) = 1.420, $p > .05$). This suggests that when the cat (contextually plausible item) was described as having been moved to the table participants anticipated the table more than the unnamed distractor. However, when the penguin (contextually implausible item) was described as having been moved to the table participants did not anticipate that location any more than they anticipated the distractor. During the final reference to 'the cat/penguin', there was no difference in looks to the region of the table between the plausible and implausible conditions (*t1* (31) < 1) (*t2* (15) <1)[10] (see figures 3.3., 3.5., and table 3.1.).

[10] The percentage of trials with saccades to the previous region of the table is quite low during the final reference to 'the cat/penguin' and this is the case in both conditions. However, 1000ms after the offset of 'the cat/penguin' the percentage of trials with saccades increases to 11-12%, suggesting a slight delay in looks to this region.

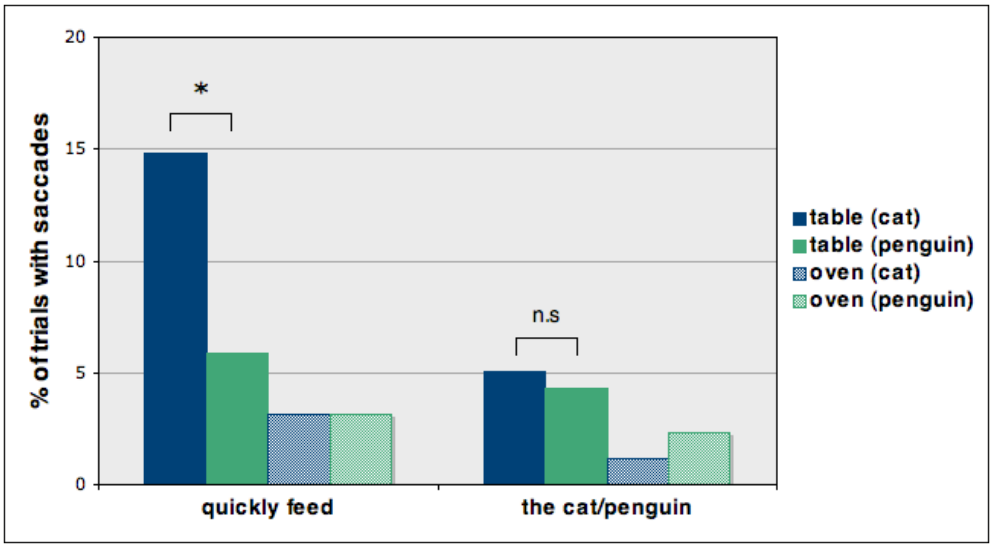


Fig 3.3. Looks toward the previous region of the table and the oven (distractor) during 'quickly feed the cat/penguin'.

Cat/penguin – plausible vs. implausible

During 'quickly feed' there was no difference in anticipatory looks to the region of the cat in the plausible condition and looks to the region of the penguin in the implausible condition (*t1* (31) < 1) (*t2* (15) <1). Likewise, during the final reference to the cat/penguin there was no difference in looks to the region of the cat in the plausible condition and looks to the region of the penguin in the implausible condition (*t1* (31) = 1.438, $p > .05$) (*t2* (15) = -1.282, $p > .05$).

However, during the first 500 milliseconds of the picture preview[11] (before the onset of spoken language) there were more looks to the region of the penguin, than there were looks to the region of the cat (*t1* (31) = -2.837, $p < .01$), although the effect was not statistically significant in the by-items analysis (*t2* (15) <1). Similarly, there were more fixations on the region of the penguin than the cat (*t1* (31) = -3.499, $p < .01$), but again the effect was not statistically significant in the by-items analysis (*t2* (15) <1) (see figure 3.4.). There was no difference in first fixation duration and gaze duration to these regions.

[11] The visual scene was presented for 5000ms and then removed before the onset of spoken language.

Time	table (impl.)	table (pl.)	oven (impl.)	oven (pl.)	penguin (impl.)	cat (pl.)
quickly feed	6	15	3	3	2	3
the cat/penguin	5	4	2	1	1	2

Table 3.1. Percentage of trials with looks to the previous location of the table, oven (distractor), penguin and cat during “quickly feed the cat/penguin”. Percentages calculated from the total number of trials.

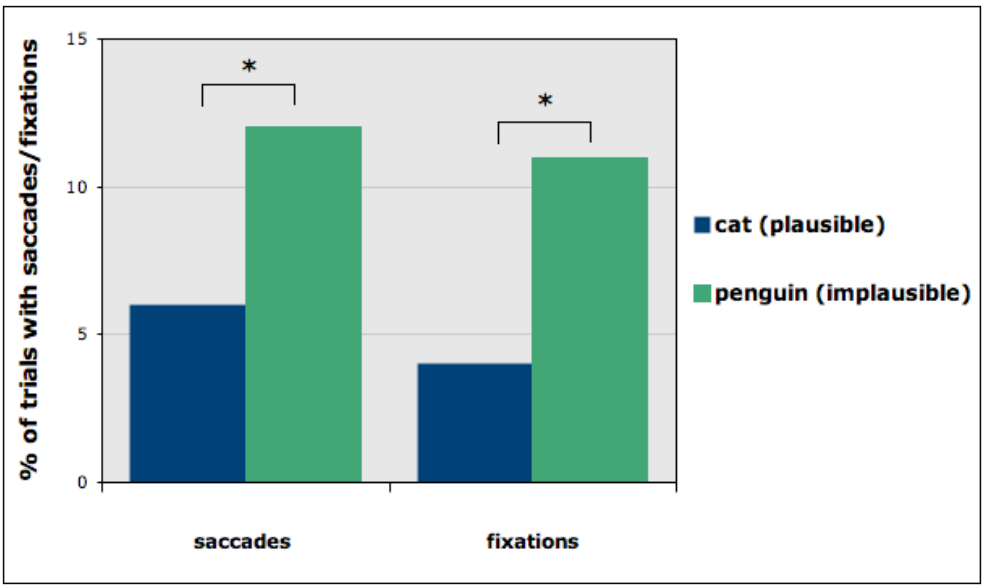


Fig 3.4. Percentage of trials with saccades to the cat and the penguin. as well as percentage of trials with fixations on the cat and the penguin during the first 500ms of the picture preview.

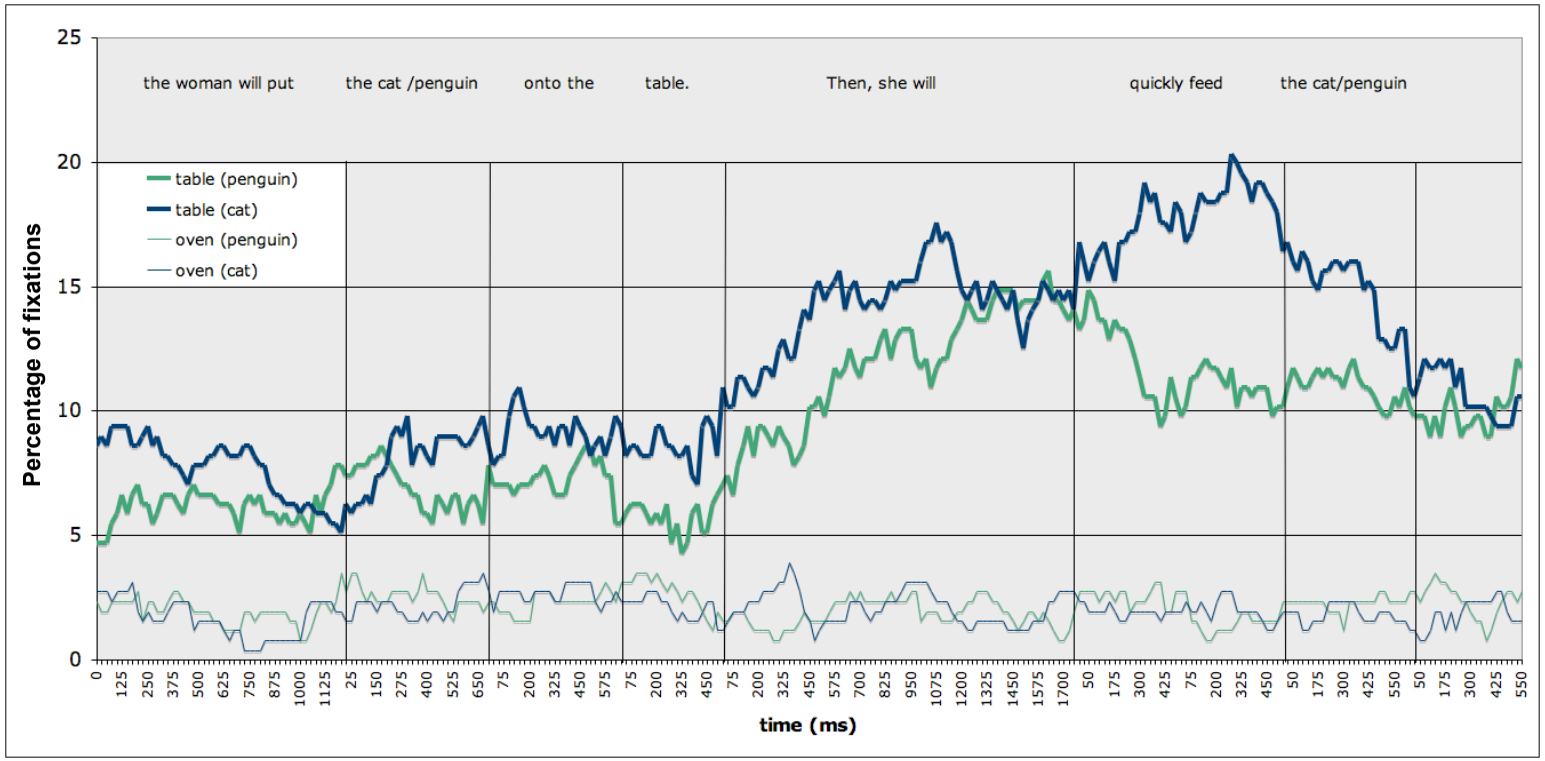


Fig 3.5. Percentage of trials with fixations toward the previous region of the table and the oven (distractor). The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the cat/penguin onto the table. Then, she will quickly feed the cat/penguin'*. The fixations were calculated every 25 ms sequentially from the synchronisation point.

3.2.3. Discussion

In line with experiment 1, plausibility influenced anticipatory eye movements to the event-specific location of the target item, but not eye movements during the final reference to 'the cat/penguin' (in experiment 1 plausibility influenced anticipatory eye movements to the event-specific location of the glass, but not eye movements during the final reference to 'the glass'). This suggests that the plausibility effect observed in experiment 1 is not exclusively related to plausibility of location, but can be further extended to contextual plausibility. During 'quickly feed' participants anticipated (and looked to) the table when the contextually plausible item (cat) was described as having been moved there but not when the contextually implausible item (penguin) had been moved there. This difference indicates that while an early visual manipulation of plausibility may have allowed participants more time to process and prepare for the implausible nature of upcoming events, this 'extra time' did not make the implausible items easier to anticipate and keep track of when referred to later in the sentences.

Interestingly, we also found a difference in anticipatory eye movements even before the onset of the verb, during 'quickly'. This suggests that anticipatory looks to the table occurred independently on any knowledge derived from verb-related information (e.g. upon hearing that the woman is about to feed the cat/penguin), instead indicating that any action the woman was about to perform (e.g. then she will quickly…) would be more likely to involve the cat, as opposed to the penguin (despite that the constraints of this part of the narrative would allow the woman to perform any number of actions not related to either the cat or the penguin). It may be the case that the difference in anticipatory eye movements reflected the number of possible actions involving the plausible and implausible items within the context of the narrative. For example, our knowledge and experience with cats informs us that cats are often kept as pets and therefore they might regularly spend time in a kitchen.

Considering this, the woman could be about to perform a number of actions involving the cat (e.g. feeding, stroking, brushing etc.). On the other hand, our experience of penguins (as a whole, or in kitchens) is for most people much more limited, as is our knowledge of the type of actions one might perform involving a penguin. This might explain why participants anticipated an event to take place on the table when the cat had been moved there, but not when the penguin had been moved there.

In experiments 1 and 2 items were always described as being moved to separate (plausible and implausible) locations, whereby differences in saliency could have made the plausible locations more accessible within participants' internal representations, or memory of the previously depicted scenes. The pattern of eye movements observed in the current experiment suggest that the plausibility effect observed in experiment 1 did not arise as a result of differences in saliency between the plausible and implausible locations since the current experiment removed any discrepancies in saliency by moving the target items to the same location. In a future study one might also manipulate the plausibility of the context in which an item is presented – for example, presenting the penguin in either a plausible context such as a zoo, or an implausible context such as a kitchen. This design would allow us to control for both the saliency of the target item, as well as the location that this item is described as being moved to. If the effect we observed in experiments 1 and 3 is purely driven by plausibility we would similarly expect to see a higher proportion of anticipatory eye movements if the penguin was to be presented in a plausible context, compared if it was presented in an implausible context.

Friedman (1979) proposed that participants' identification of unexpected items lead to more processing of local visual details because these items were more informative for later discrimination. In contrast, identification of expected items resulted in a more global level of processing – since our stereotypical knowledge about familiar

scenes already entailed the presence of such objects they were deemed less informative. Her findings supported this notion showing longer first fixations, as well as better recognition to changes to unexpected items, compared to expected items. As such, participants' recognition and discrimination of expected items seemed to rely more on prototypical knowledge, whereas later recognition and discrimination of unexpected items relied more on the visual memory of the scene itself. If this theory extends beyond identification and recognition we would expect to see longer first fixations on the contextually implausible items since participants' had to rely on internal representations or their visual memory of these scenes at the onset of spoken language. In contrast to the findings by Friedman (as well as Loftus and Mackworth, 1978), we found no difference in first fixation, or gaze durations to the plausible and implausible items. This suggests that Friedman's findings might have been specifically related to participants having to perform a subsequent discrimination and recognition task. In the current experiment participants were simply required to 'look and listen' and it may be the case that when no later recognition or discrimination is required, participants simply have less need to rely on visual memory for contextually implausible items.

We did, however see a difference in the proportion of saccades and fixations during the first 500 milliseconds of the picture preview, with more attention focused on the contextually implausible items than the plausible items. This suggests that the implausible items did initially attract participants' attention more than the plausible items, but only for a brief moment. Furthermore, this increased attention to the implausible items did not seem to make these items, or their locations more accessible when referred to later in the sentence. During the final reference to the cat/penguin there were no more looks to the region of the penguin than there were looks to the region of the cat and similarly there was no difference in looks to the table. During 'quickly feed' there was no difference in the proportion of

anticipatory looks to the region of the cat and the penguin. We did however, see more looks to the region of the table when the cat was said to have been moved there, than when the penguin was said to have been moved there. This indicates that the smaller proportion of anticipatory looks to the region of the table when the penguin had been moved there were not due to a higher proportion of looks to the region of the penguin, compared to the region of the cat. These findings suggest that the higher proportion of initial attention to the penguin did not influence the proportion of later looks to this region, or the described location of the penguin. Several factors may explain why a higher proportion of looks to the penguin did not influence the proportion of later looks to this region. Firstly, the penguin only attracted more attention than the cat for a very brief period – after 500 milliseconds there was no difference in looks to these items. It may be the case that this is simply too short a period to be able to exert any influence on accessibility and subsequent eye movements. Secondly, it is important to remember that the visual scene was replaced with a blank screen before the onset of spoken language, in effect removing any advantage of perceptual saliency the penguin may have had over the cat (however, it may have remained conceptually salient, although if that were the case, we might have expected to see more looks to the region of the table, or the penguin towards the end of the second sentence). As such, we may speculate that if the visual scene was to remain onscreen, the continuous saliency of the penguin might make it more accessible than the cat when referred to later in the sentence.

3.3. EXPERIMENT 4

The results from experiment 3 showed that participants anticipated the region of the table when the cat was said to have moved there (contextually plausible), but not when the penguin was said to have been moved there (contextually implausible). This is similar to the findings from experiment 1, which showed that participants

anticipated the region of table (plausible location), but not the lamp (implausible location). In contrast, the data from experiment 2, where the scene remained onscreen throughout the trial, showed that participants anticipated both the table and the lamp. These first two experiments suggests that the effect of plausibility is not specifically related to difficulties with processing the implausible sentences, but rather modulated somehow by the absence/presence of the visual context. In light of this difference we would similarly expect that leaving the visual scene onscreen would result in anticipatory eye movements to the table when both the cat and the penguin are described as having been moved this location.

On the other hand, certain differences in the experimental design might lead us to expect a different outcome. As mentioned earlier, reference to the plausible and implausible locations in experiments 1 and 2 was indirect since these locations were only directly referred to at the start of the narrative (e.g. the woman will move the glass onto the *table/lamp*) – any subsequent eye movements to the table and the lamp were derived from the second and final reference to the glass (e.g. then, she will pick up the bottle, and pour the wine carefully into *the glass*), whereby participants' knowledge of the location of the glass as described by spoken language lead to an increase in eye movements to this location. In contrast, the current experiments manipulated the plausibility of the referent itself (e.g. 'the woman will lift the *cat/penguin* onto the table. Then she will quickly feed the *cat/penguin*'), which allowed us to explore the extent to which the saliency of contextually plausible and implausible items might influence accessibility when referring back to these items. For example, if implausible items are more salient than their plausible counterparts they should similarly be more memorable and accessible when later referred to. While the data from experiment 3 showed a higher proportion of looks and fixations to the contextually implausible items compared to the contextually plausible items, this difference only appeared during the first 500 milliseconds of the

picture preview. Furthermore, the removal of the visual scene prior to the onset of language is likely to have minimised any influence of saliency during spoken language. If this is the case we would expect more (later) looks to the implausible items than the plausible items (and their described locations) if the visual scene remained onscreen, since the continuous saliency advantage of the implausible items should make them more accessible when referred to in the second sentence.

3.3.1. Method

Participants

Thirty-two students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

The auditory and visual stimuli were identical to those used in experiment 3.

Procedure

The procedure for this experiment was the same as in experiment 3, except that the scenes remained on the screen throughout each of the experimental trials and the visual stimulus was presented for only 1000 milliseconds before the onset of the auditory stimulus. The trials ended 11 seconds after the audio onset, lasting a total of 12 seconds. The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

We used the same four regions of interest as in experiment 3. However, since the scenes remained onscreen throughout each of the trials, we defined the regions of interest according to the outline

of the target object. This way participants' eye movements had to be directed to one of the pixels occupied by each object within the scene, as opposed to rectangular regions surrounding the objects. Firstly, we compared the proportion of saccades to the table in the plausible (cat) and implausible (penguin) conditions. This allowed us to see if the contextual plausibility of the items moved to this location had any influence on eye movements. We also compared looks to the table with looks to the oven (distractor). This provided information about whether the proportion of saccades to the table (in both the plausible and implausible conditions) was higher than the 'baseline' proportion of looks to an unnamed distractor item. Finally, we compared the proportion of saccades to the cat (plausible item) and the penguin (implausible item). This comparison allowed us to see if the unexpected nature of the contextually implausible items would attract a higher proportion of looks than the contextually plausible items.

3.3.2. Results

Table – plausible vs. implausible

During 'quickly feed' there was no difference in anticipatory looks to the table between the plausible (cat) and implausible (penguin) conditions (*t1* (31) = 1.323, $p > .05$) (*t2* (15) = -1.168, $p > .05$). Likewise, before the onset of the verb, during 'quickly' there were no more looks to the table when the cat was said to have been moved there than when the penguin was said to have been moved there (*t1* (31) < 1) (*t2* (15) < 1). During the final reference to 'the cat/penguin' there were no more looks to the table when the cat was said to have been moved there than when the penguin was said to have moved there (*t1* (31) < 1) (*t2* (15) <1) (see figures 3.6. and table 3.2.).

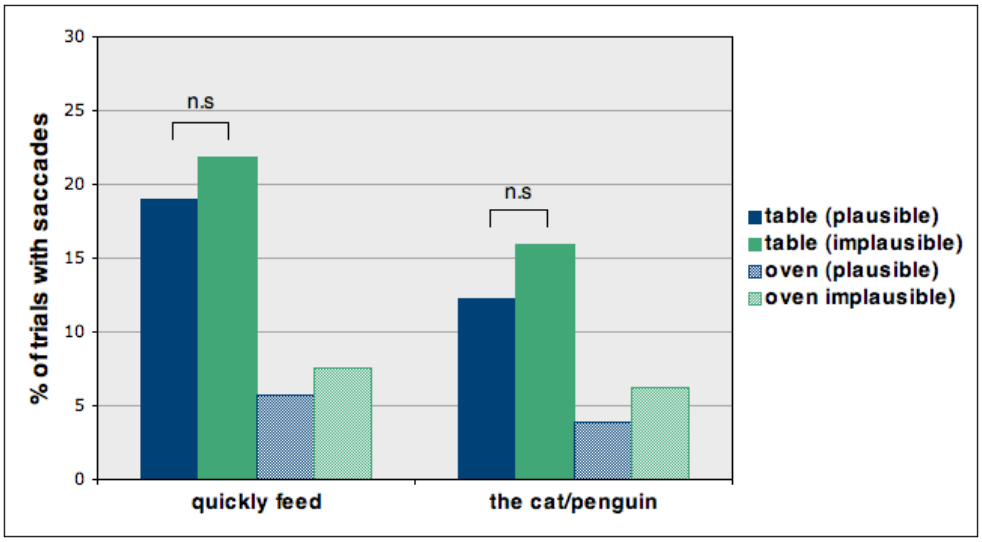


Fig 3.6. Looks toward the table and the oven (distractor) during 'quickly feed the cat/penguin'.

While these data could initially be interpreted as null results, figure 3.8. shows that compared to the baseline (at the onset of the sentence) there were both anticipatory eye movements and eye movements to the appropriate object during its reference. In other words, at the start of the sentence there were few looks to both the table and the oven (distractor) in both the plausible condition (*t1* (31) = 1.770, $p > .05$) (*t2* (15) = 2.059, $p > .05$) and the implausible condition (*t1* (31) < 1) (*t2* (15) < 1). However, at the first mention of the table and during 'quickly feed' there was a gradual increase in looks to this location, yet no increase in looks to the oven. The high proportion of fixations to the table, as well as the short timespan between the offset of 'table' and the onset of 'quickly feed' makes it difficult for the proportion of looks to the table to decrease much during this period, consequently constraining the potential increase in looks to the table during 'quickly feed' (the proportion of looks to the table did not begin to decrease until approximately 500 milliseconds before the onset of 'quickly feed'). This confound might have been prevented by including a reference to a different object before the anticipatory region of the sentence (e.g. 'the woman will put the cat/penguin onto the table. Then, she will reach for the dish, and quickly feed the cat/penguin').

With respect to the proportion of anticipatory eye movements it is further worth noting that during 'quickly feed' there were more looks to the table than the oven in both the plausible (*t1* (31) = 6.700, $p < .001$) (*t2* (15) = 4.103, $p < .01$) and implausible (*t1* (31) = 6.830, $p < .001$) (*t2* (15) = 3.414, $p < .01$) conditions. However, in the context of a blank screen (experiment 3 – see figure 3.3. on page 88) there were more anticipatory looks to the table than the oven in the plausible condition (*t1* (31) = 4.996, $p < .001$) (*t2* (15) = 2.961, $p < .05$), but not in the implausible condition (*t1* (31) = 1.158, $p > .05$) (*t2* (15) = 1.420, $p > .05$). While we cannot compare directly across the two experiments, the proportion of looks within each experiment nonetheless suggests that participants anticipated the lamp (implausible location) in the context of a concurrent scene, but not in the context of a blank screen.

Cat/penguin – plausible vs. implausible

During 'quickly feed' there was no difference in anticipatory looks to the cat in the plausible condition and looks to the penguin in the implausible condition (*t1* (31) < 1) (*t2* (15) < 1). Similarly, at the end of the second sentence during 'the cat/penguin' there was no difference in looks to the cat in the plausible condition and the penguin in the implausible condition (*t1* (31) < 1) (*t2* (15) < 1). However, during the first 500 milliseconds of the picture preview[12] there were more looks to the penguin than the cat (*t1* (31) = 2.881, $p < .01$), although the effect was not statistically significant in the by-items analysis (*t2* (15) = 1.686, $p > .05$). Similarly, there were more fixations on the region of the penguin than the cat (*t1* (31) = 2.417, $p < .05$), but again the effect was not statistically significant in the by-items analysis (*t2* (15) = 1.086, $p > .05$) (see figure 3.7.). As in experiment 3 there was no difference in first fixation duration and

[12] The visual scene was presented for 1000ms and then removed before the onset of spoken language.

gaze duration to these regions.

Time	table (impl.)	table (pl.)	oven (impl.)	oven (pl.)	penguin (impl.)	cat (pl.)
quickly feed	22	19	8	6	7	9
the cat/penguin	16	12	6	4	13	10

Table 3.2. Percentage of trials with looks to the table, oven (distractor), penguin and cat during “quickly feed the cat/penguin”. Percentages calculated from the total number of trials.

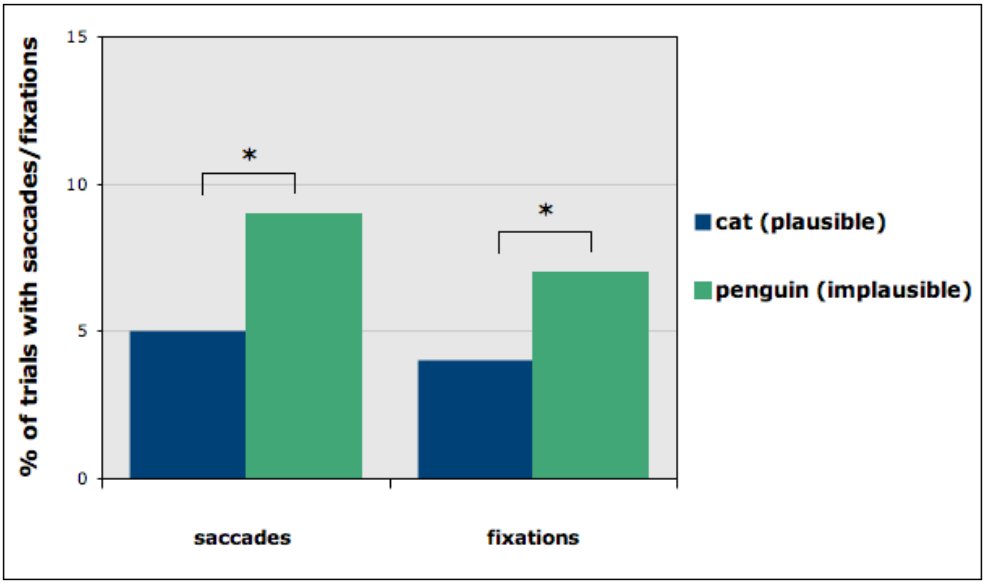


Fig 3.7. Percentage of trials with saccades to the cat and the penguin. as well as percentage of trials with fixations on the cat and the penguin during the first 500ms of the picture preview.

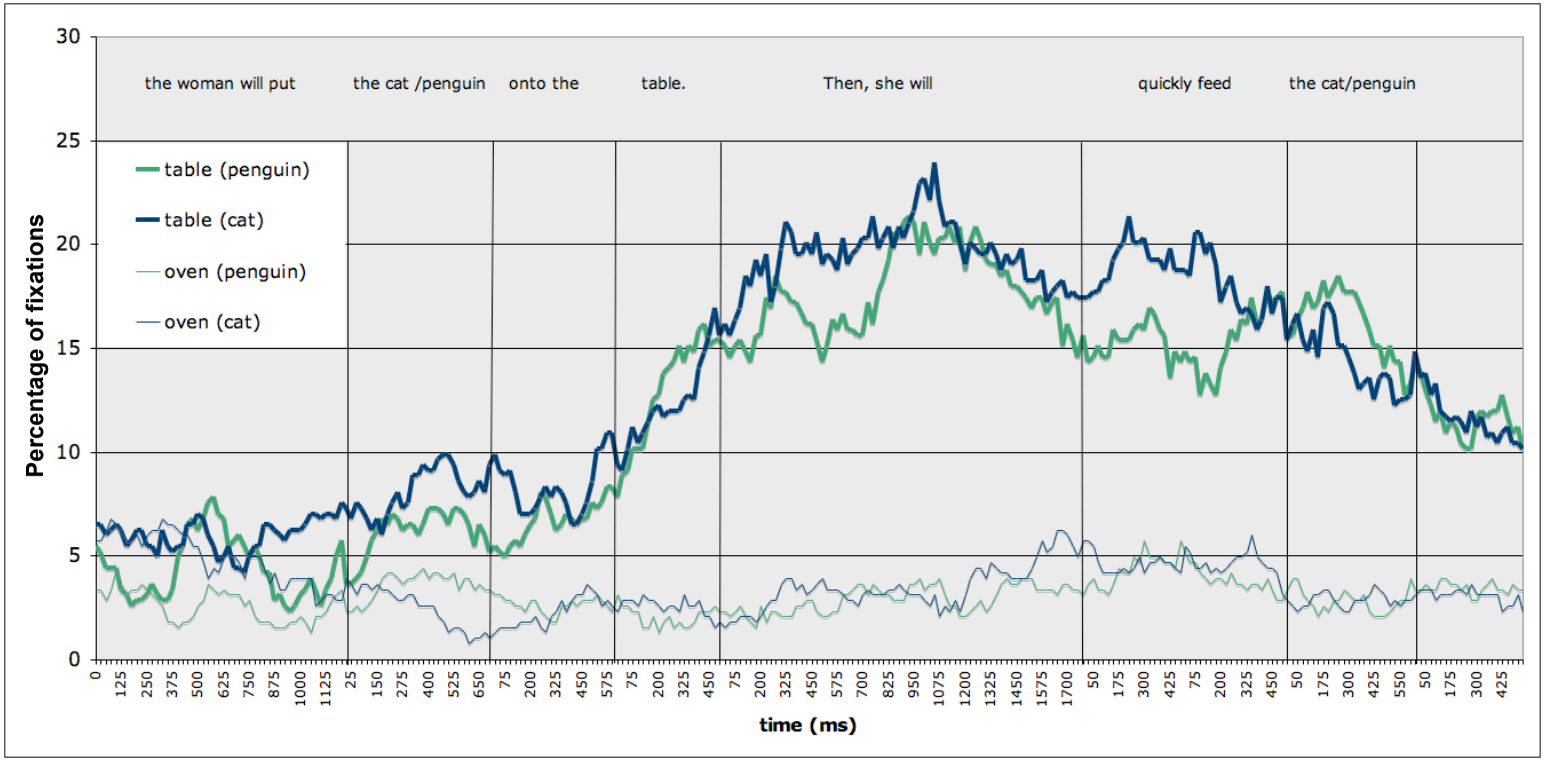


Fig 3.8. Percentage of trials with fixations toward the table and the oven (distractor). The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the cat/penguin onto the table. Then, she will quickly feed the cat/penguin'*. The fixations were calculated every 25 ms sequentially from the synchronisation point.

3.3.3. Discussion

In experiment 3 where we removed the visual scene before the onset of spoken language we found that plausibility influenced anticipatory eye movements to the event-specific location of the target item, but not eye movements during the final reference to 'the cat/penguin'. In other words, participants anticipated the location of the table when the cat (plausible item) was said to have been moved there, but not when the penguin (implausible item) was said to have been moved there. However, when the cat/penguin was directly referred to at the end of the second sentence there was no difference in looks to the region of the table. In experiment 4 the visual scene remained onscreen throughout the spoken narrative. Here we found no more anticipatory looks to the table when the cat was said to have been moved there, than when the penguin was said to have been moved there. Similarly, there was no difference in looks to the table during the final reference to 'the cat/penguin'.

A previous study by Loftus and Mackworth (1978) found that participants fixated more frequently on contextually implausible objects (e.g. an octopus in a farmyard scene) compared to contextually plausible objects (e.g. a tractor in a farmyard scene). The study further showed that first fixations to the implausible objects were longer than first fixations to the plausible objects. Loftus and Mackworth explained these findings, suggesting that people are more attentive to implausible items since these tend to be more informative when later having to discriminate between similar visual scenes. The current experiment manipulated contextual similarity in a similar fashion, but in this experiment participants were not required to perform any discrimination tasks – they were simply told to look at the visual scenes and listen to the spoken sentences. In contrast to the findings by Loftus and Mackworth we found no difference in first fixation duration to the contextually implausible and plausible items. However, during the first 500 milliseconds of the picture preview (before the onset of spoken language) we found a higher proportion

of saccades and fixations to the implausible items, suggesting that the contextually implausible items did initially attract more attention than the plausible items, even if only for a brief period of time. While Loftus and Mackworth interpreted participants' increased attention to implausible objects as a consequence of such objects being more informative for later discrimination, the current study was interested in the extent to which an increased salience of implausible items might make these items more accessible when later referred to. For example, if the implausible items attracted more attention than the plausible items (before any references were made to these items), we might expect this greater level of attention to make the implausible items more accessible, consequently leading to a higher proportion of looks when anticipated, or referred to at the end of the narrative. However, our data does not indicate that this is the case – while we did see more saccades and fixations to the implausible items during the first 500 milliseconds of the picture preview there were no more looks to these items when referred to later in the narrative (during 'quickly feed the cat/penguin'). This suggests that a continuous saliency advantage of the contextually implausible items did not make these items more accessible when leaving the visual scene onscreen during the spoken narrative.

The current experiment show a similar pattern of eye movements to that observed in experiment 2 (where we similarly left the visual scene onscreen during spoken language). While experiment 2 showed as many looks to the implausible locations (e.g. lamp) as to the plausible locations (e.g. table), the data from the current experiment showed that participants looked as much to the location of the implausible items as they looked to the location of the plausible items. Together these findings suggest that the effect of plausibility observed in experiments 1 and 3 (where we removed the visual scene before the onset of language) is not specifically related to difficulties with processing references to either implausible locations, or implausible items, but rather modulated somehow by the

absence/presence of the visual context.

3.3.4. Summary & questions

The findings from experiment 3, where we removed the visual scene before the onset of spoken language, showed that after having heard 'the woman will move the cat/penguin onto the table' participants looked to the location of the table during 'quickly feed' when they anticipated the cat (contextually plausible), but not when they anticipated the penguin (contextually implausible). During the final reference to 'the cat/penguin' participants looked to the location of the table regardless of whether the cat, or the penguin had been moved there. In experiment 4 where the visual scene remained onscreen during the spoken narrative we found that participants looked to the table both when they anticipated the cat and when they anticipated the penguin. As in experiments 3, during the final reference to 'the cat/penguin' there were as many looks to the table irrespective of whether the cat or the penguin was said to have been moved there.

The findings from experiment 3 and 4 show a similar pattern of eye movements to experiments 1 and 2, where an identical target item was moved to either a plausible or an implausible location. In these experiments items were described as being moved to either a plausible or an implausible location (e.g. 'the woman will move the glass onto the table/lamp. Then, she will pick up the bottle and pour the wine carefully into the glass.'). The data showed that in the context of a blank screen (experiment 1) participants anticipated the location of the table, but not the lamp. However, in the context of a concurrent visual scene (experiment 2) participants anticipated both the table and the lamp as the upcoming location for pouring the wine into. There was no influence of plausibility during the final reference to the glass in either experiment 1 or 2. Taken together, these findings suggest that the influence of plausibility observed in experiment 1 is not exclusively related to plausibility of location (or

any difference in saliency between the plausible and implausible locations), but may be extended to the contextual plausibility of the referent itself.

The data from experiments 3 and 4 further indicate that while an early visual introduction of the implausible items might have given participants more time to prepare for the implausible nature of upcoming events, this 'extra time' did not make the implausible items more accessible, or in other words, easier to anticipate when later referred to in the context of a blank screen. Likewise, the higher proportion of 'initial' eye movements to the more salient implausible items did not result in a higher proportion of looks to either the items themselves, or the described locations of these items. However, it is nonetheless important to keep in mind that any differences in eye movements only appeared during the first 500 milliseconds of the picture preview, whereby it is possible that any advantage of saliency for the implausible items might have been too brief to be able to exert any influence on later eye movements.

Why do we see this difference in anticipatory looks depending on the presence/absence of a visual scene? In the previous chapter we proposed an explanation based on the differences in constraints afforded by a blank or concurrent visual scene and we will return to this account later in the thesis. In the following chapter we explore a different theory based on the notion that in the context of a blank screen the proportion of previous looks toward a specific location matters. For example, in experiment 1 there were always more (and earlier) looks to the region of the table than the region of the lamp when first mentioned ('the woman will move the glass onto the table/lamp'), simply because of the plausibility (and anticipation) of this location. In other words, when hearing 'the woman will move the glass onto the…' the most likely place to put the glass (within the visual scene) is the table and this is why participants begin to look to the region of the table even before they hear the spoken location of the glass. It is possible that this higher proportion of early looks to the

region of the table could have subsequently made this more accessible than the location of the lamp (i.e. during the anticipatory region in ‘then, she will pick up the bottle, and pour *the wine carefully into* the glass’). We explore this theory in the following chapter by manipulating the proportion of looks to the region of the table and the lamp when these first are mentioned in the narrative.

CHAPTER 4

MANIPULATING INITIAL EYE MOVEMENTS

In the previous experiments we explored how plausibility can influence the accessibility of objects within event-representations and further how this might be modulated by the absence or presence of a visual context. Our findings showed that plausibility did not influence looks to the described location of an object once this object was directly referred to. However, participants made few anticipatory eye movements to locations when these were either implausible (e.g. a glass on a lamp), or when a contextually implausible item was described as being in this location (e.g. a penguin on a kitchen worktop). Interestingly, this pattern of eye movements only occurred in the context of a blank screen – when the visual scene remained onscreen participants anticipated these locations regardless of plausibility.

In experiment 5 we investigated whether the proportion of previous eye movements to specific locations might explain these findings. Going back to the data from experiment 1 we found that there were always more (and earlier) looks to the regions of the plausible locations compared to the implausible locations when these were first mentioned. For example, when hearing 'the woman will move the glass onto the table/lamp' participants looked more to the previous region of the table than the lamp because world knowledge informed them that this was the most likely location to put a glass.

This likelihood further allowed participants to anticipate the table, during 'move the glass', as the upcoming location for the glass even before they heard where the glass was moved. As such, their expectations turned out to be correct when the glass was described as being moved to the table, but not when the glass was described as being moved onto the lamp. In this case, participants needed to hear that the glass would be moved onto the lamp in order to redirect their eye movements from the region of the table to the region of the lamp. This resulted in a slower increase in looks to the region of the lamp, compared to the region of the table, as well an overall smaller proportion of looks by the time a new referring expression (in this case 'the bottle') directed eye movements to a different location[13] (see figure 4.1.). Taking this into account it is possible that the higher proportion of looks to the region of the table could have made the location of the table more accessible later in the sentence, resulting in more anticipatory looks to that location (during 'then, she will pick up the bottle, and pour *the wine carefully into* the glass')[14]. In experiment 5 we attempted to manipulate (and equalise) the proportion of looks to the region of the table and the lamp when these locations were first mentioned at the beginning of the narrative.

[13] In the by-subjects analysis there were more looks to the region of the table than the region of the lamp ($t1$ (47) = 2.324, $p < .05$) between the offset of the first mention of 'the glass' and the onset of 'pour'. However, this difference was not statistically significant in the by-items analysis ($t2$ (15) < 1).

[14] We performed a contingency analysis of the data from experiment 1 in order to separate the data according to whether the region of the table/lamp had been fixated prior to the onset of the second sentence. The results did not indicate that participants who looked to the region of the table/lamp when first mentioned were more likely to look to these locations later in the narrative during "the wine carefully into the glass". However, as the proportion of eye movements to these regions was quite low any relationship between early and later looks might have been difficult to establish using this type of analysis.

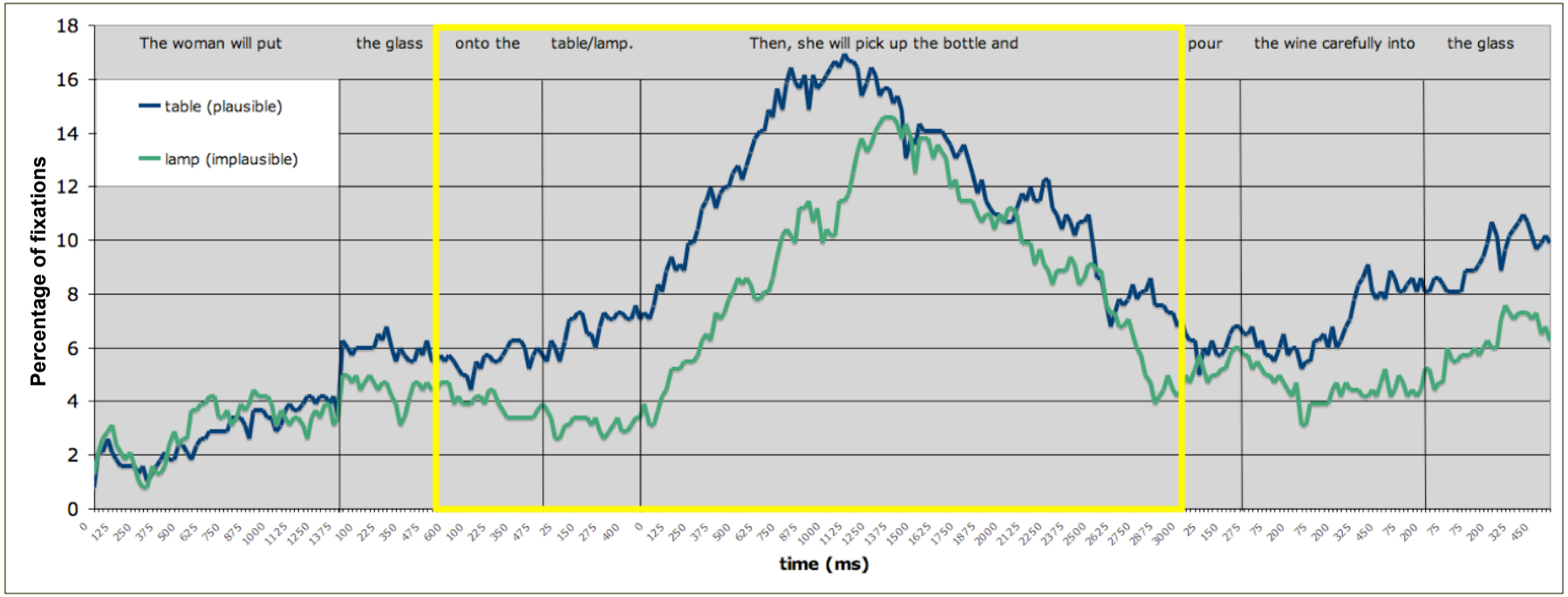


Fig 4.1. Percentage of trials with fixations toward the previous region of the table and the lamp in the plausible and implausible conditions. The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

4.1. Accessibility and spatial memory

A number of findings suggest that we associate spoken information with certain locations on the screen and then use these associations as spatial cues during recall (e.g. Bourlon, Oliviero, Wattiez, Pouget & Bartolomeo, 2011; Brandt & Stark, 1997; Johansson, Holsanova & Holmqvist, 2005; Laeng & Theodorescu, 2002; Spivey & Geng, 2001). As mentioned earlier (chapter 2) Richardson and Spivey (2000), showed participants a sequence of visual scenes that were divided into four equally sized quadrants. Each scene depicted a different face in a different quadrant and each face would deliver a fact of general knowledge (*e.g. "the Pyrenees is a mountain range separating France and Spain"*) whereafter it would disappear from the screen. Next, participants heard a second statement, which was related to one of the four previously provided facts (e.g. *"the Pyrenees is a mountain range"*) and were required to say if the statement was either true or false. The findings showed that when participants were formulating their answers, they were more likely to look at the quadrant that had previously contained the person who delivered the fact in question. Interestingly, this occurred even though the spatial information was irrelevant to the task, or in other words even though the blank region of the screen did not provide any visual information that would be able to directly help participants answer the questions correctly. Furthermore, and perhaps more relevant for the present study, Richardson and Spivey did not find a relationship between participants' proportion of looks to the relevant region of the screen and the rate of accurate answers, which suggests that eye movements to the region associated with certain information did not facilitate participants' memory of the previous information.

A subsequent study that we previously discussed in chapter 2 is by Hoover and Richardson (2008) who similarly found no correlation between memory for linguistic information and eye movements to the associated locations on the screen. In this experiment participants

watched animals burrowing underground and then emerging from a molehill. For example, in one condition a rabbit would first emerge in one location during which the participants would hear a piece of information, for example about Cleopatra. The rabbit would then descend the molehill and a moment later an identical rabbit would emerge from a different location (this time no information was given) and descend shortly afterwards. Participants were then asked a (yes/no) question about the previous information. The second condition was similar to the first, except this time the rabbit would descend the molehill whereafter burrowing would begin in a different off-screen location. This presentation sequence suggested that the rabbit associated with the provided information was different from the second rabbit to emerge. In contrast, the on-screen burrowing in the first condition connected the appearances of the rabbit, suggesting that the second rabbit to emerge was the same as the first. When answering the questions participants looked at both locations when the rabbit appeared to be the same. However, when there appeared to be two separate rabbits, participants only looked to the first location. As in the previous study by Richardson and Spivey (2000) participants associated the spoken information with certain locations on the screen and used this as a spatial cue when later recalling the information. When what appeared to be the same rabbit emerged in two different locations, participants kept track of these locations and relied on (and looked to) both locations when answering questions. In contrast, when what appeared to be two different rabbits emerged in two locations participants only relied on the location that had been associated with the spoken information. In this case, participants were able to keep track of each individual rabbit as it moved around the screen and as such constrain the spatial cues to those exclusively associated with the spoken information. However, like in the previous experiment by Richardson and Spivey (2000) there was no relationship between eye movements to the locations of the rabbit associated with the spoken information and the proportion of

accurately answered questions. This suggests that while participants might have used the locations of the rabbit as spatial cues during recall, looks to these locations did not enhance performance when having to recall information associated with these locations.

On the other hand, Laeng and Theodorescu (2002) conducted a slightly different experiment exploring the function of eye movements during imagery and showed a correlation between eye movements during imagery and memory of visual information. In the first (perceptual) part of the experiment participants were shown an image displaying a tropical fish in one of the four corners of the screen and told to try and remember the image. In the second (imagery) part participants were asked to imagine, or construct an internal image of the previous display whilst keeping their eyes open. When participants indicated that they had done so they were asked a question concerning a property or physical attribute of the fish (e.g. whether the tail was yellow). Critically, participants in one condition were free to move their eyes during both the perceptual and the imagery part of the experiment. Participants in the second condition were free to move their eyes during the perceptual phase, but not during the imagery phase. In contrast, participants in the third condition were required to fixate on the center of the screen during the perception phase, but were free to move their eyes during the imagery phase. Findings showed that participants' scan patterns, or in other words, sequence of eye movements during the perceptual phase correlated with the eye movements made during the imagery phase. In short, if participants had looked at the fish during the perceptual phase, they also looked to the previous location of the fish during the imagery phase. Similarly, participants who had looked at the center of the screen during the perceptual phase, continued to look at the center of the screen during the imagery phase and this occurred even when no constraints had been imposed on eye movements during the imagery phase. Furthermore, data from the memory test showed that participants who were free to move their

eyes during both the perception and imagery phase responded more accurately than participants who were free to move their eyes during the perception phase, but restricted to look at the center of the screen during the imagery phase. Laeng and Theodorescu proposed that eye movements to the previous location of the fish served to enhance participants' mental image of the fish and that this lead to a higher proportion of accurate answers when participants were subsequently required to recall the physical properties of the fish. These findings indicate that (during imagery) the proportion and specific location of eye movements in the context of a blank screen might serve a functional role, specifically in terms of assisting the encoding of information for later recall.

Going back to experiment 1, the findings showed that in the context of a blank screen participants only anticipated the location of the glass if this was plausible – if the glass had been moved to the lamp there were few anticipatory looks to this location. This difference in eye movements suggests that the implausible location is a less viable candidate than the plausible location when it comes to guessing, or anticipating where the wine will be poured. In contrast, plausibility did not influence looks to the described location of an object, once this object was directly referred to. In other words, during the final reference to 'the glass' there were as many looks to the table as there were to the lamp, suggesting that participants had no problem retrieving the implausible location once the glass was directly referred to. While these findings seems to indicate that the plausibility effect is not directly related to problems with retrieving the implausible locations, it doesn't rule out the notion that fewer previous eye movements to the lamp could have made participants' internal representation of the lamp less accessible, consequently leading to a *delay* in anticipatory looks to this region. This notion relates to the findings by Laeng and Theodorescu (2002) (referred to above) suggesting that a higher proportion of previous looks to the region of the table may have served to enhance participants'

representation of the table, thereby making it a more prominent and viable candidate when having to anticipate where the wine will be poured.

4.2. EXPERIMENT 5

In the following experiment we explored whether a higher proportion of looks to the plausible locations could have made these locations more accessible and therefore easier to anticipate when later referred to. In order to manipulate the proportion of early looks to the table and lamp we enhanced the stimuli to include a preliminary sentence, which aimed to attract an equal number of looks to the table and lamp when first mentioned. For example, participants would either hear, at the start of each vignette, " the woman will wipe the table" (condition 1), or "the woman will dust the lamp" (condition 2). In both cases the vignettes then continued as in Experiments 1 and 2, with the woman moving the glass to the table/lamp and subsequently pouring the wine into the glass. Since wiping the table and dusting the lamp are both plausible we would expect an equal proportion of looks to the table and the lamp during this early part of the narrative. In a third condition the preliminary sentence referred to an otherwise unnamed distractor (e.g. "the woman will look at the books"). Apart from the preliminary sentence this condition was identical to condition 2. As such, condition 3 provided us with a 'baseline' proportion of looks to the lamp, thereby allowing us to compare 'later' looks to the lamp when it had either been referred at the beginning of the narrative, or not. If the proportion of 'initial' eye movements to a specific location is related to the proportion of later (anticipatory) looks to this location we would expect more anticipatory looks to the region of the lamp after initially hearing " the woman will dust the lamp" than after hearing "the woman will look at the books". The intention, then, is to use the wiping/dusting actions to equalize the proportion of initial eye movements to the table and to the lamp, with as many looks to the region of the lamp after hearing "the

woman will dust the lamp", as looks to the region of the table after hearing "the woman will wipe the table".

4.2.1. Method

Participants

Forty-eight students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

Twenty-four experimental scenes (the same used in experiments 1 and 2) were matched with three conditions:

1. The woman will wipe the table. Then, she will put the glass onto the table, pick up the bottle, and pour the wine carefully into the glass.

   (Plausible – table)

2. The women will dust the lamp. Then, she will put the glass onto the lamp, pick up the bottle, and pour the wine carefully into the glass.

   (Implausible – lamp)

3. The women will look at the books. Then, she will put the glass onto the lamp, pick up the bottle, and pour the wine carefully into the glass.

   (Implausible – books)

In all three conditions the first sentence always described the agent performing a plausible action related to one of the objects within the scene: In condition 1 the action referred to the plausible location (e.g. table), whereas in condition 2 the action referred to the subsequently implausible location (e.g. lamp).

By inserting these references we hoped to increase the proportion of initial looks to the implausible locations, making them equal to the proportion of initial looks to the plausible locations.[15] In condition 3 the action referred to a distractor item (e.g. the books), which were not otherwise referred to. This provided us with a 'baseline' proportion of later looks to the implausible locations allowing us to compare looks to the e.g. the lamp when this had either been referred to at the beginning of the sentence, or not. In addition to the 24 experimental scenes, 24 sentence-picture pairs were taken from the set of fillers previously used in experiments 1 and 2.

Procedure

The procedure for the experiment was the same as in the previous blank screen experiments. The visual images were presented for five seconds and then replaced by a grey screen. The onset of the audio occurred one second after the scene had been removed and the trials finished 14 seconds after the audio onset. As such, each trial lasted a total of 20 seconds. The total duration of the experiment was approximately 45 minutes. The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

We used the same four regions of interest as in experiment 1. To begin with, we compared the proportion of saccades to the region of the table in the plausible condition with the proportion of saccades to the region of the lamp in the implausible conditions. Secondly, we compared the proportion of saccades to the region of the lamp after participants had initially heard either 'the woman will dust the lamp',

[15] In experiment 1 there were always more (and earlier) looks to the regions of the plausible locations, compared to the implausible locations when these were first mentioned. For example, participants looked to the previous region of the table earlier than the region of the lamp because the table was the most likely location to put a glass.

or ‘the woman will look at the books’. These comparisons allowed us to explore whether a similar proportion of ‘initial’ eye movements to the implausible and plausible locations made the implausible regions more accessible when later anticipated and referred to. Finally, we performed a contingency analysis in order to separate the data according to whether the plausible and implausible regions had been fixated prior to their second mention. In this analysis we compared the proportion of saccades (which were initiated after the onset of ‘pour the wine carefully into the glass’) to the region of the table/lamp/books depending on whether participants had either initially looked to these locations, or not. These comparisons allowed us to explore the extent to which the proportion of early looks to the region of the table and the lamp are related to the later proportion of looks to those regions.

4.2.2. Results

Initial eye movements

Between the first mention of ‘the table/lamp’ and the onset of ‘the glass’ (e.g. the woman will wipe *the table. Then, she will put* the glass onto the table…) there was a similar increase in looks to both the region of the table in the plausible condition and the region of the lamp in the implausible condition (see figure 4.2.), whereby the manipulation of participants’ ‘initial’ eye movements lead to an equal proportion of early looks to these locations ($t1$ (47) = -1.037, $p > .05$) ($t2$ (23) < 1). In contrast, participants looked more to the region of the lamp after hearing ‘the woman will dust the lamp’ than after hearing ‘the woman will look at the books’ ($t1$ (47) = -5.746, $p < .001$) ($t2$ (23) = 7.201, $p < .001$).

It is further worth noting that during this region there were more looks to the previous region of the table than the region of the books (distractor) in the plausible condition, where participants heard ‘the woman will wipe the table’ ($t1$ (47) = 5.934, $p < .001$) ($t2$ (23) = 6.488, $p < .001$). Similarly, there were more looks to the previous region of

the lamp than the region of the books in the implausible condition, where participants heard ‘the woman will dust the lamp’ (*t1* (47) = 4.277, $p < .001$) (*t2* (23) = 5.275, $p < .001$). In contrast, when participants heard ‘the woman will look at the books’ there were more looks to the region of the books, than there were looks to the region of the lamp (*t1* (47) = -6.575, $p < .001$) (*t2* (23) = -5.013, $p < .001$).

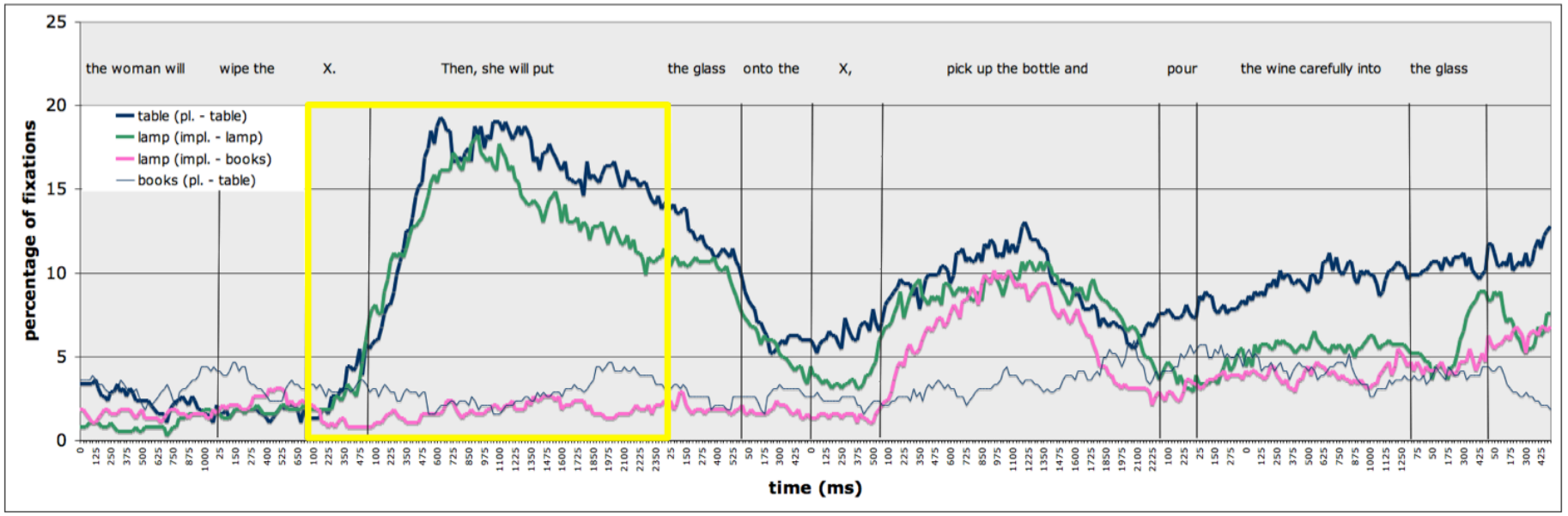


Fig 4.2. Percentage of trials with fixations toward the previous region of the table and the lamp in the plausible and implausible conditions, as well as the books (distractor) in the plausible condition. The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms

Plausible – table vs. implausible – lamp

During ‘the wine carefully into’ there were more anticipatory looks to the region of the table in the plausible condition (the woman will wipe the table) than there were looks to the region of the lamp in the implausible condition (the woman will dust the lamp), however, this difference was not statistically significant (*t1* (47) = -1.325, *p* > .05) (*t2* (23) = 1.756, *p* > .05). During ‘the glass’ there were no more looks to the region of the table in the plausible condition than there were looks to the region of the lamp in the implausible condition (*t1* (47) < 1) (*t2* (23) < 1) (see figure 4.3. and table 4.1.).

During ‘the wine carefully into’ there were more looks to the previous region of the table than the region of the books (distractor) in the plausible condition, where participants heard ‘the woman will wipe the table’ (*t1* (47) = 2.810, *p* < .01) (*t2* (23) = 4.045, *p* < .01). In the implausible condition, where participants heard ‘the woman will dust the lamp’ there were also more looks to the previous region of the lamp than the region of the books (*t1* (47) = 2.144, *p* < .05) (*t2* (23) = 2.912, *p* < .01). In contrast, when participants initially heard ‘the woman will look at the books’ there were no more anticipatory looks to the region of the lamp, than there were looks to the region of the books (*t1* (47) < 1) (*t2* (23) < 1). During the final reference to ‘the glass’, there were more looks to the region of the table than the books in the plausible condition (*t1* (47) = 2.810, *p* < .01) (*t2* (23) = 2.828, *p* < .05) and similarly more looks to the region of the lamp than the books in the implausible condition (*t1* (47) = 2.144, *p* < .05) (*t2* (23) = 2.731, *p* < .05). In contrast, when participants initially heard ‘the woman will look at the books’ there were no more looks to the region of the lamp, than there were looks to the region of the books (*t1* (47) < 1) (*t2* (23) < 1).

Lamp – implausible (lamp) vs. implausible (books)

During ‘the wine carefully into’ participants looked no more to the region of the lamp after initially hearing ‘the woman will dust the

lamp', than after initially hearing 'the woman will look at the books', (*t1* (47) = 1.095, *p* > .05) (*t2* (23) = 1.565, *p* > .05). Similarly, during 'the glass' there was no difference in looks to the region of the lamp between the two conditions (*t1* (47) = 1.618, *p* > .05), although this difference was marginally significant in the by-items analysis (*t2* (23) = 1.949, *p* > .06) (see figure 4.3. and table 4.1.).

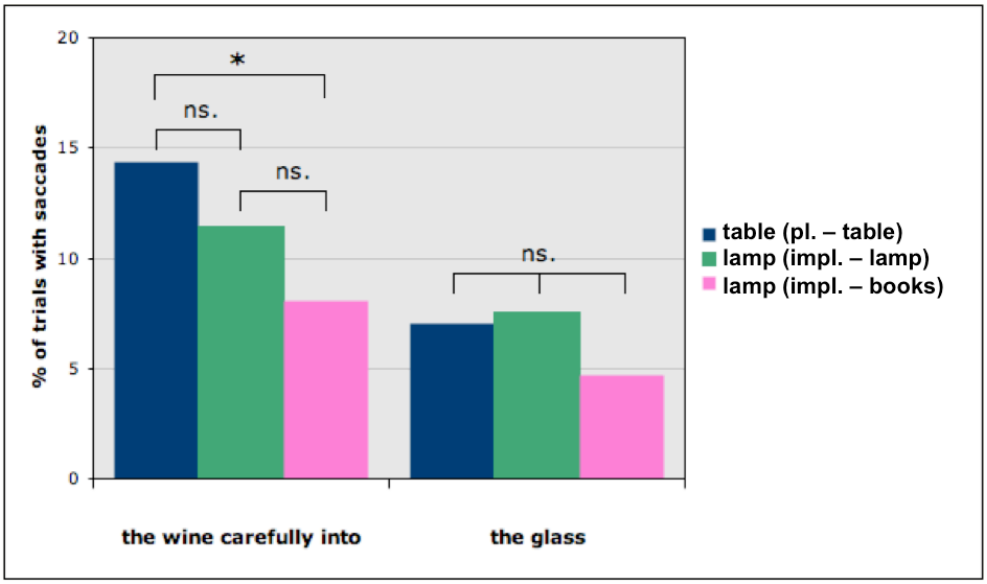


Figure 4.3. Looks to the previous location of the lamp (implausible location) and the table (plausible location) during "the wine carefully into the glass".

Time	table (pl. – table)	lamp (impl. – lamp)	lamp (impl. – books)
the wine carefully into	14	11	8
the glass	7	8	5

Table 4.1. Percentage of trials with looks to the previous location of the lamp and the table during "the wine carefully into the glass". Percentages are calculated from the total number of trials.

This data above show that by adding a preliminary sentence to the narratives we were able to equalise the proportion of 'initial' looks to the plausible (e.g. table) and implausible (e.g. lamp) locations. The findings also show that in contrast to experiment 1 there was no difference between the proportions of anticipatory looks to the region

of the table (plausible location) and looks to the region of the lamp (implausible location). However, there was also no difference in the proportion of anticipatory looks to the region of the lamp when participants had initially heard either ‘the woman will dust the lamp’, or ‘the woman will look at the books’. This data suggests that an early mention of the lamp (as opposed to an early mention of the books) did not make its location more accessible when anticipating the location of the glass later in the narrative. In order to separate the data according to whether the plausible and implausible regions had been fixated prior to their second mention we performed a contingency analysis.

Contingency analysis

The contingency analysis served to separate the data according to whether the region of the table and the lamp had been fixated prior to the second mention of those locations (e.g. between the first mention of ‘the table/lamp’ and the onset of ‘the glass’ – the woman will wipe *the table. Then, she will put* the glass onto the table, pick up the bottle and pour the wine carefully into the glass). This type of analysis allowed us to explore the extent to which the proportion of early looks to the region of the table and the lamp is related to the later proportion of looks to those regions. In order to avoid baseline differences due to participants lingering on the regions of the table and the lamp after their first mention we eliminated trials in which eye movements started before the anticipatory point in the narrative. In other words, we only included eye movements that were initiated after the onset of ‘pour the wine carefully into the glass’. As such, the proportion of later looks was calculated to start from zero at the onset of ‘pour’ (see figure 4.4.).

Table – plausible (table)

Participants who had looked to the previous location of the table during the early part of the narrative (e.g. between the first mention of

'the table/lamp' and the onset of 'the glass' – the woman will dust *the lamp. Then, she will put* the glass onto the lamp…) were no more likely to return to that location later in the narrative (e.g. at the offset of 'into' after hearing 'Then, she will pick up the bottle, and pour the wine carefully into…the glass') compared to participants who had not looked to the previous location of the table during the early part of the narrative (*t1* (47) < 1) (*t2* (23) = 1.818, *p* > .05) (see figures 4.4., and 4.5.). This suggests that participants' early looks to the region of the table did not make this location any more accessible later on.

Lamp – implausible (lamp)

In contrast, participants who had looked to the previous location of the lamp during the early part of the narrative (e.g. between the first mention of 'the table/lamp' and the onset of 'the glass' – the woman will dust *the lamp. Then, she will put* the glass onto the lamp…) were more likely to return to that location later in the narrative (e.g. at the offset of 'into' after hearing 'Then, she will pick up the bottle, and pour the wine carefully into…the glass') compared to participants who had not looked to the previous location of the lamp during the early part of the narrative (*t1* (47) = 2.225, *p* < .05) (*t2* (23) = 3.039, *p* < .01) (see figures 4.4., and 4.5.). This suggests that participants' early looks to the region of the lamp made that location more accessible later on.

Table (plausible) vs. lamp (implausible)

Participants who had initially looked to the previous location of the table (after hearing 'the woman will wipe the table) did not return to that location later in the narrative (e.g. at the offset of 'into' after hearing 'Then, she will pick up the bottle, and pour the wine carefully into…the glass') any more than participants who had initially looked to the previous location of the lamp (after hearing 'the woman will dust the lamp') returned to the region of the lamp (*t1* (47) = 1.188, *p* > .05) (*t2* (23) = 1.247, *p* > .05). In contrast, participants who had

initially looked to the previous region of the table returned more to that location, compared to participants who had not initially looked to the previous region of the lamp returned to the region of the lamp (*t1* (47) = 2.934, $p < .01$) (*t2* (23) = 4.065, $p < .001$). However, participants who had initially looked to the previous region of the lamp did not return to that location any more than participants who had not initially looked to the previous region of the table returned to the region of the table (*t1* (47) < 1) (*t2* (23) < 1) (see figures 4.4., and 4.5.). This suggests that the region of the table remained as accessible as the region of the lamp, even when participants did not initially look to the table.

Lamp – implausible (lamp) vs. implausible (books)

Participants who initially heard 'the woman will dust the lamp. Then she will put the glass onto the lamp, pick up the bottle and pour the wine carefully into the glass' were no more likely to return to the location of the lamp later in the narrative (e.g. at the offset of 'into' after hearing 'Then, she will pick up the bottle, and pour the wine carefully into…the glass'), compared to participants who initially heard 'the woman will look at the books. Then she will put the glass onto the lamp, pick up the bottle and pour the wine carefully into the glass' (*t1* (47) = 1.578, $p > .05$) (*t2* (23) = 1.966, $p > .05$). Furthermore, participants who had initially looked to the previous region of the lamp (after hearing 'the woman will dust the lamp') did not return to that location, any more than participants who had not initially looked to the previous region of the lamp (after hearing 'the woman will look at the books') (*t1* (47) < 1) (*t2* (23) = 1.241, $p > .05$). Similarly, participants who had initially looked to the previous region of the lamp (after hearing 'the woman will look at the books') did not return to that location any more than participants who had not initially looked to the previous region of the lamp (after hearing 'the woman will dust the lamp') (*t1* (47) < 1) (*t2* (23) < 1) (see figures 4.4., and 4.5.). These findings indicate that the proportion of later looks to the

lamp is not driven solely by whether participants initially looked to that region, or not.

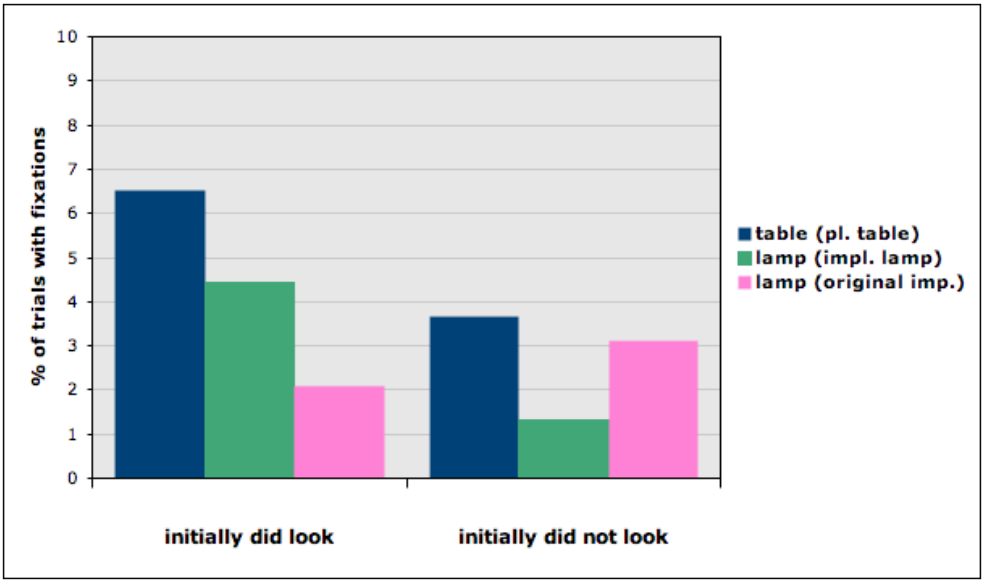


Figure 4.4. Fixations on the previous location of the lamp (implausible location) and the table (plausible location) when participants had either initially looked to these locations or not initially looked to these locations.

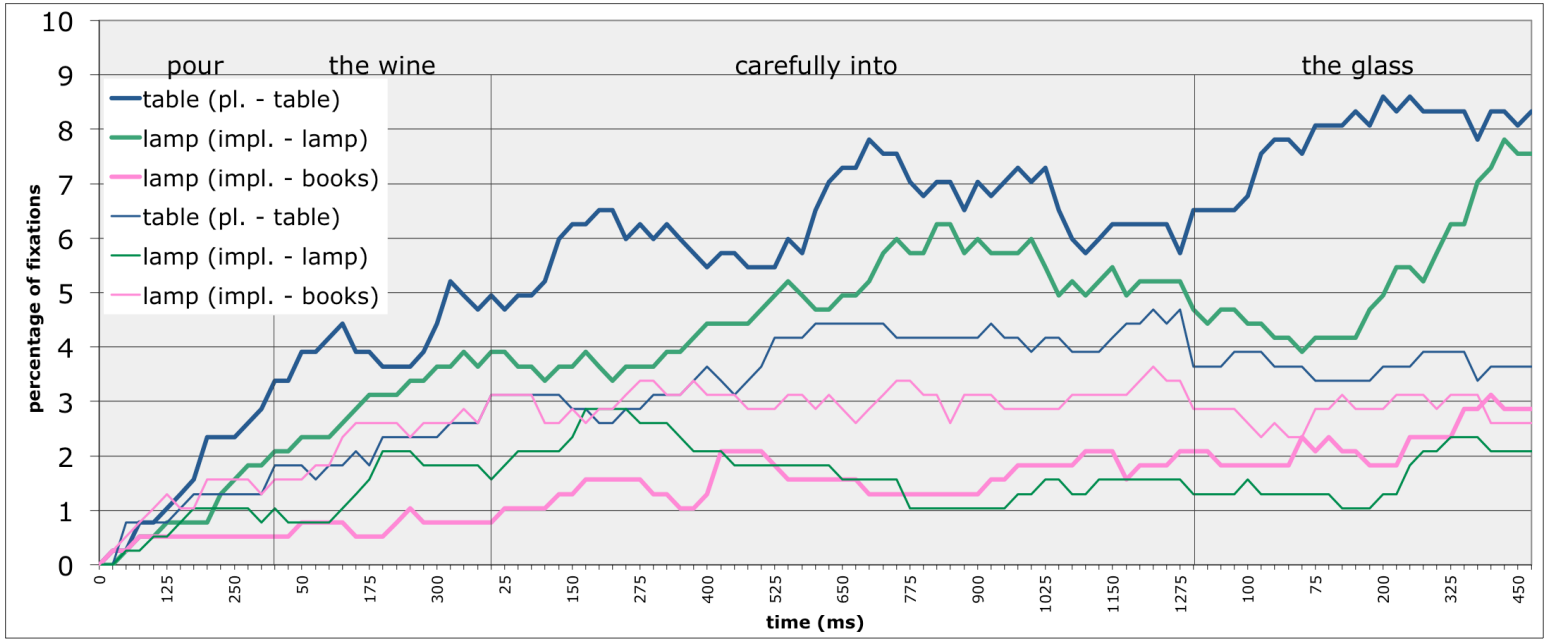


Fig 4.5. Percentage of trials with fixations toward the previous region of the table after initially hearing 'wipe the table', the lamp after initially hearing 'dust the lamp', and the lamp after initially hearing 'look at the books'. The percentages show the proportion of trials on which participants fixated on each region of interest during *'pour the wine carefully into the glass.* The thick lines represent trials were participants initially looked to the target region and the thin lines represent trials were participants did not initially look to the target region. The fixations were calculated every 25 ms sequentially from the synchronisation point.

4.2.3. Discussion

In the current experiment we explored the extent to which a higher proportion of 'initial' eye movements to the region of the plausible locations (e.g. the table) might have made this location more accessible later in the sentence thereby resulting in a higher proportion of anticipatory looks to this region. The findings showed that by adding a preliminary sentence to the narratives we were able to equalise the proportion of 'initial' looks to the plausible (e.g. table) and implausible (e.g. lamp) locations. In order to separate the data according to whether the region of the table and the lamp had been fixated prior to the second mention of those locations we performed a contingency analysis.

The findings from this analysis showed that participants who initially looked to the previous location of the lamp during the early part of the narrative were more likely to anticipate that location later in the narrative, compared to participants who did not initially look to the previous region of the lamp. These data suggests that when participants initially looked to the region of the lamp, the proportion of 'early' looks served to enhance the accessibility of this location, thereby making it easier to anticipate later in the narrative. However, it is interesting to note that this relationship only occurred between early and later looks to the region of the lamp. Going back to the beginning of this chapter we speculated that the higher proportion of looks to the region of the table in experiment 1 could have made that location more accessible and therefore easier to anticipate later in the narrative. If this were the case we would expect to find a relationship between the proportion of early and later looks to the region of the table, as well as to the lamp. However, the contingency analysis showed that participants who initially looked to the previous location of the table were just as likely to anticipate that location later in the narrative as participants who had not initially looked to the region of the table. This indicates that participants' early looks to the region of the table did not make this location any more accessible

later on. The discrepancy between contingent looks to the region of the table and the lamp makes it difficult to draw any firm conclusions as to whether the pattern of eye movements observed in experiment 1 (where we found anticipatory looks to the region of the table, but not to the region of the lamp) can be explained by a higher proportion of 'initial' eye movements to the region of the table.

A further analysis of eye movements to the table in condition 1 (where participants initially heard "the woman will wipe the table") and condition 2 ("the woman will dust the lamp") allowed us to compare whether the proportion of early looks to the region of the table and the lamp is related to the later proportion of looks to those regions. The results showed that participants who had initially looked to the previous location of the table (after hearing 'the woman will wipe the table') were no more likely to anticipate that location than participants who had initially looked to the previous location of the lamp (after hearing 'the woman will dust the lamp'). This indicates that 'initial' looks to the region of the table/lamp made these regions equally accessible later in the narrative. Furthermore, participants who had initially looked to the previous region of the table returned more to that location, compared to cases where participants who had not initially looked to the previous region of the lamp returned to the region of the lamp. On the other hand, participants who had initially looked to the previous region of the lamp returned as much to that location as participants who had not initially looked to the previous region of the table returned to the region of the table. This suggests that the region of the table remained as accessible as the region of the lamp, even when participants did not initially look to the location of the table.

The findings also showed that participants who initially looked to the previous region of the lamp (after hearing 'the woman will dust the lamp') returned as much to that location as participants who had initially looked to the previous region of the books (after hearing 'the woman will look at the books'). This indicates that the proportion of

later looks to the lamp is not driven solely by whether participants initially looked to that region, or not. Taken together the data suggests that the higher proportion of anticipatory eye movements to the plausible locations (observed in experiment 1) is not solely dependent on the higher proportion of early looks to these locations.

This relates to previous studies by Hoover and Richardson (2008) and Richardson and Spivey (2000) who found no relationship between participants' proportion of eye movements to regions associated with specific information and their memory for linguistic information. In these experiments participants were more likely to look at the (blank) region of the screen associated with the information related to the current question, suggesting that they used the location as a spatial cue when required to recall information. However, in both experiments there was no correlation between participants' proportion of looks to the relevant region of the screen and the rate of accurate answers. This indicates that a higher proportion of eye movements to the region associated with certain information did not serve to enhance participants' performance when recalling information associated with these locations.

In contrast, findings by Laeng and Theodorescu (2002) showed that participants who were free to move their eyes while viewing an item (e.g. a fish) and thereafter imaging that item responded more accurately to subsequent questions about the physical properties of the item compared to another group of participants who were free to move their eyes when viewing the item, but not when imagining the item. These findings suggest that the proportion of eye movements in the context of a blank screen might serve a functional role, specifically in terms of assisting the encoding of information for later recall. However, it is important to keep in mind that participants in this experiment were required to recall information concerning the visual properties of the previous image. In contrast, Richardson and Spivey asked participants questions about previously encountered spoken information and this study showed no relationship between eye

movements and accuracy during recall. This inconsistency between the findings from these two studies may be related to the type of information that participants were asked to recall. In other words, it may be the case that while eye movements in the context of a blank screen are able to facilitate recall for visual information, there is no facilitation when having to recall spoken information. One way to explore this notion further could be to ask questions about both the visual properties of a previously seen image (e.g. was the bottle in the picture green?), as well as the spoken information associated with a previously seen image (e.g. did the woman pick up a bottle?). Such an experiment would allow us to directly compare the extent to which eye movements are able to enhance recall of both visual and linguistic information.

4.2.4. Summary & questions

In experiment 1 we explored how plausibility is able to influence the accessibility of objects within our representations of a described event. The findings from this experiment showed that in the context of a blank screen plausibility did not influence looks to the described location of an object once this object was directly referred to (e.g. 'the woman will move the glass onto the table/lamp. Then, she will pick up the bottle, and pour the wine carefully into *the glass'*). However, during 'the wine carefully into' participants made few anticipatory eye movements to locations when these were implausible (e.g. the lamp) compared to when an item was described as having been moved to a plausible location (e.g. the table).

In the current experiment we investigated whether the higher proportion of 'initial' eye movements to the plausible locations could have made these locations more accessible when later having to anticipate where the wine would be poured. The findings showed that participants who looked to the previous location of the lamp during the early part of the narrative were more likely to anticipate that location later in the narrative, compared to participants who did not

initially look to the previous region of the lamp. However, participants who initially looked to the previous location of the table were no more likely to anticipate that location later in the narrative than participants who had not initially looked to the region of the table. This suggests that the proportion of initial looks only made the implausible regions (e.g. the lamp) more accessible later in the narrative, making it difficult to explain the pattern of eye movements observed in experiment 1 (where we found anticipatory looks to the region of the table, but not to the region of the lamp) in terms of enhanced accessibility due to the higher proportion of 'initial' eye movements to the region of the table. The data further indicated that participants who initially looked to the previous region of the lamp (after hearing 'the woman will dust the lamp') returned as much to that location as participants who had initially looked to the previous region of the books (after hearing 'the woman will look at the books'). Taken together the data suggests that the higher proportion of anticipatory eye movements to the plausible locations (observed in experiment 1) is not solely dependent on the higher proportion of early looks to these locations.

So how might we explain this difference and why plausibility only influences anticipatory eye movements in the context of a blank screen? As mentioned earlier, the effect of plausibility might be related to the way we process and anticipate upcoming discourse when language refers to something outside our visual environment, as opposed to when language refers to entities within our immediate visual proximity. When a visual context is available we automatically assume that participants in a described event will be drawn from within this context (e.g. Altmann & Mirkovic, 2009, Cooper, 1974; Tanenhaus et al., 1995). In terms of anticipation the concurrent visual context alongside language provides a stronger constraint, which in turn makes implausible object-representations more accessible and therefore easier to anticipate. However, in the context of a blank screen the anticipatory activation of appropriate representations is no

longer restricted by the visual context, but may now be influenced by real world knowledge. This experiential knowledge provides additional information concerning event-plausibility, which could lead us to consider a number of alternative options during anticipation.

In everyday language an item is rarely introduced without having some relevance to the upcoming topic of conversation. Likewise, if an object is introduced into a narrative we normally assume that this object is both informative and relevant to the current topic (e.g. Grice, 1975) and as such, is likely to be referred to again. According to this assumption, if we were to linguistically introduce an object into the context of a previously seen visual scene we might similarly anticipate that this object will be referred to later in the narrative. As such, our expectations of an upcoming reference to this recently introduced object might provide a greater level of constraint, making it easier to anticipate the appropriate reference to an object (as well as its location). In the following experiment we compared looks to the described location of the glass having either removed it from the visual scene, thus introducing it purely through language, or having first introduced it within the visual scene. This manipulation allowed us to investigate the extent to which a purely linguistic representation is able to enhance anticipation towards the implausible locations in the context of a blank screen.

CHAPTER 5

INTRODUCING 'THE GLASS' PURELY THROUGH LANGUAGE

The previous experiments showed that in the context of a blank screen we only anticipate locations when these are either plausible, or when the items moved to these locations are contextually plausible. However, in the context of a concurrent visual scene we anticipate upcoming locations regardless of plausibility. Furthermore, the findings from experiment 5 suggest that participants' early looks to the plausible locations (in experiment 1) did not make these locations more accessible later in the narrative. Together these findings suggest that the absence/presence of a visual context provides a quantitative difference in how we process and anticipate implausible events. In this chapter we compared eye movements to the described locations of a target item (e.g. the glass), having either removed this item from the visual scene (thus introducing it purely through language), or first introducing the item within the context of the visual scene. This manipulation allowed us to investigate the extent to which a purely linguistic representation is able to enhance anticipation towards the implausible locations in the context of a blank screen.

5.1. Assuming future reference

In everyday language an item is rarely introduced without having some relevance to the upcoming topic of conversation. Likewise, if an object is introduced into a narrative we normally assume that this object is both informative and relevant to the current topic (e.g. Grice, 1975) and as such, is likely to be referred to again. For example, when reading a sentence such as “Jenny was searching for her passport…” we expect the passport to be relevant to the upcoming discourse and therefore likely to be referred to again (e.g. “Jenny was searching for her passport…She remembered seeing it in her study a couple of days ago”). On the other hand, if the passport turns out not to be relevant to the upcoming discourse we would expect the sentence to be followed by some type of transitional cue indicating a shift in, for example topic, time or space (e.g. Zwaan et al., 1995a, 1995b) (e.g. “Jenny was searching for her passport…Suddenly the phone rang so she went into the kitchen to find out who was ringing her”).

A study by Altmann (1999) further explored the circumstances under which we assume reference to previously mentioned entities and how this may be related to information derived from specific verbs. In this experiment participants read a series of scenarios such as the one below:

> A car was driving downhill when it suddenly veered out of control. In its path were some dustbins and a row of bollards. It injured／missed...

The sentences were presented word-by-word and after each word participants were asked to indicate whether the sentence continued or stopped making sense. After the verb ‘injured’ participants’ reading times were longer than after the verb ‘missed’ and there were more indications that the sentence stopped making sense. However, when the scenario was slightly changed (as below) there were no differences between the two verbs.

> A car was driving downhill when it suddenly veered out of control. In its path were some pigeons and a row of bollards. It injured／missed...

These findings suggest that during the verb participants anticipated that the upcoming action (e.g. injured/missed) would refer to one of the entities previously introduced into the narrative (e.g. dustbins/pigeons). In the first scenario there is no antecedent that the verb 'injured' can refer to without violating selectional restrictions since both items are inanimate. As such, when reading the verb participants either indicated that the sentence stopped making sense, or they spent more time reading the verb. In the second scenario on the other hand, both an inanimate, as well as an animate antecedent had been introduced, allowing participants to plausibly assume that the verb 'injured' would refer back to the animate antecedent. This shows that when encountering a verb we evaluate it in terms of how it may plausibly refer to any previously introduced entities within the current scenario. The data further illustrates our tendency to anticipate that verb-based information refers to previously introduced entities even when this might not necessarily be the case, as demonstrated in the example below:

> A car was driving downhill when it suddenly veered out of control. In its path were some pigeons and a row of bollards. It injured several people who were standing nearby.

In this example our assumption that the verb 'injured' will refer to a previously mentioned entity turns out to be wrong. This example shows that the verb doesn't necessarily have to refer to any of the previously mentioned entities, so why do participants nonetheless indicate that the sentence stops making sense as soon as the verb cannot plausibly refer to any of the previously introduced entities? Altmann (1999) proposed that when encountering the verb 'injured' we automatically evaluate the extent to which it can plausibly refer to

any of the previously introduced entities. If a plausible entity has been mentioned we anticipate that the verb will refer to this entity. However, if it turns out that no plausible entities have been introduced, we cannot anticipate the referent and as such we either take longer to read the verb, or we judge that the sentence has stopped making sense.

5.2. EXPERIMENT 6

The study by Altmann (1999) illustrates how during anticipation we evaluate the extent to which the language we encounter is likely to refer to any previously mentioned entities with a narrative. It further showed that when a plausible entity has been introduced earlier in the narrative we tend to assume that subsequent language (or in this case verbs) is going to refer back to this entity. In the following experiment we explored the effect of plausibility when removing the target item (e.g. the glass) from the visual scene and introducing it purely through language. In other words, we compared looks to the described location of the glass in situations where we had either removed it completely from the visual scene, or when the glass was first presented within the visual scene and thereafter referred to by language. In condition 1 the target items (e.g. the glass) were always presented in the visual scene (see figure 5.1.). After the visual scene was removed these items were further referred to by spoken language (e.g. “the woman will put the glass onto the lamp. Then she will pick up the bottle, and pour the wine carefully into the glass”). However, in condition 2 the target items (e.g. the glass) were absent from the visual scene and therefore introduced purely through language (see figure 5.2.).

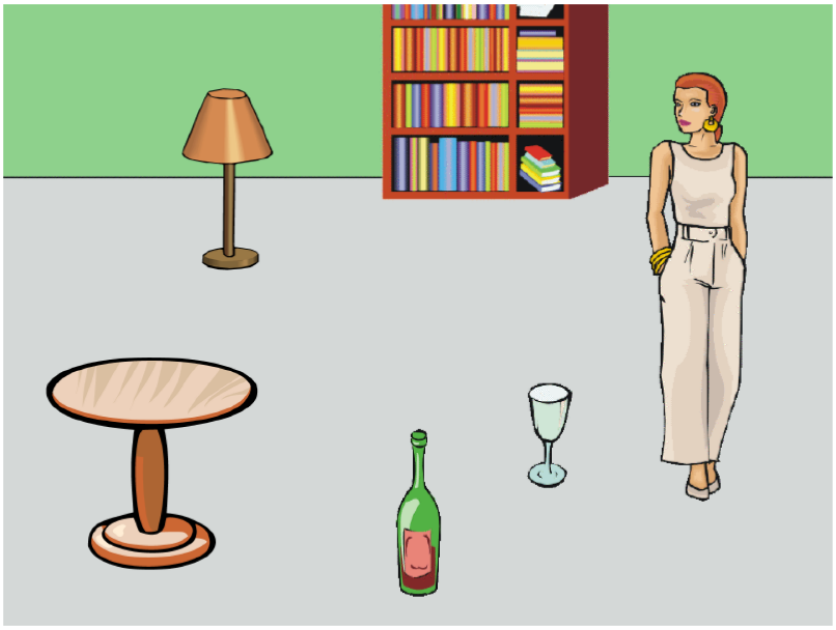


Figure 5.1. Example of one of the visual scenes used in condition 1.

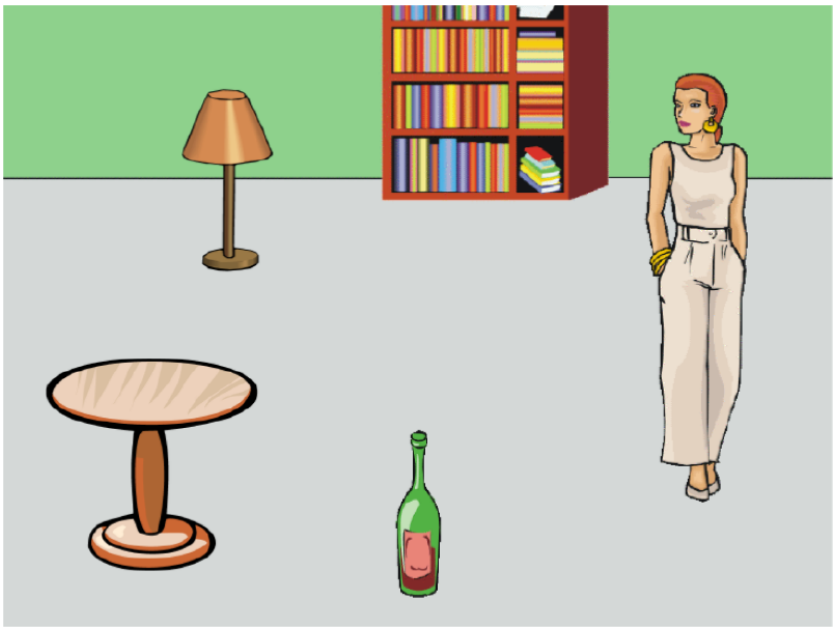


Figure 5.2. Example of one of the visual scenes used in condition 2.

This manipulation allowed us to investigate the notion of competition (e.g. Altmann and Kamide, 2009[16]), which might lead to participants' memory of the visual instantiation of the glass (in condition 1 the glass is always depicted on the floor) to compete with their internal representation of the linguistically introduced glass (describing the glass as having been moved onto the lamp). In other words, in condition 1 participants' visual memory of the glass being on the floor conflicts with their representation of the glass being on the lamp. This conflict might lead to a lower proportion of looks to the region of the lamp since participants' memory of the visual location of the glass is forced to compete with the described location of the glass. In condition 2, on the other hand, the glass is never depicted. In this case there is only one (linguistic) representation of the glass and therefore no competition between a visual and described location of the glass. As such, the absence of the glass in the visual scene should serve to minimise any interference from the depicted location of the glass and this might increase sensitivity, thereby allowing more subtle differences to emerge.

Another issue worth considering is the extent to which either a purely linguistic object-representation, or a visual/linguistic object-representation is able to enhance anticipation towards the implausible locations in the context of a blank screen. As mentioned earlier, if an object is introduced into a narrative we normally assume that this object is both informative and relevant to the current topic (e.g. Grice, 1975) and as such, is likely to be referred to again. Having previously introduced the glass in the visual scene (condition 1) any subsequent references to 'the glass' are interpreted in this context. In contrast, when the glass is absent from the visual scene

[16] Altmann and Kamide (2009) showed that the salience of a concurrent visual scene (depicting the glass on the floor) lead to increased competition and therefore fewer looks to the described location of the glass (table), compared to the visually depicted location of the glass. In contrast, when the visual scene had been removed prior to the onset of spoken language competition decreased, now showing fewer looks to the (previous) visual location of the glass, compared to the described location of the glass.

(condition 2) it is introduced purely through language and therefore we assume that it must be relevant to the upcoming topic of conversation (why else introduce it into the narrative?). In other words, we might expect that there is a specific reason for why 'a glass' is suddenly being brought into a narrative, which otherwise corresponds perfectly to a prior visual context. This might lead participants to expect that the recently introduced glass will be referred to again, consequently making it more prominent and accessible when anticipating where the wine will be poured. As in the case of competition (see above) we would similarly expect to see a higher proportion of anticipatory eye movements to the described location of the glass when it has been introduced solely through language, compared to when it has first been introduced within the visual scene and then referred to by language.

5.2.1. Method

Participants

Thirty-two students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

Twenty-four experimental scenes were matched with two conditions. In condition 1 we used the same experimental scenes used in experiment 1 (e.g. depicting the glass on the floor). In condition 2 the visual scenes were identical except now the target items (e.g. the glass) had been removed from the visual scenes. In both conditions the filler items were the same as the fillers used in experiment 1.

1. The woman will put the glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Original implausible)

2. The woman will put a glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

(No glass implausible)

Procedure

The procedure for this experiment was the same as in the previous blank screen experiments – the visual scenes were presented for five seconds and then replaced by a grey screen. The onset of the audio occurred one second after the scene had been removed and the trials finished 11 seconds after the audio onset. As such, each trial lasted a total of 17 seconds. The total duration of experiment was approximately 45 minutes. The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

In condition 1 we used the same regions of interest as in experiment 1 (e.g. the region of the glass, table, lamp and books). In condition 2 the regions of interest were the same, except for the region corresponding to the location of the glass. This region of interest was excluded since the target items (e.g. the glass) were never depicted in this condition.

To begin with, we compared the proportion of saccades to the region of the lamp (implausible location) when glass had either been depicted in the visual scene or not been depicted. This comparison allowed us to explore whether a purely linguistic introduction of the glass is able to enhance anticipation towards the implausible locations in the context of a blank screen. Secondly, we compared the proportion of saccades to the region of the lamp with the proportion of saccades to the region of the books (distractor). This provided information about whether the proportion of saccades to the described location of the glass was higher than the 'baseline'

proportion of saccades to an unnamed distractor item.

5.2.2. Results

Lamp – original implausible vs. no glass implausible

During 'the wine carefully into' there was no difference in anticipatory looks to the region of the lamp between the original implausible (e.g. the woman will put the glass onto the lamp…) and the no glass implausible (the woman will put a glass onto the lamp…) conditions (*t1* (31) = -1.109, $p > .05$) (*t2* (23) = -1.293, $p > .05$). Similarly, during 'the glass' there was no difference in looks to the region of the lamp between conditions (*t1* (31) = 1.305, $p > .05$) (*t2* (23) = 1.596, $p > .05$) (see figures 5.3., 5.4., and table 5.1.).

Lamp vs. distractor

During 'the wine carefully into' there was no difference in anticipatory looks to the region of the lamp and the region of the unnamed distractor item (in this case the books). This occurred both in the original implausible condition (*t1* (31) = 1.441, $p > .05$) (*t2* (23) = 1.513, $p > .05$) and the no glass implausible condition (*t1* (31) < 1) (*t2* (23) < 1). This suggests that participants did not anticipate the lamp any more than they anticipated the unnamed distractor and this occurred both when the glass was present or absent in the visual scene. Likewise, during 'the glass' there was no difference between looks to the region of the lamp and the distractor region in the original implausible condition (*t1* (31) < 1) (*t2* (23) < 1). This was also the case in the no glass implausible condition (*t1* (31) = 1.274, $p > .05$) (*t2* (23) = 1.901, $p > .05$).

Time	lamp (glass)	lamp (no glass)	books (glass)	books (no glass)
the wine carefully into	9	6	5	7
the glass	3	5	3	2

Table 5.1. Percentage of trials with looks to the previous location of the lamp and books (distractor) during “the wine carefully into the glass”. Percentages calculated from the total number of trials.

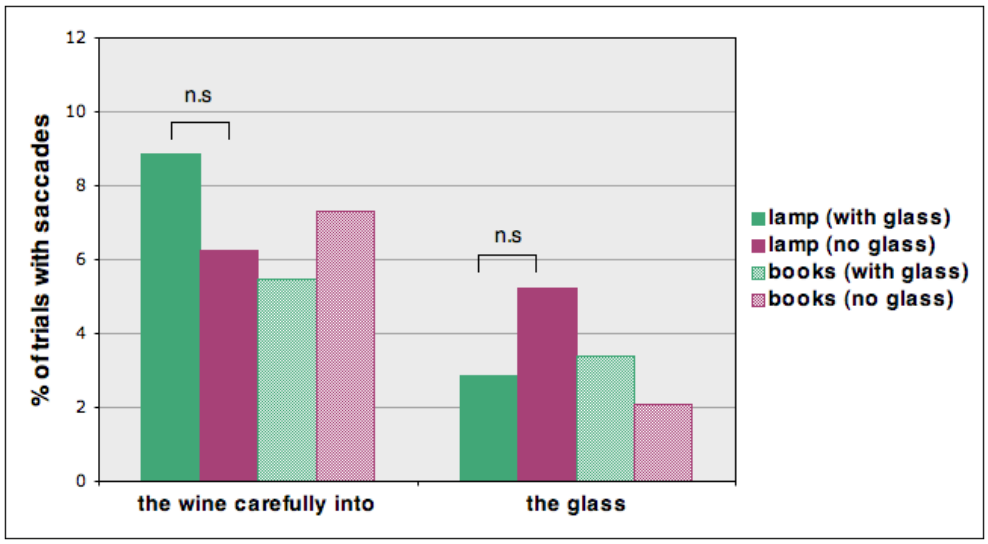


Fig 5.3. Looks toward the lamp and the books (distractor) during ‘the wine carefully into the glass’.

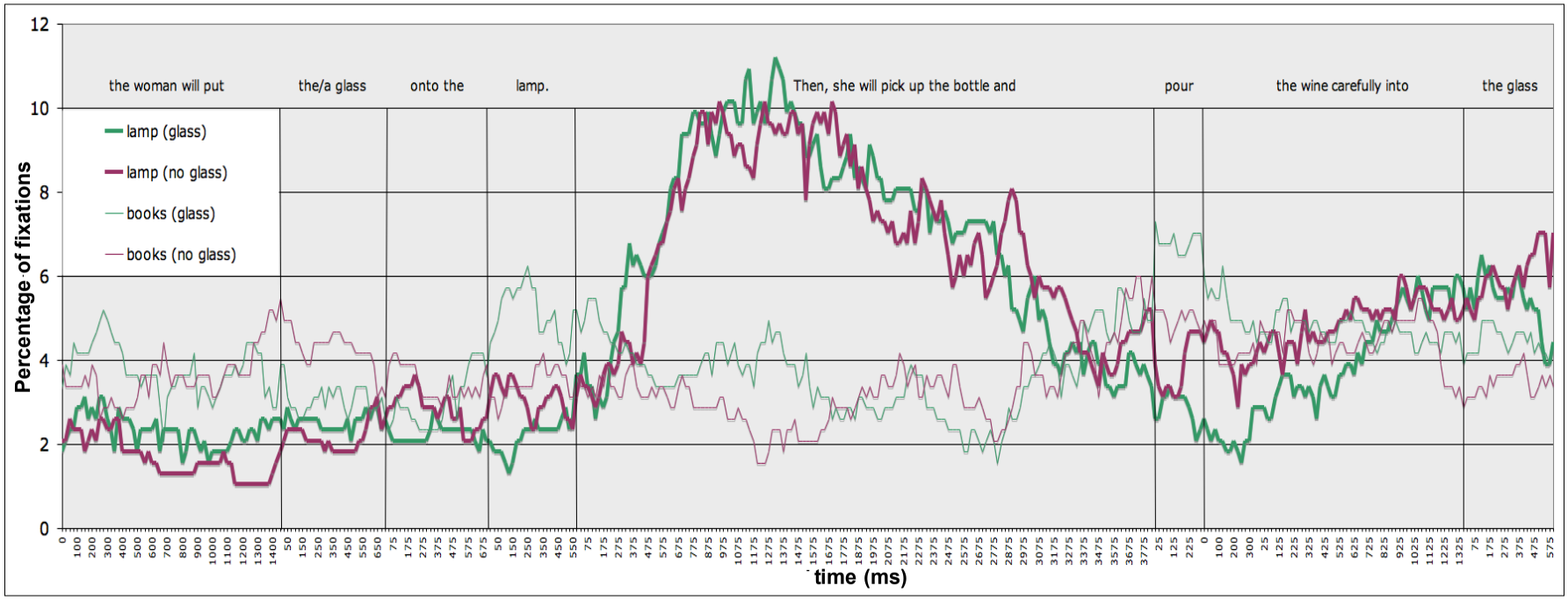


Fig 5.4. Percentage of trials with fixations toward the previous region of the lamp and the books (distractor). The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

5.2.3. Discussion

In this experiment we explored the effects of plausibility when removing the target item (e.g. the glass) from the visual scene and introducing it purely through language. We compared looks to the lamp in one condition where we had either removed the glass completely from the visual scene, or in a second condition where the glass was first presented within the visual scene and thereafter referred to by language. This manipulation allowed us to investigate the extent to which either a purely linguistic representation of the glass, or a visual/linguistic representation of the glass would be able to enhance anticipation towards the implausible locations in the context of a blank screen.

The findings showed that during 'the wine carefully into' there were slightly more anticipatory eye movements to the region of the lamp in condition 1 (where the glass had previously been presented in the visual scene), compared to condition 2 (where the glass had been removed from the visual scene). However, this difference was not statistically significant. Similarly, there was no difference between conditions in looks to the region of the lamp during the final reference to the glass. These data suggests that a purely linguistic representation of 'a glass' did not enhance the accessibility and anticipation of its described location any more than a visual/linguistic representation of 'the glass'. In other words, the linguistic introduction of 'a glass' did not to seem to emphasise the future relevance of this item enough to make it's described location more accessible and therefore easier to anticipate.

The findings further suggest that looks to the region of the lamp were not influenced by different levels of competition, in terms of whether participants' representation of the linguistically introduced glass (describing the glass as having been moved onto the lamp) had to compete with their memory of the visual instantiation of the glass (depicting the glass on the floor). In condition 1 participants' visual memory of the glass being on the floor conflicted with their

representation of the glass having been moved to the lamp and we speculated that this conflict might have forced the two instantiations of the glass to compete, thereby resulting in a lower proportion of looks to the region of the lamp, compared to when participants only had one (linguistic) representation of the glass. Altmann and Kamide (2009) showed that the salience of a concurrent visual scene (depicting the glass on the floor) lead to increased competition and therefore fewer looks to the described location of the glass (table), compared to the visually depicted location of the glass. In contrast, when the visual scene had been removed prior to the onset of language the degree of competition decreased, resulting in fewer eye movements to the (previous) visual location of the glass, compared to the described location of the glass. So why did we not see any influence of competition in the current experiment? One reason could be that once the visual scene had been removed the salience of participants' visual memory-representation of the glass declined to the extent that the prior instantiation of the glass was not prominent enough to compete with participants' linguistic representation of the glass. A future version of the current experiment might investigate this notion further by measuring eye movements to the lamp in the context of a concurrent visual scene. In this case the stronger salience of a concurrent visual representation of the glass (as well as a competing linguistic representation of the glass) might lead to fewer looks to the lamp than a purely linguistic representation of the glass.

Finally, it is important that we highlight one potential problem with the data from the current experiment. When we compared looks to the region of the lamp (implausible location) and the region of the books (distractor) we found that during the critical part of the sentence (the wine carefully into the glass) there were no more looks to the region of the lamp than there were looks to the region of the books, and this was the case for both conditions. The distractor regions always corresponded to items that were not named in the narrative, thus providing us with a 'baseline' of the proportion of eye

movements to any unnamed items. Since there were no more looks to the region of the lamp than the region of the distractor it is possible that participants' looks to the region of the lamp were not derived from any previous mention of the/a glass being moved to this location, but rather occurred due to the same levels of chance that lead to a certain proportion of looks to an unnamed distractor region. Alternatively, it may be the case that the proportion of looks to the regions of the distractor did not differ from the proportion of looks to the region of the lamp because of the low likelihood of this location.

In the next experiment we attempted to control for this potential confound by comparing looks to plausible (e.g. table) and implausible (e.g. lamp) locations, having previously introduced the target item (e.g. the glass) solely through language. Whereas in the previous experiment the glass was either present, or absent in the visual scene, this experiment never depicted the glass in either of the conditions. For example, having removed the glass from the visual scene participants would hear either 'the woman will put *a glass* onto *the table*. Then, she will pick up the bottle, and pour the wine carefully into the glass' (plausible condition), or they would hear 'the woman will put *a glass* onto *the lamp*. Then, she will pick up the bottle, and pour the wine carefully into the glass' (implausible condition). Like in the previous experiment we speculate that a purely linguistic introduction of the glass will lead to a strong expectation of its future relevance, in which case we would expect anticipatory looks to both the plausible and implausible location of the glass. Furthermore, if looks to the region of the lamp in experiment 6 were not derived from language, but purely due to chance we would similarly expect to see no difference in looks to the region of the table and the region of the books (unnamed distractor).

5.3. EXPERIMENT 7

5.3.1. Method

Participants

Thirty-two students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

Twenty-four experimental scenes were matched with two conditions. We used the same experimental scenes as in experiment 1, but this time we removed the target items (e.g. the glass) from the visual scenes for both the experimental trials (see figure 5.5), as well as the filler trials (see appendix 8 for an example of the filler items).

1. The woman will put a glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Plausible moved)

2. The woman will put a glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Implausible moved)

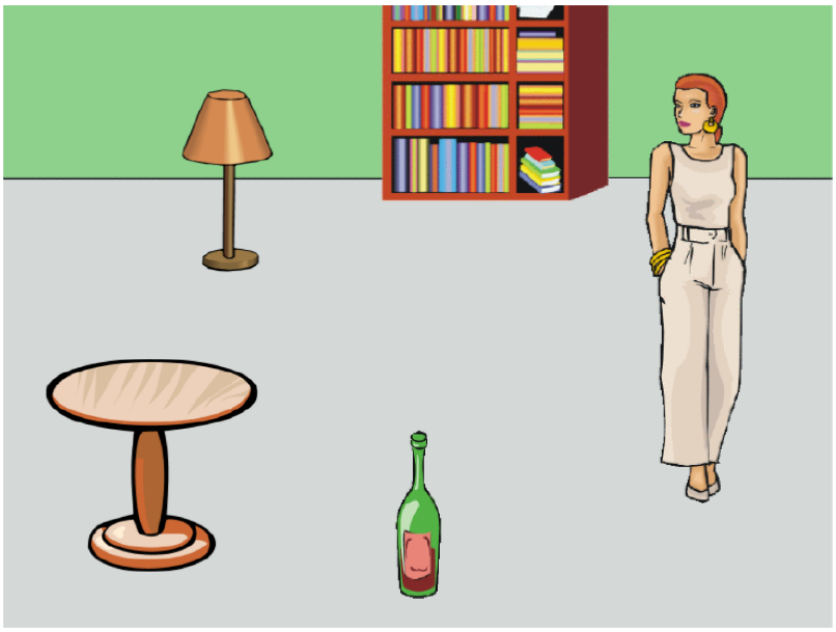


Figure 5.5. Example of one of the visual scenes used in conditions 1 and 2.

Procedure

The procedure for this experiment was the same as in the previous blank screen experiments. The scenes were presented for five seconds and then replaced by a grey screen. The onset of the audio occurred one second after the scene had been removed and the trials finished 11 seconds after the audio onset. As such, each trial lasted a total of 17 seconds. The total duration of the experiment was approximately 45 minutes. The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

We used the same regions of interest as in experiment 1 (e.g. the table, lamp and books) with the exception of the region corresponding to the previous location of the target item (e.g. the glass). First, we compared the proportion of saccades to the region of the table (plausible locations) and the region of the lamp (implausible locations). This allowed us to see if the plausibility of these locations had any influence on eye movements when the glass was not depicted in the visual scene, but introduced purely through

language. Secondly, we compared looks to the region of the table/lamp with looks to the region of the books (distractor). These comparisons provided information about whether the proportion of saccades to the plausible and implausible locations was higher than the 'baseline' proportion of looks to an unnamed distractor item.

5.3.2. Results

Table (no glass plausible) vs. lamp (no glass implausible)

During 'the wine carefully into' there were more anticipatory looks toward the region of the table in the plausible condition (the woman will move a glass onto the table), than there were looks to the region of the lamp in the implausible condition (the woman will move a glass onto the lamp) (*t1* (31) = -3.173, $p < .01$) (*t2* (23) = -2.209, $p < .05$). However, during the final reference to the glass there were no more looks to the region of the table in the plausible condition, than there were looks to the region of the lamp in the implausible condition (*t1* (31) = -1.461, $p > .05$) (*t2* (23) < 1) (see figures 5.6., 5.7., and table 5.2.). This is similar to the pattern of eye movements previously observed in experiment 1.

No glass implausible – lamp vs. distractor

During 'the wine carefully into' there were no more anticipatory looks to the region of the lamp than the region of the books (distractor) (*t1* (31) = 1.488, $p > .05$) (*t2* (23) = -1.829, $p > .05$. During the final reference to the glass there were more looks to the region of the lamp than the region of the books (*t1* (31) = 3.589, $p < .01$) (*t2* (23) = -3.704, $p < .01$).

No glass plausible – table vs. distractor

During 'the wine carefully into' there were more anticipatory looks to the region of the table than the region of the books (distractor) (*t1* (31) = 4.611, $p < .001$) (*t2* (23) = -5.939, $p < .001$). This suggests that participants anticipated the table (plausible locations) more than the

unnamed distractor, but they did not anticipate the lamp (implausible locations) any more than they anticipated the distractor. During the final reference to the glass there were more looks to the region of the lamp than the region of the books (*t1* (31) = 4.088, $p < .001$) (*t2* (23) = -3.788, $p < .01$).

Time	lamp (impl.)	table (pl.)	books (impl.)	books (pl.)
the wine carefully into	13	21	9	5
the glass	9	12	2	3

Table 5.2. Percentage of trials with looks towards the previous location of the lamp, table and books (distractor) during "the wine carefully into the glass". Percentages calculated from the total number of trials.

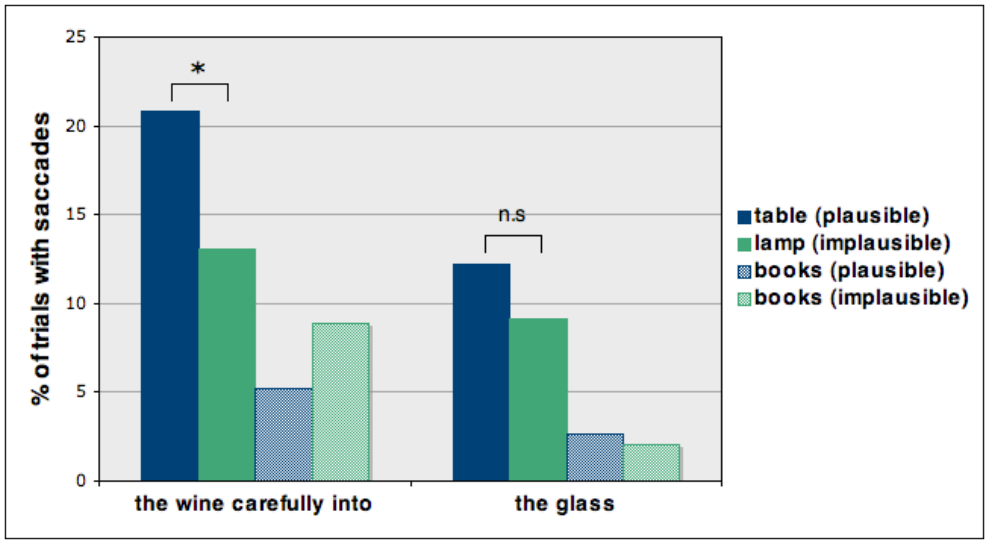


Fig 5.6. Looks toward the lamp/table during 'the wine carefully into the glass'.

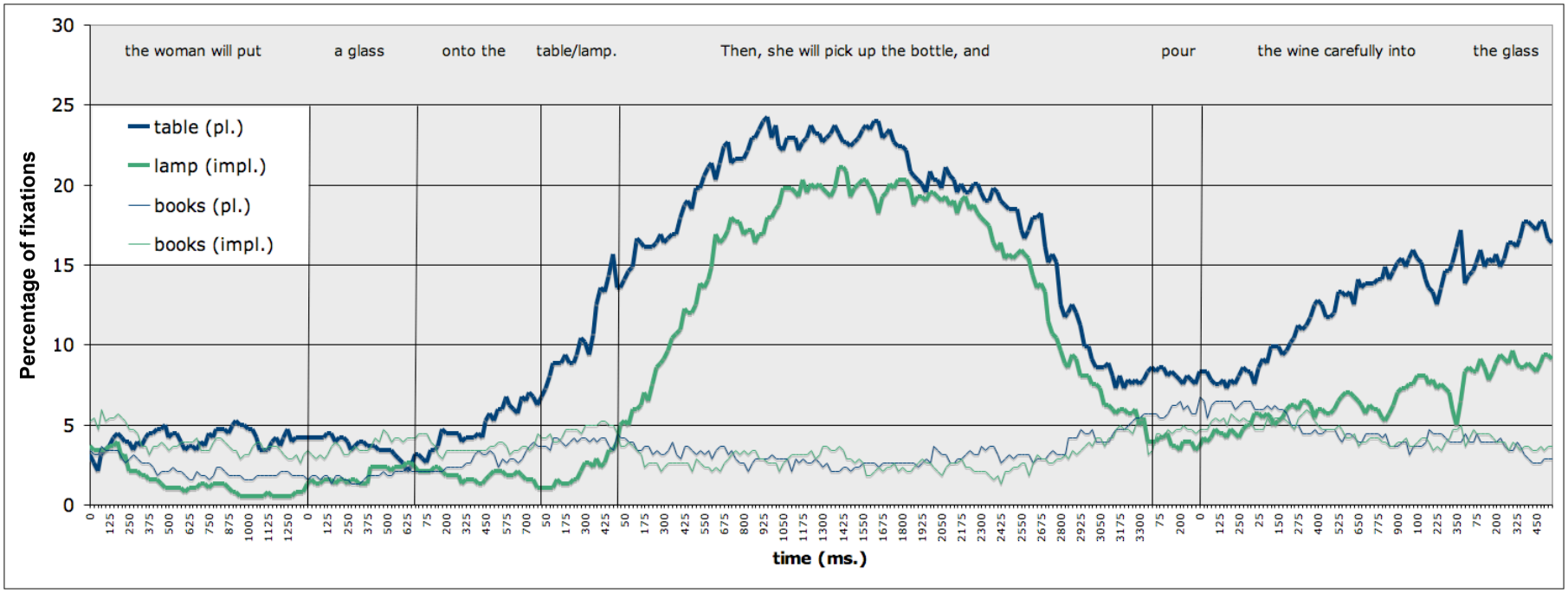


Fig 5.7. Percentage of trials with fixations toward the previous region of the table (plausible condition), the lamp (implausible condition) and the books (distractor). The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

5.3.3. Discussion

In the current experiment we removed the target item (e.g. the glass) from the visual scene and introduced it purely through language (e.g. “the woman will put *a glass* onto the table/lamp…”). This allowed us to explore the extent to which a purely linguistic representation of the glass might be able to enhance anticipation towards implausible locations in the context of a blank screen. The results showed that participants anticipated the location of the glass when it was said to have been moved to a plausible location, but not when it was said to have been moved to an implausible location. However, when the glass was directly referred to at the end of the narrative participants looked to its described location regardless of plausibility.

This suggests that a purely linguistic representation of the glass did not lead to an increase in the proportion of anticipatory looks to the implausible locations. Rather, the findings were similar to the pattern of eye movements we observed in experiment 1 – in this experiment the target items (e.g. the glass) were always presented in the visual scene whereafter they were introduced by spoken language (e.g. “the woman will put *the glass* onto the table/lamp…”). While we cannot compare eye movements directly across these two experiments, the similarity of the observed patterns further indicates that a purely linguistic representation of the glass did not enhance participants’ anticipation of the implausible locations any more than a visual/linguistic representation of the glass. In other words, the linguistic introduction of ‘a glass’ did not appear to emphasise the future relevance of the glass (or increase its accessibility) any more than when that same glass had first been presented within the visual scene. Like in experiment 1, when ‘a glass’ was described as being moved to the lamp (implausible location) we found little evidence of anticipatory eye movements to the previous region of the lamp. In contrast, when ‘a glass’ was described as being moved to the table (plausible location), we found that participants anticipated and looked to this region. During the final reference to the glass participants

looked as much to the region of the lamp as they looked to the region of the table.

In experiment 6 we compared looks to implausible locations (e.g. the region of the lamp) in situations where we had either removed the target items (e.g. the glass) completely from the visual scene, or when the target items were first presented within the visual scene and thereafter referred to by language. When we compared looks to the region of the lamp and the region of the books (distractor) we found that during the critical part of the sentence (the wine carefully into the glass) there were no more looks to region of the lamp than there were looks to the unnamed books and this was the case for both conditions. As such, it is possible that the proportion of looks to the region of the lamp were not derived from any previous mention of the/a glass being moved to an implausible location, but rather occurred due to the same level of chance that lead to a certain proportion of looks to an unnamed distractor region. In the current experiment we attempted to control for this confound by comparing looks to plausible (e.g. table) and implausible (e.g. lamp) locations when the target items (e.g. the glass) were absent from the visual scene in both conditions. During 'the wine carefully into' we found no difference in the proportion of anticipatory looks to the region of the lamp (implausible locations) and looks to the region of the books (unnamed distractors). In contrast, there were more anticipatory looks to the region of the table (plausible locations) than the region of the books. During the final reference to the glass there were more looks to both the region of the table and the lamp, compared to the region of the books. This suggests that the proportion of looks to the plausible and implausible locations were not due to levels of chance, but rather derived from internal representations of the glass as described by spoken language. In other words, during 'the wine carefully into' participants looked more to the region of the table than the region of the books because they anticipated the table as the location where the wine would be poured. However, when the glass

was described as having been moved to the lamp, participants did not anticipate this location (because of its unlikelihood) and therefore looked no more to the region of the lamp than the region of the books.

To sum up, the pattern of eye movements observed in the current experiment was similar to that observed in experiment 1, showing that plausibility influenced anticipatory eye movements to the event-specific location of the glass, but not eye movements during the final reference to the glass. How might we explain this pattern of eye movements? Going back to Altmann's (1999) experiment, participants indicated that a sentence stopped making sense when the verb (e.g. injured) could not plausibly refer to any of the previously introduced (inanimate) entities, suggesting that as soon as encountering the verb, participants evaluated the extent to which it could plausibly refer to any of the previously introduced entities. If a plausible entity had been previously introduced participants anticipated that the verb would refer to this entity. However, if no plausible entity had been introduced, participants took longer to read the verb and were more likely to indicate that the sentence no longer made any sense. In the current experiment we found that after hearing the verb 'pour' participants anticipated the plausible location of the glass (table), but not the implausible location (lamp). It may be that when hearing 'pour the wine carefully into…' participants evaluated the extent to which 'pour' might refer to any of the previously introduce entities, correctly determining that the glass was the only entity that afforded 'pouring into'. Therefore when the glass was described as having been moved to the table participants anticipated and looked to the previous location of the table. However, when the glass was described as having been moved to the lamp the unlikelihood of this location might have lead participants to no longer consider the glass a plausible referent and this might explain why it is we see anticipatory eye movements to the plausible locations, but not the implausible locations.

5.3.4. Summary & questions

In experiment 6 we compared looks to the lamp in situations where the glass had either been removed from the visual scene (and therefore introduced solely through language), or when the glass was first presented within the visual scene and then referred to by language. This manipulation allowed us to investigate the extent to which a purely linguistic representation of the glass would be able to enhance anticipation towards implausible locations in the context of a blank screen. We found no difference in the proportion of looks to the region of the lamp, suggesting that a purely linguistic representation of 'a glass' did not enhance the accessibility and anticipation of the its described location any more than a visual/linguistic representation of 'the glass'. However, since there was no difference in the proportion of eye movements to the region of the lamp and the region of an unnamed distractor item it is possible that looks to the region of the lamp during the critical part of the narrative were not related to spoken language, but rather occurred due to levels of chance.

In experiment 7 we attempted to control for this potential confound by comparing looks to the region of the table (plausible locations) with looks to the region of the lamp (implausible locations), after removing the glass from the visual scene in both conditions. The results showed a similar pattern of eye movements to experiment 1 (where the glass was first presented in the visual scene and then referred to by spoken language). In both these experiments participants anticipated the location of the glass when it had been moved to a plausible location, but not when it had been moved to an implausible location. In contrast, there was no influence of plausibility during the final reference to the glass. Together these findings suggest that a purely linguistic representation of the glass did not increase the proportion of anticipatory eye movements to the implausible locations. The data further showed that during 'the wine carefully into' participants anticipated the table as the location where the wine would be poured, more than they anticipated the location of

the distractor. However, when the glass was said to have been moved to the lamp participants did not anticipate (and look to) this location any more than they anticipated the location of the distractor. During the final reference to the glass there were more looks to both the region of the table and the lamp, compared to the region of the distractor. These findings suggest that the proportion of looks to the event-specific location of the glass were derived from language-mediated representations, rather than levels of chance.

Previous data by Altmann (1999) indicate that the difference in anticipatory eye movements may be related to the likelihood of the verb (e.g. pour) referring to one of the previously introduced entities. As such, it may be the case that when the glass is described as having been moved onto the table it presents a plausible referent for 'pouring the wine carefully into'. In contrast, when the glass is described as having been moved onto the lamp, the unlikelihood of this location makes it a much less plausible referent. This account might explain why participants anticipated the location of the table, but not the lamp. However, it doesn't explain why plausibility only influenced eye movements in the context of a blank screen – recall that in experiment 2 where the visual scene remained onscreen, participants anticipated both the plausible and implausible locations. In the next chapter we outline a theory attempting to explain why plausibility only affects anticipatory eye movements in the context of a blank screen.

CHAPTER 6

ANTICIPATION, EVENT-PLAUSIBILITY AND SCENE CONSTRAINTS

Experiments 1 and 2 explored the extent to which plausibility of location is able to influence the accessibility of objects within event-representations and further, how this may be modulated by the nature of the visual context (as present or absent). We found that in the context of a blank screen (experiment 1) participants only anticipated plausible locations. When items had been moved to implausible locations there was little or no evidence of anticipatory eye movements. However, in the context of a concurrent visual scene (experiment 2) participants anticipated both the plausible and implausible locations. In both experiments there was no influence of plausibility during the final reference to the target item.

Experiments 3 and 4 investigated whether the pattern of eye movements observed in experiments 1 and 2 could be generalised and extended to contextual plausibility. As opposed to moving objects to either plausible or implausible locations, we now moved contextually plausible (e.g. a cat in a kitchen) and implausible items (e.g. a penguin in a kitchen) to the same location. The findings were similar to the pattern of eye movements we observed in experiments 1 and 2. In other words, plausibility only influenced anticipatory eye movements in the context of a blank screen (experiment 3). When

the visual scene remained onscreen (experiment 4) participants anticipated the described location, both when a contextually plausible, or implausible item was said to have been moved there. During the final reference to the target item there was no influence of plausibility in either of the experiments.

In experiment 5 we investigated the extent to which a higher proportion of initial eye movements to the plausible locations (observed in experiment 1) could have enhanced the accessibility of those locations during later anticipation. The findings showed that while the proportion of initial looks made the implausible regions more accessible later in the narrative, this was not the case for the plausible locations. The data further indicated that participants who initially looked to the previous region of the lamp returned as much to that location as participants who had initially looked to the previous region of the books. Taken together the data suggests that the higher proportion of anticipatory eye movements to the plausible locations (observed in experiment 1) is not solely dependent on the higher proportion of early looks to these locations.

In experiments 6 and 7 we explored the extent to which a purely language-mediated representation of the target item (e.g. a glass) would be able to increase the proportion of anticipatory looks towards the implausible locations in the context of a blank screen. The idea here was that introducing new entities into the language generally indicates that they will be referred to subsequently (whereas referring to previously introduced entities, such as “the glass” is not necessarily accompanied by subsequent references). The findings indicated that a linguistic representation of the target item did not enhance the accessibility and anticipation of its described location.

Together these data suggest that the absence/presence of a visual context provides a quantitative difference in how we process and anticipate implausible events. In other words, the constraint that is drawn from a previous visual scene is weaker than the constraint drawn from a concurrent visual scene and this allows real-world

knowledge to exert a greater influence on the probabilistic activation of likely candidates. This probabilistic approach means that the underlying mechanism is the same, but that quantitative differences in the strength of constraints leads to what appear to be qualitative differences in behaviour. In the context of a blank screen we only anticipate locations when these are either plausible, or when the items moved to these locations are contextually plausible. However, in the context of a concurrent visual scene we anticipate upcoming locations regardless of plausibility. As such, it seems that the absence of a visual scene somehow renders implausible items, or locations less viable in terms of guessing or anticipating any upcoming actions involving these items or locations. Furthermore, it is important to note that both in the context of a blank, or concurrent visual scene the data showed no influence of plausibility when the target item was referred to at the end of the narrative. This suggests that once the target item was directly referred to participants had no difficulty retrieving its (implausible) location.

In the previous chapter we proposed that the difference in anticipatory eye movements (in the context of a blank screen) could be related to the likelihood of the verb referring to one of the previously introduced entities (e.g. Altmann, 1999). For example, when hearing “the woman will move the glass onto the table/lamp. Then she will pick up the bottle, and pour the wine carefully into…” participants evaluated the extent to which the verb ‘pour’ could plausibly refer to any of the objects previously introduced within the narrative, correctly determining that the glass was the only object that could afford ‘pouring into’. When the glass had been moved onto the table participants judged it to be a plausible referent and therefore they anticipated and looked to the location of the table. However, when the glass was described as having been moved to the lamp, the unlikelihood of this location could have resulted in participants no longer considering it to be a plausible referent when having to anticipate where the wine would be ‘poured into’. Only when the

glass was directly referred to at the end of the narrative (pour the wine carefully into…the glass) did it become clear that the glass was the correct referent, whereby participants looked to its location irrespective of plausibility. While this account might explain why participants only anticipated the glass when it had been moved to a plausible location, it doesn't explain why plausibility only influenced eye movements in the context of a blank screen – in experiments 2 and 4 where the spoken narratives were presented in the context of a concurrent visual scene we found no effect of plausibility.

In this chapter we propose an account explaining why plausibility affects anticipatory eye movements in the context of a blank screen, but not in the context of a concurrent visual scene. Our explanation is based on the assumption that there is a quantitative difference in how we process and anticipate upcoming discourse when language refers to something outside our visual environment, as opposed to when language refers to entities within our immediate visual proximity. During language comprehension we assume that participants in an event will be drawn from our visual environment (e.g. Altmann & Mirkovic, 2009; Cooper, 1974; Tanenhaus et al., 1995) and this occurs even when characters are depicted performing a non-stereotypical action (e.g. Knoeferle & Crocker, 2006). However, in the absence of a visual context we have to rely on our memory, or internal representation of the visual scene. When this is the case the constraints of the visual scene may be weakened since our memory trace of the scene is inevitably weaker than the actual source of that trace. In effect, we are no longer constrained as strongly as in the concurrent case, to anticipate that the object about to be referred to is an object that has been previously introduced.

6.1. Visual and linguistic constraints

Several studies have shown that during language comprehension we very rapidly establish reference to objects within our visual environment (e.g. Allopenna, Magnuson & Tanenhaus 1998;

Eberhard, Spivey-Knowlton, Sedivy & Tanenhaus, 1995). Interestingly, this also happens when we are not explicitly instructed to look towards the named items (Altmann, 2011) and even when we are actively encouraged to ignore spoken language (Salverda & Altmann, 2011). Early research by Cooper (1974) showed that eye movements tend to be rapidly directed towards items when these are directly referred to by spoken language. In this study participants viewed a visual scene consisting of nine different items while they listened to short stories containing words that directly referred to the depicted items. The findings showed that when hearing the name of one of the depicted items (for example, a zebra) participants looked more toward that item, compared to the unnamed items.

Later research by Tanenhaus et al. (1995) demonstrated that during language comprehension we sequentially establish reference to objects within our visual environment and that this occurs at the earliest moment in time. In this study participants were shown a visual display containing several different objects and asked to perform a set of instructions related to one of those objects. The findings showed that participants looked sequentially to each potential object until the instructions uniquely identified which specific object was referred to. As soon as participants had enough information to identify which of the objects was referred to, they looked to this object, whereafter they carried out the instructions. These studies show a rapid and sequential integration between language comprehension and the visual world. In other words, as language unfolds spoken information serves to constrain the potential number of upcoming referents within the visual scene.

6.2. Anticipatory constraints

Altmann and Kamide (1999) further demonstrated how language (in this case semantic information derived from verbs) is able to rapidly direct our attention towards anticipated objects within a specific visual context. When shown a visual scene depicting a boy, a toy

train set, a toy car, a balloon, and a birthday cake and hearing either "*the boy will eat the cake"* or *"the boy will move the cake"* participants looked more towards the cake upon hearing the verb 'eat', compared to when they heard the verb 'move'. These findings suggest that participants used verb-related information (i.e. the verb eat implied that the upcoming referent must be something edible) in order to anticipate which of the items in the visual scene would most likely be referred to next. In other words, by integrating verb-related information with the features and affordances of the items presented in the visual context, participants were able to restrict the number of possible upcoming referents and thereby correctly anticipate the most appropriate antecedent before it was directly referred to. As such, the concurrent visual context alongside linguistic information served to constrain the number of potential participants consequently making it easier for participants to anticipate which item would be referred to next.

Kamide et al. (2003) extended this research by investigating whether anticipatory eye movements are exclusively related to semantic information derived from the verb, or if we further rely on real-world knowledge when anticipating the most likely outcome of an event in which the verb (on its own) allows for several outcomes. Participants viewed a concurrent visual scene depicting a man, a young girl, a motorbike and a carousel. They then heard either *"the man will ride the motorbike"* or *"the girl will ride the carousel"*. In this event the verb 'ride' could refer to both the carousel and the motorbike, whereby verb-related information alone did not allow participants to determine whether 'ride' referred to the depicted motorbike, or the carousel. When this was the case participants had to further rely on experiential knowledge to provide a stronger degree of constraints in terms of plausibility. In other words, while the man could be intending to ride both the motorbike and the carousel, real-world knowledge would suggest that a man would be more likely to ride a motorbike, whereas a young girl would be more likely to ride a

carousel. The data reflected this likelihood showing that upon hearing "*the man will ride the...*" participants made more anticipatory eye movements toward the motorbike. In contrast, when hearing "*the girl will ride the...*" there were more anticipatory looks toward the carousel. This shows that in situations where verb-related information may equally apply to several objects within a concurrent visual context, participants had to integrate linguistic information with real-world knowledge in order to anticipate the most plausible referent. In other words, participants' experiential knowledge increased the degree of constraint by informing participants about the plausibility of each possible outcome.

6.3. Constraints in the absence of a visual context

While language is often used to refer to items within our visual proximity, we also use language to refer to objects when these are absent from our visual context. The blank screen paradigm allows us to explore how we map language onto our internal representation of a previously encountered visual scene, as opposed to the visual scene itself. In such cases we might expect the constraints derived from the visual scene to be weakened, since our memory trace of a visual scene is inevitably weaker than a concurrent visual scene (e.g. Averbach & Coriell, 1961; Sperling, 1960). In other words, in the context of a blank screen the anticipatory activation of appropriate representations is no longer bound exclusively by the visual context, but may now be further influenced by real-world knowledge. This notion of 'visual constraints' relates to a reading study by Rayner et al. (2004), which investigated the effects of reading plausible (e.g. John used a knife to chop the large carrots for dinner), implausible (e.g. John used an axe to chop the large carrots for dinner) and anomalous (e.g. John used a pump to inflate the large carrots for dinner) sentences. Rayner and colleagues found an earlier pattern of disruption in response to the anomalous sentences compared to the implausible sentences, which only showed disruption at a later point in time. In contrast, the visual world studies by Altmann and Kamide

(1999) and Kamide et al. (2003) showed early plausibility effects, specifically in terms of anticipating upcoming references. Rayner et al. proposed that this difference between experiments might be related to the stronger degree of constraint afforded by visual scenes, which made it easier to for participants to anticipate the most plausible referents. In contrast, during reading participants are not restricted by a visual context and are therefore less able to anticipate upcoming referents based on their likelihood. According to this theory, we might similarly expect a concurrent visual scene (as opposed to a blank screen) to provide a stronger degree of constraint, resulting in an earlier (anticipatory) influence of plausibility.

Altmann (2004) explored how language is able to direct our attention towards anticipated objects when the visual scene was removed before the onset of the spoken language. Participants were shown a visual image depicting a man, a woman, a newspaper and a cake. After five seconds the image was removed and replaced with plain white screen and participants heard either "*the man will eat the cake"*, or *"the woman will read the newspaper"*. Like in the previous experiment by Altmann and Kamide (1999) participants looked more towards the previous region of cake upon hearing the verb 'eat' than when hearing the verb 'read'. In contrast, there were more anticipatory looks to the previous region of the newspaper upon hearing the verb 'read'. These findings suggest that anticipatory eye movements are not necessarily dependent on seeing a concurrent visual scene – in the absence of a visual context, participants simply mapped spoken language onto an internal representation of the previously encountered scene. Furthermore, while the constraints derived from participants' memory of the visual scene may have been weaker than the constraints provided by a concurrent visual scene, linguistic constraints derived from the verb, alongside real-world knowledge allowed participants to anticipate the most appropriate object even in the absence of a concurrent visual

context. In this example the cake was the only item in the previous visual scene that fulfilled the selectional restrictions of the verb ‘eat’. Similarly, real-world knowledge informed participants that cakes are not only edible, but also tasty and this in turn made the cake a more plausible referent than the newspaper.

But what might have happened if, instead of a cake the visual scene had depicted something that fulfilled the selectional restrictions of the verb ‘eat’ (e.g. a tarantula, or a poisonous toadstool), but that real-world knowledge informed us was not something most people would be likely to eat? Under these circumstances real-world knowledge differs from the selectional restrictions in terms of what we can eat and what we are likely to eat and this discrepancy might lead us to anticipate a number of alternative options that are not directly related to the previous visual context.

6.4. Explaining the plausibility effect

Going back to experiments 1 and 2 we found that in the context of a concurrent visual scene (experiment 2) participants anticipated the location of the glass regardless of whether it had been moved to the table (plausible location), or the lamp (implausible location). However, in the context of a blank screen (experiment 1) participants only anticipated the location of the glass when it had been moved to the table – when the glass had been moved to the lamp there was little or no evidence of anticipatory eye movements. In the context of a concurrent visual scene participants assumed that referred items would be drawn from within this context. Looking back at the visual scene (see figure 6.1.) it is clear that the glass is the only item within this scene that affords ‘pouring into’. Therefore, when hearing ‘the woman will move the glass onto the table/lamp. Then, she will pick up the bottle, and pour *the wine carefully into* the glass’ participants anticipated and looked to the described location of the glass, even when it had been moved to the lamp. In other words, if participants assumed that the upcoming reference would be drawn from within

the visual scene and the glass was the only item within the scene that could afford ‘pouring into’, the visual context constrained them to anticipate the event-specific location of the glass, even when this location was implausible.

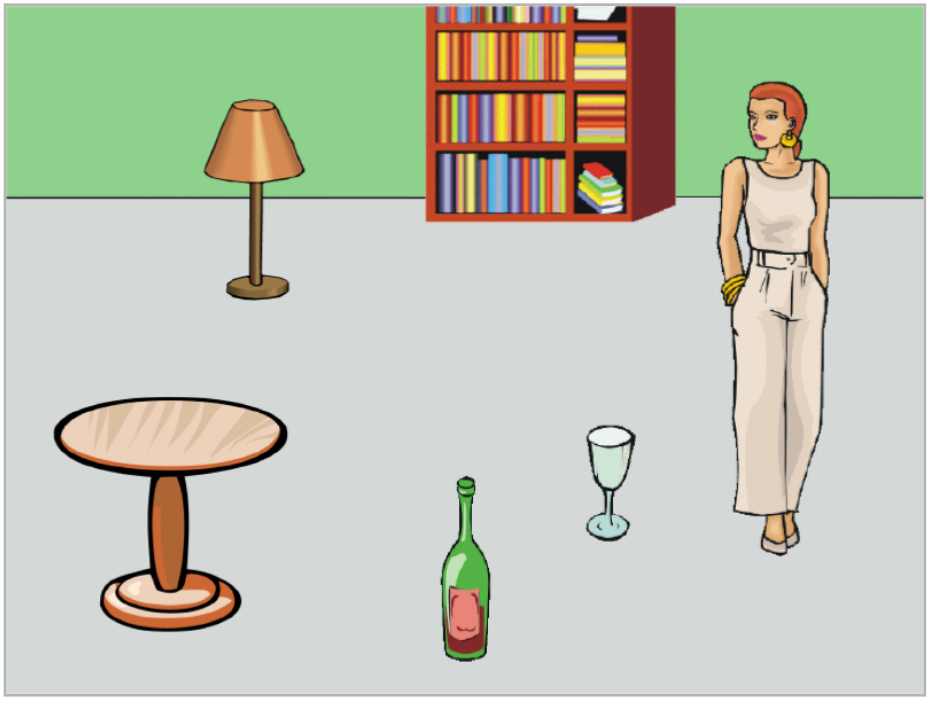


Figure 6.1. Example of one of the visual scenes from experiment 1 and 2.

However, in the context of a blank screen participants had to rely on a visual memory, or internal representation of the scene. As such, the constraints derived from the visual scene were weakened since anticipation was no longer so strongly bound by the visual context. Rather, in the absence of a visual scene participants relied on linguistic information, as well as real-world knowledge and this could have rendered the upcoming reference much less specific. When the glass was said to have been moved to the table real-world knowledge would have informed participants that this was a plausible and appropriate location to put a glass. Based on this information participants anticipated and looked to the table as the upcoming location for pouring the wine into. However, when the glass was described as having been moved to the lamp experiential knowledge would have informed participants that this was an unlikely and

inappropriate location to put a glass. It is possible that the unlikelihood of this location changed participants' perception of the glass (making it less affordable) and this may have lead them to consider and anticipate a number of alternative options for 'pouring the wine into'. In other words, in the absence of a visual context participants' anticipation was *no longer exclusively constrained by the visual scene* and as such the wine needn't be poured into that one specific glass, but could instead be poured into a completely different location. For example, if we were to redirect our attention from e.g. the table to the kitchen counter, we would no longer be able to see any objects that were on the table (e.g. the glass), but we would be able to see other affordable objects situated on the kitchen counter (e.g. another glass, the sink, or a carafe). Only during the final reference to 'the glass' did it become clear that the wine was about to be poured into the previously mentioned glass, whereby participants looked to its described location regardless of plausibility.

In the following chapter we tested this account, firstly by manipulating the number of affordable items within the context of a concurrent visual scene (experiment 8) and secondly by linguistically restricting the number of affordable items within the context of a blank screen (experiment 9). These manipulations allowed us to explore the extent to which both visual and linguistic constraints are able to guide anticipatory eye movements to appropriate (yet implausible) locations.

CHAPTER 7

MANIPULATING THE NUMBER OF AFFORDABLE ITEMS

In the previous chapter we proposed an account explaining why plausibility affects anticipatory eye movements in the context of a blank screen, but not in the context of a concurrent visual scene. Our explanation is based on the assumption that there is a quantitative difference in how we process and anticipate upcoming discourse when language refers to something outside our visual environment, as opposed to when language refers to entities within our immediate visual proximity. During language comprehension we assume that participants in an event will be drawn from the visual context if one is available (e.g. Altmann & Mirkovic, 2009; Cooper, 1974; Tanenhaus et al., 1995). However, in the absence of a visual context we have to rely on our memory, or internal representation of the visual scene. When this is the case the constraints of the visual scene are weakened since our memory trace of the scene is inevitably weaker than the actual source of that trace (e.g. Averbach & Coriell, 1961; Sperling, 1960).

In the context of a concurrent visual scene the referred target (e.g. the glass) was the only item that could afford the described action (e.g. pouring into). As such, the concurrent visual context alongside language served to constrain the number of potential upcoming

referents and this would explain why we anticipate the event-specific location of the target referent regardless of plausibility. However, in the context of a blank screen we have to rely on our memory/internal representation of a visual scene. In this case the anticipatory activation of appropriate representations is no longer strongly bound by the visual context, but may now be influenced by real-world knowledge. When the referred target is described as being moved to a plausible location (e.g. the table) real-world knowledge informs us that this is an appropriate location and therefore we anticipate (and look) to this location. However, when the referred target is described as being moved to an implausible location (e.g. the lamp) real-world knowledge similarly informs us that this is an inappropriate and unlikely location. This unlikelihood may lead us to consider and anticipate a variety of options. Only when the target is directly referred to at the end of the narrative does it become clear that spoken language refers to the previously mentioned item, whereby we look its event-specific location irrespective of whether this is plausible or not.

In this chapter we present two experiments that aimed to explore the extent to which both visual and linguistic constraints are able to guide anticipatory eye movements to appropriate (yet implausible) locations. In experiment 8 we manipulated the number of affordable items within the context of a concurrent visual scene. The aim of this experiment was to investigate the extent to which the inclusion of a second (un-referred) affordable item might be able to influence eye movements in the context of a concurrent visual scene. In the previous experiments the glass was always the only item that could afford ‘pouring into’ and we proposed that this is why participants anticipated its (implausible) location when the narrative was presented in the context of a concurrent visual context. As such, we might speculate that adding a second glass to the visual scene would decrease the weighting of constraints derived from object-affordances by providing an alternative option for ‘pouring the wine

into'. In contrast, experiment 9 attempted to increase the weighting of constraints (in the context of a blank screen) by linguistically restricting the number of affordable items within the context of the narrative. By doing so, we hoped to linguistically equalise the constraints provided by a concurrent and a blank screen context. If language constrains the context of the narrative so that there is only one possible glass we would expect to see anticipatory eye movements to the appropriate location of the glass, even when implausible.

7.1. EXPERIMENT 8

7.1.1. Affordance-based constraints

In the previous chapter we proposed that the constraints of a concurrent visual scene guides anticipatory eye movements to the event-specific location of the glass even when this location is implausible. In other words, since the glass is the only item in the scene that affords 'pouring into', the visual context (and the object-affordances within this context) alongside language constrains anticipation, whereby participants look the location of the glass regardless of plausibility.

A number of studies have explored how affordance-based constraints are able to influence language comprehension (e.g. Chambers, Tanenhaus, Eberhard, Filip & Carlson, 2002; Glenberg et al., 2009; Kaschak & Glenberg, 2000), leading Glenberg and Robertson (2000) to suggest that our internal representations of language are representations of an event (as well as the affordances of this event), as opposed to representations of language itself. Glenberg (1997) further suggested that language comprehension often depends on the combination of affordances, whereby implausible combinations of affordances in terms of underlying goals can serve to constrain comprehension. Such affordance-based constraints are demonstrated in sentences like *"Marissa forgot to bring her pillow on her camping trip, as a substitute for her pillow she*

filled up an old sweater with water" (Glenberg et al., 2000, p. 385), whereby the combination is incompatible with the goal since a sweater cannot afford the containment of liquids. This in turn, makes it difficult for people to form a realistic mental representation of the described event.

Chambers, Magnuson and Tanenhaus (2004) investigated the extent to which participants' behavioural goals and the affordances of presented objects were able to influence eye movements when listening to a set of syntactically ambiguous instructions. In this experiment participants were seated in front of a display of real objects and instructed to either; *"pour the egg in the bowl over the flour"* (temporarily ambiguous instructions), or *"pour the egg that's in the bowl over the flour"* (unambiguous instructions). Two versions of the display were used (see figure 7.1.), each containing four object types: one target referent (egg in bowl – top right), one referential competitor (egg in glass – top left), one true goal (flour – bottom right), and one false goal (empty bowl – bottom left).

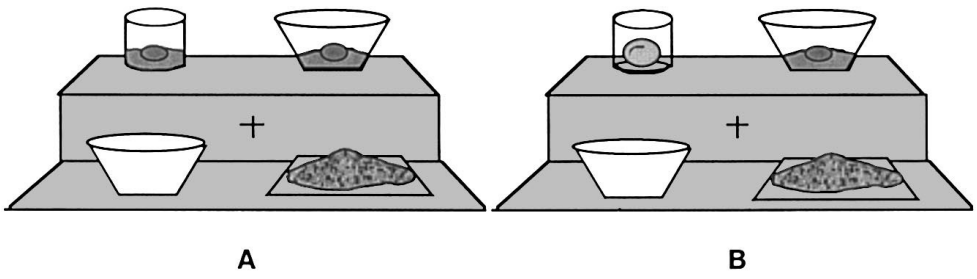


Figure 7.1. Example display (taken from Chambers et al., 2004, p. 689).

In display A the target referent and the referential competitor were both displayed in liquid form, thus both being compatible with the action requested in the instructions (pour the egg). However, in display B the target referent was displayed in liquid form, whereas the referential competitor was presented in solid form. As such, the target referent was the only item in the display that was compatible

with the instructions.

The results showed that when both eggs were presented in liquid form participants initially found it difficult to determine which of the two eggs the instruction referred to. In other words, since both of the eggs could afford pouring participants interpreted the phrase 'in the bowl' as determining which of the eggs was referred to. In contrast, when only one egg was presented in liquid form participants were able to immediately determine which of the eggs the instructions referred to, whereby they interpreted the phrase 'in the bowl' as referring to the intended location for the pourable egg. Chambers et al. (2004) suggested that when one egg was displayed in liquid form and the other egg in solid form the affordances of each of the eggs, combined with verb-related information (pour) influenced syntactic ambiguity resolution by directing participants' attention to the most plausible goal.

Other studies have further shown how the combination of objects and action-based affordances are able to guide anticipatory eye movements to the most appropriate item within a visual scene (e.g. Altmann & Kamide, 1999; Kamide, Scheepers & Altmann, 2003). Altmann & Kamide (2007) explored how a combination of the temporal aspects of a verb and the affordances of objects are able to influence anticipatory eye movements. They found that when presented with an image depicting, a man, a table with a full glass of beer, an empty wine glass, some cheese, and some Christmas crackers on the floor, participants looked more towards the full glass of beer upon hearing *'the man will drink all of…'* whereas when hearing *'the man has drunk all of…'* participants looked more to the empty wine glass. This study shows that participants' knowledge about past and future events activated the objects' affordances (i.e. the full glass afforded a future drinking event, whereas the empty glass could only afford a past drinking event), which then constrained the number of possible upcoming referents, enabling participants to correctly anticipate the appropriate reference.

But how do we anticipate upcoming references in situations where there is more than one affordable item in the visual context? Kamide et al. (2003) showed that in such circumstances we have to rely on real-world knowledge in order to anticipate the most likely outcome of an event. When hearing sentences such as *'the man will ride the…'*, (whilst viewing a scene depicting a man, a girl, a motorbike, and a carousel), participants made more anticipatory eye movements towards the motorbike compared to the carousel. In contrast, when hearing or '*the girl will ride the…'* there were more anticipatory looks to the carousel. In this example both the motorbike and the carousel were compatible with the verb (ride) and as such, participants had to further rely on real-world knowledge in order to correctly anticipate the most plausible referent. In other words, participants' experiential knowledge about these types of events strengthened the weighting of constraint by informing them about the plausibility of each possible outcome.

In the current experiment we similarly manipulated the number of affordable items within a concurrent visual scene and this allowed us explore the extent to which the addition of a second (un-referred) affordable item might influence eye movements when the first affordable item is described as being moved to either a plausible, or implausible location. In the previous concurrent experiment (experiment 2) the glass was the only item in the visual scene that could afford 'pouring into' and we proposed that this is why participants anticipated its location regardless of plausibility. Taking this into account we might speculate that the inclusion of a second (affordable) glass to the visual scene will weaken the affordance-based constraints, resulting in a higher influence of plausibility, compared to conditions where only one glass is depicted in the visual scene (see figure 7.2.). In other words, the addition of an alternative glass for 'pouring the wine into' might lead participants to consider this option more when the alternative glass is in an implausible

location, compared to when it is in a plausible location.

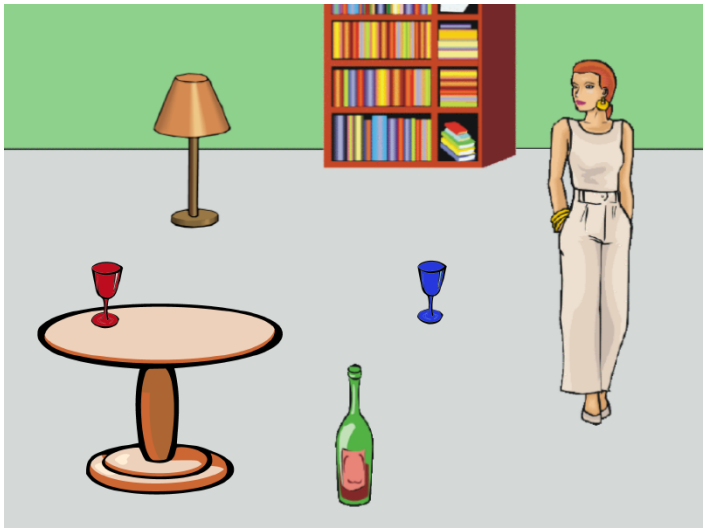


Figure 7.2. Example of one of the visual scenes depicting two affordable items.

7.1.2. Method

Participants

Sixty-four students from the University of York participated in this study, receiving either course credit or a payment of £4. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

Sixty-four experimental scenes (see figures 7.3. and 7.4.) were matched with four conditions. In conditions 1 and 2 we presented two affordable target items. For example, one blue glass was depicted on the floor and one red glass was depicted on the table. In conditions 3 and 4 we presented one affordable item and one unrelated item that could not afford the described action. In this example one blue glass was depicted on the floor and one red book was depicted on the table. This design allowed us to compare looks to the table (plausible location) with looks to the lamp (implausible location), as well as looks to the red glass (un-referred affordable item) between the plausible and implausible conditions. In addition to the experimental

trials, 64 sentence-picture pairs were included as fillers (see appendix 9 for examples of the filler items). Sixteen of the filler items depicted two differently coloured versions of the same un-referred item and a further 16 filler items depicted one version of a named coloured item. As such, 16 (filler) trials in each condition depicted two differently coloured versions of an un-named item and 16 (experimental) trials depicted two differently coloured versions, naming one of those items. Another 16 (filler) trials and 16 (experimental) trials referred to one version of a colour-specific item (e.g. the purple flower). The remaining 32 filler items referred to an item without specifying its colour (e.g. the flower).

1. The woman will move the blue glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Two glasses plausible)

2. The woman will move the blue glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Two glasses implausible)

3. The woman will move the blue glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (One glass plausible)

4. The woman will move the blue glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (One glass implausible)

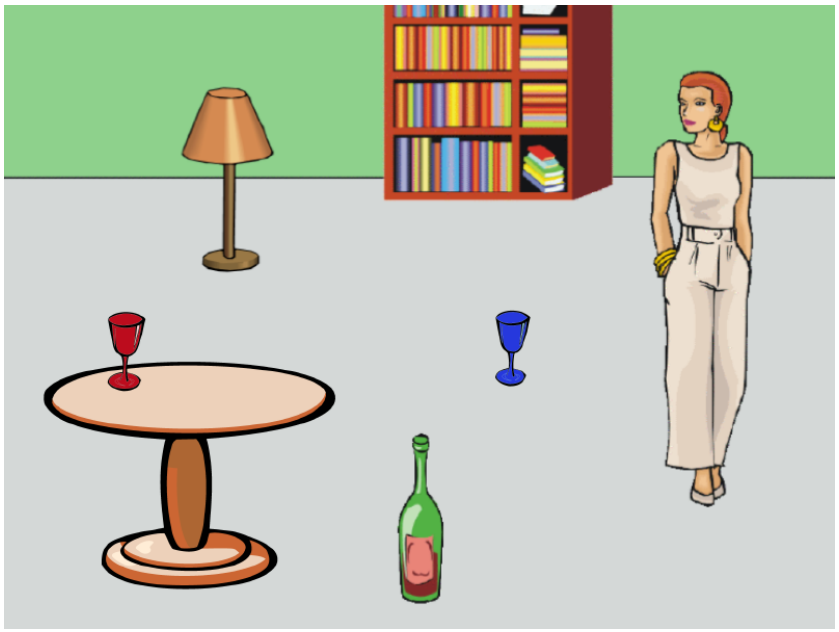


Figure 7.3. Example of one of the visual scenes paired with sentences: 1) and 2).

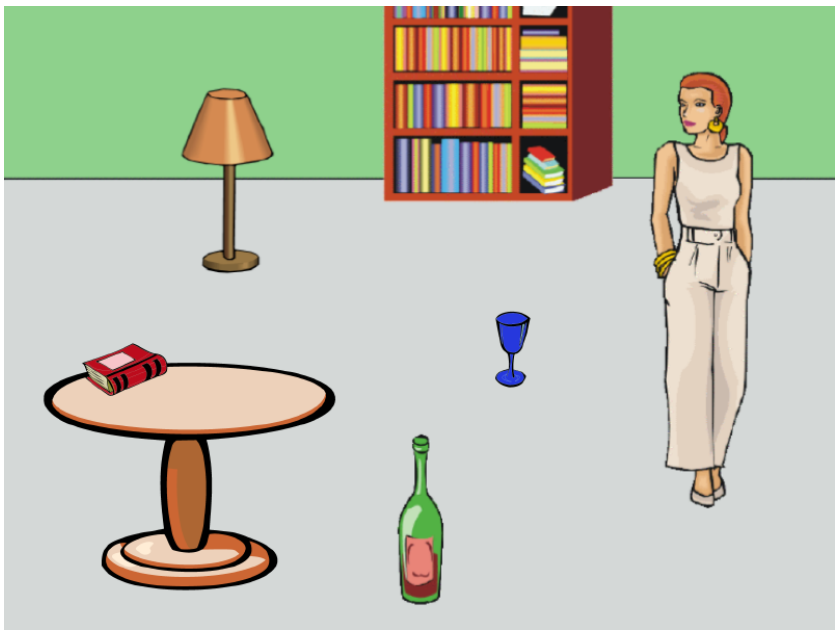


Figure 7.4. Example of one of the visual scenes paired with sentences: 3) and 4).

Procedure

The procedure for this experiment was the same as in the previous concurrent screen experiments – the visual scenes were presented for 1000 milliseconds, followed by the onset of the auditory stimuli. The trials ended 11 seconds after the audio onset, lasting a total of 12 seconds. The experiment was run on an Eyelink II head-mounted

eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

We defined four regions of interest within each scene; one area corresponded to the plausible location of the moved glass (e.g. the table, excluding the red glass/book), another corresponded to the implausible location of the moved glass (e.g. the lamp), and a third area corresponded to the unmoved location of the glass (e.g. the blue glass). Finally, we included a fourth area that either corresponded to the un-referred affordable item (e.g. the red glass), or the un-referred unaffordable item (e.g. the book). Since the visual scenes remained onscreen throughout each of the trials, we defined the regions of interest according to the outline of each target object. As such, participants' eye movements had to be directed to one of the pixels occupied by each object within the scene. Participants' eye movements were examined during certain time points in the spoken sentences with the percentage of saccades as the dependent measure. As in the previous experiments these critical time points occurred during 'the wine carefully into' (anticipatory eye movements), and during the final noun phrase 'the glass'.

To begin with, we compared the proportion of saccades to the table/lamp when either one glass, or two glasses were presented in the visual scene. These comparisons allowed us to see if the inclusion of a second (un-referred) affordable item would be able to decrease the weighting of constraints and thereby influence eye movements to the plausible and implausible locations when the first glass had been moved there. Secondly, we compared the proportion of saccades to the region of the (un-referred) glass when the other glass had been moved to the table or the lamp. This allowed us to explore whether participants would anticipate and look to the alternative (un-referred) glass more when the first glass had been moved to an implausible locations, compared to a plausible location.

7.1.3. Results

Table – one glass plausible vs. two glasses plausible

During 'the wine carefully into' there were more anticipatory looks to the table in the 'one glass' plausible condition than there were looks to the table in the 'two glasses' plausible condition ($t1$ (63) = 2.622, $p < .05$), although this difference was not statistically significant in the by-items analysis ($t2$ (31) = -1.445, $p > .05$). During the final reference to 'the glass' there were no more looks to the table in the 'one glass' plausible condition than there were looks to the table in the 'two glasses' plausible condition ($t1$ (63) = -1.010, $p > .05$) ($t2$ (31) < 1) (see figures 7.5., 7.6., and table 7.1.).

Lamp – one glass implausible vs. two glasses implausible

During 'the wine carefully into' there were no more anticipatory looks to the lamp in the 'one glass' implausible condition than there were looks to the lamp in the 'two glasses' implausible condition ($t1$ (63) < 1) ($t2$ (31) < 1). Similarly, during the final reference to 'the glass' there were no more looks to the lamp in the 'one glass' implausible condition than there were looks to the lamp in the 'two glasses' implausible condition ($t1$ (63) < 1) ($t2$ (31) < 1) (see figures 7.5., 7.6., and table 7.1.). This suggests that the inclusion of a second (affordable) glass did not weaken affordance-based constraints sufficiently to show an influence of plausibility.

Table vs. lamp

During 'the wine carefully into' there was no difference in anticipatory looks to the table in the 'one glass' plausible condition and looks to the lamp in the 'one glass' implausible condition ($t1$ (63) = 1.681, $p > .05$) ($t2$ (31) < 1). There was also no difference in anticipatory looks to the table in the 'two glasses' plausible condition and the lamp in the 'two glasses' implausible condition ($t1$ (63) < 1) ($t2$ (31) < 1). During the final reference to 'the glass' there were no more looks to the table in the 'one glass' plausible condition than there were looks

to the lamp in the 'one glass' implausible condition (*t1* (63) = 1.298, *p* > .05) (*t2* (31) = 1.266, *p* > .05). Similarly, there was no difference in looks to the table in the 'two glasses' plausible condition and looks to the lamp in the 'two glasses' implausible condition (*t1* (63) = 1.089, *p* > .05) (*t2* (31) < 1) (see figures 7.5., 7.6., and table 7.1.).

Time	table (1 glass)	table (2 glasses)	lamp (1 glass)	lamp (2 glasses)
the wine carefully into	34	28	30	29
the glass	12	13	16	17

Table 7.1. Percentage of trials with looks to the table and the lamp during "the wine carefully into the glass". Percentages calculated from the total number of trials.

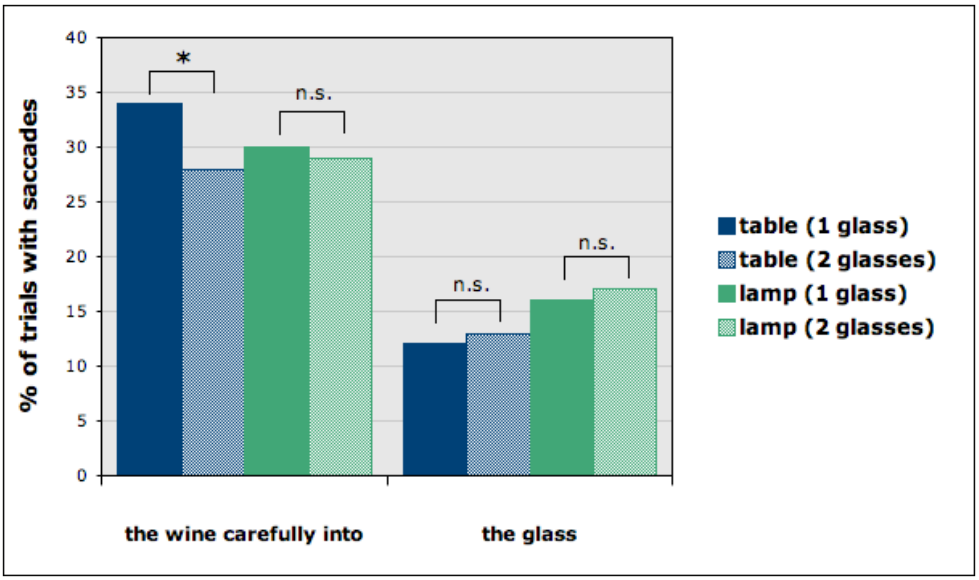


Figure 7.5. Looks to the table (plausible location) and the lamp (implausible location) during "the wine carefully into the glass".

Figure 7.6. illustrates the percentage of trials with fixations toward the table and the lamp in the plausible and implausible conditions. The relatively low increase in looks to the table and the lamp during the final reference to 'the glass' is presumably due to competition from the visual instantiation of the referred blue glass depicted on the floor

(see Altmann & Kamide, 2009[17]). Also, note the difference in the proportion of saccades in figure 7.5. and the proportion of fixations in figure 7.6., specifically during 'the wine carefully into'. This dissociation is best explained by the dependent measures used to plot each of the figures – figure 7.5. shows the proportion of saccades during the critical part of the narrative and figure 7.6. illustrates the proportion of fixations during each moment of the narrative. While the proportion of saccades and the proportion of fixations are necessarily related (fixations provide information about when and on which region the eyes are lingering, saccades inform us about when, and to which region the eyes are being directed) these dependent measures have nonetheless been shown to disassociate since the likelihood of fixating on a region during a specific time frame does not account for when those fixations began (see Atmann & Kamide, 2004, pages 380-382; Altmann, 2011, pages 994-995). Looking at the discrepancy during 'the wine carefully into' there is a relatively high proportion of saccades to the table and the lamp in both conditions, yet a much lower proportion of fixations to those regions during the same period. This difference arises because there are almost as many looks away from the table/lamp as there are looks towards the those regions. And because different participants across different trials launch saccades towards, or away from, these regions at different times within the extended temporal region of interest, the net result is that across the temporal regions of interest there is only a small net increase in fixations, while there are many more saccades towards (and almost as many away from) these regions.

[17] Altmann and Kamide (2009) showed that the salience of a concurrent visual scene (depicting the glass on the floor) lead to increased competition and therefore fewer looks to the described location of the glass (table), compared to the visually depicted location of the glass. In contrast, when the visual scene had been removed prior to the onset of spoken language competition decreased, now showing fewer looks to the (previous) visual location of the glass, compared to the described location of the glass.

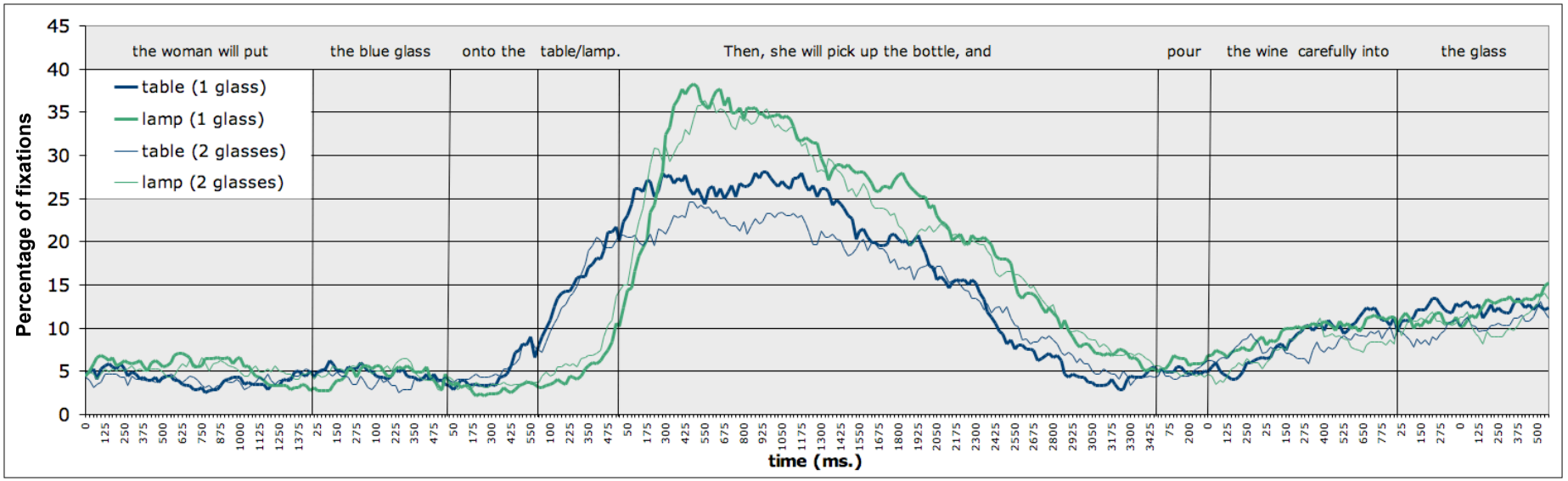


Fig 7.6. Percentage of trials with fixations toward the table and the lamp in the plausible and implausible conditions. The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

Red glass – two glasses plausible vs. two glasses implausible

During 'the wine carefully into' there were no more anticipatory eye movements to the un-referred affordable item (red glass) in the 'two glasses' plausible condition (where the blue glass was said to have been moved to the table) than there were looks to the un-referred affordable item in the 'two glasses' implausible condition (where the blue glass was said to have been moved to the lamp) (*t1* (63) < 1) (*t2* (31) = 1.761, $p > .05$). Similarly, during the final reference to 'the glass' there was no difference in looks to the un-referred affordable item (red glass) between these two conditions (*t1* (63) < 1) (*t2* (31) < 1) (see figures 7.7., 7.8., and table 7.2.). This suggests that when the blue glass was described as having been moved to the lamp participants did not consider the un-referred alternative red glass any more than when the blue glass was described as having been moved to the table.

Time	book (pl.)	book (impl.)	red glass (pl.)	red glass (impl.)
the wine carefully into	18	19	25	22
the glass	10	5	9	9

Table 7.2. Percentage of trials with looks to the un-referred book and the un-referred red glass during "the wine carefully into the glass". Percentages calculated from the total number of trials.

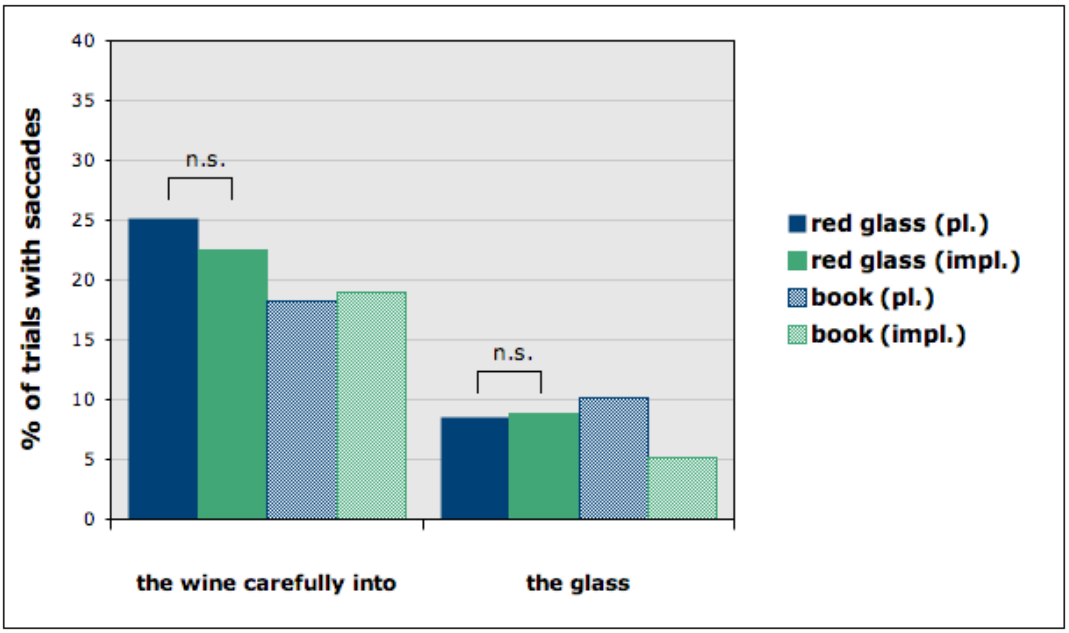


Figure 7.7. Looks to the red glass (un-referred affordable item) and the book (un-referred unaffordable item) during “the wine carefully into the glass”.

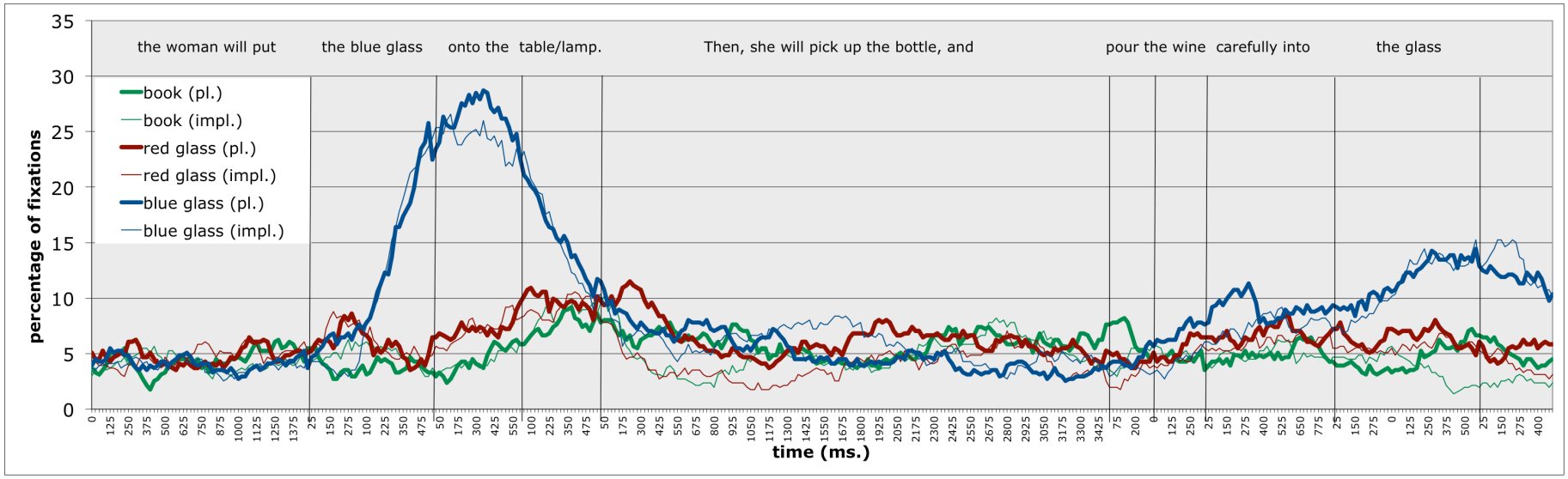


Fig 7.8. Percentage of trials with fixations toward the book, the red and the blue glass in the plausible and implausible conditions. The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

7.1.4. Discussion

In this experiment we manipulated the number of affordable items within the context of a concurrent visual scene and this allowed us to explore the extent to which the addition of a second (un-referred) glass might influence eye movements when the first glass had been moved to either the table (plausible location), or the lamp (implausible location). The findings showed that during 'the wine carefully into' there were more anticipatory eye movements to the table in the 'one glass' plausible condition than there were looks to the table in the 'two glasses' plausible condition. In contrast, there were no more anticipatory looks to the lamp in the 'one glass' implausible condition than there were looks to the lamp in the 'two glasses' implausible condition. During the final reference to 'the glass' there was no difference in the proportion of looks to the table/lamp when either one, or two glasses were depicted in the scene. There was no difference in the proportion of anticipatory looks to the un-referred red glass when the blue glass had either been moved to the table or the lamp. This was also the case during the final reference to 'the glass'.

These findings suggest that the inclusion of a second (affordable) glass did not weaken affordance-based constraints sufficiently to show an influence of plausibility. In other words, the addition of an alternative glass for 'pouring the wine into' did not lead participants to consider this option any more when the other glass was in an implausible location, compared to when it was in a plausible location[18]. One possible explanation for these results may be related to the notion of 'advantage of first mention' (Gernsbacher, 1990), whereby participants assume that the wine will be poured into the blue glass (regardless of plausibility), since the reference to this

[18] We also ran two other versions of this experiment (see appendices 10 and 11 for methods and results). Like in the current experiment we found no influence of plausibility when including a second affordable item in the visual scene.

particular glass renders it more prominent and accessible than the un-referred red glass. One way to get around this problem in future studies would be to simplify the structure of the narratives, making them similar to the sentences used by Kamide et al. (2003). In this experiment participants heard sentences such as *'the man will ride the motorbike',* or '*the girl will ride the carousel'* (whilst viewing a scene depicting a man, a girl, a motorbike, and a carousel). Upon hearing 'the man will ride…' participants anticipated and looked to the motorbike, whereas when hearing 'the girl will ride…' there were more anticipatory looks to the carousel. In this example both the motorbike and the carousel could afford to be ridden, but in contrast to the current experiment, anticipatory eye movements were measured before any of these items were directly referred to. The findings showed that in scenarios where two different items are equally compatible with the verb, participants relied on real-world knowledge in order to correctly anticipate the most plausible referent. However, taking into account the notion of scene constraints we might further speculate that if the carousel was the only item in the scene to afford riding, affordance-based constraints would lead participants to anticipate that both the girl, as well as the man would ride the carousel. Likewise, if the motorbike was the only item that could afford riding, affordance-based constraints should lead participants to anticipate that both the man and the girl would ride the motorbike. In other words, if the motorbike (or the carousel) is the only item that affords riding, the constraints of the visual scene should guide anticipatory eye movements to this item irrespective of who is going to ride it. In contrast, when both the motorbike and carousel are depicted in the visual scene participants rely on real-world knowledge, which then guides anticipatory eye movements to the most plausible upcoming reference.

7.2. EXPERIMENT 9

7.2.1. Linguistic constraints in a blank screen context

In chapter 6 we proposed that the constraints of a concurrent visual scene guide anticipatory eye movements to the event-specific location of the glass, even when this location is implausible. However, when we have to rely on our memory/internal representation of the visual scene these 'visual' constraints are weakened since our memory trace of the scene is inevitably weaker than the actual source of that trace. In the context of a blank screen anticipation is no longer strongly bound by the visual context and when this is the case we have to rely more on real-world knowledge when anticipating upcoming references. In experiment 8 we attempted to induce an influence of plausibility in the context of a concurrent visual scene by including a second affordable item. In contrast, in the current experiment we aimed to eradicate the influence of plausibility in the context of a blank screen, by providing a linguistic context in which there is only one affordable item. In other words, in the current experiment we tried to encourage anticipatory eye movements to the implausible locations by linguistically constraining the context so that it equalled the weighting of constraint provided by a concurrent visual scene.

Several studies have showed that constraints derived from linguistic information (alongside real-world knowledge) are able to influence language comprehension in the absence of a concurrent visual context (e.g. Altmann, 1999; Boland et al., 1995; McRae, Ferretti & Amyote, 1997; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985; Vu, Kellas, Petersen & Metcalf, 2003). Such studies suggest that when visual information is not available we rely on linguistic information when having to anticipate upcoming references. The previous studies presented in this thesis show that participants anticipate implausible locations in the context of a concurrent visual scene, but not in the context of a blank screen. These findings suggest that there is a quantitative difference in how

we process and anticipate upcoming discourse when language refers to something outside our visual environment, as opposed to when language refers to entities within our immediate visual proximity. Rayner et al. (2004) noted that their reading study showed a later influence of plausibility compared to visual world studies by Altmann and Kamide (1999) and Kamide et al. (2003). They proposed that this difference might be related to the stronger level of constraints afforded by a concurrent visual scene, which made it easier for participants to anticipate the most plausible referent amongst a limited set of options. In contrast, during the reading study the number of possible upcoming references was not restricted by a visual context and as such participants were less able to anticipate specific upcoming referents based on their likelihood.

In experiment 2 the glass was the only item in the concurrent visual scene that could afford 'pouring into' and we proposed that this is why participants anticipated its location regardless of plausibility. However, in the absence of a visual context we may require a more constraining linguistic context in order to achieve the same weighting of constraints provided by a concurrent visual scene. In experiment 9 we attempted to linguistically restrict the number of affordable items within the context of a blank screen so that it matched the constraint provided by a concurrent visual scene. In this experiment participants would hear a short narrative like this one below:

> The woman wanted some wine, *but she could only find one glass*. She was desperate for a drink so she put the glass on the table/lamp. Then, she picked up the bottle, and poured the wine carefully into the glass.

If the difference in anticipatory eye movements is related to the weighting of constraints provided by the presence or absence of a visual context, we would expect this more constraining linguistic context (that unambiguously restricts the number of affordable items in the context of a blank screen) to guide anticipatory eye movements to the appropriate, yet implausible location of the glass.

In other words, by constraining the context of the narrative so that there is only one possible glass we would expect to see a similar pattern of eye movements to that observed in experiment 2 (concurrent screen) where participants anticipated the plausible location of the glass (table), as much as they anticipated the implausible location of the glass (lamp).

7.2.2. Method

Participants

Thirty-two students from the University of York participated in this study, receiving either course credit or a payment of £4. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

Twenty-four experimental scenes (see figure 7.9.) were matched with two conditions. Due to the increased length of the narrative we decided to use a quadrant design (displaying one item in each of the four quadrants of the display), as opposed to arranging the items within a visual scene like in the ones used in the previous experiments. We hoped that this design would increase the proportion of eye movements in the context of a blank screen, thereby preventing a floor effect due to the length of the narrative. The positions of the items representing the plausible locations (e.g. the table), the implausible locations (e.g. the lamp), the location of the item referred to after the plausible/implausible locations (e.g. the bottle), and the distractor locations (e.g. the bookcase) were counterbalanced to ensure that the location of each item varied across the full set of experimental scenes. In conditions 1 and 2 the first sentence always explained why the target item was the only one of its kind in the 'world' within which the event could unfold. In other words, the linguistic constraints of the context implied that the introduced item was the only accessible item of its kind. In the

second sentence the target item was described as being moved to either a plausible, or an implausible location. Like in the previous experiments the third and final sentence was the same for both conditions (see appendices 13 and 14 for a full list of the experimental sentences and images). In addition to the experimental trials, 48 sentence-picture pairs were included as fillers (see appendix 15 for an example of the filler items).

1. The woman wanted some wine, but she could only find one glass. She was desperate for a drink so she put the glass on the table. Then, she picked up the bottle, and poured the wine carefully into the glass.

   (Plausible)

2. The woman wanted some wine, but she could only find one glass. She was desperate for a drink so she put the glass on the lamp. Then, she picked up the bottle, and poured the wine carefully into the glass.

   (Implausible)

Figure 7.9. Example of one of the visual scenes paired with sentences: 1) and 2).

Procedure

The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye. The procedure for this experiment was the same as in the previous blank screen experiments – the visual scenes were presented for five seconds and then replaced by a grey screen. The onset of the auditory stimuli occurred one second after the scene had been removed and the trials ended 19 seconds after the audio onset. As such, each trial lasted a total of 25 seconds.

Analysis

We defined four identically sized regions of interest within each scene; one area corresponded to the plausible location of the moved glass (the table), another corresponded to the implausible location of the moved glass (the lamp), and a third area corresponded to the location of the item referred to after the plausible/implausible locations (the bottle). Finally, we included a fourth area that corresponded to an un-referred and unrelated distractor item (the dress). This item was included instead of the target item (the glass) since we needed to provide a 'baseline' for the proportion of looks to an un-named item[19]. This design allowed us to compare the proportion of looks to previously named locations (plausible and

[19] In experiment 1 the target item (e.g. the glass) was always depicted in the visual scene and the results showed that participants anticipated the glass when this was described as having been moved to a plausible location (e.g. the table), but not when it was described as having been moved to an implausible location (e.g. the lamp). During the final reference to the glass participants looked to the event-specific location of the glass regardless of plausibility. In experiment 7 we removed the target items from the visual scene and the results from this experiment replicated the plausibility effect seen in experiment 1. This suggests that the exclusion of the target items has no influence on the effect of plausibility observed in experiment 1. On the basis of these findings the absence of the target items in the current experiment should have no implications when relating the pattern of eye movements from the current experiment to that observed in experiment 1.

implausible) with the proportion of looks to an un-referred item, thereby ensuring that participants' eye movements to the table/lamp were not due to levels of chance. Participants' eye movements were examined during certain time points in the spoken sentences with the percentage of saccades as the dependent measure. As in the previous experiments these critical time points occurred during 'the wine carefully into' (anticipatory eye movements), and during the final noun phrase 'the glass'.

To begin with, we compared the proportion of saccades to the region of the table (plausible locations) and the region of the lamp (implausible locations). This allowed us to see if the plausibility of these locations had any influence on eye movements. Secondly, we compared looks to the region of the table/lamp with looks to the region of the dress (distractor). This provided information about whether the proportion of saccades to the plausible and implausible locations was higher than the 'baseline' proportion of looks to an unnamed distractor item.

7.2.3. Results

Plausible vs. implausible

During 'the wine carefully into' there were no more anticipatory looks toward the previous location of the table in the plausible condition (the woman will move the glass onto the table…) than there were looks to the previous location of the lamp in the implausible condition (the woman will move the glass onto the lamp…) (*t1* (31) < 1) (*t2* (23) < 1). Similarly, during 'the glass' there were no more looks towards the previous location of the table in the plausible condition than there were looks to the previous location of the lamp in the implausible condition (*t1* (31) = -1.635, *p* > .05) (*t2* (23) = -1.844, p > .05) (see figures 7.10., 7.11., and table 7.3.). This is different to the pattern of eye movements observed in experiments 1 and 7, which both showed a higher proportion of anticipatory eye movements to the plausible locations compared to the implausible locations.

Plausible – table vs. dress

During 'the wine carefully into' there were more anticipatory looks toward the previous location of the table in the plausible condition than there were looks to the previous location of the dress (un-referred distractor) (*t1* (31) = 2.900, $p < .01$) (*t2* (23) = -2.986, $p < .01$). Similarly, during 'the glass' there were more looks towards the previous location of the table in the plausible condition than there were looks to the previous location of the dress (*t1* (31) = 3.566, $p < .01$) (*t2* (23) = -4.252, $p < .001$).

Implausible – lamp vs. dress

During 'the wine carefully into' there were more anticipatory looks toward the previous location of the lamp in the implausible condition than there were looks to the previous location of the dress (un-referred distractor) (*t1* (31) = 2.690, $p < .05$) (*t2* (23) = -3.728, $p < .01$). During 'the glass' there were more looks towards the previous location of the lamp in the implausible condition than there were looks to the previous location of the dress (*t1* (31) = 2.619, $p < .05$) (*t2* (23) = -2.807, $p < .05$). These differences suggest that the proportion of looks to the table and the lamp were not due to levels of chance. In other words, during the critical part of the narrative participants were more likely to look at the table/lamp than the unnamed distractor.

Time	table (pl.)	lamp (impl.)	dress (pl.)	dress (impl.)
the wine carefully into	20	20	11	11
the glass	15	12	6	5

Table 7.3. Percentage of trials with looks to the table, lamp, and dress (distractor) during "the wine carefully into the glass". Percentages calculated from the total number of trials.

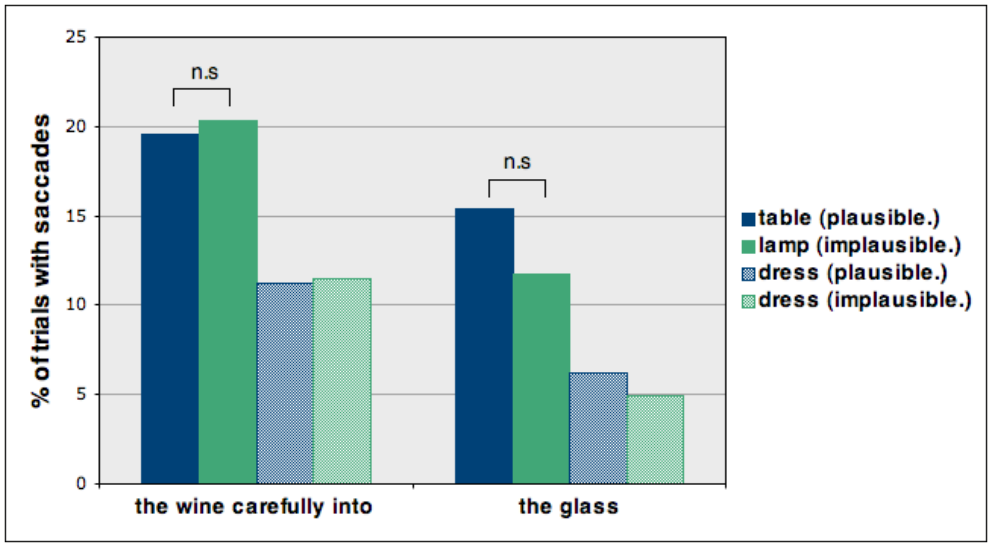


Figure 7.10. Looks to the table (plausible location), lamp (implausible location) and the dress (distractor) during “the wine carefully into the glass”.

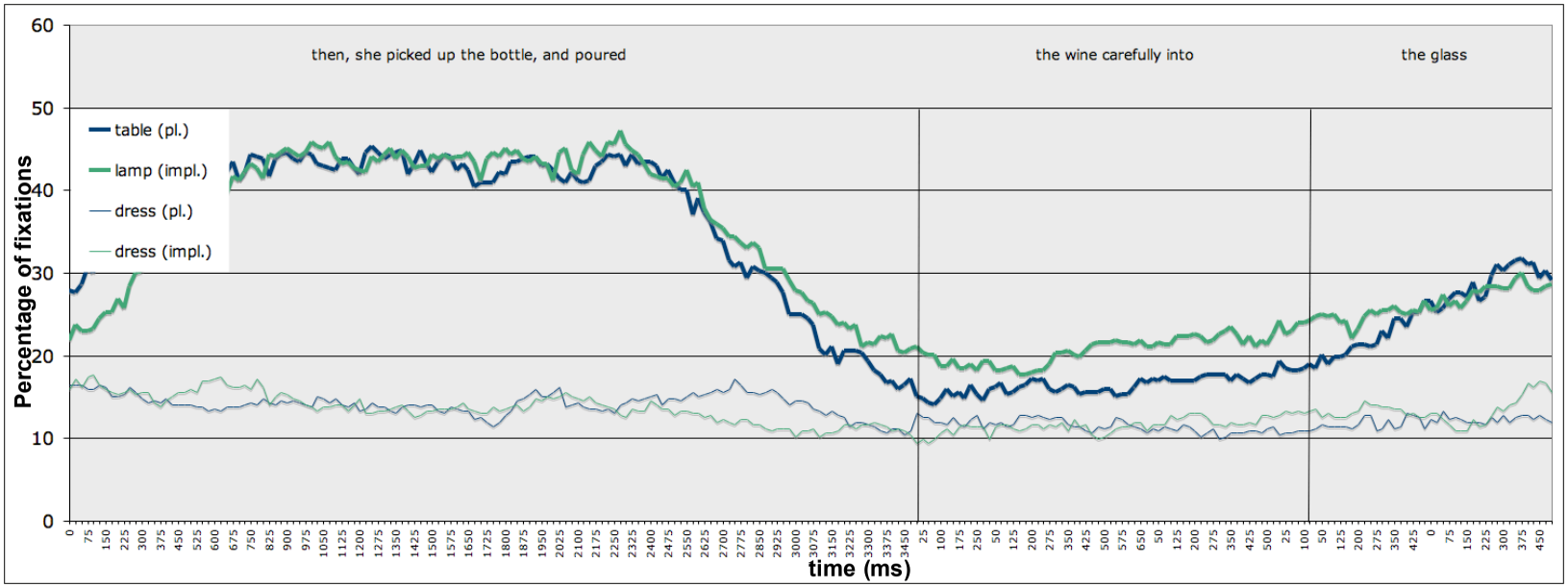


Fig 7.11. Percentage of trials with fixations toward the previous region of the table (plausible location), the lamp (implausible location) and the dress (distractor). The percentages show the proportion of trials on which participants fixated on each region of interest during *'then, she picked up the bottle, and poured the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

7.2.4. Discussion

In the current experiment we linguistically restricted the number of affordable items within the context of a blank screen so that it matched the constraint provided a concurrent visual scene. This design allowed us to explore the extent to which the linguistic constraints within a narrative would be able to guide anticipatory eye movements to implausible locations within the context of a blank screen. The findings showed that during 'the wine carefully into' there were as many anticipatory eye movements to the previous region of the lamp (implausible location) when the glass was described as having been moved there, as there were looks to the previous region of the table (plausible location) when the glass was described as having been moved there. During the final reference to 'the glass' there was similarly no difference in the proportion of looks to the plausible and implausible locations. This is different to the pattern of eye movements observed in experiments 1 and 7, which both showed a higher proportion of anticipatory eye movements to the plausible locations compared to the implausible locations. Together, these findings suggest that when the narrative restricts the number of affordable items, the linguistic context is able to guide anticipatory eye movements to the appropriate, yet implausible location of the glass.

While the current experiment provided a more constraining linguistic context, we also changed the style of the visual images (from scenes to quadrant) and this change of design could have lead to potential problems. In experiments 1 and 7 participants were presented with visual scenes that were subsequently removed before the onset of language. In contrast, due to the increased length of the narrative in the current experiment we decided to use a quadrant design as we hoped that this would increase the proportion of eye movements during the latter part of the spoken narrative, thereby preventing a floor effect. While the quadrant design did increase the proportion of eye movements, it could also appear to be the case that

participants anticipated the implausible locations as much as the plausible locations simply because the quadrant design made it easier to remember the regions of the implausible locations. However, going back to experiments 1 and 7 we found that during the final reference to ‘the glass’ participants looked as much to the region of the lamp as they looked to the region of the table – they just didn’t anticipate the region of the lamp. If the plausibility effect was due to difficulties with recalling the implausible location we would have expected to see fewer looks to the lamp compared to the table both during ‘the wine carefully into’ as well as during ‘the glass’.

Another potential problem with the quadrant design is that the proportion of looks to the plausible and implausible locations could have been due to level of chance. In other words, if there were no difference in the proportion of looks to the plausible/implausible locations and the unnamed distractor this would suggest that participants’ eye movements were not related to the spoken language and this might explain why there were no more anticipatory looks to the plausible locations, compared to the implausible locations. However, when comparing the proportion of looks to the plausible and implausible locations with looks to the un-named distractor we find that during the critical part of the sentence (the wine carefully into the glass) there were always significantly more looks to both the table and the lamp than there were looks to the distractor. This difference suggests that the proportion of looks to the table and the lamp were not due to levels of chance. Furthermore, going back to experiments 1 and 7 we find that during ‘the wine carefully into’ there were more looks to the plausible locations compared to the distractors, but no difference in looks to the implausible locations and the distractors. In contrast, during the final reference to ‘the glass’ there were more looks to both the plausible and implausible locations than there were looks to the distractors. This shows that when the linguistic context does not explicitly restrict the number of affordable items, participants anticipate the plausible

locations more than the distractors. However, they do not anticipate the implausible locations any more than they anticipate the locations of the un-named distractors.

In line with previous studies (e.g. Altmann, 1999; Schwanenflugel & LaCount, 1988; Vu, Kellas, Petersen & Metcalf, 2003) the data from the current experiment demonstrates that in the absence of a concurrent visual context we rely on linguistic information in order to anticipate upcoming references. The findings from the previous studies suggest that there is a quantitative difference in how we process and anticipate upcoming discourse depending on whether the visual scene is available, or removed before the onset of spoken language. During language comprehension our anticipation of upcoming references is constrained by the visual context. However, in the absence of a visual context these constraints are weakened, encouraging us to rely more on real-world knowledge when anticipating upcoming references. In the concurrent visual scene the glass is the only item that affords 'pouring into' and this is why we anticipate its location whether it is plausible or not. In the context of a blank screen on the other hand, anticipation is no longer strongly bound to the visual context and therefore upcoming references are much less specific. The current experiment shows that a linguistic context, which unambiguously restricts the number of affordable items, is able to guide anticipatory eye movements to the appropriate, yet implausible location of the glass. In other words, when we constrain the context of the narrative so that there is only one possible glass we find a similar pattern of eye movements to that observed in experiment 2 (concurrent screen) where participants anticipated the plausible location of the glass (table), as much as they anticipated the implausible location of the glass (lamp). This suggests that in the absence of a visual context we require a more constraining linguistic context in order to achieve the same weighting of constraints provided by a concurrent visual scene.

7.2.5. Summary

In experiment 8 we attempted to encourage an influence of plausibility similar to that observed in the context of a blank screen. The findings showed that the inclusion of a second (affordable) glass did not weaken affordance-based constraints sufficiently to show an influence of plausibility. In other words, the addition of an alternative glass for 'pouring the wine into' did not lead participants to consider this option any more when the other glass was in an implausible location, compared to when it was in a plausible location.

In experiment 9 we attempted to increase the weighting of constraint (in the context of a blank screen) by linguistically restricting the number of affordable items within the context of the narrative. By doing so, we hoped to linguistically create an equal weighting of constraint to that provided by a concurrent visual context. The findings showed that in contrast to experiments 1 and 7, there were no more anticipatory eye movements to the previous region of the table (plausible location) than there were looks to the previous region of the lamp (implausible location). This shows that when spoken language unambiguously restricts the number of affordable items, the linguistic context is able to guide anticipatory eye movements to the appropriate, yet implausible location of the glass. In other words, when we constrain the context of the narrative so that there is only one possible glass we find a similar pattern of eye movements to that observed in experiment 2 where the visual scene was presented alongside spoken language. This suggests that in the absence of a visual context we require a more constraining linguistic context in order to achieve the same weighting of constraints provided by a concurrent visual scene. In the final chapter we provide an overview of the experimental findings and discuss the implications of how differently weighted constraints are able to influence our anticipation of upcoming events.

CHAPTER 8

IMPLICATIONS & CONCLUSIONS

8.1. Background & aims

In everyday life we regularly use language to refer to items and people within our immediate proximity and several studies have shown that during language comprehension we very rapidly establish reference to objects within our visual environment (e.g. Allopenna et al., 1998; Cooper, 1974; Eberhard et al., 1995; Sedivy et al., 1999). Furthermore, as language unfolds, spoken information serves to constrain the potential number of upcoming referents within the visual scene and this makes it easier to anticipate the most plausible role-fillers within a described event (e.g. Altmann & Mirkovic, 2009; Tanenhaus et al., 1995 Rayner et al., 2004). However, while language is often used to refer to items within our visual proximity, we may also use language to refer to items, people, and events when these are absent from our visual context. When language refers to something outside our visual environment the anticipatory activation of appropriate representations is no longer bound so strongly to the visual context, but is now constrained purely by language. When this is the case the nature of the upcoming reference is much less specific and in such situations, any number of plausible role-fillers may be referred to.

The research presented in this thesis aimed to explore the mapping of language onto internal representations of visually available scenes, as well as previously viewed scenes. Firstly, we were interested in how event-plausibility is able to influence our internal representations of described events and secondly, how these representations might be modulated by the nature of the visual context (as present or absent).

8.2. Summary of experimental findings

In experiments 1 and 2 participants listened to spoken narratives where an item was either moved to a plausible or an implausible location within the visual scene (e.g. "*the woman will move the glass onto the table/lamp. Then she will pick up the bottle, and pour the wine carefully into the glass*"). In experiment 1 (blank screen) we found that during 'the wine carefully into' participants only anticipated plausible locations. When items had been moved to implausible locations there was little or no evidence of anticipatory eye movements. However, in the context of a concurrent visual scene (experiment 2) participants anticipated both the plausible and implausible locations. In both experiments there was no influence of plausibility during the final reference to the glass. In other words, once the glass was directly referred to participants had no difficulty retrieving its location regardless of whether this was plausible or implausible. This suggests that the difference in anticipatory eye movements between the concurrent and blank screen experiments is not specifically related to problems retrieving the implausible locations, but rather associated with being able to retrieve the most likely candidate (the glass) into which the pouring will take place. As such, the effect of plausibility could be a reflection of the weaker constraint drawn from a previous visual scene (compared to the constraint drawn from a concurrent visual scene), which then allows real-world knowledge to exert a greater influence on the probabilistic

activation of likely candidates.

Experiments 3 (blank screen) and 4 (concurrent screen) explored the extent to which this pattern of eye movements could be generalised and extended to contextual plausibility. Instead of moving objects to different locations we now moved contextually plausible (e.g. a cat in a kitchen scene) and implausible items (e.g. a penguin in a kitchen scene) to the same location. We found a similar pattern of eye movements to that observed in experiments 1 and 2. These findings suggest that the influence of plausibility observed in experiment 1 was not exclusively related to plausibility of location, but may be extended to the contextual plausibility of the referent itself.

In experiment 5 we attempted to uncover why plausibility only affects anticipatory eye movements in the context of a blank screen. More specifically, we investigated the extent to which a higher proportion of initial eye movements to the plausible locations (in experiment 1) could have made these locations more accessible and therefore easier to anticipate later in the narrative. The findings showed that while the proportion of initial looks made the implausible regions more accessible later in the narrative, this was not the case for the plausible locations. The data further indicated that participants who initially looked to the previous region of the lamp returned as much to that location as participants who had initially looked to the previous region of the books. Taken together the data suggests that the higher proportion of anticipatory eye movements to the plausible locations (observed in experiment 1) is not solely dependent on the higher proportion of early looks to these locations.

In experiments 6 and 7 we explored the implications of introducing an item purely through language (for example, the woman will move *a* glass onto the table/lamp). The results showed a similar pattern of eye movements to experiment 1 (where the glass was first presented in the visual scene, the visual scene was removed, and then the glass was referred to by spoken language). In other words,

participants anticipated the location of the glass when it had been moved to a plausible location, but not when it had been moved to an implausible location. Like in experiment 1, there was no influence of plausibility during the final reference to the glass. In other words, once the glass was directly referred to participants had no difficulty retrieving its location regardless of whether this was plausible or implausible. These findings suggest that a purely linguistic representation of the glass did not increase the proportion of anticipatory eye movements to the implausible locations.

In chapter 6 we outlined an account explaining why plausibility affects anticipatory eye movements in the context of a blank screen, but not in the context of a concurrent visual scene. We proposed that there is a quantitative difference in how we process and anticipate upcoming discourse depending on whether the visual scene is available, or removed immediately before the onset of spoken language. In other words, the constraint that is drawn from a previous visual scene is weaker than the constraint drawn from a concurrent visual scene and this allows real-world knowledge to exert a greater influence on the probabilistic activation of likely candidates. This probabilistic approach means that the underlying mechanism is the same, but that quantitative differences in the strength of constraints leads to what appear to be qualitative differences in behaviour.

During language comprehension we assume that participants in an event will be drawn from our visual environment (e.g. Altmann & Mirkovic, 2009; Cooper, 1974; Tanenhaus et al., 1995). In the visual scene the referred target is the only item that affords the described action and this is why we anticipate its location regardless of plausibility. In other words, if participants assume that the upcoming reference will be drawn from within the visual scene and the glass is the only item within the scene that can afford ‘pouring into’, the visual context constrains them to anticipate the event-specific location of the glass, even when this location is implausible. However, in the context of a blank screen participants have to rely on a visual

memory, or internal representation of the scene. As such, the constraints derived from the visual scene are weakened since anticipation is no longer so strongly bound by the visual context. Rather, in the absence of a visual scene participants were further influenced by real-world knowledge and this could have rendered the upcoming reference much less specific.

The final two experiments explored the extent to which both concurrent scene constraints and linguistic constraints are able to guide anticipatory eye movements to appropriate (yet implausible) locations. In experiment 8 we attempted to *decrease* the weighting of constraints derived from object-affordances (in a concurrent visual scene) by providing an alternative option for 'pouring the wine into'. The findings showed that the inclusion of a second (affordable) item did not weaken affordance-based constraints sufficiently to show an influence of plausibility. In contrast, experiment 9 attempted to *increase* the weighting of constraints (in the context of a blank screen) by linguistically restricting the number of affordable items within the context of the narrative (e.g. the woman really wanted some wine, but she could only find one glass…). By doing so, we hoped to linguistically equalise the constraints provided by a concurrent and a blank screen context. In contrast to experiments 1 and 7, there were as many anticipatory eye movements to the implausible locations as there were looks to the plausible locations. This shows that when spoken language unambiguously restricts the number of affordable items, the linguistic context is able to guide anticipatory eye movements to the appropriate, yet implausible location of the glass.

Taken together, the findings suggest that that the absence/presence of a visual context provides a quantitative difference in how we process and anticipate implausible events. In the absence of a visual context we require a more constraining linguistic context in order to achieve the same degree of constraints provided by a concurrent visual scene. Thus, our memory, or internal

representation of a visual scene, is not always an equal substitute for the scene itself, since different constraints come into play.

8.3. Theoretical implications

A number of studies have shown that when language refers to the current communicative situation we very rapidly establish reference to objects within our visual environment (e.g. Allopenna et al. 1998; Altmann, 2011; Cooper, 1974; Eberhard et al., 1995; Salverda & Altmann, 2011; Sedivy et al., 1999). During language comprehension we often rely on real-world knowledge and experience in order to anticipate upcoming events before they unfold within a narrative (e.g. Kamide, 2008; Kamide et al., 2003; Matsuki et al., 2011; McRae & Matsuki, 2009; Van Berkum et al. 2005) and visual world studies have shown that semantic information derived from language allows us to anticipate the most appropriate item within a visual context before being directly referred to (e.g. Altmann & Kamide, 1999, Kamide et al., 2003). In other words, when language unfolds, linguistic information (as well as real-world knowledge) serves to constrain the potential number of upcoming referents within our immediate visual environment. Such 'visual' constraints are particularly useful during anticipation, since they allow us to integrate linguistic information with the features and affordances of the items within our current visual context, thus making it easier to correctly anticipate the most likely upcoming referent.

While language often occurs in conjunction with a visual context, modular theories of language comprehension (e.g. Coltheart, 1999; Fodor, 1983) assume that early stages of language processing are guided solely by syntactic constraints, whereby a single grammatical interpretation is selected or ranked on the basis of the features of the unfolding linguistic composition (Frazier, 1987; Frazier & Clifton, 1996). According to this approach non-linguistic constraints such as real-world knowledge and visual information are not employed until later stages of processing. Consequently, research subscribing to

modular theories of language processing has largely neglected the importance of non-linguistic sources and the extent to which visual information is able to influence on-line language comprehension.

In contrast, constraint-based approaches (e.g. MacDonald, Pearlmutter & Seidenberg, 1994; Tanenhaus, Spivey-Knowlton & Hanna, 2000; Tanenhaus & Trueswell, 1995) assume that people continuously evaluate multiple syntactic options, employing a range of constraints, which may be derived from both linguistic and non-linguistic sources of information. According to these accounts, any available and relevant information is continuously assimilated in order to determine the most suitable interpretation of language as it unfolds. In this way constraints derived from both syntactic, as well as non-linguistic information, are able to influence language comprehension from the earliest stages of processing. However, while constraint-based approaches promotes the notion of non-linguistic constraints, studies have yet to demonstrate how information derived from a concurrent visual scene is weighted against that from our memory, or internal representation of that same scene (but see Altmann & Kamide, 2009, for an initial exploration of such issues).

Several studies have explored how visual information can influence comprehension when language refers to items within a concurrent visual context (e.g. Chambers et al., 2002, 2004; Knoeferle, Crocker, Scheepers & Pickering, 2005; Magliano, Dijkstra & Zwaan, 1996; Tanenhaus et al., 1995). Most notably, Knoeferle and Crocker (2006) investigated the influence and interaction of real-world knowledge and visual information by measuring participants' eye movements as they listened to descriptions of accompanying visual scenes. The data showed that when verb-related information was stereotypically related to one of the characters (depicted performing an unrelated action) and visually related to another non-stereotypical character (depicted performing the described action) participants looked more to the non-stereotypical character when

anticipating the upcoming agent. This suggests that participants relied more on information associated with the depicted event, than knowledge associated with the description of the event. In contrast, when neither of the agents were depicted performing the described action participants looked more to the stereotypical character, in this case relying more on real-world knowledge derived from the description of the event. These findings may further be related to the notion of 'concurrent scene constraint', whereby the visual depiction of the characters takes 'precedence' over participants' real-world knowledge of the stereotypical actions of the characters. On the other hand, when the characters' actions are not constrained by the visual scene, participants draw on real-world knowledge in order to anticipate which of the characters language refers to.

On the basis of these findings Knoeferle and Crocker (2006) proposed an account, which emphasises the coordinated interplay of the information derived from the visual scene, linguistic information and real-world knowledge during language comprehension. According to this account language comprehension occurs in an incremental fashion, whereby we interpret and anticipate upcoming language gradually as each word unfolds. Thereby, unfolding language guides eye movements to named and anticipated items, or events within a visual scene, which is then able to influence comprehension. According to this account, online comprehension relies on multiple sources of available information, which each serves to constrain our interpretation of upcoming language. Furthermore, in the context of a concurrent visual context this tight temporal interaction between different types of information leads us to visually inspect named items/events at the earliest possible moment (during anticipation before the item is named) and this increases their salience leading to a greater reliance on information derived from the depicted event, compared to real-world knowledge associated with the described event.

The experimental data presented in this thesis corresponds well

with constraint-based theories, as well as the coordinated interplay account of situated language comprehension, showing that we rely on multiple sources of information to constrain comprehension. More specifically, we have demonstrated how a concurrent visual context is able to influence language comprehension by constraining our anticipation of upcoming references. This suggests that when language refers to the current communicative situation we rely on a combination of visual information (and the affordances of items within this visual context), linguistic information and real-world knowledge in order to most efficiently constrain our interpretation of upcoming language. Our findings further suggests that there is a quantitative difference in how we process and anticipate implausible events, whereby in the absence of a visual context we require a more constraining linguistic context in order to achieve the same weighting of constraints provided by a concurrent visual scene. This in turn provides more detailed information in terms of the weighting of constraints provided by a concurrent, or recently encountered visual context and the extent to which we rely on visual information to constrain our interpretation of upcoming language.

8.4. Implications for reading studies

The discrepancy between the weighting of constraints in the absence/presence of a visual context may also be useful when looking at the influence of plausibility on reading. For instance, Rayner et al. (2004) found an earlier pattern of disruption in response to anomalous sentences compared to implausible sentences, which only showed disruption at a later point in time. In contrast, visual world studies by Altmann and Kamide (1999) and Kamide et al. (2003) have showed early plausibility effects, specifically in terms of anticipating upcoming references. Rayner et al. proposed that this difference between experiments might be related to the higher degree of constraints afforded by the visual context, which made it easier for participants to anticipate the most plausible referents within

the visual scene. On the other hand, participants in the reading experiment were not constrained by a visual context and therefore less able to anticipate upcoming referents based on their likelihood.

This difference in constraints allows us to speculate that a visual world version of the study by Rayner et al. would lead to an earlier influence of plausibility, compared to the effect observed during reading. For example, if participants listened to the same plausible and implausible sentences used in the reading study (e.g. John used a knife/an axe to chop the large carrots for dinner) whilst viewing a visual scene depicting a man, a knife, an axe, some carrots, we would expect the visual constraints of the scene to result in anticipatory looks to the carrots (upon hearing the verb 'chop') regardless of whether John used a knife or an axe to chop them. This is similar to the findings by Rayner et al., which showed no difference in reading times between the plausible and implausible sentences – in this study only anomalous sentences (e.g. John used a pump to inflate the large carrots for dinner) lead to an early pattern of disruption. However, if we were to include a pile of wood in the visual scene we might expect to see an early influence of plausibility (similar to that observed by Kamide et al., 2003) with more anticipatory looks to the carrots when 'John used a knife to chop…' and more looks to the pile of wood when 'John used an axe to chop…', compared to when 'John used a knife to chop…'. This scenario (although hypothetical) illustrates how the constraints provided by a concurrent visual context might influence comprehension differently compared to when reading about the same event in the absence of any visual information. It further illustrates how a more complex visual environment (in which several affordable items are present) might encourage us to rely more on real-world knowledge (just like in the blank screen) in order to correctly determine the most likely outcome on an event.

In a broader context is it also worth considering how visually grounded representations of items within our concurrent visual

environment at times have to compete with linguistic representations of the same item (see Altmann & Kamide, 2009). In such circumstances the constraints from our visual environment may interfere with comprehension since we have to keep track of separate (visual and linguistic) representations of an item. Often during descriptive events the location of the characters may change regularly, objects are moved to new locations, different events occur, new characters are introduced and so on (Zwaan, Magliano & Graesser, 1995b; Zwaan & Madden, 2004) and such changes require us to continuously update our mental representations of the described situation as it unfolds, in order to gain an up-to-date full representation of the described event. When relating a described event to a static visual scene we must interpret whether the scene corresponds to the beginning, the middle, or the end of that event. Using the previous example of 'moving the glass', the depicted glass on the floor corresponds to the beginning of the described event (e.g. the woman will move the glass…). However, once the glass is described as having been moved to the table the visual scene no longer corresponds to this part of the event. In this subsequent part of the narrative the depicted version of the glass (on the floor) interferes with our internal representation of the glass (described as being on the table) and this discrepancy might make it more difficult to update our mental representation of the unfolding event. In contrast, during reading there is no interference from a visual context and this should make event-representations more dynamic and therefore easier to update.

8.5. Conclusions

The studies presented in this thesis explored the influence of event-plausibility on internal representations of described events and how these representations may be modulated by the nature of the visual context. Our findings suggest that the conceptual representations activated during language processing in a concurrent visual context

are quantitatively different from those activated when the visual context to which that language applies is absent; our memory, or internal representation of a visual scene is not a surrogate for the scene itself since different constraints come into play.

While language is often used to refer to items, people, and events when these are absent from our visual context, we may also use language to refer to items within our visual proximity and the findings from our experiments highlight the importance of visual information for understanding and interpreting language in those situations where language does refer to our immediate visual context. However, in the absence of a visual context our anticipation is no longer restricted by visual information and in such situations we may require a more constraining linguistic context to attain the same degree of constraint as that provided by a concurrent visual scene.

These findings correspond well with constraint-based theories, as well as the coordinated interplay account of situated language comprehension, by demonstrating our reliance on multiple sources of available information when interpreting unfolding events. Specifically, the studies presented in this thesis provide more detailed information in terms of the weighting of constraints provided by a concurrent, or recently encountered visual context and the extent to which we rely on visual information to constrain our interpretation of upcoming language.

Appendix 1: Experimental sentences – experiments 1 & 2

1.
   a. The woman will put the bread onto the plate. Then, she will take some butter, and spread it sluggishly onto the bread.

   b. The woman decided not to put the bread onto the plate. She will take some butter, and spread it sluggishly onto the bread.

   c. The woman will put the bread onto the bottle. Then, she will take some butter, and spread it sluggishly onto the bread.

2.
   a. The office worker will move the dustbin right in front of the fan. Then, he will grab the can, and chuck it violently into the dustbin.

   b. The office worker has just moved the dustbin away from the fan. Now, he will grab the can, and chuck it violently into the dustbin.

   c. The office worker will move the dustbin onto the printer. Then, he will grab the can, and chuck it violently into the dustbin.

3.
   a. The woman will lift the pet carrier onto the table. Then, she will take hold of the cat, and put it carefully into the pet carrier.

   b. The woman has just lifted the pet carrier down from the table. Now, she will take hold of the cat, and put it carefully into the pet carrier.

   c. The woman will lift the pet carrier onto the flowers. Then, she will take hold of the cat, and put it carefully into the pet carrier.

4.
   a. The businessman will put the computer onto the desk. Then, he will pick up the disk, and insert it gently into the computer.

   b. The businessman was unable to put the computer onto the desk. But, he will pick up the disk, and insert it gently into the computer.

   c. The businessman will put the computer onto the briefcase. Then, he will pick up the disk, and insert it gently into the computer.

5.
   a. The secretary will move the folder right next to the lamp. Then, she will look at the documents, and file them efficiently in the folder.

   b. The secretary has moved the folder away from the lamp. She will look at the documents, and file them efficiently in the folder.

   c. The secretary will move the folder onto the sandwich. Then, she will look at the documents, and file them efficiently in the folder.

6.
   a. The woman will place the pan on the cooker. Then, she will reach for the bowl, and transfer the egg swiftly into the pan.

   b. The woman will soon place the pan on the cooker. But first, she will reach for the bowl, and transfer the egg swiftly into the pan.

   c. The woman will place the pan on the pepper mill. Then, she will reach for the bowl, and transfer the egg swiftly into the pan.

7.
   a. The housewife will move the vase onto the sideboard. Then, she will pick up the flowers, and arrange them delicately in the vase.

   b. The housewife is too tired to move the vase onto the sideboard. But, she will pick up the flowers, and arrange them delicately in the vase.

   c. The housewife will move the vase onto the Hoover. Then, she will pick up the flowers, and arrange them delicately in the vase.

8.
   a. The chef will place the pan on the cooker. Then, she will notice the lid, and place it quickly onto the pan.

   b. The chef will check the pan and the cooker. Then, she will notice the lid, and place it quickly onto the pan.

   c. The chef will place the pan on the potatoes. Then, she will notice the lid, and place it quickly onto the pan.

9.
   a. The woman will move the jewellery box right next to the coffee. Then, she will admire the necklace, and hide it quickly inside the jewellery box.

   b. The woman will examine the jewellery box as she drinks the coffee. Then, she will admire the necklace, and hide it quickly inside the jewellery box.

   c. The woman will move the jewellery box right onto the grapefruit. Then, she will admire the necklace, and hide it quickly inside the jewellery box.

10.
   a. The man will shift the box onto the worktop. Then, he will lift up the pizza, and put it carefully into the box.

   b. The man has just shifted the box off the worktop. Now, he will lift up the pizza, and put it carefully into the box.

   c. The man will shift the box onto the ice cream. Then, he will lift up the pizza, and put it carefully into the box.

11.
   a. The man will put the gramophone onto the sideboard. Then, he will clean the record, and place it carefully on the gramophone.

   b. The man will soon put the gramophone onto the sideboard. But first, he will clean the record, and place it carefully on the gramophone.

   c. The man will put the gramophone onto the candles. Then, he will clean the record, and place it carefully on the gramophone.

12.
   a. The girl will suspend the hanger on the rail. Then, she will reach for the shirt, and hang it cheerfully onto the hanger.

   b. The girl has taken the hanger off the rail. Now, she will reach for the shirt, and hang it cheerfully onto the hanger.

   c. The girl will suspend the hanger on the plant. Then, she will reach for the shirt, and hang it cheerfully onto the hanger.

13.
   a. The man will move the cup onto the table. Then, he will reach for the teapot, and pour the tea slowly into the cup.

   b. The man has taken the cup off the table. Now, he will reach for the teapot, and pour the tea slowly into the cup.

   c. The man will move the cup onto the figurine. Then, he will reach for the teapot, and pour the tea slowly into the cup.

14.
   a. The man will move the chair next to the girl. Then, he will lift up the teddy bear, and sit it affectionately on the chair.

   b. The man will look at the chair and then at the girl. Then, he will lift up the teddy bear, and sit it affectionately on the chair.

   c. The man will move the chair onto the Christmas tree. Then, he will lift up the teddy bear, and sit it affectionately on the chair.

15.
   a. The woman will move the mug onto the trolley. Then, she will reach for the bottle, and tip the water quickly into the mug.

   b. The woman has taken the mug off the trolley. Now, she will reach for the bottle, and tip the water quickly into the mug.

   c. The woman will move the mug onto the Wellies. Then, she will reach for the bottle, and tip the water quickly into the mug.

16.
   a. The woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   b. The woman is too lazy to put the glass onto the table. Instead, she will pick up the bottle, and pour the wine carefully into the glass.

   c. The women will put the glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

17.
   a. The boy will put the boat into the pool. Then, he will pick up the dummy, and place it cautiously inside the boat.

   b. The boy has just taken the boat out of the pool. Now, he will pick up the dummy, and place it cautiously inside the boat.

   c. The boy will put the boat on the clothesline. Then, he will pick up the dummy, and place it cautiously inside the boat.

18.
   a. The woman will move the suitcase onto the bed. Then, she will take the shoes, and put them hurriedly into the suitcase.

   b. The woman cannot be bothered to move the suitcase onto the bed. Instead, she will take the shoes, and put them hurriedly into the suitcase.

   c. The woman will move the suitcase onto the stereo. Then, she will take the shoes, and put them hurriedly into the suitcase.

19.
   a. The girl will shortly put the glass onto the tray. Then, she will grab the jug, and pour some lemonade attentively into the glass.

   b. The girl will shortly put the glass onto the tray. But first, she will grab the jug, and pour some lemonade attentively into the glass.

   c. The girl will shortly put the glass onto the mannequin. Then, she will grab the jug, and pour some lemonade attentively into the glass.

20.
   a. The man will lift the briefcase onto the desk. Then, he will reach for the notepad, and place it hesitantly inside the briefcase.

   b. The man refuses to lift the briefcase onto the desk. Rather, he will reach for the notepad, and place it hesitantly inside the briefcase.

   c. The man will lift the briefcase onto the water cooler. Then, he will reach for the notepad, and place it hesitantly inside the briefcase.

21.
   a. The girl will fasten the collar on the dog. Then, she will pick up the leash, and attach it cheerfully onto the collar.

   b. The girl will soon fasten the collar on the dog. But first, she will pick up the leash, and attach it cheerfully onto the collar.

   c. The girl will fasten the collar on the lamp. Then, she will pick up the leash, and attach it cheerfully onto the collar.

22.

a. The man will lift the toolbox onto the shelf. Then, he will pick up the hammer, and drop it sloppily into the toolbox.

b. The man doesn't have time to lift the toolbox onto the shelf. Instead, he will pick up the hammer, and drop it sloppily into the toolbox.

c. The man will lift the toolbox onto the drill. Then, he will pick up the hammer, and drop it sloppily into the toolbox.

23.

a. The man will put the basket onto the ladder. Then, he will pick some apples, and place them gently in the basket.

b. The man is too exhausted to put the basket onto the ladder he will pick some apples, and place them gently in the basket.

c. The man will put the basket onto the donkey. Then, he will pick some apples, and place them gently in the basket.

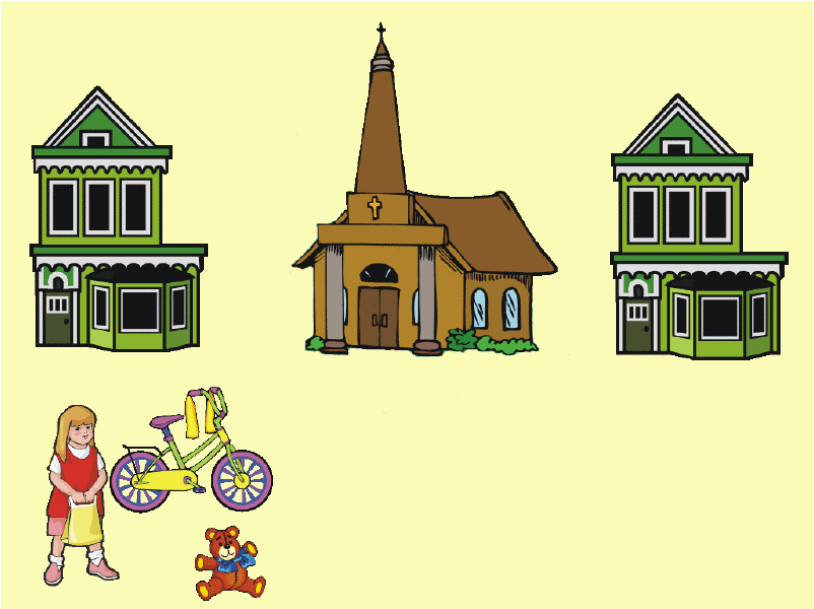
24.

a. The boy will put the jar onto the stool. Then, he will take the biscuits, and transfer them attentively into the jar.

b. The boy will soon put the jar onto the stool. But first, he will take the biscuits, and transfer them attentively into the jar.

c. The boy will put the jar onto the birdcage. Then, he will take the biscuits, and transfer them attentively into the jar.

Appendix 2: Example of filler items – experiments 1 & 2



a. The girl will push the bicycle past the church. And then, she will quickly go into the house.

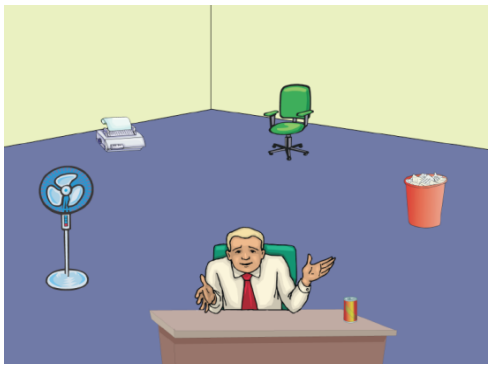
b. The girl will chuck the teddy past the church. And then, she will quickly go into the house.

c. The girl will skip past the church. And then, she will quickly go into the house.

d. The girl will look into her bag. And then, she will quickly go into the house.

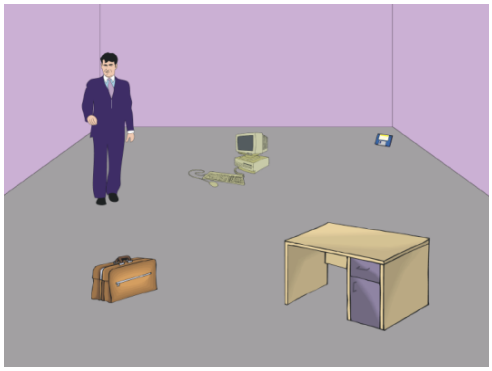
Appendix 3: Experimental scenes – experiments 1 & 2

01



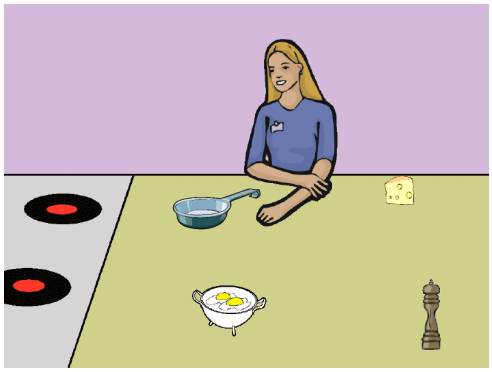
02

03

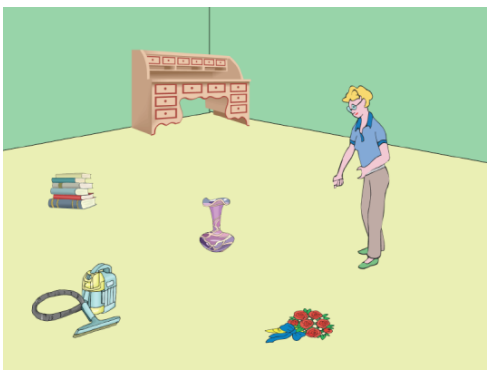


04

05

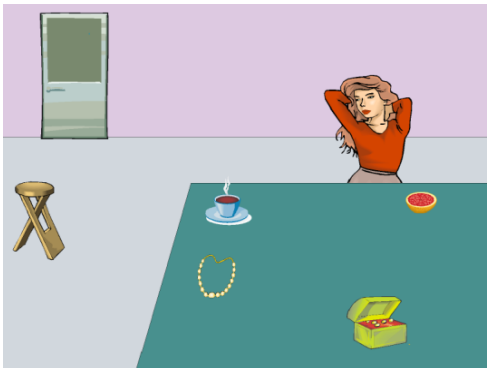


06

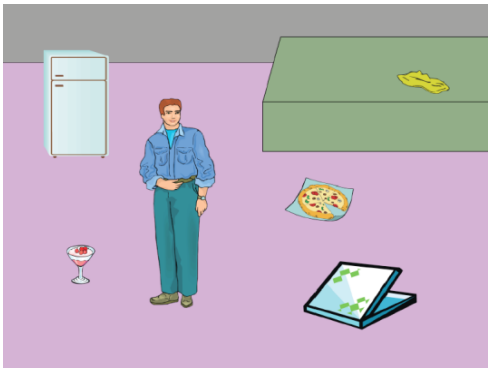


07

08

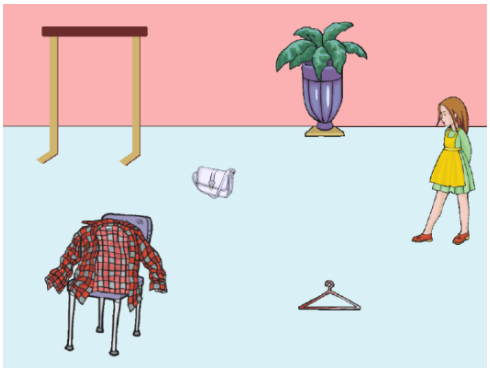


09

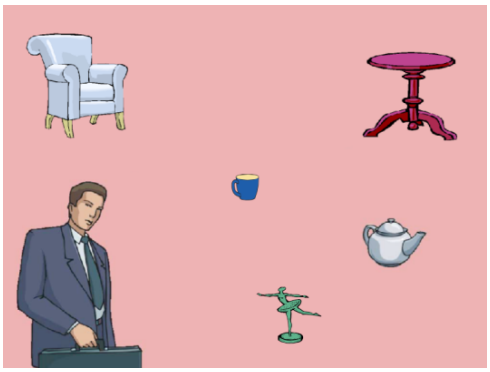


10

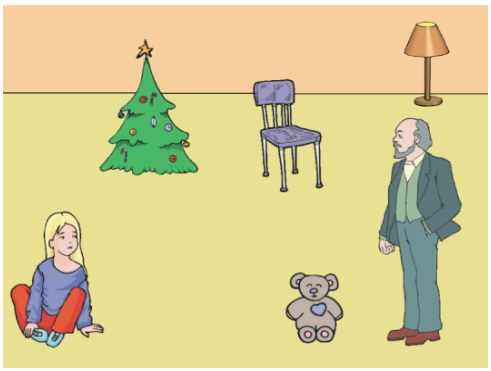
11



12

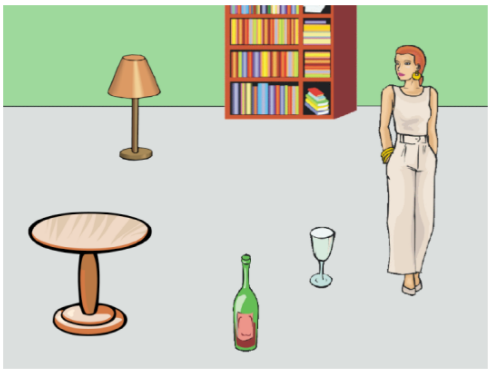


13



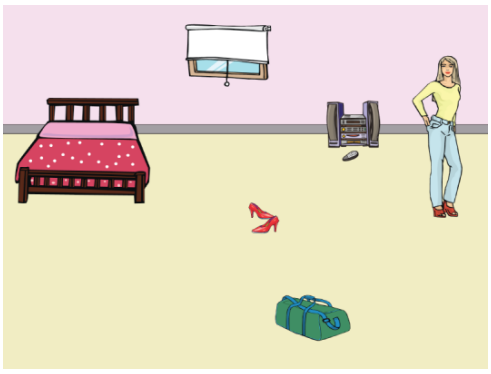
14

15

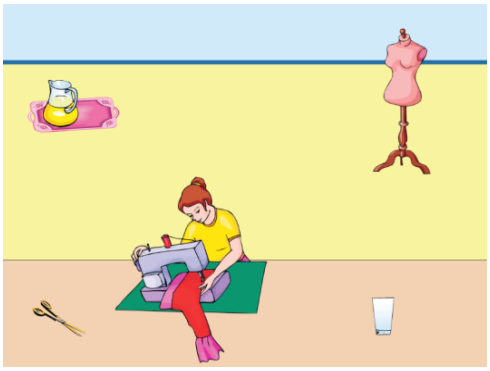


16

17

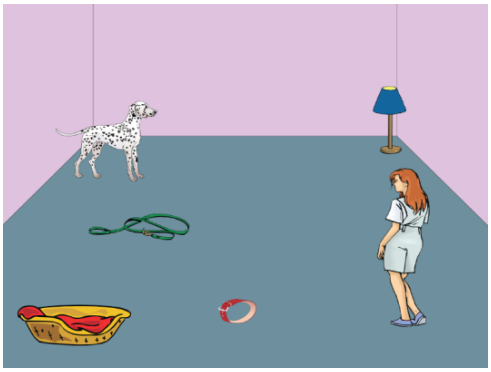


18



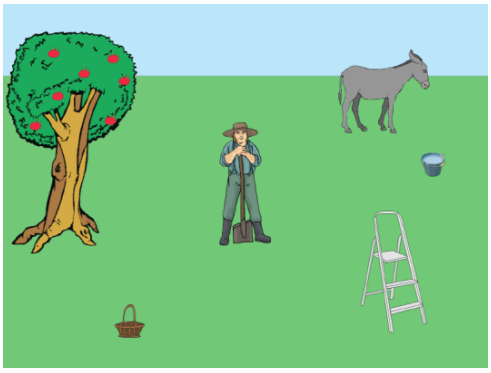
19

20

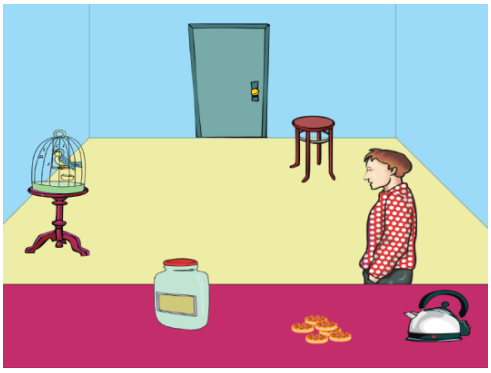


21

22



23



24

Appendix 4: Replicating the plausibility effect

Method

Participants

Thirty-two students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

Twenty-four experimental scenes were matched with two conditions. We used the same experimental scenes as in experiment 1, but since this design only consisted of two conditions (as opposed to the three conditions used in experiment 1) we increased the number of filler trials in order to equalise the distribution of implausible to plausible trials to that of experiment 1. As such, out of a total of 72 trials, twelve experimental trials described the target item as being moved to an implausible condition and twelve experimental trials described the target item as being moved to a plausible location. The remaining 48 trials were made up of filler items, which always described an item being moved to a plausible location (see appendix 7 for an example of the filler items).

1. The woman will move the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Plausible moved)

2. The woman will move the glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Implausible moved)

Procedure

The procedure for this experiment was the same as in experiment 1. The scenes were presented for five seconds and then replaced by a grey screen. The onset of the audio occurred one second after the scene had been removed and the trials finished 11 seconds after the audio onset. As such, each trial lasted a total of 17 seconds. The total duration of experiment was approximately 45 minutes. The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

We used the same regions of interest as in experiment 1. We compared the proportion of saccades to the region of the table (plausible locations) and the region of the lamp (implausible locations). This allowed us to see if the plausibility of these locations had any influence on eye movements.

Results

Plausible (table) vs. implausible (lamp)

During 'the wine carefully into' there were more anticipatory looks toward the previous location of the table in the plausible condition (the woman will move the glass onto the table), than there were looks to the previous location of the lamp in the implausible condition (the woman will move the glass onto the lamp) ($t1$ (31) = -2.642, $p < .05$) ($t2$ (23) = -2.584, $p < .05$). However, during 'the glass' there were no more looks toward the previous location of the table in the plausible condition than there were looks to the previous location of the lamp in the implausible condition ($t1$ (31) = -1.845, $p > .05$) ($t2$ (23) = -1.443, $p > .05$) (see figures 1., 2., and table 1.). This replicates the pattern of eye movements we observed in experiment 1.

Time	lamp (impl.)	table (pl.)	books (impl.)	books (pl.)
the wine carefully into	9	16	6	7
the glass	8	9	4	2

Table 1. Percentage of trials with looks to the previous location of the lamp, table and books (distractor) during “the wine carefully into the glass”. Percentages calculated from the total number of trials.

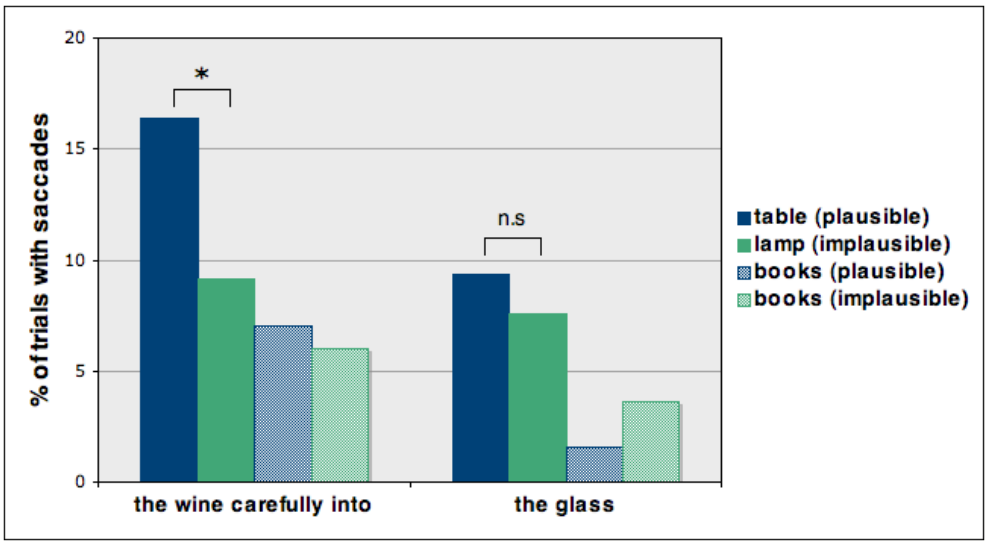


Fig. 1. Looks toward the lamp/table: during ‘the wine carefully into’, and during ‘the glass’.

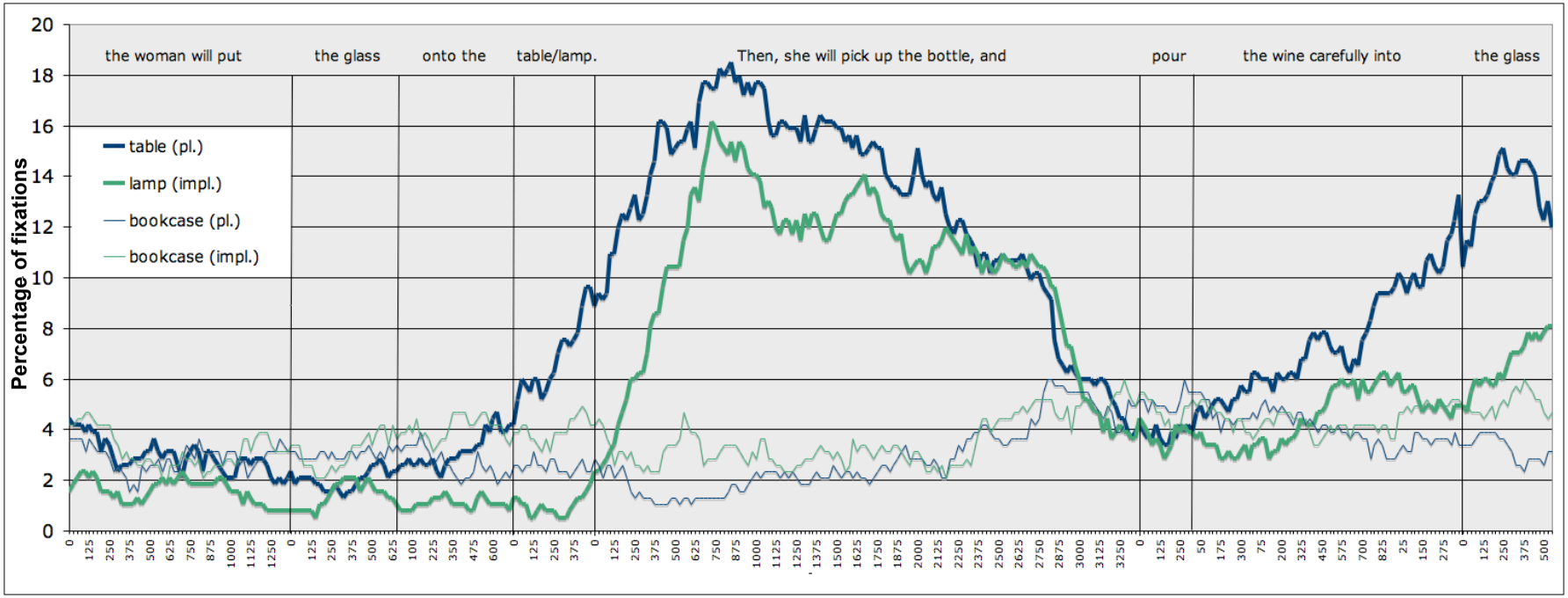


Fig 2. Percentage of trials with fixations toward the previous region of the table (plausible condition), the lamp (implausible condition) and the books (distractor). The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

Appendix 5: Experimental sentences – experiments 3 & 4

1.
   a. The woman will lift the cat onto the table. Then, she will quickly feed the cat.

   b. The woman will lift the penguin onto the table. Then, she will quickly feed the penguin.

2.
   a. The diver will put the shell into the boat. Then, he will carefully open the shell.

   b. The diver will put the umbrella into the boat. Then, he will carefully open the umbrella.

3.
   a. The student will move the pencil onto the desk. Then, he will casually sharpen the pencil.

   b. The student will move the axe onto the desk. Then, he will casually sharpen the axe.

4.
   a. The woman will place the wine bottle on the table. Then, she will hurriedly open the wine bottle.

   b. The woman will place the oxygen tank on the table. Then, she will hurriedly open the oxygen tank.

5.
   a. The girl will put the dog into the bathtub. Then, she will thoroughly wash the dog.

   b. The girl will put the pig into the bathtub. Then, she will thoroughly wash the pig.

6.
   a. The flight attendant will put the blanket onto the trolley. Then, she will cheerfully hand the girl the blanket.

   b. The flight attendant will put the rocket onto the trolley. Then, she will cheerfully hand the girl the rocket.

7.
   a. The lumberjack will put the chainsaw onto the tree stump. Then, he will rigorously clean the chainsaw.

   b. The lumberjack will put the sewing machine onto the tree stump. Then, he will rigorously clean the sewing machine.

8.
   a. The girl will lift the blender onto the table. Then, she will merrily turn on the blender.

   b. The girl will lift the leaf blower onto the table. Then, she will merrily turn on the leaf blower.

9.
   a. The girl will put the teddy onto the blanket. Then, she will gleefully photograph the teddy.

   b. The girl will put the coffee machine onto the blanket. Then, she will gleefully photograph the coffee machine.

10.
   a. The man will move the fish into the canoe. Then, he will swiftly clean the fish.

   b. The man will move the candelabra into the canoe. Then, he will swiftly clean the candelabra.

11.
   a. The boy will drop the mushroom into the basket. Then, he will carefully inspect the mushroom.

   b. The boy will drop the stapler into the basket. Then, he will carefully inspect the stapler.

12.
   a. The woman will move the handbag onto the dressing table. Then, she will slowly open the handbag.

   b. The woman will move the poison onto the dressing table. Then, she will slowly open the poison.

13.
   a. The man will move the watering can right next to the hose. Then, he will lazily gaze at the watering can.

   b. The man will move the floppy disk right next to the hose. Then, he will lazily gaze at the floppy disk.

14.
    a. The girl will put the present inside the rucksack. Then, she will sneakily look at the present.

    b. The girl will put the hand grenade inside the rucksack. Then, she will sneakily look at the hand grenade.

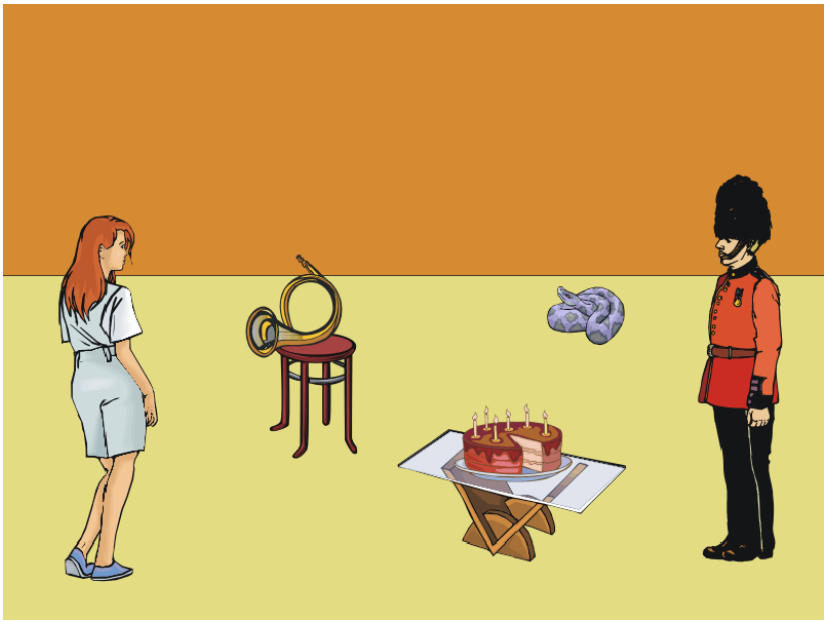
15.
    a. The woman will put the candle onto the table. Then, she will cautiously light the candle.

    b. The woman will put the dynamite onto the table. Then, she will cautiously light the dynamite.

16.
    a. The man will place the bottles right in front of the fireplace. Then, he will cautiously dust off the bottles.

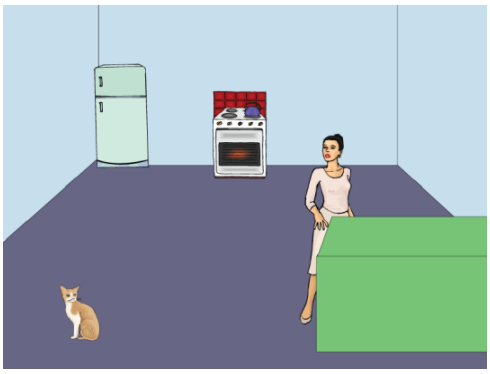
    b. The man will place the robot right in front of the fireplace. Then, he will cautiously dust off the robot.

Appendix 6: Example of filler items – experiments 3 & 4

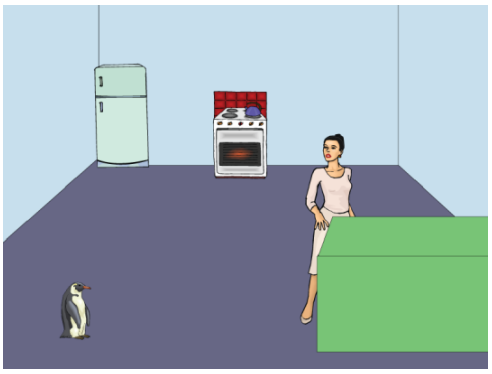


The guard will pick up the horn. Then, he will skilfully play a tune for the girl.

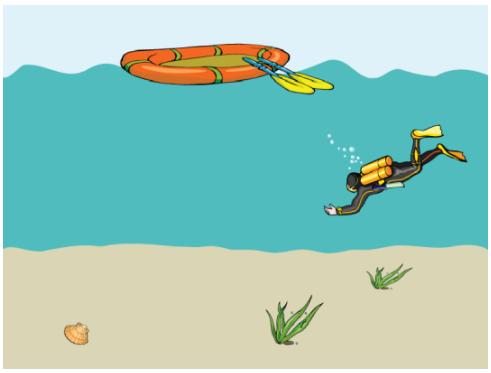
Appendix 7: Experimental scenes – experiments 3 & 4



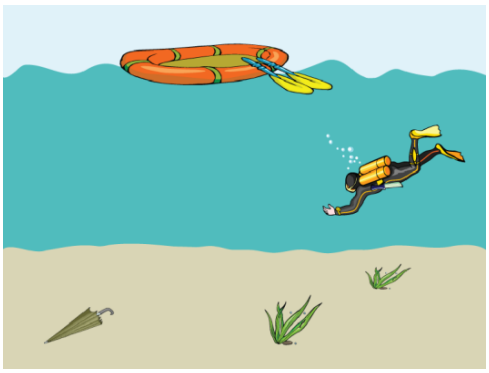
1a



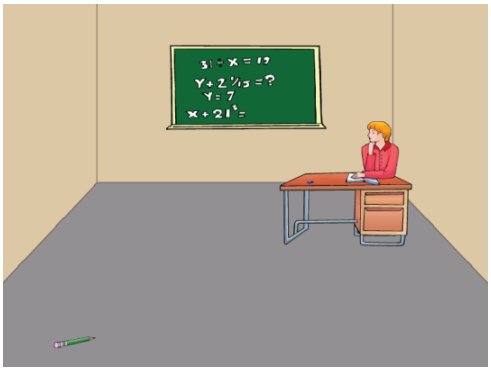
1b



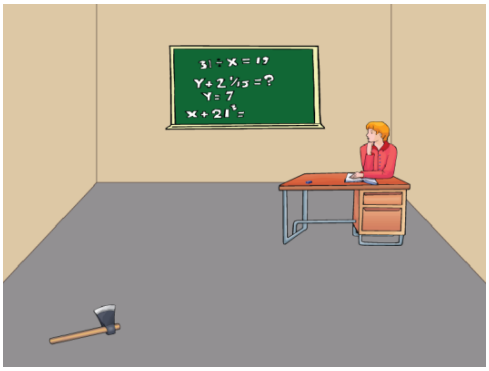
2a



2b



3a

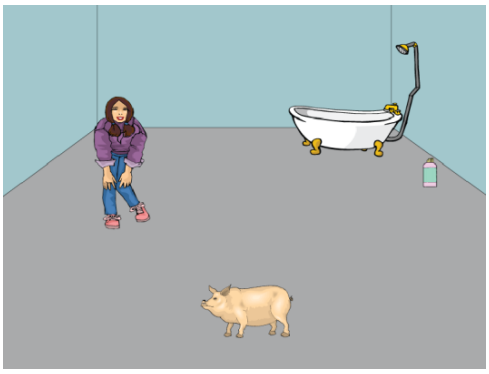


3b

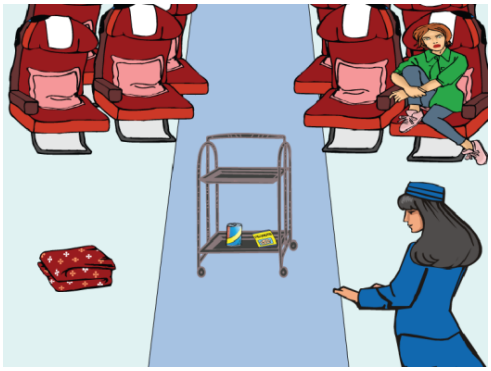
4a

4b

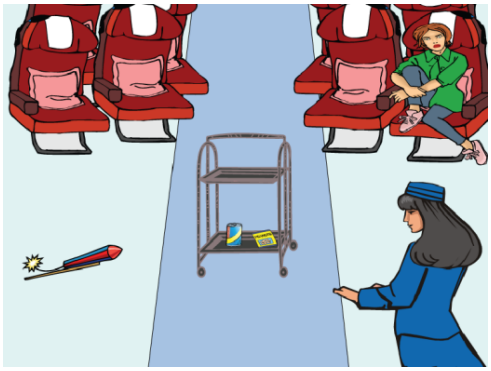
5a



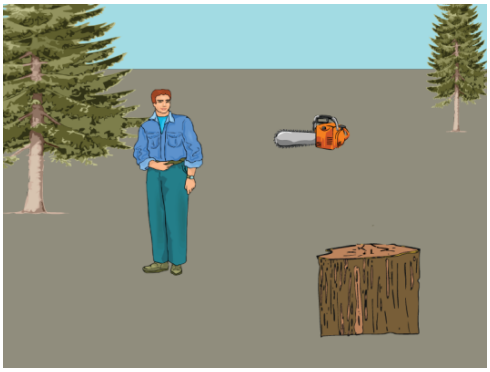
5b



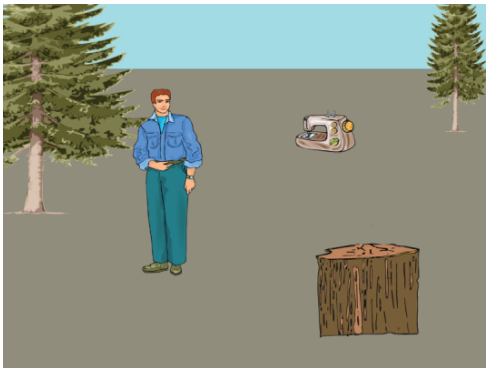
6a



6b



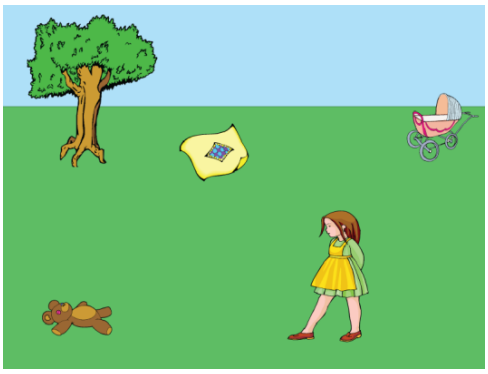
7a



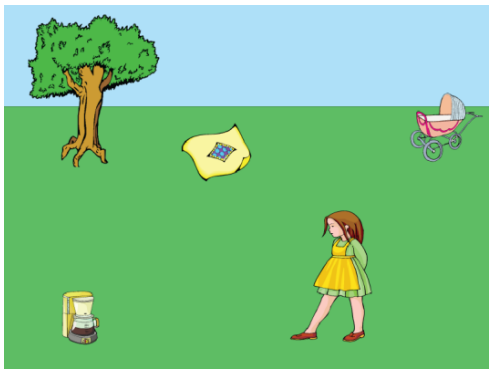
7b

8a

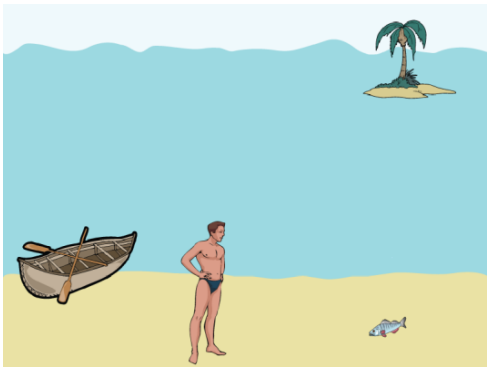
8b



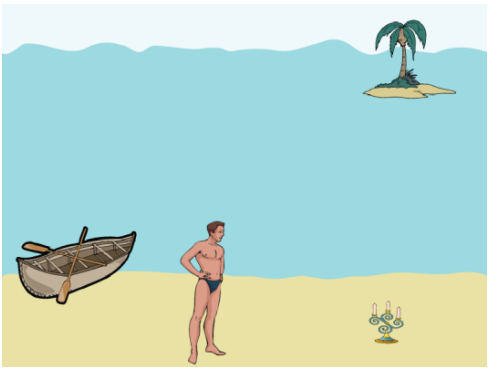
9a



9b



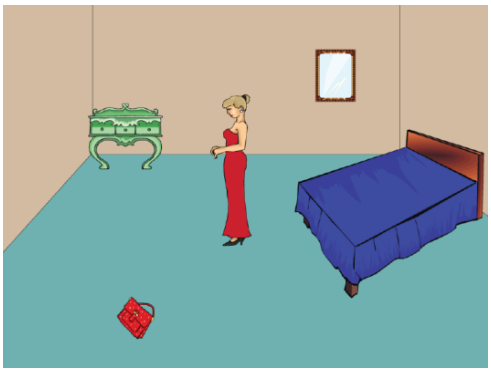
10a



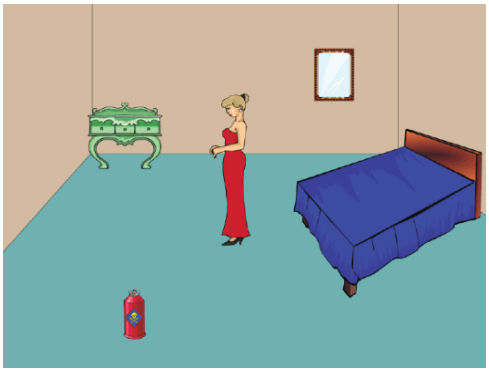
10b

11a

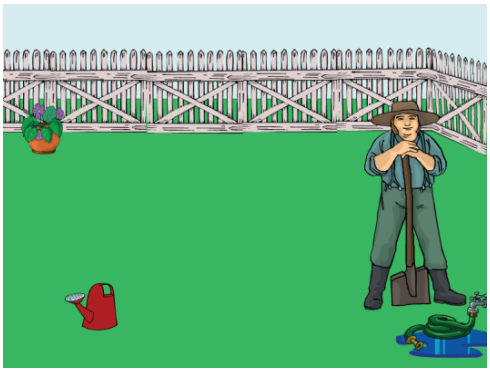
11b



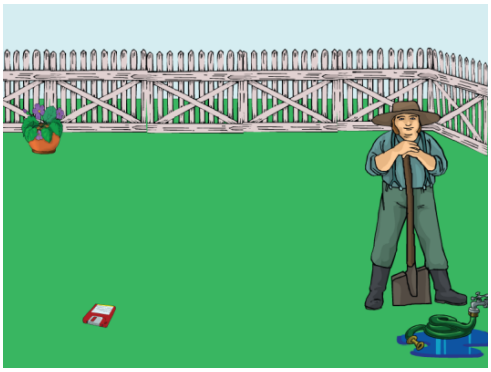
12a



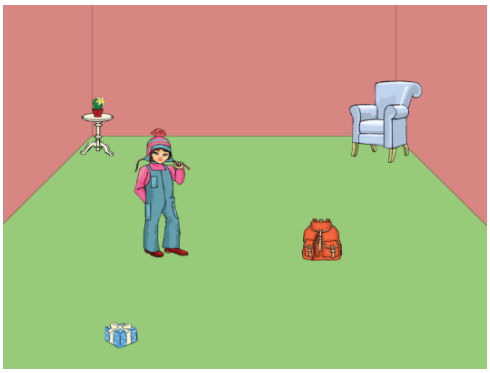
12b



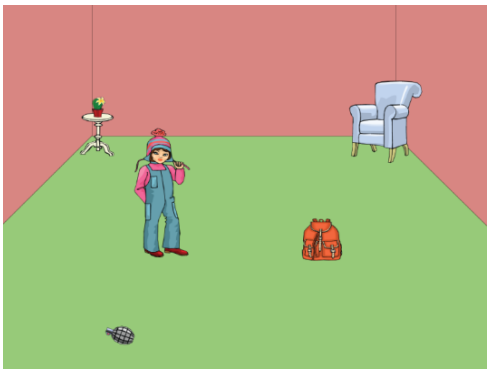
13a



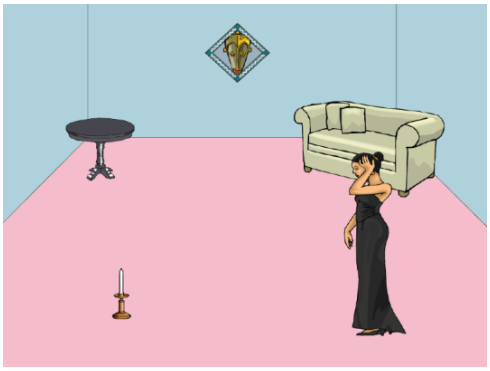
13b



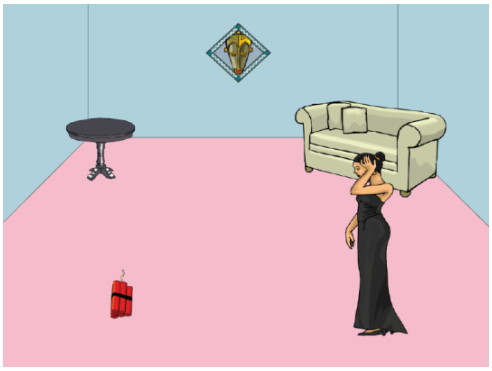
14a



14b

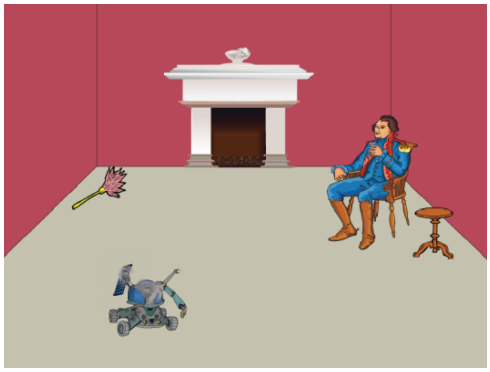


15a



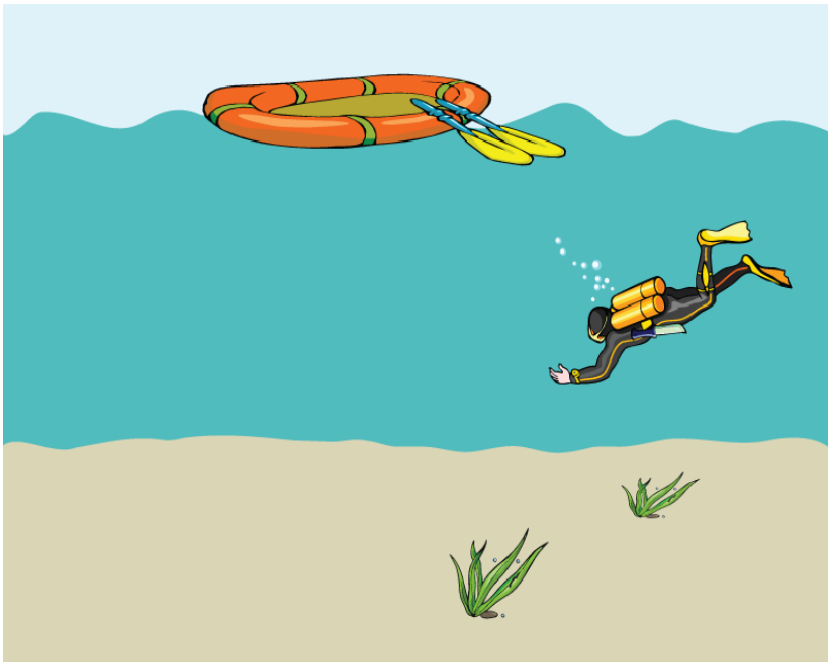
15b

16a



16b

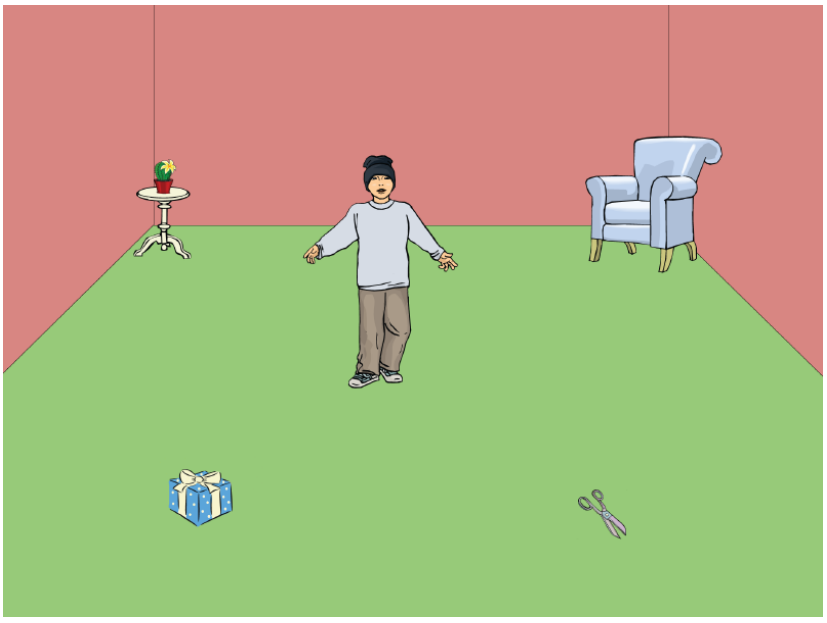
Appendix 8: Example of filler items – experiment 7



The diver will put a shell into the boat. Then, he will grab the knife and carefully open the shell.

Appendix 9: Example of filler items – experiment 8

The boy will kick the football into the goal. Then, he will take off his jacket, and cheerfully fetch the football.



The boy will lift the present onto the armchair. Then, he will grab the scissors, and excitedly unwrap the present.

Appendix 10: Experiment 8 (version 1)

Method

Participants

Forty-eight students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

Twenty-four experimental scenes (see figures 1. and 2.) were matched with three conditions. In conditions 1 and 2 we presented two affordable target items. For example, one blue glass and one red glass were depicted on the floor. In condition 3 we presented one affordable item (e.g. one glass on the floor). This design allowed us to compare looks to the table in the two glasses plausible condition with looks to the lamp in the two glasses implausible condition, as well as looks to the red glass (un-referred affordable item) between each condition. In addition to the experimental trials, 24 sentence-picture pairs were included as fillers (see appendix 2 for an example of the filler items).

1. The woman will move the blue glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Two glasses plausible)

2. The woman will move the blue glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Two glasses implausible)

3. The woman will move the glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the glass.

   (Original implausible)

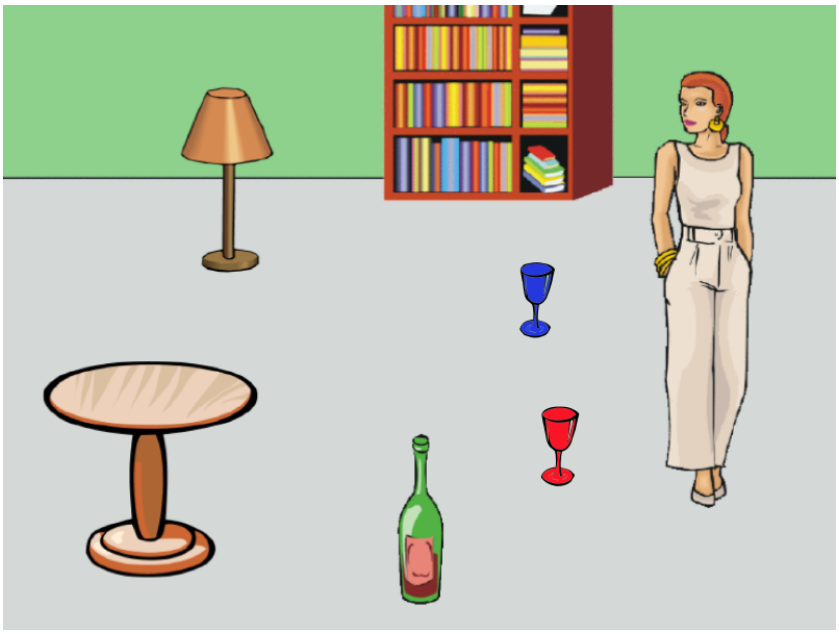


Figure 1. Example of one of the visual scenes paired with sentences: 1) and 2

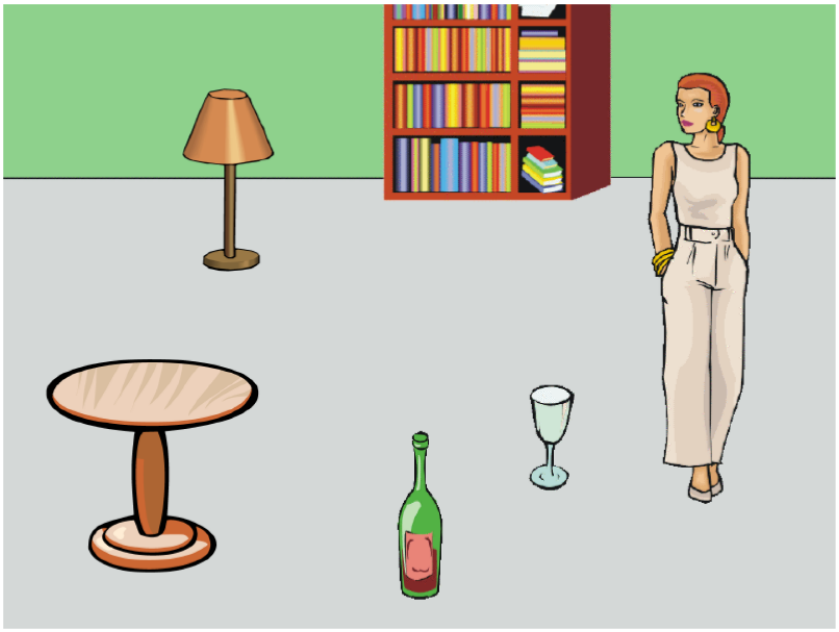


Figure 2. Example of one of the visual scenes paired with sentence: 3).

Procedure

The procedure for this experiment was the same as in the previous concurrent screen experiments – the visual scenes were presented for 1000 milliseconds, followed by the onset of the auditory stimuli. The trials ended 11 seconds after the audio onset, lasting a total of

12 seconds. The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

We defined four Regions of interest within each scene; one corresponding to the plausible location of the moved glass (e.g. the table), another corresponding to the implausible location of the moved glass (e.g. the lamp), and a third area corresponding to the original location of the moved glass (e.g. the blue glass). Finally, we included a fourth area that corresponded to the un-referred affordable item (e.g. the red glass). As the scenes remained onscreen throughout each of the trials, we defined the regions of interest according to the outline of each target object. As such, participants' eye movements had to be directed to one of the pixels occupied by each object within the scene. Participants' eye movements were examined during certain time points in the spoken sentences. As in the previous experiments these critical time points occurred during 'the wine carefully into' (anticipatory eye movements), and during the final noun phrase 'the glass'.

To begin with, we compared the proportion of saccades to the table/lamp when either one glass, or two glasses were presented in the visual scene. These comparisons allowed us to see if the inclusion of a second (un-referred) affordable item would be able to decrease the weighting of constraints and thereby influence eye movements to the plausible and implausible locations when the first glass had been moved there. Secondly, we compared the proportion of saccades to the region of the (un-referred) glass when the other glass had been moved to the table or the lamp. This allowed us to explore whether participants would anticipate and look to the alternative (un-referred) glass more when the first glass had been moved to an implausible locations, compared to a plausible location.

Results

Table/lamp – two glasses plausible vs. two glasses implausible

During 'the wine carefully into' there were no more anticipatory looks to the table in the 'two glasses' plausible condition than there were looks to the lamp in the 'two glasses' implausible condition (*t1* (47) < 1) (*t2* (23) < 1). Similarly, during the final reference to 'the glass' there were no more looks to the table in the 'two glasses' plausible condition than there were looks to the lamp in the 'two glasses' implausible condition (*t1* (47) < 1) (*t2* (23) < 1) (see figures 3., 4. and table 1.).

Lamp – one glass implausible vs. two glasses implausible

During 'the wine carefully into' there were no more anticipatory looks to the lamp in the 'one glass' implausible condition than there were looks to the lamp in the 'two glasses' implausible condition (*t1* (47) = -1.123, $p > .05$) (*t2* (23) = 1.606, $p > .05$). During the final reference to 'the glass' there were also no more looks to the lamp in the 'one glass' implausible condition than there were looks to the lamp in the 'two glasses' implausible condition (*t1* (47) < 1) (*t2* (23) = -1.588, $p > .05$) (see figures 3., 4. and table 1.).

Time	table (2 glasses)	lamp (2 glasses)	lamp (1 glass)
the wine carefully into	23	20	23
the glass	11	12	11

Table 1. Percentage of trials with looks to the table and the lamp during "the wine carefully into the glass". Percentages calculated from the total number of trials.

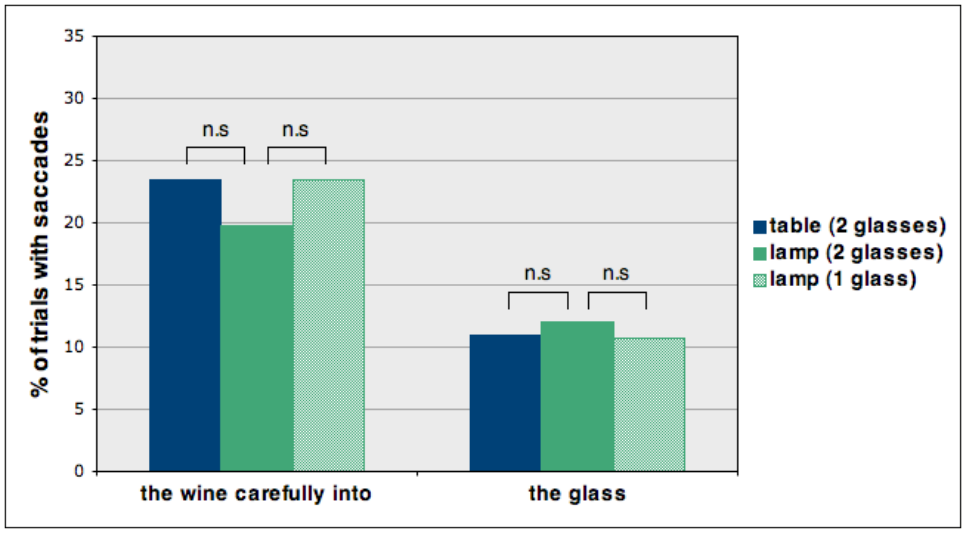


Figure 3. Looks to the table (plausible location) and the lamp (implausible location) during “the wine carefully into the glass”.

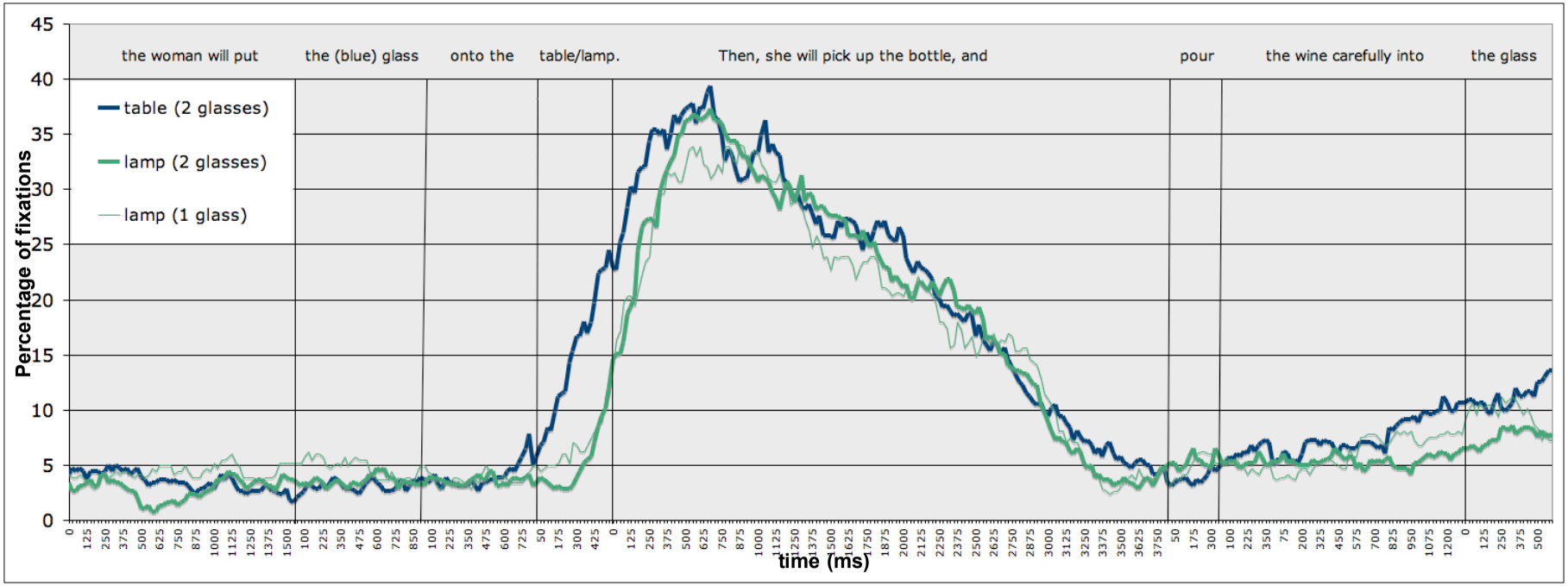


Fig 4. Percentage of trials with fixations toward the table and the lamp in the plausible and implausible (two glasses) conditions. The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

Red glass – two glasses plausible vs. two glasses implausible

During 'the wine carefully into' there were no more anticipatory looks to the un-referred red glass in the 'two glasses' plausible condition than there were looks to the un-referred red glass in the 'two glasses' implausible condition (*t1* (47) < 1) (*t2* (23) = -1.556, *p* > .05). Similarly, during the final reference to 'the glass' there was no difference in the proportion of looks to the un-referred red glass between these two conditions (*t1* (47) < 1) (*t2* (23) = -1.099, *p* > .05) (see figures 5., 6., and table 2.).

Time	red glass (pl.)	red glass (impl.)
the wine carefully into	23	20
the glass	11	12

Table 2. Percentage of trials with looks to the un-referred red glass during "the wine carefully into the glass". Percentages calculated from the total number of trials.

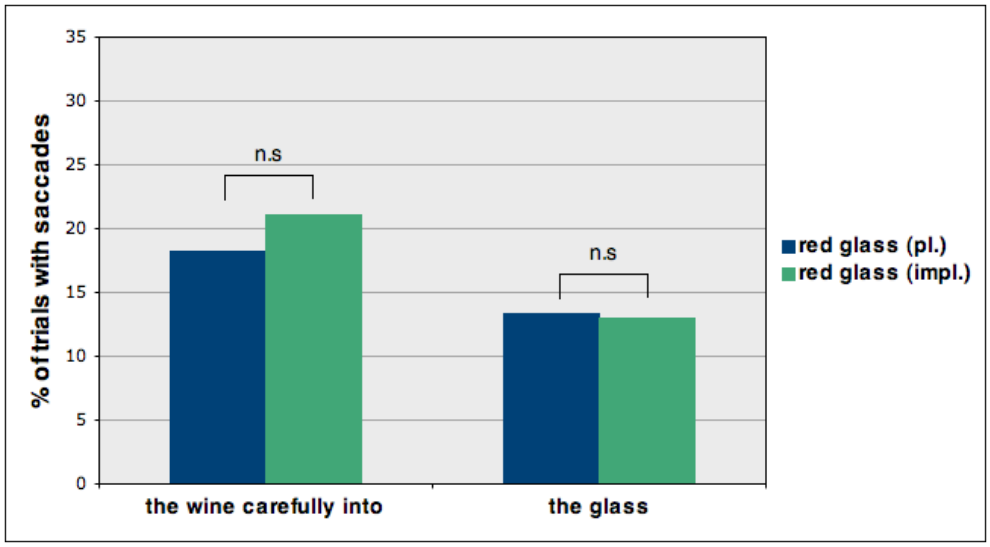


Figure 5. Looks to the red glass (un-referred affordable location) during "the wine carefully into the glass".

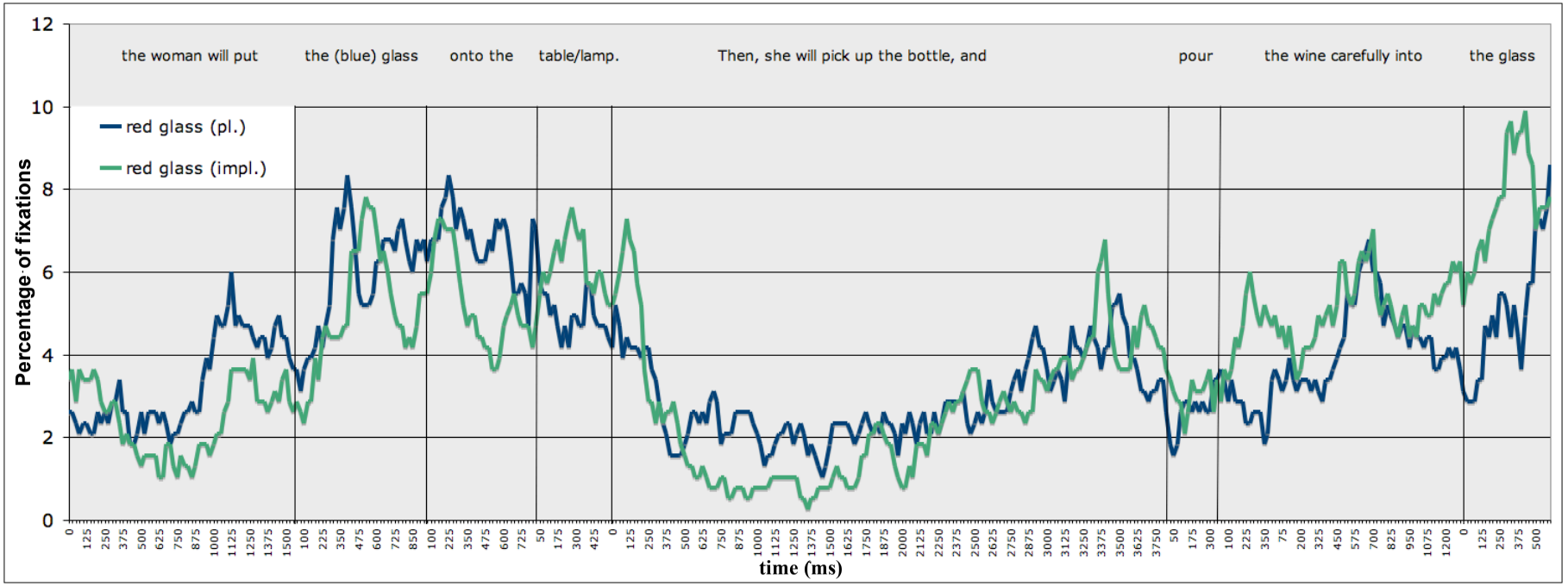


Fig 6. Percentage of trials with fixations toward the un-referred red glass in the plausible and implausible (two glasses) conditions. The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

Appendix 11: Experiment 8 (version 2)

Method

Participants

Thirty-two students from the University of York participated in this study, receiving either course credit or a payment of £3. All participants were native English speakers and had either uncorrected vision or corrected-to-normal vision.

Stimuli

Twenty-four experimental scenes (see figure 1.) were matched with two conditions. In both conditions we presented two affordable target items. For example, one blue glass and one red glass were depicted on the floor. In condition 1 the blue glass was described as being moved to the table (plausible location 1) and the red glass was described as being moved to the stool (plausible location 2). In this condition the first mentioned item was always referred to at the end of the narrative. In condition 2 the blue glass was described as being moved to the table (plausible location 1) and the red glass was described as being moved to the lamp (implausible location). In this condition the first mentioned item was subsequently referred to in 50% of the trials and the second mentioned item was referred to in 50% of the trials. This design allowed us to compare the proportion of anticipatory looks to the lamp in the two glasses implausible condition with the proportion of anticipatory looks to the table and the stool in the two glasses implausible condition. In addition to the experimental trials, 48 sentence-picture pairs were included as fillers (see appendix 12 for examples of the filler items). Twelve of the filler items referred to the first mentioned item (e.g. blue glass) at the end of the narrative and 36 of the filler items referred to the second mentioned item (e.g. red glass) at the end of the narrative. As such, 50% of the trials always referred to the first mentioned items and

50% of the trials always referred to the second mentioned item.

1. The woman will move the blue glass onto the table and the red glass onto the stool. Then, she will pick up the bottle, and pour the wine carefully into the blue glass.

   (Plausible)

2. The woman will move the blue glass onto the table and the red glass onto the lamp. Then, she will pick up the bottle, and pour the wine carefully into the blue/red glass.

   (Implausible)

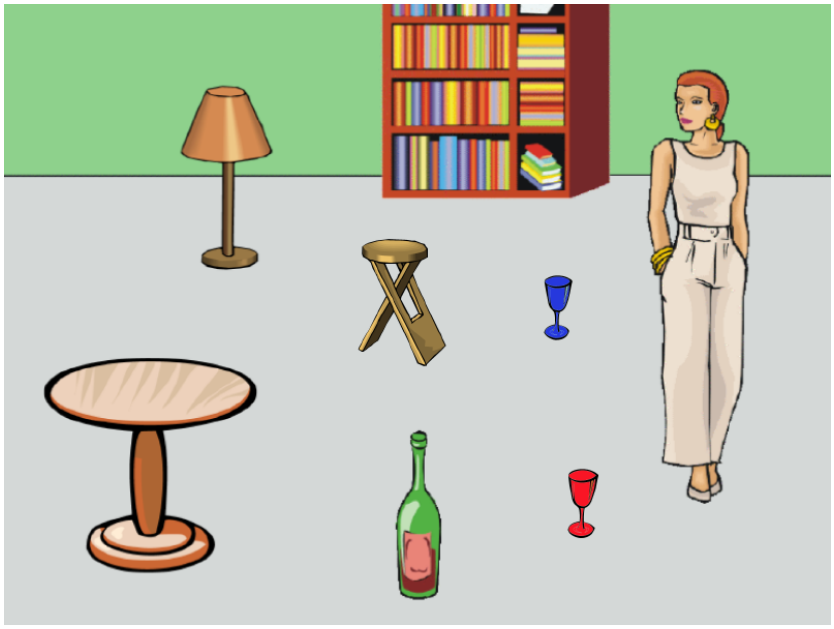


Figure 1. Example of one of the visual scenes paired with sentences: 1) and 2).

Procedure

The procedure for this experiment was the same as in the previous concurrent screen experiments – the visual scenes were presented for 1000 milliseconds, followed by the onset of the auditory stimuli. The trials ended 14 seconds after the audio onset, lasting a total of 15 seconds. The experiment was run on an Eyelink II head-mounted eye-tracker, which sampled at 250 Hz from the right eye.

Analysis

We defined five regions of interest within each scene; one corresponding to the plausible location of the moved blue glass (e.g. the table), another corresponding to the plausible location of the moved red glass (e.g. the stool), and a third area corresponding to the implausible location of the moved red glass (e.g. the lamp). We included a fourth area that corresponded to the original location of the blue glass and a fifth area corresponding to the original location of the red glass. As the scenes remained onscreen throughout each of the trials, we defined the regions of interest according to the outline of each target object. As such, participants' eye movements had to be directed to one of the pixels occupied by each object within the scene. Participants' eye movements were examined during certain time points in the spoken sentences. As we were solely interested in anticipatory eye movements, the critical time points occurred during 'the wine carefully into'.

To begin with, we compared the proportion of saccades to the table/lamp when the blue glass had been moved there. We also compared the proportion of saccades to the stool with the proportion of looks to the table/lamp. These comparisons allowed us to explore whether participants would anticipate and look to the stool (alternative location) more when the red glass had been moved to an implausible location, compared to a plausible location.

Results

Table vs. lamp

During 'the wine carefully into' there were no more anticipatory looks to the table in the plausible condition than there were looks to the lamp in the implausible condition (*t1* (31) < 1) (*t2* (23) < 1). There was also no difference in the proportion of looks to the table in the implausible condition and the lamp in the implausible condition (*t1* (31) < 1) (*t2* (23) < 1) (see figures 2., 3., and table 1.).

Stool vs. lamp/table

During 'the wine carefully into' there were no more anticipatory looks to the stool in the plausible condition than there were looks to the lamp in the implausible condition (*t1* (31) < 1) (*t2* (23) = -1.529, *p* > .05). Similarly, there was no difference in the proportion of looks to the table in the plausible condition and the stool in the plausible condition (*t1* (31) = 1.725, *p* > .05) (*t2* (23) = -1.818, *p* > .05) see figures 2., 3., and table 1.).

Time	table (pl.)	stool (pl.)	table (pl.)	lamp (impl.)
the wine carefully into	15	9	14	11

Table 1. Percentage of trials with looks to the table, stool and the lamp during "the wine carefully into". Percentages calculated from the total number of trials.

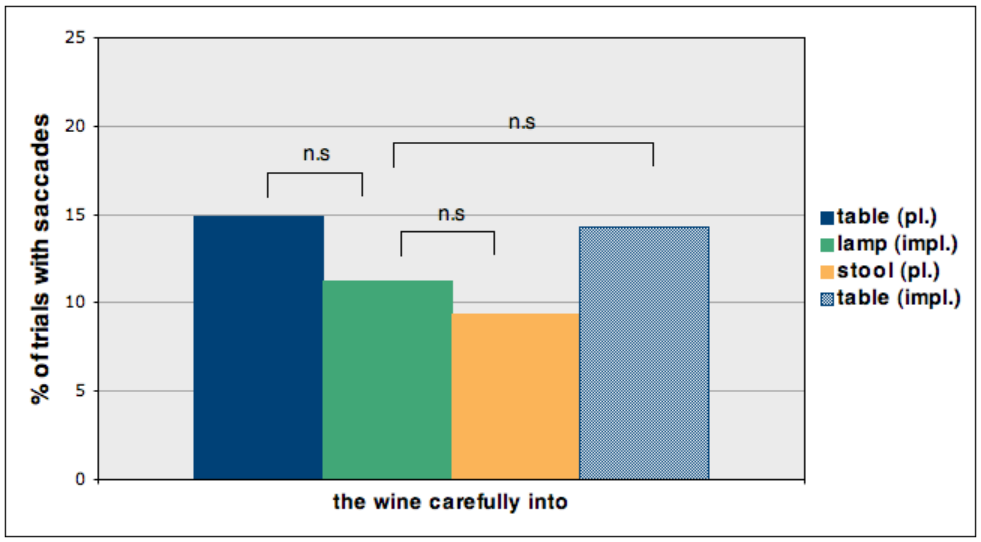


Figure 2. Looks to the table (plausible location 1), the stool (plausible location 2) and the lamp (implausible location) during "the wine carefully into".

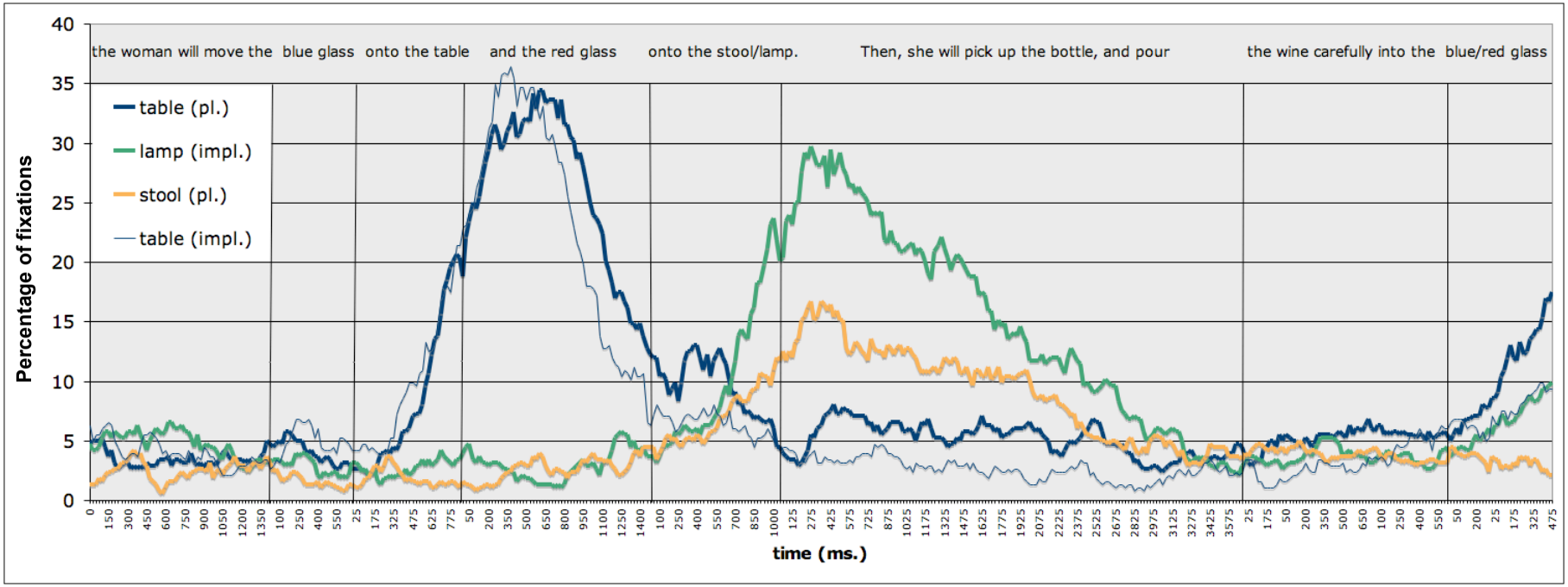
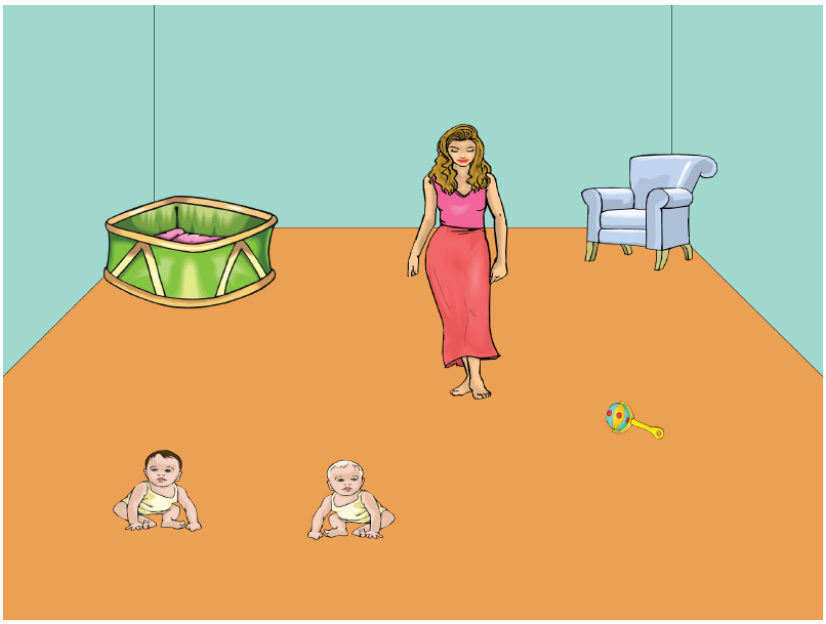


Fig 3. Percentage of trials with fixations toward the table, stool and the lamp in the plausible and implausible conditions. The percentages show the proportion of trials on which participants fixated on each region of interest during *'the woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.* The fixations were calculated every 25 ms sequentially from the synchronisation point.

Appendix 12: Example of filler items – experiment 8 (version 2)

The farmer will move the small pot over to the shovel and the large pot next to the watering can. Then, he will fetch the flower, and place it gingerly inside the small pot.



The woman will sit the blonde baby in the chair and the brunette baby in the playpen. Then, she will get the rattle, and hand it lovingly to the brunette baby.

Appendix 13: Experimental sentences – experiment 9

1.
   a. The woman was very hungry, but she only had one slice of bread left. Quickly, she put it onto the bottle. Then, she grabbed the butter, and spread it sluggishly onto the bread.

   b. The woman was very hungry, but she only had one slice of bread left. Quickly, she put it onto the plate. Then, she grabbed the butter, and spread it sluggishly onto the bread.

2.
   a. The man had been clearing his desk and he had nearly filled the only dustbin in the office. He wanted to tidy up so he moved the dustbin onto the printer. Then, he grabbed the can, and chucked it violently into the dustbin.

   b. The man had been clearing his desk and he had nearly filled the only dustbin in the office. He wanted to tidy up so he moved the dustbin right in front of the fan. Then, he grabbed the can, and chucked it violently into the dustbin.

3.
   a. The woman didn't have a pet carrier, but she had just been to borrow one from her friend. She was in a bit of a hurry so she lifted the pet carrier onto the flowers. Then, she picked up the cat, and put it carefully into the pet carrier.

   b. The woman didn't have a pet carrier, but she had just been to borrow one from her friend. She was in a bit of a hurry so she lifted the pet carrier onto the table. Then, she picked up the cat, and put it carefully into the pet carrier.

4.
   a. The businessman had an important meeting and the only computer he had access to was running low on battery. To prepare for the meeting, he put the computer on the cactus. Then, he picked up the disk, and inserted it quickly into the computer.

   b. The businessman had an important meeting and the only computer he had access to was running low on battery. To prepare for the meeting, he put the computer on the desk. Then, he picked up the disk, and inserted it quickly into the computer.

5.

a. The woman had been instructed to hand over her expense claims and she was given a special envelope to put them in. She was very annoyed and angrily put the envelope right on the sandwich. Then, she reached for the documents, and put them hesitantly into the envelope.

b. The woman had been instructed to hand over her expense claims and she was given a special envelope to put them in. She was very annoyed and angrily put the envelope right next to the lamp. Then, she reached for the documents, and put them hesitantly into the envelope.

6.

a. The student had just moved away from home and she had to share the only pan in the new house with her housemates. She was late for class and quickly put the pan on the pepper mill. Then, she reached for the bowl, and transferred the eggs swiftly into the pan.

b. The student had just moved away from home and she had to share the only pan in the new house with her housemates. She was late for class and quickly put the pan on the cooker. Then, she reached for the bowl, and transferred the eggs swiftly into the pan.

7.

a. The housewife only had one vase because her husband had broken the rest. She was tidying up the living room so she put it on the Hoover. Then, she picked up the flowers, and arranged them happily in the vase.

b. The housewife only had one vase because her husband had broken the rest. She was tidying up the living room so she put it on the sideboard. Then, she picked up the flowers, and arranged them happily in the vase.

8.

a. The chef was being filmed for a new cookery show and she was only allowed to use one pan. She had 30 minutes to prepare a meal so she placed the pan on the potatoes. Then, she noticed the lid, and placed it carefully onto the pan.

b. The chef was being filmed for a new cookery show and she was only allowed to use one pan. She had 30 minutes to prepare a meal so she placed the pan on the cooker. Then, she noticed the lid, and placed it carefully onto the pan.

9.

a. The woman only had one jewellery box and she used it to store her most precious belongings. She was expecting company so she put the jewellery box right on the grapefruit. Then, she admired the necklace, and hid it quickly inside the jewellery box.

b. The woman only had one jewellery box and she used it to store her most precious belongings. She was expecting company so she put the jewellery box right next to the coffee. Then, she admired the necklace, and hid it quickly inside the jewellery box.

10.

a. The man wanted to cut down some trees, but his only chainsaw was running out of petrol. He was trying to finish before it got dark so he put the chainsaw onto the axe. Then, he picked up the jerry can, and poured some petrol quickly into the chainsaw.

b. The man wanted to cut down some trees, but his only chainsaw was running out of petrol. He was trying to finish before it got dark so he put the chainsaw onto the tree stump. Then, he picked up the jerry can, and poured some petrol quickly into the chainsaw.

11.

a. The man wanted to listen to some music, but he only had an old gramophone player. He especially liked classical music so he put the gramophone on the candles. Then, he cleaned the record, and placed it delicately on the gramophone.

b. The man wanted to listen to some music, but he only had an old gramophone player. He especially liked classical music so he put the gramophone on the sideboard. Then, he cleaned the record, and placed it delicately on the gramophone.

12.

a. The girl had been asked to tidy up her room, but she could only find one hanger. She wanted to please her mother so she suspended the hanger on the plant. Then, she reached for the shirt, and hung it cheerfully onto the hanger.

b. The girl had been asked to tidy up her room, but she could only find one hanger. She wanted to please her mother so she suspended the hanger on the rail. Then, she reached for the shirt, and hung it cheerfully onto the hanger.

13.
    a. Even though the businessman had booked a double room he could only find one cup. He was very thirsty so he put the cup on the figurine. Then, he reached for the teapot, and poured the tea slowly into the cup.

    b. Even though the businessman had booked a double room he could only find one cup. He was very thirsty so he put the cup on the table. Then, he reached for the teapot, and poured the tea slowly into the cup.

14.
    a. The man had just moved into his new apartment and so far he only had one chair. He wanted to relax the atmosphere so he put the chair on the Christmas tree. Then, he lifted up the teddy bear, and sat it affectionately on the chair.

    b. The man had just moved into his new apartment and so far he only had one chair. He wanted to relax the atmosphere so he put the chair right next to the girl. Then, he lifted up the teddy bear, and sat it affectionately on the chair.

15.
    a. After the meeting finished the woman had put all the used mugs in the dishwasher except for one. Quickly, she put the mug on the Wellies. Then, she reached for the bottle, and tipped the water quickly into the mug.

    b. After the meeting finished the woman had put all the used mugs in the dishwasher except for one. Quickly, she put the mug on the trolley. Then, she reached for the bottle, and tipped the water quickly into the mug.

16.
    a. The woman wanted some wine, but she could only find one glass. She was desperate for a drink so she put the glass on the lamp. Then, she picked up the bottle, and poured the wine carefully into the glass.

    b. The woman wanted some wine, but she could only find one glass. She was desperate for a drink so she put the glass on the lamp. She was desperate for a drink so she put the glass on the table. Then, she picked up the bottle, and poured the wine carefully into the glass.

17.
   a. The boy wanted to play in the garden, but all he could find was a small plastic boat. He inspected it for a while and decided to put it on the clothesline. Then, he picked up the dummy, and placed it curiously inside the boat.

   b. The boy wanted to play in the garden, but all he could find was a small plastic boat. He inspected it for a while and decided to put it in the pool. Then, he picked up the dummy, and placed it curiously inside the boat.

18.
   a. The woman was going on holiday, but she only had one suitcase. Her flight was leaving in two hours so she put the suitcase on the stereo. Then, she took her shoes, and put them quickly into the suitcase.

   b. The woman was going on holiday, but she only had one suitcase. Her flight was leaving in two hours so she put the suitcase on the bed. Then, she took her shoes, and put them quickly into the suitcase.

19.
   a. The girl wanted to offer her friend a drink, but she could only find one glass. She finished the blouse and put the glass on the mannequin. Then, she grabbed the jug, and poured some lemonade attentively into the glass.

   b. The girl wanted to offer her friend a drink, but she could only find one glass. She finished the blouse and put the glass on the tray. Then, she grabbed the jug, and poured some lemonade attentively into the glass.

20.
   a. The man had just started a fancy new job, but he only had a shabby old briefcase. He studied the graph and lifted the briefcase onto the water-cooler. Then, he reached for the notepad, and placed it hesitantly inside the briefcase.

   b. The man had just started a fancy new job, but he only had a shabby old briefcase. He studied the graph and lifted the briefcase onto the desk. Then, he reached for the notepad, and placed it hesitantly inside the briefcase.

21.

a. The girl had just bought a new dog collar because the dog had chewed up the old one. She wanted to try it out and decided to fasten the collar on the lamp. Then, she picked up the leash, and attached it cheerfully on the collar.

b. The girl had just bought a new dog collar because the dog had chewed up the old one. She wanted to try it out and decided to fasten the collar on the dog. Then, she picked up the leash, and attached it cheerfully on the collar.

22.

a. The handyman had been told to tidy up his equipment, but he had only brought a very small toolbox. He was in a rush, so he lifted the toolbox onto the drill. Then, he picked up the hammer, and dropped it sloppily into the toolbox.

b. The handyman had been told to tidy up his equipment, but he had only brought a very small toolbox. He was in a rush, so he lifted the toolbox onto the shelf. Then, he picked up the hammer, and dropped it sloppily into the toolbox.

23.

a. The man was out gathering fruit but he had only brought one basket. He looked around and decided to put the basket on the donkey. Then, he picked some apples, and placed them gently inside the basket.

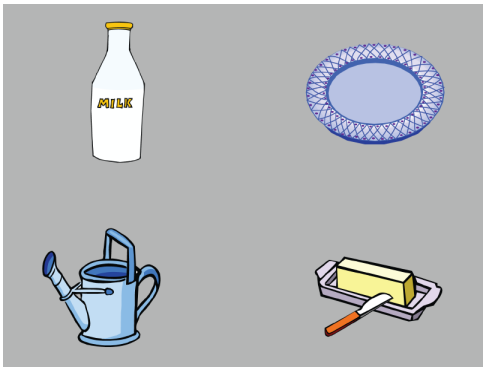
b. The man was out gathering fruit but he had only brought one basket. He looked around and decided to put the basket on the ladder. Then, he picked some apples, and placed them gently inside the basket.

24.

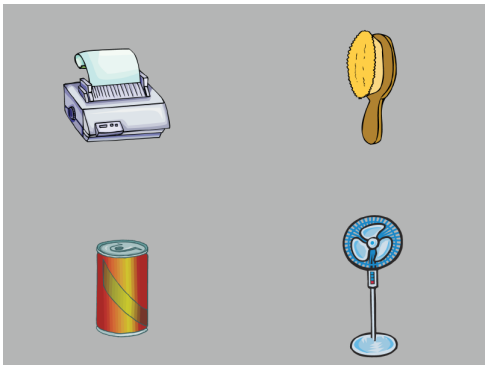
a. The boy only had one cookie-jar because his brother had broken the other one. He was late for school so he put the jar on the birdcage. Then, he took the biscuits, and transferred them attentively into the jar.

b. The boy only had one cookie-jar because his brother had broken the other one. He was late for school so he put the jar on the stool. Then, he took the biscuits, and transferred them attentively into the jar.

Appendix 14: Experimental scenes – experiment 9

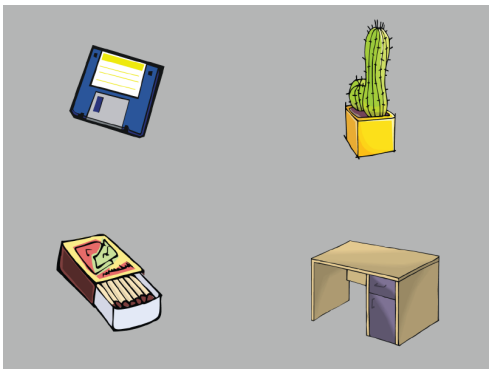


01

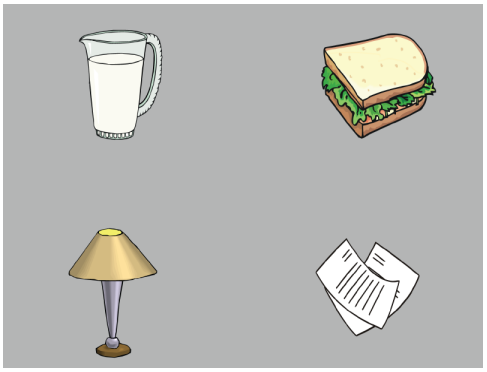


02

03

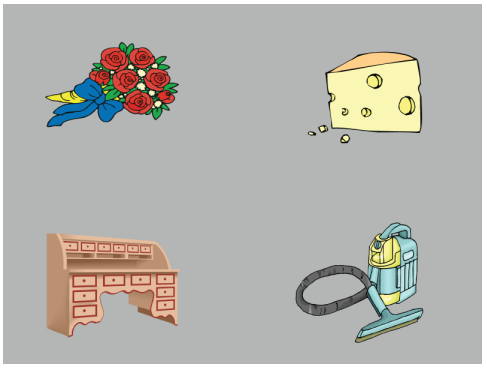


04

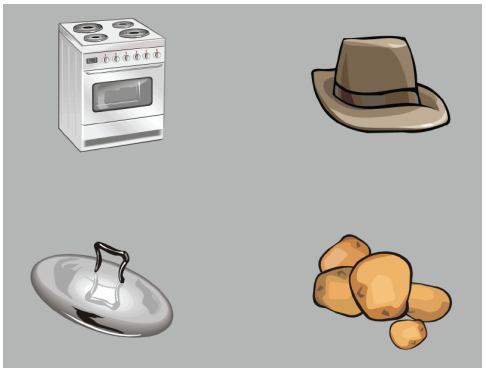


05

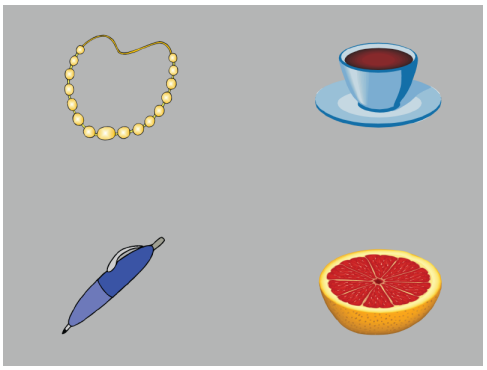
06



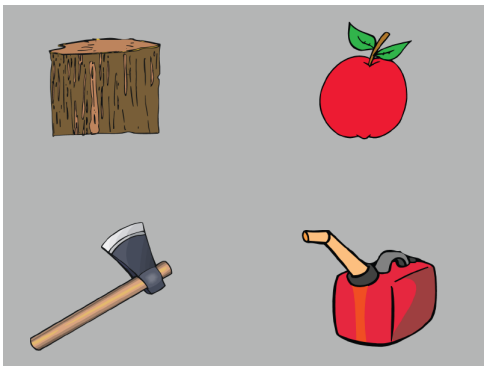
07



08



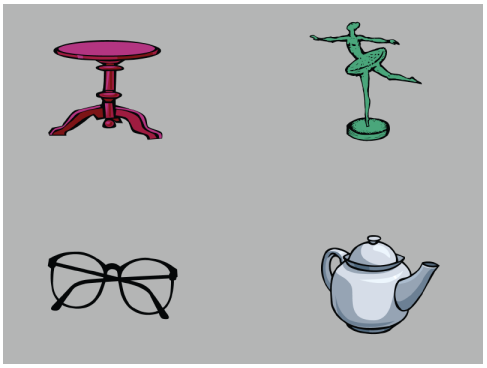
09



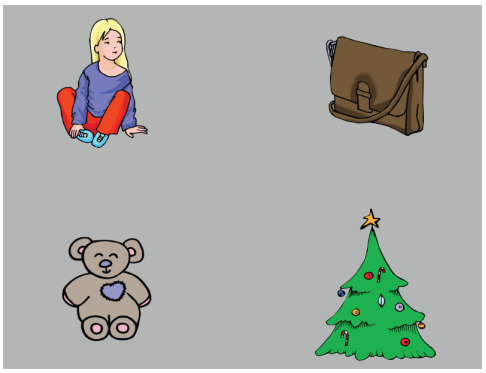
10

11

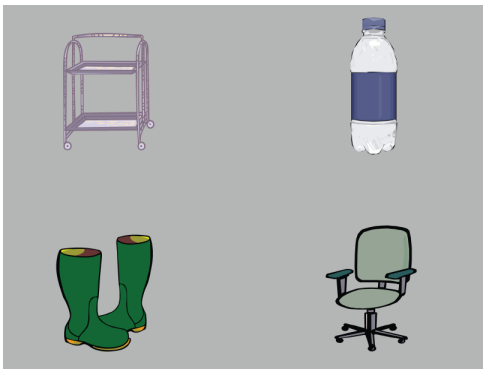
12



13

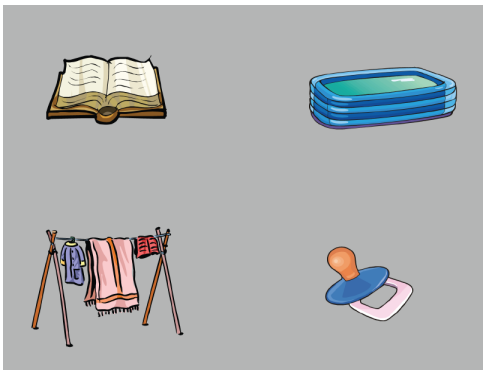


14



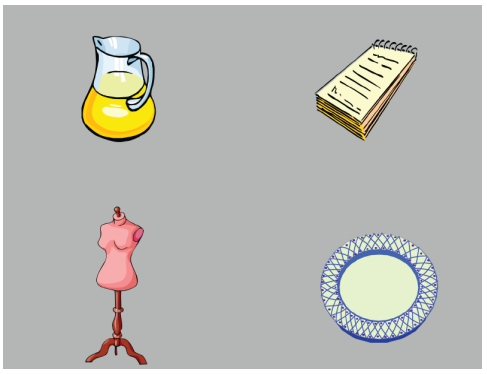
15

16



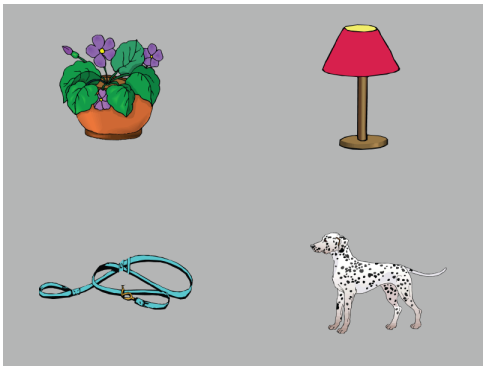
17

18

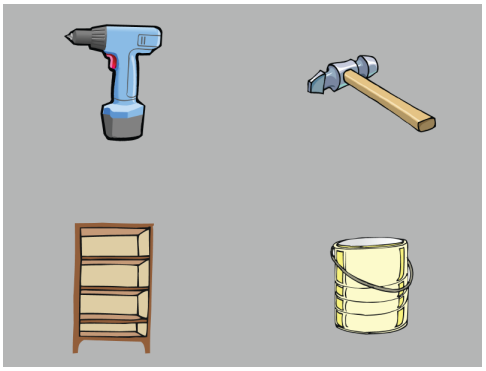


19

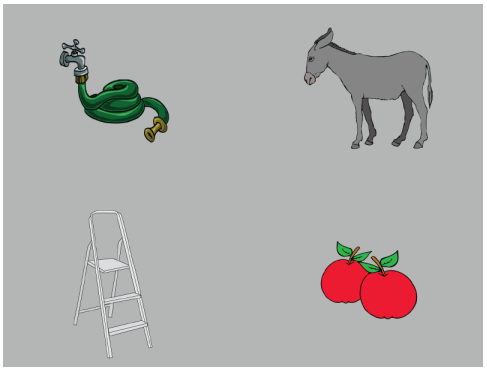
20



21



22



23

24

Appendix 15: Example of filler items – experiment 9

The gardener had imported a very rare orchid from the Amazon. He was eager to study it so he put the orchid onto the tray. Then, he picked up the watering can, and poured the water liberally over the orchid.

REFERENCES

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419-439.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank' screen paradigm. *Cognition, 93*, 79-87.

Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory and Language, 41*, 124–145.

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica, 137,* 190-200.

Altmann, G. T. M. (2011). The mediation of eye movements by spoken language. In S.P. Liversedge, I. D. Gilchrist, & S. Everling (eds.), *The Oxford handbook of eye movements* (pp. 979-1004). Oxford, UK: Oxford University Press.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347-386). New York: Psychology Press.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57,* 502-518.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111,* 55-71.

Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cognitive Science, 33*, 583-609.

Anderson, A., Garrod, S. C., & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly Journal of Experimental Psychology, 35A*, 427-440.

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell Systems Technical Journal, 40,* 309-328.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577-660.

Barsalou, L.W., & Prinz, J.J. (1997). Mundane creativity in perceptual symbol systems. In T. B. Ward, S. M. Smith, & J. Vaid, (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 267–307). Washington, DC: American Psychological Association.

Boland, J. D., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34*, 774–806.

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory and Cognition, 32,* 863-873.

Bourlon, C., Oliviero, B., Wattiez, N., Pouget, P., & Bartolomeo, P. (2011). Visual mental imagery: What the head's eye tells the mind's eye. *Brain Research, 1367*, 287-297.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9*, 27-38.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31,* 25–44.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language, 56,*103–128.

Carreiras, M., Carriedo, N., Alonso, M. A., & Fernandez, A. (1997). The role of verbal tense and verbal aspect in the foregrounding of information in reading. *Memory & Cognition, 25*, 438-446.

Chafe, W. L. (1979). The flow of thought and the flow of language. In T. Givón (Ed.), *Syntax and semantics, Vol. 12: Discourse and syntax* (pp. 159-181). New York: Acacemic Press.

Chambers, C. G., Magnuson, J. S., & Tanenhaus, M. K. (2004). Actions and affordances in syntactic ambiguity resolution, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 687-696.

Chambers, C. G., Tanenhaus, M. K. Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47,* 30-49.

Clark, H. H. (1973). Space, time, semantics and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27-63). New York: Academic Press.

Coll-Florit, M., & Gennari, S. P. (2011). Time in language: Event duration in language comprehension. *Cognitive Psychology, 62,* 41-79.

Collins, A., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science, 13*, 1-49.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences, 3,* 115–120.

Connell, L., & Keane, M. Y. (2004). What affects plausibility? Concept coherence and distributional word coherence as factors influencing plausibility judgments. *Memory and Cognition, 32,* 185-197.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6,* 84-107.

Dahan, D., & Tanenhaus, M. K. (2002). Activation of conceptual representations during spoken word recognition. *Abstracts of the Psychonomic Society, 7*, 14.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121.

Denis, M., & Cocude, M. (1989). Scanning visual images generated from verbal descriptions. *European Journal of Cognitive Psychology, 1*, 293–307.

Eberhard, K. M., Spivey-Knowlton, M. J. Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24,* 409-436.

Erlich, K., & Johnson-Laird, P. (1982). Spatial descriptions and reference continuity. *Journal of Verbal Learning and Verbal Behaviour, 21,* 296-306.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469–495.

Ferguson, H.J., Sanford, A.J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language, 58*, 609–626.

Ferguson, H. J., Scheepers, C., & Sanford, A. J. (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes, 25,* 297-346.

Filik, R. (2008). Contextual override of pragmatic anomalies: Evidence from eye movements. *Cognition, 106,* 1038–1046.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 601–681). Hillsdale, NJ: Erlbaum.

Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge, MA: MIT Press.

Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics, 33, 575-581.*

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108,* 316-355.

Garnham, A. (1996). The other side of mental models: Theories of language comprehension. In J. Oakhill & A. Garnham (Eds.), *Mental models in cognitive science* (pp. 35-52). Hove, UK: Psychology Press.

Gernsbacher, M. A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Erlbaum.

Glenberg, A. M. (1997). What memory is for. *Behavioral & Brain Sciences, 20,* 1-55.

Glenberg, A. M., Becker, R., Koltzer, S., Kolanko, L., Muller, S., & Rink, M. (2009). Episodic affordances contribute to language comprehension. *Language and Cognition, 1,* 113–135.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review, 9,* 558-565.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31,* 129-151.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69-83.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43,* 397-401.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371-395.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Vol. 3: Speech Acts* (pp. 41-58). New York: Academic Press.

Haenggi, D., Kintsch, W., & Gernsbacher, M. A. (1995). Spatial situation models and text comprehension. *Discourse Processes, 19*, 173–199.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438–441.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 210-228.

Hoover, M. A., & Richardson, D. C. (2008). When facts go down the rabbit hole: Contrasting features and objecthood as indexes to memory. *Cognition, 108*, 533-542.

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing “frog”: How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology, 64,* 122-145.

Johansson, R., Holsanova, J., Holmqvist, K. (2005). What do eye movements reveal about mental imagery? Evidence from visual and verbal elicitations. In B. G., Bara, L., Barsalou, M., Bucciarelli, (Eds.), *Proceedings of the 27th Annual Conference of the cognitive science society* (pp. 1054-1059). Mahwah, NJ: Erlbaum.

Johnson-Laird, P. (1981). Comprehension as the construction of mental models. *Philosophical Transactions of the Royal Society of London, 295*, 353-374.

Johnson-Laird, P. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, U.K.: Cambridge University Press.

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children’s and adults’ processing of anomaly and implausibility during reading: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology, 61,* 708-723.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass, 2/4,* 647-670.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133-156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32,* 37-55.

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language, 43*, 508–529.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., Zwaan, R. A. (2005). Perception of motion affects langiage processing. *Cognition, 94*, 79-89.

Kaschak, M. P, Zwaan, R. A., Aveyard, M., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science, 30,* 733-744.

Kelter, S., Claus, B., & Kaup, B. (2004). Representing a described sequence of events: A dynamic view of narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 451-464.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163-182.

Kintsch, W. (1992). A cognitive architecture for comprehension. In H. L. Pick, Jr., P. van den Broek & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 143-164). Washington, DC: American Psychological Association.

Knoeferle, P., Crocker, M.W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition, 95*, 95–127.

Knoeferle, P., & Crocker, M.W. (2006).The coordinated interplay of scene, utterance,and world knowledge: Evidence from eye-tracking. *Cognitive Science, 30*, 481–529.

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language, 57*, 519–543.

Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence processing: Insights from ERP's and picture-sentence verification. *Psychophysiology, 48*, 495-506.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15,* 272-293.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161-163.

Laeng, B., & Teodorecu, D. (2002). Eye scanpaths during visual imagery re-enact those of perception of the same visual scene. *Cognitive Science, 26*, 207-231.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565-572.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes, 22,* 199-224.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology, 15*, 533-545.

Marmolejo-Ramos, F., Elosu´a de Juan, M. R., Gygax, P., Madden, C., Mosquera, S. (2009). Reading between the lines: The activation of embodied background knowledge during text comprehension. *Pragmatics & Cognition 17*, 77–107.

Martarelli, C. S., & Mast, F. W. (2011). Preschool children's eye movements during pictorial recall. *British Journal of Developmental Psychology, 29,* 425-436.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory and Cognition, 32,* 1389-1400.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 913-934.

McCawley, J. D. (1971). Where do noun phrases come from? In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics* (pp. 217-231). Cambridge: CUP.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes, 12*, 137-176.

McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass, 3/6,* 1417-1429.

Meteyard, L., Bahrami, B, & Vigliocco, G. (2007). Motion detection and motion verbs. *Psychological Science, 18,* 1007-1013.

Miller, G. A., & Johnson-Laird, P. (1976). *Language perception.* Cambridge, MA: Harvard.

Mishra, R. K, & Singh, N. (2010). Online fictive motion understanding: An eye-movements study with Hindi. *Metaphor and Symbol, 25,* 144-161.

Morrow, D. G. (1994). Spatial models created from text. In H. Van Oostendorp & R. A. Zwaan (Eds.), *Naturalistic text comprehension* (pp. 57-78). Norwood, NJ: Ablex.

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1990). Situation-based inferences during narrative comprehension. In A. C. Graesser & G. H. Bower (Eds.), *The psychology of learning and motivation: Vol. 25. Inferences and text comprehension* (pp. 123-135). San Diego, CA: Academic Press.

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Verbal Learning and Verbal Behaviour, 28,* 292-312.

Morrow, D. G., & Clark, H. H. (1989). Interpreting words in spatial descriptions. *Language and Cognitive Processes, 3*, 275-291.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language, 26*, 165-187.

Murray, W. S. (2006). The nature and time course of pragmatic plausibility effects. *Journal of Psycholinguistic Research, 35*, 79-99.

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research, 27,* 515–540.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18*, 1098–1111.

O'Brien, E. J., & Albrecht, J.E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 777–784.

Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research, 1153*, 166–177.

Radvansky, G. A., Zwaan, R. A., Federico, T., & Franklin, N. (1998). Retrieval from temporally organized situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1224-1237.

Rayner, K., Warren, T., Juhasz, B. J.,& Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1290–1301.

Richardson, D., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition, 102,* 129-138.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood squares: Looking at things that aren't there anymore. *Cognition, 76*, 269-295.

Richter, T., & Zwaan, R. A. (2009). Processing of color words activates color representations. *Cognition, 111,* 383-389.

Richter, T., & Zwaan, R. A. (2010). Integration of perceptual information in word access. *The Quarterly Journal of Experimental Psychology, 63,* 81-107.

Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory and Cognition, 28,* 1310-1320.

Rinck, M., Hähnel, A., Bower, G. H., & Glowalla, U. (1997). The metrics of spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 622-637.

Rinck, M., Williams, P., Bower, G. H., & Becker, E. S. (1996). Spatial situation models and narrative understanding: Some generalizations and extensions. *Discourse Processes, 21*, 23-55.

Salverda, A. P., & Altmann, G. T. M. (2011). Attentional capture of objects referred to by spoken language. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1122-1133.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 344-354.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language, 24*, 232-252.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109-147.

Shepard, R. N., & Hurwits, S. (1984). Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition, 18,* 161-194.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74,* 1-29.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research, 65*, 235-241.

Spivey, M., & Richardson, D. (2009). Language embedded in the environment. In P. Robbins, & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 383–395). Cambridge, UK: Cambridge University Press.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12,* 153–156.

Sundermeier, B. A., van den Broek, P., & Zwaan, R. A. (2005). Causal coherence and the availability of locations and objects during narrative comprehension. *Memory and Cognition, 33*, 462-470.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632-1634.

Tanenhaus, M. K., Spivey-Knowlton, M. J., & Hanna, J. E. (2000). Modelling discourse context effects: A multiple constraints approach. In M. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 90–118). Cambridge: Cambridge University Press.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Speech, language and communication* (pp. 217-262). San Diego, CA: Academic Press.

Tversky, B. (1991). Spatial semantic models. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory (Vol. 27)* (pp. 109-145). New York: Academic Press.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology, 59,* 1931–1949.

Urbach, T. P., & Kutas, M. (2010). Quantifiers more or less quantify online: ERP evidence for partial incremental interpretation. *Journal of Memory and Language, 63,* 158-179.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Haoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERP's and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 443-467.

van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Van Gompel, R.P.G., & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition, 90,* 255-264.

Vu, H., Kellas, G., Petersen, E., & Metcalf, K. (2003). Situation-evoking stimuli, domain of reference, and the incremental interpretation of lexical ambiguity. *Memory and Cognition, 31*, 1302-1315.

Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye movements in reading*. Psychonomic Bulletin & Review, 14,* 770–775.

Wassenburg, S. I., & Zwaan, R. A. (2010). Reader's routinely represent implied object rotation: The role of visual experience. *The Quarterly Journal of Experimental Psychology, 63*, 1665-1670.

Winter, B., & Bergen, B. (2009). Language-induced mental simulation of distance. Poster presented at the 15th Annual Conference on Architectures and Mechanisms for Language Processing, Barcelona, Spain, 2009.

Yang, J., Wang, S., Chen, H., & Rayner, K. (2009). The time course of semantic and syntactic processing in Chinese sentence comprehension: Evidence from eye movements. *Memory & Cognition, 37*, 1164-1176.

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition, 195*, 229-236.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1196-1207.

Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science, 8,* 15-18.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *Psychology of Learning and Motivation: Vol. 44* (pp. 35-62). San Diego, CA: Academic Press.

Zwaan, R. A. (2009). Mental simulation in language comprehension and social cognition. *European Journal of Social Psychology, 39*, 1142-1150.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995a). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6,* 292-297.

Zwaan, R. A., & Madden, C. J. (2004). Updating situation models. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30,* 283-288.

Zwaan, R. A., Madden, C. J., & Whitten, S. N. (2000). The presence of an event in the narrated situation affects its availability to the comprehender. *Memory & Cognition, 28*, 1022-1028.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science, 28,* 611-619.

Zwaan, R. A., Magliano, P., & Graesser, A. C. (1995b). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 386-397.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162-185.

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Contructing multidimensional situation models during reading. *Scientific Studies of Reading, 2,* 199-220.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13,* 168-171.

Zwaan, R. A., & Yaxley, R. H. (2003). Hemispheric differences in semantic- relatedness judgments. *Cognition, 87*, B79–B.