Language Comprehension

Examining the dynamic changes in activation of mental simulations

Lara N. Hoeben Mannaert

Language Comprehension:
Examining the dynamic changes in activation of mental simulations

Lara N. Hoeben Mannaert

Colophon

Cover design: Lara N. Hoeben Mannaert

Layout: Lara N. Hoeben Mannaert

Printed by: Ridderprint | www.ridderprint.nl

Language Comprehension:
Examining the Dynamic Changes in Activation of Mental Simulations

Taalbegrip: een onderzoek naar de dynamische veranderingen in activering van mentale simulaties

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Friday 13 March 2020 at 11:30

by

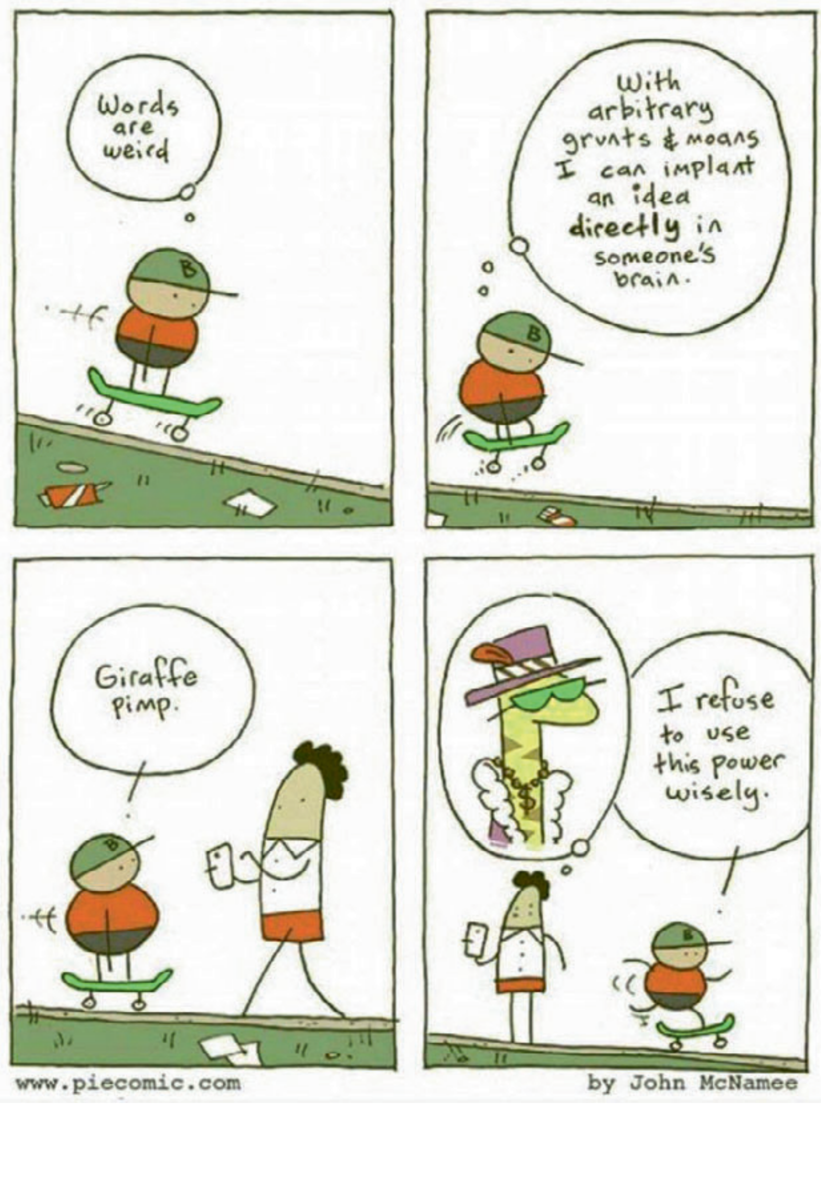
Lara Natasha Hoeben Mannaert

Born in Australia

Erasmus University Rotterdam

Doctoral Committee

Promotor:	Prof.dr. R.A. Zwaan
Other members:	Prof.dr. P.W. van den Broek
	Dr. H.K. Tabbers
	Dr. P.P.J.L. Verkoeijen
Copromotor:	Dr. K. Dijkstra


Words are weird
With arbitrary grunts & moans I can implant an idea directly in someone's brain.
Giraffe pimp.
I refuse to use this power wisely.
www.piecomic.com
by John McNamee

Contents

1

General Introduction

"That's what books are for… to travel without moving an inch." (Lahiri, 2019, p.16). Many people have experienced it: they are reading a book and feel as though they have been transported into another world, seeing the events unfold as though they were watching a movie. Indeed, there are many studies showing that readers report experiencing spontaneous mental imagery while they are reading (e.g., Long, Winograd, & Bridge, 1989; Sadoski, 1983; 1985; Sadoski, Goetz, Kangiser, 1988). The idea that perception and language comprehension are linked has been around for millennia (Barsalou, 1999). Since the 20th century, scientists have actively begun testing and theorizing how language is represented in the brain. Although there are many theories of how language is processed, two main branches exist that encompass many of these theories, namely the *traditional views of cognition* and *grounded (*or *embodied*) *cognition.*

Traditional views of cognition

Especially in the 20th century, research on psycholinguistics was focused on the belief that the processing of language makes use of amodal symbols, a trend that was strongly influenced by the developments in mathematics, computer science, and formal logic. What is meant by "amodal" here is that the symbols contain no direct link to the sensorimotor systems of the brain. For example, if one were to learn the concept "tree", this could happen as follows: a child sees a tree for the first time while a parent provides a label for the object, so that the arbitrary sound "tree" is linked to its referent. During encoding, the sensory information (e.g., what a tree looks like, the sound of the word "tree") is first transduced into a new type of representational language, before it is stored in larger structures (e.g., feature lists, frames, schemata, semantic networks, etc.), where it can be manipulated for various cognitive processes (Fodor, 1983; Pylyshyn, 1984). The result of this transduction, however, is that the created symbols do not retain any of the sensory information present during the real-world experiences. Cognition is therefore viewed as a system separated from perception, containing only arbitrary links to the sensorimotor systems of the brain, which are not required for the actual cognitive processes occurring within the brain.

Although using amodal symbol systems as an explanation for cognition has its advantages (e.g., it can easily explain the processing of abstract concepts), the largest problem with this theory is the *symbol grounding problem* (Harnad, 1990). Harnad argued that a system that is based on the manipulation of meaningless symbols cannot be grounded in anything other than more meaningless symbols. He provides the example of trying to learn Chinese as a first language from a Chinese/Chinese dictionary. A person looks up one symbol in the dictionary, only to be faced with multiple other symbols. In turn, when those symbols are looked up, more Chinese symbols are provided. In essence, it is impossible to derive meaning from any of the symbols without connecting them up to the real world.

Grounded cognition

Theories of grounded cognition (also referred to as "embodied cognition") are based on the idea that cognition has to be grounded in perception (Barsalou, 1999; 2008). The way this works is through the use of mental simulations, which are defined as the "reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind" (Barsalou, 2008, p.618). This differs from traditional accounts of cognition in one very important way: rather than transducing perceptual states into some meaningless symbols that are manipulated by an independent system, these perceptual states are captured and partially reactivated whenever this is required for a cognitive process. In our previous "tree" example, when the boy sees the tree, there is activation across the sensorimotor systems of the brain. The way the tree looks is captured by the visual system, the sound of the rustling leaves is captured by the auditory system, the smell of the blooming flowers is captured by the olfactory system, and so forth. Whenever the concept “tree” has to be accessed again (e.g., when coming across the word “tree” in a book), this pattern of activation is partially reactivated and integrated into a multimodal simulation. Barsalou (1999) in his Perceptual Symbol Systems (PSS) theory suggests that not the exact same pattern of activation that occurs during the actual experience is reactivated, but that only a subset of this pattern is stored as a perceptual symbol, which is used for these simulations.

Not much evidence exists for the traditional view of cognition, and this is mostly because amodal symbols can be used to explain all experimental findings, making the theory unfalsifiable. The benefits of the grounded cognition theories are that they allow for specific predictions to be made regarding the activation of the brain or the behavior of participants in cognitive experiments. For instance, if the sensorimotor systems are required for language comprehension, then we would expect these systems to become activated while performing linguistic tasks. Indeed, a functional magnetic resonance imaging (fMRI) study by Hauk, Johnsrude, and Pulvermüller (2004) illustrated that the same areas of the brain become activated when a verb signifying an action is read as when that action is actually performed. Moreover, the specificity of action verbs (e.g., *to clean* versus *to wipe*) can modulate the blood oxygen level dependent (BOLD) response in the bilateral inferior parietal lobule, a region associated with representing action plans (Van Dam, Rueschemeyer, & Bekkering, 2010). This activation of motor areas while reading action verbs has not only been found in adults, but in children as well (James & Maouene, 2009). Moreover, Pulvermüller, Hauk, Nikulin, and Ilmoniemi (2005) conducted a transcranial magnetic stimulation (TMS) experiment to find a causal link between the motor areas and the processing of action words. They found that stimulating arm areas led to faster responses on a lexical decision task when words associated with arm movements were shown. Furthermore, when the leg areas were stimulated, participants responded faster to the words associated with leg movements. Together, these studies provide ample evidence for the existence of a link between the motor cortex and action-related language.

Aside from action verbs relating to the motor regions of the brain, neuroimaging studies have also shown that the same areas of the brain activate when you produce speech sounds as when you hear them (Pulvermüller et al., 2006; Wilson, Saygin, Sereno, Iacoboni, 2004). Furthermore, reading words associated with particular odors (e.g., garlic, jasmine) also leads to activation of the primary olfactory cortex (González et al., 2006). Similarly, gustatory words (e.g. chocolate, mustard) produce activation in the primary and secondary gustatory cortices (Barrós-Loscertales et al., 2012). Due to this consistent link being found between language and the sensorimotor areas of the brain, many researchers agree that those systems are involved in some way in language comprehension.

Interestingly, the degree of this involvement is still a current topic of debate. Meteyard, Cuadrado, Bahrami, and Vigliocco (2012) conducted a theoretical review of the relevant theories, placed them on a continuum of embodiment, and reviewed the evidence for the different viewpoints. They concluded that the strongly embodied theories, which argue that cognition requires a complete dependence on sensorimotor systems, have insufficient scientific support as they would predict that the primary sensorimotor cortices would be active during all semantic tasks, and this is not supported by the neuroimaging data. Similarly, they found no evidence for a completely unembodied view of cognition, which argues that there is absolutely zero overlap between language and sensorimotor areas of the brain and that the activations of those systems would be due to indirect pathways. This leaves us with the two remaining views on the continuum: weak embodiment versus secondary embodiment.

Theories of weak embodiment suggest that semantic representations contain sensorimotor information, but that this information is abstracted to some extent (ibid). What this means is that information is taken from the primary modalities and potentially integrated in areas adjacent to these primary cortices. This integrated modal information then forms the basis of semantic content. As the information is abstracted away from the primary cortices, this can no longer be called strong embodiment, as this would no longer constitute a complete simulation of the real-world (ibid). Note that Barsalou's (1999) PSS theory could also be interpreted as weak embodiment as that theory similarly argues only a subset of the original activation pattern is stored to be used for cognitive processes later. Evidence for weak embodiment comes from studies showing that the motor cortex become activated very quickly (roughly 200ms) after encountering action words (Hauk, Shtyrov, & Pulvermüller, 2008; Meteyard et al., 2012).

Theories of secondary embodiment argue that there is an amodal core processor which contains non-arbitrary connections to the sensorimotor systems. Semantic content is therefore amodal in nature, and when activated leads to the passive activation of the sensorimotor systems via a spreading activation mechanism (Meteyard et al., 2012). According to Mahon and Caramazza (2008), language impairments caused by damage to the sensorimotor systems can be explained by concepts becoming isolated.

To sum up, the main differences between theories of secondary embodiment and weak embodiment are whether semantic content is sensorimotor or amodal in nature. Secondary embodiment suggests that there is an independent semantic hub that contains non-arbitrary connections to the sensorimotor systems, while weak embodiment suggests there are multiple zones (e.g., convergence zones, see Damasio, 1989) where modal information that is abstracted away from the primary sensorimotor cortex is integrated. Given that both theories can potentially explain the findings published in the embodied cognition literature, more research is needed to make strong conclusions regarding the actual nature of semantic representations.

Language and situated simulation

The theories described above suggest that knowledge can only be stored via one type of representation. Barsalou, Santos, Simmons, and Wilson (2008) instead propose that both linguistic forms from the brain's language systems and mental simulations from the modalities are combined to represent knowledge. According to their proposed framework, the language and situated simulation theory (LASS), when a word is perceived, the linguistic system is activated first, which immediately activates word associations (e.g., 'chicken' activates 'egg'), which can provide superficial strategies for certain cognitive tasks without requiring the retrieval of deeper conceptual knowledge. Once the linguistic system recognizes the word, associated simulations become active, which make use of sensorimotor and introspective areas of the brain in order to represent the meaning of the concept.

The authors explain how, when nonwords in a lexical decision task are easily distinguishable from words, only the linguistic system needs to be accessed to complete the task, as the word's meaning does not need to be activated to provide an accurate response. However, when nonwords are phonologically and orthographically similar to actual words, the simulation system needs to be recruited to complete the task as deeper conceptual information needs to be accessed to provide an accurate response.

Evidence for LASS comes from a study by Santos, Chaigneau, Simmons, and Barsalou (2014), who used a property generation task, where participants received a cue word and had to come up with as many properties of that word as they could think of. Their results illustrated that the properties that were produced earlier on tended to be related linguistically to the word cue (i.e., via word associations, for example: BEE → sting), while the properties that were generated later on tended to describe objects and situations (e.g., BEE → flowers). These findings were further supported in an fMRI study by Simmons, Hamann, Harenski, Hu, and Barsalou (2008), who found that activations early in conceptual processing overlapped with activations for word associations, and that activations for later conceptual processing overlapped with activations for situated generation (where participants had to think of a situation where a certain word would appear and how to describe it). Together these findings provide support for the idea that language systems and simulation systems interact to support language comprehension.

Mental simulations

Given the clear importance of the simulation system in language comprehension, it is important to gain a clear understanding of both the content and the underlying mechanisms of mental simulations. Several behavioral studies have been conducted over the past two decades to provide us with an idea of the contents of mental simulations. In an experiment by Stanfield and Zwaan (2001), participants read sentences that implied a particular orientation (e.g., “John put the pencil in the cup” implies that the pencil is standing upright) and performed a sentence-picture verification task. Important to note here is that, according to the traditional theories of cognition, no inferences could be made here regarding the pencil’s orientation (ibid). Conversely, the PSS theory would predict that a simulation of the event described by the sentence would be created and thus should lead to faster responses when the picture shown matches the implied orientation. Indeed, participants responded significantly faster to the pictures that matched the implied orientation compared to the pictures that mismatched, suggesting that object orientation is actively simulated during language comprehension when orientation is relevant.

This "match effect" has been found for many different object properties, namely shape (Pecher & Zwaan, 2012; Zwaan, Stanfield, & Yaxley, 2002), visibility (Yaxley & Zwaan, 2007), motion (Zwaan, Madden, Yaxley, & Aveyard, 2004), sound (Brunyé, Ditman, Mahoney, Walters, & Taylor, 2010) and has also been found in children (Engelen, Bouwmeester, de Bruin, & Zwaan, 2011) and elderly populations (Dijkstra, Yaxley, Madden, & Zwaan, 2004). Combined, these studies provide support for the idea that mental simulations are involved in language comprehension, and that these are modal in nature.

The role of color in mental simulations

Interestingly, for the object property color, conflicting findings exist in the literature. A study by Connell (2007) found that participants responded significantly faster to pictures that mismatched the color implied by the sentence compared to those that matched. When Pecher and Zwaan (2012) conducted a replication of that study, they did find a significant match effect. This led to the question: in what capacity is color present in mental simulations? In **Chapter 2**, we investigated the role of color in mental simulations by performing a conceptual replication of the Connell and Pecher and Zwaan studies. The main difference between our study and those that came before is that we used an improved stimulus set. In the original studies, there were several items that could change shape as well as color (e.g., a steak that is cooked has a different shape than a raw steak), and items that do not describe color (e.g., a polar bear is white). As such, the first goal of our study was to examine the role of color in mental simulations. Given the previous literature on mental simulations, we expected to find a significant match effect for color.

The second goal of our study was to examine how much sensory information is captured by mental simulations, which was done by lowering the saturation of the pictures. When lowering the saturation of pictures, there are two possible outcomes. Firstly, it is possible that the facilitation provided by the matching color is eliminated, leading to a slower response in the match condition compared to when the picture is shown in full saturation. Secondly, it is possible that, as a result of the saturation being lowered, there is a less vivid difference between the mismatching picture and the mental simulation (i.e., less interference), thus leading to a faster response in the mismatch condition. We expected that the match advantage would be decreased when saturation is lowered.

The final goal of this study was to examine whether a match advantage still exists when pictures are shown completely in grayscale. We expected to find no significant match effect here as this would provide contradictory evidence for the grounded cognition approach. Conducting this experiment was of interest as studies overall have shown that color aids in object recognition (e.g. Bramão, Reis, Petersson, & Faísca, 2011), so we expected that participants would respond significantly faster when the pictures are shown in color compared to when they are shown in grayscale.

The updating of mental simulations

Many studies on mental simulations have provided us with the knowledge of which object properties are actively simulated during language comprehension, but researchers are now more and more interested in how these mental simulations unfold across texts. For example, studies have shown that mental simulations can be reactivated a later point in time (Pecher, Van Dantzig, Zwaan, & Zeelenberg, 2009), and can remain activated longer when an ongoing situation is described compared to a situation that has already occurred (Madden & Therriault, 2009; Madden & Zwaan, 2003; Magliano & Schleich, 2000). Most studies examining this, however, have examined updating within the context of the situation model, not mental simulations.

For language comprehension, three levels of representation have been identified: the surface text form, which refers to the exact words and syntax used in the text; the textbase, which refers to the abstract representation of the ideas in the text; and the situation model, which is the mental representation of the events described in a text (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). When the situation model was first introduced as a construct for language comprehension, it was believed that the mental representations used to create the situation model were amodal in nature. More recently, however, researchers have thought that those mental representations are perceptual in nature (Zwaan, 2016), and thus can also be called mental simulations.

Given the role of mental simulations in building the situation model that supports language comprehension, it is important to gain a clear understanding of how this works. Many studies on situation model updating have used reading times as a measure for updating, but this tells us nothing about the actual activation patterns of mental simulations. Given that mental simulations can remain active over longer periods of time or can reactivate if the context requires it, how do these simulations behave in a changing discourse context? This question was the focus of **Chapter 3** in this dissertation.

In this chapter, we were interested in examining whether implying a change in shape, over the course of two or four sentences, would lead to the simultaneous activation of both the initial and the final shape, or whether the initially implied shape would deactivate. Based on the fact that mental simulations can remain activated for a while, we hypothesized that when implying one shape directly after the other (i.e., in two sentences), that both shapes would remain activated in the mental simulation. We further expected that when the final shape is emphasized using three sentences (out of four sentences in total) that the initial shape would become deactivated as more emphasis was placed on the final shape.

The updating of situation models in older adults

Much of the research that has been done on cognitive aging over the past several decades has focused on the deterioration of cognitive functions. More recently, however, there has been a shift to examining the preservation of functions in older adults. One such an example is the preservation of situation model updating in older adults. In **Chapter 4**, we present a review of the prominent theories on situation model updating, the evidence for preservation of this function in an aging population, and how this compares to young adults.

The completeness of mental simulations

As mentioned previously, many studies have illustrated that various object properties are activated in mental simulations. Importantly, however, these studies have only ever examined the presence of a single object in mental simulations, and never whether multiple objects can be combined. If the building of a comprehensive situation model is required for language comprehension to occur, then the situation model should contain a complete representation of the events described by the text, but this has not been studied before. As such, in **Chapter 5**, the research question of interest was: do we combine multiple objects in a mental simulation in order to construct a comprehensive situation model, and is this influenced by task instructions?

In this chapter, we present two experiments that had participants read sentences describing animals using a tool in a particular way. After reading the sentence, participants had to press a button indicating whether the pictured cartoon animal (experiment 1) or tool (experiment 2) was mentioned in the previous sentence. The pictured cartoon animal either completely matched (i.e., both the correct animal and tool were shown), partially matched or mismatched (i.e., either the correct animal or tool was shown), or completely mismatched (i.e., neither the correct animal nor tool was shown) the event described by the previous sentence. If task instructions can influence the contents of a mental simulation, then this has significant consequences for the relevance of mental simulations in language comprehension. However, if a complete mental simulation of the events described in a text is required for language comprehension to progress smoothly, then we would expect both the animal and the tool in this study to be simulated. We expected that a complete mental simulation would be created by reading the sentence, and thus would lead to significantly faster responses in the complete match condition compared to the partial match condition, regardless of task instructions.

The deactivation of mental simulations

While reading a text, many dimensions are tracked throughout the narrative, such as the objects, goals, locations, events, and actions described (Zwaan, Langston, & Graesser, 1995; Zwaan, Radvansky, Hilliard, & Curiel, 1998), and subsequently are integrated into a situation model. Given that many events are tracked throughout a narrative, does this mean every time a particular dimension is involved that all of the associated information is reactivated? For instance, when one starts reading a novel, usually a lot of information regarding the looks of the protagonist and the side characters are described. Once a mental simulation of this character is built, does this reactivate every time this character is mentioned?

In **Chapter 6**, we investigated whether color is continuously activated in mental simulations across a changing discourse context. Based on the findings that color is actively simulated during language comprehension, we were interested in how this object property evolves when changes occur in a narrative. For example, when reading the sentences "*The boy rode on the red bicycle to the station. At the station he stepped off of his bicycle*.", would the color red become reactivated when only the object is mentioned the second time?

Contradictory findings exist in the literature regarding the continued activation of perceptual information when events change. Swallow et al. (2009), for instance, found that perceptual details can be cleared from active memory if the target object is not present at event boundaries, while Pecher et al. (2009) found that perceptual information can be retained for long periods of time. In order to fully understand the role of mental simulations in language comprehension, it is important to know how the perceptual content of these simulations are affected by a narrative.

In the current study, participants read stories made of one (experiment 1), two (experiment 2), or five sentences (experiment 3). When the story was one sentence long, this sentence could either make a reference to a color or make no reference to a color. When the story was two or five sentences long, either the first or last sentence made a reference to a color. Participants performed a sentence-picture verification task where they judged whether the pictured object was mentioned in the previous sentence or not. These pictures were shown either in full color or in grayscale. Based on the findings from **Chapter 2**, we expected that participants would respond significantly faster to colored pictures when a color was mentioned, compared to pictures shown in grayscale. In the second experiment, we expected that color would deactivate when the second sentence did not refer to a color, as this would not be necessary for language comprehension. Based on the findings from the second experiment, we expected that, in the third experiment, color would continue to be activated, even at the end of a five-sentence story.

The importance of preregistration

There are several commonalities across the chapters in this dissertation. The first of which is that the method used to study mental simulations is the sentence-picture verification paradigm. This paradigm is often used in psycholinguistics research because it allows us to examine whether the pictures shown match what is activated in mental simulations, and thus allows us to gain insights into the underlying mechanisms of language comprehension.

The second commonality in the chapters of this dissertation is that all the experiments from each chapter were preregistered online prior to data collection. Unfortunately, questionable research practices (QRPs) have been rife in psychology. Researchers perform post-hoc analyses and report them as planned, they run many different experiments or analyses and only report the ones that show a significant p-value, they choose to stop data collection when a significant p-value has been reached, or choose which outliers to remove from the dataset to ensure a significant finding. These are just some of the examples of QRPs that have been reported in psychology, a problem which led to the Reproducibility Project, a project that set out to replicate 100 studies in psychology, and found that only about a third of the effects could be replicated (Nosek et al., 2015). One of the solutions to the replication crisis in psychology is for researchers to preregister their hypotheses, data collection plan, analysis plan, and materials online, prior to data collection. This way, a reader can be certain that the results and conclusions of those studies did not come about via QRPs. The URLs of all the preregistrations of the current dissertation can be accessed on: https://osf.io/qezt6.

2

Is Color an Integral Part of a Rich Mental Simulation?

This chapter has been published as:

Hoeben Mannaert, L. N., Dijkstra, K., & Zwaan, R. A. (2017). Is color an integral part of a rich mental simulation? *Memory & Cognition*, *45*(6), 974-982. doi:10.3758/s13421-017-0708-1

Abstract

Research suggests that language comprehenders simulate visual features such as shape during language comprehension. In sentence-picture verification tasks, whenever pictures match the shape or orientation implied by the previous sentence, responses are faster than when the pictures mismatch implied visual aspects. However, mixed results have been demonstrated when the sentence-picture paradigm was applied to color (Connell, 2007; Zwaan & Pecher, 2012). One of the aims of the current investigation was to resolve this issue. This was accomplished by conceptually replicating the original study on color, using the same paradigm but a different stimulus set. The second goal of this study was to assess how much perceptual information is included in a mental simulation. We examined this by reducing color saturation, a manipulation that does not sacrifice object identifiability. If reduction of one aspect of color does not alter the match effect, it would suggest that not all perceptual information is relevant for a mental simulation. Our results did not support this: We found a match advantage when objects were shown at normal levels of saturation, but this match advantage disappeared when saturation was reduced, yet still aided in object recognition compared to when color was entirely removed. Taken together, these results clearly show a strong match effect for color, and the perceptual richness of mental simulations during language comprehension.

Keywords: Color, language comprehension, perception, mental simulation

Introduction

Many empirical studies have supported theories of grounded cognition, which suggest that we use the same sensorimotor regions in the brain during activity as during cognitive processes, through the use of mental simulations (Barsalou, 1999, 2008). It has been argued that activation of perceptual areas in the brain during language comprehension are not merely epiphenomenal but that language can, in addition to communication, serve as a control mechanism to shape mental content (Lupyan & Bergen, 2015). One such experiment examined whether we create mental simulations of an object's orientation when the orientation is implied in the sentence (Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012). The study showed that when the implied orientation matches the orientation of the object shown in an object-verification task, that reaction times are shorter than when they mismatch, suggesting that we create mental simulations during sentence comprehension. This match advantage has also been found for visual aspects such as shape (Zwaan, Stanfield, & Yaxley, 2002), visibility (Yaxley & Zwaan, 2006), and motion (Zwaan, Madden, Yaxley, & Aveyard, 2004); has been found for children (Engelen, Bouwmeester, de Bruin, & Zwaan, 2011) as well as for the elderly population (Dijkstra, Yaxley, Madden, & Zwaan, 2004); that spoken words also rapidly activate visual representations that affect our ability to recognize objects (Ostarek & Huettig, 2017); and the shape of an object becomes activated during encoding, and not simply during retrieval (Zeng, Zheng, & Mo, 2016).

However, mixed results have been found when this sentence-picture paradigm was applied to color. For instance, Connell's (2007) study illustrated an advantage in the mismatch condition. Connell (2007) suggested that color may be represented differently than other visual features because it is one of the few object properties that is unimodal, (i.e., it can only be perceived with the visual modality) and has been shown to be less vital to object identification than shape (Tanaka, Weiskopf, & Williams, 2001) or orientation (Harris & Dux, 2005). Thus, it should be easier for participants to ignore mismatching color information and focus on a stable object property such as shape than to ignore the matching color as it aids in solving the task demands and requires processing. Zwaan and Pecher (2012), however, conducted six replication experiments to investigate this match advantage in greater detail for object orientation, shape, and color, and found a match advantage for all three object properties. Moreover, the match advantage for color had a larger effect size than those for shape and orientation. Another study also appeared to support a match advantage for color, as reading words in a color (e.g., white ink) matching the color implied by a previous sentence (e.g., *Joe was excited to see a bear at the North Pole*) facilitated reading times (Connell & Lynott, 2009).

These contradictory findings in studies examining color as part of mental simulations prompt further questions into how we process color during language comprehension and how much sensory information we include in these simulations. One possibility is that color is an unstable visual feature in mental simulations, as the color of an object can change without eliminating the ability to recognize the object, and therefore may play a less present role in mental simulations.

One of the goals of the current investigation was to address the potential problem of color instability caused by the stimulus set used in the original study (Connell, 2007) and in the replications (Zwaan & Pecher, 2012). To address this issue, we created a stimulus set that met more stringent criteria with regard to the visual features than the earlier stimulus sets did. For example, there were some items in the previous study in which features other than color could vary (i.e., a steak that is cooked has a different shape than a steak that is raw). This problem does not occur for more carefully chosen, less variable, items, such as a red or green tomato. Therefore, in the current investigation, all potentially problematic items were removed and replaced with stimuli that could undergo a color change while their shape remained unaltered. Another difference in our stimulus set was that full-color photographs were used rather than line drawings, to allow for a more realistic representation of the described objects (Holmes & Wolff, 2011).

The second goal of the study was to examine how much sensory information is captured in a mental simulation. Color is a useful tool for exploring this, as it is the only visual feature that can be decomposed into different dimensions, namely hue, saturation, and brightness (Palmer, 1999). This decomposition is solely a color aspect manipulation as the decomposition process still allows for the object to be recognized (i.e., there is no change in shape, size, or orientation). For instance, a tomato without hue will simply become a gray tomato, maintaining its shape and preserving all other visual features. At the same time, however, changes in color, saturation or brightness affect the richness of the visual stimulus, as these dimensions alter what is typical about the visual properties of the stimulus. Thus, if these dimensions affect the richness of the visual stimulus, is it necessary to represent them in a mental simulation? When one processes a sentence implying a certain color, is information regarding the saturation of the color stored? For example, when reading about a ripe tomato, would a simulation include a bright red, or would this not be as vital to the simulation as other sensory information?

Our current study explored how much sensory information is included in mental simulations by conducting four experiments, using the same experimental paradigm as Connell (2007) and Zwaan and Pecher (2012) where sentences are used to imply a certain color, followed by an object-verification task. For example, the sentence *The driving instructor told Bob to stop at the traffic lights* is used to imply a red traffic light, rather than explicitly stating *The driving instructor told Bob to stop at the red light*. After reading a sentence implying a certain color, participants see either a matching (e.g., red light) or mismatching picture (e.g., green light) and have to press a button on the keyboard verifying whether the pictured object was mentioned in the previous sentence, where the correct answer to experimental items always required a “yes” response.

The first experiment was conducted as a conceptual replication of Connell’s (2007) and Zwaan and Pecher’s (2012) experiments on color, to resolve which of the contradicting findings has more empirical support. Given the previous literature, we predicted to find a significant match advantage. Experiment 2a and 2b addressed the question of how much perceptual information is included in a mental simulation. This was accomplished by lowering the saturation of the pictures used in Experiment 1 to the lowest level at which the hue could still be recognized. It is possible that by reducing the level of saturation in the picture there is less of an overlap with what is currently being simulated, which could lead to there being less facilitation of a response in the match condition under low levels of saturation. A further possibility is that rather than the match condition acting as a facilitatory mechanism, the match effect exists due to there being a vivid difference between what is simulated and what is pictured in the mismatching condition. Reducing the level of saturation would then reduce the disparity between the picture and the simulation, leading to faster responses in the mismatch condition. In other words, there would be less interference. Experiment 3 examined whether a match advantage still exists when objects are shown completely in grayscale. This is of interest for several reasons. First, if a match advantage does appear under low levels of saturation, then it should disappear when the pictures are shown in grayscale. Second, studies have shown that color does aid in object recognition (Bramão, Reis, Petersson, & Faísca, 2011). With this in mind, we expect that participants' response times (RTs) in Experiments 1 and 2 will, overall, be faster than in Experiment 3, where no color is present.

Ethics statement

The participation in all four experiments and in the norming studies was voluntary. The participants subscribed to the experiments online. They were briefed with the content of each study but obtaining a written consent was not required by the Ethics Committee of Psychology at the Erasmus University Rotterdam, The Netherlands, who approved the project, because the experiments were non-invasive and the data collection and analysis were anonymous.

Preregistration

The predictions, exclusion criteria, design, methods, analyses, and materials of all the experiments reported in this article were preregistered in advance of data collection and analysis on an online research platform—Open Science Framework (OSF; see Nosek & Bar-Anan, 2012; Nosek, Spies, & Motyl, 2012, for a detailed discussion on replications and preregistration). This ensured that confirmatory procedures (hypotheses testing) were conducted according to a priori criteria. In the current article, a clear distinction between confirmatory and explanatory analyses was made, as suggested by De Groot (1956/2014). The post hoc analyses are included in the Exploratory Analyses section

Experiment 1

Method

Participants. Two hundred and five participants were recruited via Amazon's Mechanical Turk[1] (87 males, mean age 37.78 years, range: 20–87 years). The participants were paid $1.50 for their participation.

Materials. The experimental flow was programmed in Qualtrics Survey Software. It allowed for an automatic collection of information such as Browser Type, Browser Version, Operating System, Screen Resolution, Flash Version, Java Support, and User Agent for each participant.

Pictures. Thirty-two pictures were selected as experimental items and 16 as filler items. The pictures were obtained from the internet (Google image search engine). Picture size was unified across the trials: none of the pictures exceeded 300 × 300 pixels (approximately 7.9 × 7.9 cm onscreen). The objects depicted in the images had one dominant color (e.g., green in the green traffic light picture). The experimental items formed 16 pairs of objects, and pictures within a pair differed in color (i.e., red traffic light vs. green traffic light). The pictures of the objects within a pair were matched in terms of size and shape to ensure that neither shape nor size could be a confounding variable.

Sentences. There were 48 sentences constructed in total: 32 experimental and 16 filler sentences. Similar to the pictures, experimental sentences also formed pairs, with one sentence implying one color of an experimental and the other implying the color of the remaining item of the pair (see Figure 1). Participants viewed 16 experimental sentences and 16 filler sentences. Eight comprehension questions were added to half of the fillers to ensure that

[1] Amazon's Mechanical Turk is an Internet marketplace that enables businesses/researchers to recruit participants for surveys/social science experiments.

participants did not simply “skim” through a given sentence but read and understood it. Additionally, six sentence-picture pairs were used as practice trials.

Design and procedure. Design and procedure were almost identical to Connell (2007). There were four picture-sentence combinations, so four lists were created so that each group was presented with one of the possible combinations (see Figure 1). Each list contained the same proportion of experimental and filler sentences, and the various colors present in the pictures were spread evenly across groups. Thus, the experiment was a 2 (sentence version: Type 1, Type 2) × 2 (picture type: match, mismatch) × 4 (lists) design, with sentence version and picture type as within-subjects variables and lists as a between-subjects variable.

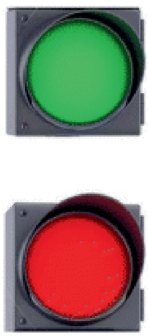
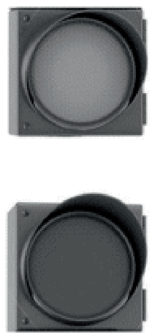
Sentence	Color (Exp.1)	Saturation (Exp.2)	Grayscale (Exp.3)
The driving instructor told Bob to **stop/go** at the traffic lights.			

Figure 1. Example of stimuli material used in each experiment. A matching picture illustrates that color was implied by the sentence (i.e., red when asked to stop at a traffic light), and a mismatching color illustrates that this color was not implied by the sentence.

Participants were instructed to read the sentence and press the spacebar when they had understood it. They were informed that each sentence would be followed by a picture, and their task was to decide whether the depicted object was mentioned in the preceding sentence. Participants were asked to respond as quickly and accurately as possible by pressing the *L* key for *yes* and the *A* key for a *no* answer. The responses were collected and saved automatically by the Qualtrics Survey Software. The instructions presented to the participants warned them that occasionally they would receive a question to test their comprehension of the previous sentence, to which they would either agree (by pressing the *L* key) or disagree (by pressing the *A* key). The trial sequence was as follows: a left aligned vertically centered fixation cross appeared on the screen for 1,000 ms followed by the sentence. After a spacebar press, a fixation cross was presented in the middle of the screen for 500 ms followed by a picture. When a yes/no decision was made, a blank screen appeared for 500 ms, after which another trial began (see Figure 2).

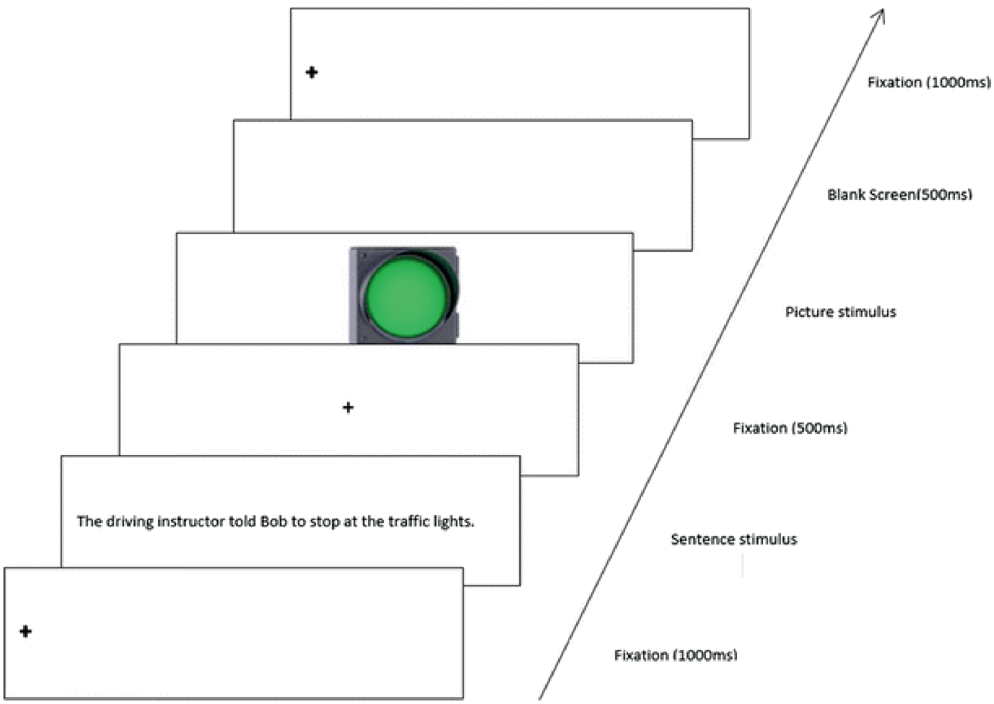


Figure 2. Example of an experimental trial sequence.

All experimental items required a *yes* response, and all filler items required a *no* response. As participants received six practice trials, it was clear for participants when a *yes* and *no* response was required.

Results and Discussion

The data from 42 participants were discarded from further analysis: five participants were not native English speakers, six participants reversed the response keys (which was indicated by accuracy scores at or below 21%), and 31 had accuracy scores lower than 80%. The drop-out rates were not equally spread across the four lists. To make the cells equal and enable parametric tests to be run, the required number of participants who were at the bottom of each list was removed (total of 25). After the exclusion process, each list included 34 participants (136 participants in total). For the analysis, we collapsed participants across lists as list was not a factor in our preregistered plan of analysis. Finally, one item was removed from the analysis as the average item accuracy was below 80%. This would indicate that participants did not believe the pictured object belonged to the preceding sentence.

Earlier research using the picture verification paradigm has used the median instead of mean reaction times (e.g., Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012; Zwaan et al., 2002). An advantage of using medians compared to means is that their use does not necessitate further decisions regarding outlier removal (e.g., whether to use cutoffs based on standard deviations, absolute RTs, or a combination thereof).

A paired-samples t test was conducted to investigate whether there was a match advantage for accuracy and RTs. For the RT analysis, only RTs of correct responses were included in the analysis. The participants showed significantly higher accuracy rates in the match condition (M= .96, SD = .06) than in the mismatch condition (M = .90, SD = .11), $t(135)$ =

5.36, $p < .001$, $d = 0.46$, BF10 = 33380.05. The match advantage was also found in the RTs: the match condition was 104 ms faster than the mismatch condition ($M = 1{,}230$ ms, $SD = 568$ ms and $M = 1{,}334$ ms, $SD = 676$ ms, respectively). This difference was significant, $t(135) = 3.00$, $p = .003$, $d = 0.26$, BF10 = 6.88. Participants' accuracy when responding to the comprehension questions was high ($M = 0.79$, $SD = 0.20$).

These findings support the results of Zwaan and Pecher (2012), rather than those of Connell (2007), and suggest that color, like shape and orientation, is an object property that is simulated during language comprehension.

Experiment 2A

The results of Experiment 1 served to illustrate that sentences implying color are represented in mental simulations but makes no conclusions as to how rich these simulations are. If color is not present in mental simulations, then reducing color saturation should not affect the match advantage. If we do simulate color, however, and do so vividly, then showing a mismatching pictures in full color should lead to a larger disparity between the two conditions than when saturation of the color is reduced.
Experiment 2 examined this problem by reducing color saturation to the lowest level at which the hue can still be distinguished to test whether a match effect would still appear, and whether it would be smaller than in Experiment 1.

Norming Study

A norming study was conducted in order to determine the lowest saturation level possible at which a certain hue could still be recognized using the same pictures as in Experiment 1. Twenty-four subjects were shown six different saturation levels per picture and were asked to choose the picture that had the lowest level of saturation while they could still perceive the associated hue.

Picture saturation was adjusted using Microsoft Office Picture Manager's Color Enhancement Tool (where −100 is a black and white/grayscale picture and 100 is a very intense, color-rich picture). The pictures that were selected by the majority of the participants as having the least amount of color while still being able to recognize the hue were used in the experiment.

Method

Participants. Two hundred and eight participants (99 males, mean age 37.93 years, range: 22–71 years) took part in this Mechanical Turk experiment. The participants were paid $1.50 for their participation.

Materials. The stimuli used in the current experiment were adapted from Experiment 1, and the levels of saturation chosen for the stimuli were determined by the norming study described above (see Figure 1). The sentences remained unchanged.

Design and Procedure. The design and procedure of Experiment 2 were identical to that of Experiment 1.

Results and Discussion

Sixty-eight participants were excluded from the analysis (five were not native English speakers; four appeared to have reversed the keys; 14 had accuracy below 80%; and 45 participants were excluded from the bottom of the lists to achieve equal numbers of subjects per list), leading to a total of 140 participants being included in the analysis.

A paired-samples *t* test was conducted to investigate whether there was a match advantage for accuracy and RTs. The results indicated no difference in accuracy rates between the match ($M = .96$, $SD = .07$) and mismatch condition ($M = .95$, $SD = .08$), $t(139) = 0.98$, $p = .331$, $d = 0.08$, BF10 = 0.15. There was also no difference in the RTs between the match ($M = 1{,}156$ ms, $SD = 558$ ms) and the mismatch conditions (*M* = 1,165 ms, *SD* = 639 ms), $t(139) = 0.25$, $p = .801$, $d = 0.02$, BF10 = 0.10. Comprehension accuracy was high (*M* = 0.81, *SD* = 0.19).

Experiment 2B

There was some concern that Experiment 2 could not be accurately tested using Mechanical Turk as there would be no way to control for the brightness of participants' computer monitors. To cope with this limitation, we replicated Experiment 2 in the lab at Erasmus University Rotterdam, using International Psychology students who participated for course credit.

Method

Participants. As the current experiment was run in the lab, we were constrained in the number of participants we could recruit (to a greater extent than on Mechanical Turk), and therefore we aimed to include 80 participants in the analysis. Ninety participants (23 male, mean age 20.02 years, range: 17–29 years) were recruited from the first year International Bachelor of Psychology students at the Erasmus University Rotterdam, where their English proficiency had to be sufficient, as determined by having a TOEFL grade above 80. Participants were tested in the lab, which is equipped with 22-in. TFT screens with a resolution of 1920 × 1200 and a ratio of 16:10.

Materials. The same materials as in Experiment 2A were used.

Design and Procedure. The design and procedure were identical to Experiments 1 and 2A, except that participants were tested in the lab.

Results and Discussion

Ten participants were excluded from further analysis: three appeared to have reversed the keys and seven performed below the 80% accuracy cutoff. Like the other experiments, use of these exclusion criteria were preregistered on the

OSF before data collection began. Eighty participants were included in the analysis. Furthermore, one item was removed from the analysis as an item analysis revealed an accuracy below our 80% cutoff. Fifteen experimental item pairs remained in the analysis.

A paired-samples *t* test was conducted to investigate whether there was a match advantage for accuracy and RTs using a stimulus set low in contrast, with saturation levels reduced to a point at which the hue was still recognizable. The results indicated no significant difference in accuracy scores between the match (M = .96, SD = .07) and mismatch conditions (M = .95, SD= .08), $t(79)$ = 0.62, p = .534, d = 0.07. Participants produced faster responses in the match than in the mismatch condition (M = 846 ms, SD = 355 ms and M = 926 ms, SD = 548 ms, respectively), but this did not reach statistical significance, $t(79)$ = −1.78, p = .080, d = 0.20, BF_{10} = 0.55. Comprehension accuracy was high (M = 0.82, SD = 0.14).

The results from Experiment 2B also support the results from Experiment 2A, as neither experiment found conclusive evidence for a match effect.

Experiment 3

To further determine the effects of reduced saturation on the match advantage, Experiment 3 was run using the same pictures as Experiment 1 and 2, except they were shown in grayscale. As no hue is present in grayscale photos, no significant difference between the match and mismatch condition is expected.

Method

Participants. Two hundred and twenty-two participants (98 males, mean age 38.64 years, range: 19–71 years) took part in the current study, and were recruited from Mechanical Turk and paid $1.50 for their participation.

Materials. The pictures used in this experiment were adapted from those used in Experiment 1 such that they were depicted in gray shades (see Figure 1). The gray shades were achieved by changing the pictures to black and white by using Paint.NET software. The sentences remained unchanged.

Design and Procedure. The design and procedure were identical to that of Experiments 1 and 2.

Results and Discussion

Forty-two participants were excluded from further analysis: Two reported that English was not their first language, seven appeared to have reversed the keys, 12 performed below the 80% accuracy cutoff, and 21 last-run participants were removed to equate the number of subjects per list. One hundred and eighty participants were included in the analysis.

A paired samples *t* test was conducted to investigate whether there was a match advantage for accuracy and RTs using pictures portrayed in grayscale. The results indicated that accuracy rates in the match condition (M = .97, SD = .06) and in the mismatch condition (M = .96, SD = .08) did not significantly differ, $t(179) = 1.89$, $p = .06$, $d = 0.14$. In the RTs there was also no significant difference between the match (M = 1,239 ms, SD = 641 ms) and mismatch conditions (M = 1,243 ms, SD = 558 ms), $t(179) = 0.21$, $p = .834$, $d = 0.02$, $BF_{10} = 0.09$. Comprehension accuracy was high (M = 0.81, SD = 0.19).

The results of Experiment 3 suggest that, when pictures are shown completely in grayscale, there is no significant match advantage present.

Exploratory Analyses

We were interested in examining exactly how the match and mismatch conditions differed from each other across experiments. As such, we conducted several exploratory analyses to gain a better appreciation of the processes that are occurring.

We conducted a repeated-measures ANOVA over the reaction time data to examine the differences between Experiments 1, 2A, and 3, where "experiment" was the between-subjects factor, and we found that there was a significant main effect of condition, $F(1, 453) = 5.01$, $p = .026$, and a significant interaction between condition and experiment, $F(2, 453) = 3.30$, $p = .038$. No main effect of experiment was found, $F(2, 453) = 1.58$, $p = .207$. On the basis of these results we decided to run additional analyses to see whether the RTs from Experiment 2A significantly differed from Experiments 1 and 3 per condition. A simple contrast revealed that the RTs in the mismatch condition were significantly faster in Experiment 2A, $t(274) = -2.26$, $p = .024$, than in Experiment 1. No further significant interactions were found (see Figure 3).

General Discussion

Previous research on the presence of color in mental simulations has come up with some contradictory findings (e.g., Connell, 2007; Zwaan & Pecher, 2012). One of the aims of the current study was to conclusively establish whether color is simulated or not. A second aim was to discover how much perceptual information is present in a mental simulation. Many object features have been studied in the past, but color is the only feature that can be decomposed while the object's identifiability remains unchanged. It may be argued that the match advantage exists because if a picture matches the perceptual image in the mental simulation, then response time is facilitated. When there is a mismatch, this facilitation cannot occur and may instead result in interference, leading to longer response times.

2

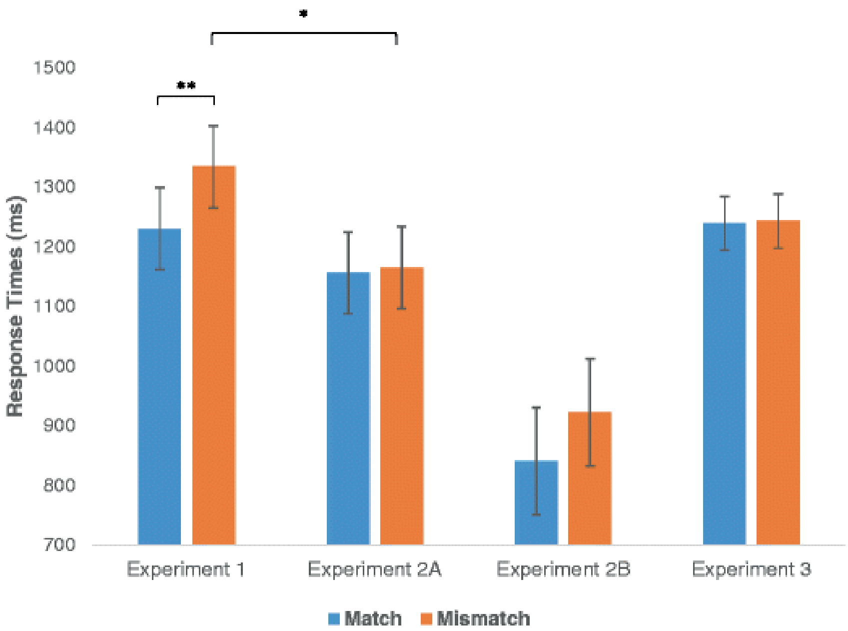


Figure 3. Size of match advantage in Experiment 1, when pictures were shown at normal levels of saturation; Experiments 2A and 2B, when saturation was reduced; and Experiment 3, when all pictures were shown in grayscale. $^{**}p < .01$. $^{*}p < .05$.

In order to successfully complete our first aim, the experiments used more stringent criteria for the stimuli compared to what was used by Connell (2007) and Zwaan and Pecher (2012), as they included some items that could change shape as well as color. Furthermore, the median reaction time rather than the mean was used, such as in the replication by Zwaan and Pecher (2012) and other studies using a similar paradigm, as it allowed for less data to be discarded and was in line with the methods of previous research.

Experiment 1 found a significant match advantage of 104 ms, which supports the hypothesis that color is indeed present in mental simulations and supports the results by Zwaan and Pecher (2012) and Connell and Lynott (2009). In order to examine the richness of mental simulations and thus address our second goal, Experiment 2 used items where the saturation of the color was reduced to the lowest point at which the hue was still recognizable. The results of this experiment found no significant difference between the match and the mismatch condition. Interestingly, however, exploratory analyses revealed that the RTs in the mismatch condition were significantly faster in Experiment 2A than in Experiment 1, while no difference was found for the match condition. The results from Experiment 2A therefore serve to illustrate two points: First, the match advantage disappears when saturation in pictures is lowered, and second, the reason it disappears is due to a speeding up of response time in the mismatch condition. These results are intriguing as they suggest that, rather than a picture being *more* of a match leading to faster response times (i.e., facilitation), it would suggest that the match effect appears due to there being a *vivid* difference between the pictured object and the simulation in the mismatch condition, leading to interference effects.

Experiment 3 provides tentative evidence in support of this hypothesis as well, as the average response times of this experiment appear to fall in between those of Experiment 1 and Experiment 2A, although this difference does not reach significance. As the average difference in reaction time between Experiment 2A and 3 is only 8 ms, it is unrealistic to expect a significant difference using a between-subjects analysis. It would be interesting for future studies to examine, using a within-subjects paradigm, at which level of saturation color can aid object recognition. Although the between-subjects comparison in our exploratory analyses were not significant, such future studies could illustrate that the mere presence of color—even at the lowest level of saturation during which the hue is still recognizable—serves to enhance performance in the object-verification task. Indeed, this is supported by the general literature stating that color aids in object recognition (Bramão et al., 2011).

In addition to finding a match advantage in the RTs of participants in Experiment 1, we also found a significant reduction in accuracy in the mismatch condition. As we removed items that had an average accuracy below 80%, this reduction cannot be explained by the pictures in that condition not matching the sentence. The match advantage in the RTs bear no relation to the accuracy scores, as only the RTs of accurate responses were used. This reduction in accuracy scores, however, could serve to explain why a match effect exists at all. We previously argued that the match effect exists due to a vivid difference occurring in the mismatch condition between the pictured object and the simulation. The task participants had to complete required them to only examine whether the actually pictured object (with no instructions mentioning color) was mentioned in the previous sentence. A strategy that could aid in the completion of such an object-verification task—in which speed is important—could be that participants simply judge whether the picture they see overlaps with what is present in the mental simulation. When there is a vivid difference, or no overlap, between the picture and the simulation, they are more likely to answer with an incorrect *no* response. It would be interesting to examine whether the removal of the instructions requiring speed would eliminate the difference in accuracy between the two conditions.

As for the "richness" of our mental simulations, we can conclude that they are rich indeed, in the sense that they include multiple object properties. We already know that color can be decomposed into different dimensions, namely hue, saturation, and brightness. If the reduction in the level of one of these dimensions (in our study: saturation) had not reduced the match advantage, we would have had to argue that color would not be present or relevant in a mental simulation. Our study, however, found that by reducing saturation, the match advantage disappears. Furthermore, we found tentative evidence that the mere presence of color—even with low levels of saturation—can aid in object recognition, compared to when color is removed entirely.

In sum, the current study found further support that color is another object property that is represented in mental simulations, in addition to shape and orientation. Furthermore, we have shown that by reducing saturation of the picture shown we can remove the match advantage as well, while still being involved in object recognition. This leads to the conclusion that, when comprehending language, we build mental simulations rich in perceptual detail.

3

How Are Mental Simulations Updated Across Sentences?

This chapter has been published as:

Hoeben Mannaert, L.N., Dijkstra, K., & Zwaan, R.A. (2019). How are mental simulations updated across sentences? *Memory & Cognition, 47*(6), 1201-1214. doi:10.3758/s13421-019-00928-2

Abstract

We examined how grounded mental simulations are updated when there is an implied change of shape, over the course of two (Experiment 1) and four (Experiment 2) sentences. In each preregistered experiment, 84 psychology students completed a sentence-picture verification task in which they judged as quickly and accurately as possible whether the pictured object was mentioned in the previous sentence. Participants had significantly higher accuracy scores and significantly shorter response times when pictures matched the shape implied by the previous sentence, than when pictures mismatched the implied shape. These findings suggest that, during language comprehension, mental simulations can be actively updated to reflect new incoming information.

Keywords: grounded cognition, mental representation, situation models, language comprehension, discourse, updating

Introduction

Imagine you are reading a story about an eagle that is flying in the sky. The eagle continues to soar for a while before eventually landing in its nest and going to sleep. If you were to make a drawing of the eagle's final shape, would the wings of the eagle be folded? Anyone with some knowledge of how birds rest in a nest would know that the answer to this question is "yes", but would not have forgotten the bird's initial shape. According to the perceptual symbol systems theory (Barsalou, 1999), during language comprehension we activate modal symbols that can be combined in a mental simulation, which may involve the same neural structures as would be used if we were to see the described event in real life. According to Sato, Schafer, and Bergen (2013), these mental simulations are formed incrementally, suggesting that, while reading this story about an eagle, we initially created a mental simulation of the eagle having spread wings as it was flying in the air, followed by the inclusion of an eagle with folded wings in the simulation later on. What remains unclear is whether or not the final object replaced the initial simulation, or whether both object states remain activated in that simulation.

Although there is still some debate as to whether mental representations require sensorimotor input or whether amodal symbols are required (e.g., Mahon & Caramazza, 2008), many cognitive psychologists believe that sensorimotor input is required in some form to support language comprehension (Barsalou, 1999; 2008). Indeed, many behavioral experiments have shown that mental simulations can include an object's shape (Zwaan, Stanfield, & Yaxley, 2002), color (Hoeben Mannaert, Dijkstra, & Zwaan, 2017; Pecher & Zwaan, 2012), orientation (Stanfield & Zwaan, 2001), size (De Koning, Wassenburg, Bos, & Van der Schoot, 2016), and movement during language comprehension (Gentilucci, Benuzzi, Bertolani, Daprati, & Gangitano, 2000; Glenberg & Kaschak, 2002).

In the past it has been difficult to tease apart whether the match effects found in word- or sentence-picture verification tasks are due to the visual system being utilized, or whether it is only the conceptual system being recruited for the task (Ostarek & Huettig, 2017). However, recent studies utilizing a technique called 'continuous flash suppression' in word-picture verification tasks, where the picture shown is rendered practically invisible by disrupting the processing of visual stimuli (Lupyan & Ward, 2013; Ostarek & Huettig, 2017), have provided evidence that spoken words also activate low-level visual representations. These findings provide support for the idea that conceptual representations also involve the visual system. Moreover, many neuroimaging studies have also illustrated that modality-specific sensory, motor, and affective systems are involved during language comprehension (Binder & Desai, 2011; Hauk, Johnsrude, & Pulvermüller, 2004; Sakreida et al., 2013; Simmons, Ramjee, Beauchamp, McRae, Martin, & Barsalou, et al., 2007), illustrating that both behavioral and neuroimaging studies support the idea that mental simulations involve sensorimotor activation.

Many studies have targeted the question of *what* is represented in a mental simulation, but more and more researchers are now focusing on *how* mental simulations unfold across texts and their relevance for language comprehension. For instance, a study by Kaup, Lüdtke, and Zwaan (2006) illustrated that responses to pictures that matched the situation described in a preceding sentence were facilitated when the sentences were affirmative (e.g. *The umbrella was open*), but only after a 750 ms delay. This facilitation was no longer present at 1500 ms, suggesting that the representation may have deactivated at that point in time. Sentences described negatively (e.g. *The umbrella was not closed*), however, only led to facilitation after a 1500 ms delay, but not after 750 ms. These findings provide evidence for the idea that mental simulations require additional processing time if sentences are complex, and that these simulations can become deactivated after a period of time.

Interestingly, it has also been shown that these simulations can be reactivated at a later point in time, if the context requires it. For example, reading sentences implying a certain object shape or orientation leads to faster responses in an object-verification task performed after a 45 minute delay, illustrating that a mental simulation formed during the reading of a sentence can be reactivated if necessary at a later point in time (Pecher, Van Dantzig, Zwaan, & Zeelenberg, 2009). Furthermore, several studies examining the influence of imperfective and perfect verb aspect have shown that mental simulations remain activated longer when there is a description of an ongoing situation (e.g., *The boy was building a doghouse*) compared to a description of a situation that has already occurred (e.g., *The boy had built a doghouse*) (Madden & Therriault, 2009; Madden & Zwaan, 2003; Magliano & Schleich, 2000). These studies support the idea that grammatical markers provide cues for how a situation model should be constructed and updated.

Radvansky and Zacks (2011; see also Zwaan & Madden, 2004) explain that this updating process can take various forms: firstly, a new situation model can be constructed (*model creation*); secondly, if new information is consistent with the current situation it can be incorporated into the existing model (*model elaboration*); thirdly, the model can be altered to accommodate new information (*model transformation*); and finally, it can merge two models into one (*model blending*). Model creation may be utilized when there is a shift to a new event in a narrative, for instance when a character moves from one location to the next (Radvansky, 2012). Model transformation occurs when, for example, contradictory information needs to be integrated into the existing situation model. For example, if a person claims to be a vegetarian, but subsequently eats a hamburger (cf. Albrecht & O'Brien, 1993), this would require the model to be transformed to accommodate this contradiction. Model elaboration is what occurs when more information about the current situation is added without requiring a structural change (e.g., reading about a jogger in a marathon and eventually noticing that his shoes are blue. Model blending happens when initially two events are perceived to be distinct, but eventually are considered to be part of the same event. For example, if a man walks into another room to pick up his coat it may initially be perceived as an event boundary – and thus a separate situation model would need to be constructed. However, once he returns to the room he came from while putting on the coat, it becomes clear that the grabbing and wearing of the coat is part of an ongoing event, and thus, the events need to be blended into one coherent situation model.

Given that there are various ways in which a situation model can be updated, how would mental simulations be affected during language comprehension? If mental simulations are simply “the reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind” (Barsalou, 2008, p.618), then would the process of updating a situation model lead to the activation of multiple states in a mental simulation, or would only the “final” state stay activated? If situation models are representations of the text that underpin language comprehension, then any changes occurring in a mental simulation – an important subcomponent of the situation model – would influence the construction of the situation model. In other words, if mental simulations are directly involved in the comprehension process, then we would expect mental simulations to update dynamically as a narrative unfolds. However, if mental simulations are not directly involved, but rather a function that activates all relevant perceptual input – which is subsequently distinguished by higher order cognitive processes – then it would make sense for all relevant perceptual input to be activated in a mental simulation. As such, in order to gain a full understanding of how mental simulations are involved in language comprehension, it is important to find out how the updating process affects mental simulations. Most studies have only focused on situation model updating by looking at the slowdown of reading as a measure for this updating process, but these studies fail to provide a complete picture of this updating process.

When reading about inflating a balloon, it is of course interesting to know that there is an increase in processing time necessary to comprehend the changes to the situation model, but what exactly happens to the representations described by that sentence? Does the initial shape of the balloon (i.e. deflated) stay activated as the final shape (i.e. inflated balloon) also activates? Or does the deflated balloon representation deactivate as the inflated balloon activates? An fMRI study by Hindy, Solomon, Altmann, and Thompson-Schill (2015) suggests that the brain does encode object state changes, using short sentence items such as “inflate the balloon” to investigate this. Their results suggested that the ventral visual cortex encodes both the initial object state (e.g., deflated balloon) and the final object state (e.g., inflated balloon). Furthermore, the authors suggest that the posterior parietal cortex may be recruited for conceptual binding, so that the distinct states are bound together in a stable representation. These findings suggest that we have to know what an initial object state is in order to comprehend that a change in state is occurring. Possibly, this could mean that when a change in an object state is described, that both object representations are activated in a mental simulation.

To our knowledge, only one study thus far has explored what happens to mental simulations of changing object states. Sato et al. (2013) were interested in finding out whether mental simulations are formed incrementally (i.e., while reading a sentence) or after all the information has been obtained (i.e., at the end of the sentence). If a mental simulation is only formed once all the information has been collected, then it would be unlikely for an initial object state to be activated. If a mental simulation is formed incrementally, however, multiple object states could become activated and potentially deactivated during the process of comprehension. To explore this, Sato et al. had participants read a Japanese sentence where an expectation of an object shape was created, but contradicted at the end of the sentence. For example, in one item participants read about a person wearing a *yukata* (a cotton kimono) to a fireworks festival, but the *yukata* had been torn apart. As Japanese is a verb-final language, this was an ideal medium by which to create an expectation of an object shape (i.e. a whole *yukata*) and examine what a subsequent contradiction would do to the mental simulation of this event. In order to test which shape would be simulated, they used a picture verification task either before or after the final verb in the sentence. Their results showed that participants responded faster to the picture matching the shape implied both before and after the final verb, suggesting that mental simulations are formed incrementally, and that an initial shape can be deactivated in a mental simulation if a person is presented with information that contradicts earlier expectations.

What remains unclear, however, from Sato et al.'s (2013) findings, is whether an initial shape activation would become deactivated in a mental simulation when no contradictory information is supplied. For example, when reading a sentence pair such as: "The eagle was moving through the air. That evening the eagle was resting in its nest." the latter sentence does not contradict the event that took place earlier on. Instead, the eagle has spread wings while it is flying, which changes to an eagle having folded wings once it is resting later on. At the end of this sentence pair, would a person still have the initial object state active in a mental simulation, or would this have been deactivated? We are interested in finding out whether implied changes in shape lead to the simultaneous activation of both object shapes, or whether only the final shape remains active in the mental simulation. As most studies examining situation model updating use reading times as a dependent variable, using a sentence-picture verification task offers a unique way to explore how these mental simulations unfold, and whether this updating process requires both visual representations to be active simultaneously, or whether the final representation becomes active as the prior one deactivates. Such a task allows us to gleam additional information regarding the nature of mental simulation updating, which cannot be done with reading time tasks alone.

The Present Study

Two experiments were conducted to examine how mental simulations are updated when changes in implied shape are described over the course of several sentences, using a sentence-picture verification task. Participants had to read either two (Experiment 1) or four (Experiment 2) sentences that described either a change in shape (change condition) or no change in shape (constant condition), followed by a picture that either matched or mismatched the shape implied by the final sentence, where they had to decide whether the pictured object was mentioned in the text.

Given that previously simulated information can be reactivated when necessary (Pecher et al., 2009; Sundermeier et al., 2005) and can continue to influence language comprehension in the future (O'Brien, Cook, & Guéraud, 2010), it is possible that, in a context where both object states are implied immediately one after the other, both would remain activated in order to update the situation model. This could potentially occur during model elaboration as opposed to model creation (Radvansky & Zacks, 2011). It is also possible, however, that due to mental simulations being formed incrementally, that the initial object state is replaced immediately when the second state is mentioned. This explanation would fall in line with what is proposed in the Event Horizon Model (Radvansky, 2012). According to the Event Horizon Model, a new model is created at event boundaries and could therefore suggest the initial object state is replaced immediately. If the newer object state is then activated in the new situation model, then response times should be shorter when the new object is displayed compared to when the first object is displayed. As such, if reading about a change in an object's shape is considered to be an event boundary, then we can expect shorter response times when the picture shown matches the changed object's shape compared to when it mismatches this shape. If instead the model is merely elaborated upon, then no differences would be expected when the picture shown matches the first- or last-mentioned object.

We hypothesized that in Experiment 1, where one shape is mentioned directly after the other, that both shapes would remain activated in the mental simulation as we expected model elaboration to occur, rather than creation. Therefore, we predicted to find no significant differences in the response times between the match and mismatch condition in the shape change condition. We did expect a match effect in the shape constant condition, as this would be consistent with findings from previous studies (e.g., Zwaan et al., 2001; Zwaan & Pecher, 2012). In Experiment 2, however, the final object shape was emphasized using the final three sentences (out of four total). In this experiment we expected that the initial shape would become deactivated as more emphasis was placed on the final shape. Specifically, we predicted a significant match advantage in the response times in both the shape change and shape constant conditions in this experiment. We made no specific predictions regarding the accuracy scores in both Experiment 1 and 2.

Ethics Statement

The participation in all experiments and in the norming study was voluntary. The participants subscribed to the experiments online via the university platform, and were told that by signing up for a study, they declare to voluntarily participate in this study. They were briefed with the content of each study but obtaining further written consent was not required by the Ethics Committee of Psychology at the Erasmus University Rotterdam, The Netherlands, who approved the project, because the experiments were non-invasive and the data collected was processed anonymously.

Preregistration

The predictions, exclusion criteria, design, methods, analyses, and materials of all the experiments reported in this article were pre-registered in advance of data collection and analysis on the Open Science Framework (OSF) to ensure confirmatory procedures were conducted according to a priori criteria.[2] Analyses that were not pre-registered are referred to in this article under the heading “Exploratory Analyses”.

Experiment 1

Method

Norming study. In the current study, participants read sentences where a certain shape is implied, rather than explicitly stated. In order to ensure that the sentences did imply the shape we intended them to, a norming study was conducted. Forty-one participants were recruited via Mechanical Turk (www.mturk.com) and were paid $3.50 for completion of the survey, which took approximately 30 minutes to complete. The participants had a mean age of 35.10 years (*SD* = 10.87) and consisted of 19 females and 22 males.

[2] The pre-registration can be viewed on https://osf.io/f7m3q/register/565fb3678c5e4a66b5582f67

Participants in the norming study read 280 sentences (seven sentences were written for 40 different items for Experiment 1 and 2) which implied a certain object shape (e.g., *The bat was entering the cave*) and had to determine which of the shown pictures best matched the sentence shown. There was also an option where participants could state that neither picture matched the sentence they read and could provide comments to elaborate on their reasoning. By doing the norming study in this manner we could ensure that the final items we would end up using would actually imply the shape we wanted it imply. For example, for the item about an eagle, 91.30% of participants agreed that our matching picture corresponded to the shape implied by the sentence "the eagle was moving through the air", while 82.61% agreed that our matching picture corresponded to the sentence "The eagle was resting in its nest". If we consider an item to be the object of the sentences, then each item would lead to seven ratings per participant (e.g. one rating for the sentence about the eagle moving through the air, another rating for the sentence about the eagle resting in the nest, and so on). For the results, we looked at the sentence of an item that contained the lowest percentage of agreement across participants in order to be conservative in our final selection of stimuli. The results of the study illustrated that the stimulus set contained several items where the minimum average agreement on what shape matched which sentence was less than 61%. Furthermore, on some of the items participants made comments suggesting the item contained an incorrect logical match between sentence and picture. To improve the quality of the stimulus set and to ensure the final set contained a number of stimuli that was a multiple of four (due to counterbalancing), 12 items were removed. This left us with 28 experimental items that had a minimum average agreement of 77.52% (*SD* = 7.73%) regarding which picture best represented which sentence.

Participants. We expected to need 84 participants to find an effect if it existed – based on a power analysis performed on the results of the Zwaan et al. (2002) study – and therefore continued data collection until this goal had been met, replacing participants that had to be excluded due to having total accuracy scores below 80%. As a result of this measure, only one participant was excluded and replaced. The final sample consisted of 84 participants (aged 17-29 years, M_{age} = 20.33, SD_{age} = 2.04, 71 females, 13 males) who were students of the International Bachelor Programme of Psychology (IBP) at the Erasmus University Rotterdam in The Netherlands. IBP students are required to have a minimum level of 80 on the Test of English as a Foreign Language (TOEFL), or a 6.0 on the International English Language Testing System (IELTS), or be native speakers of English. Participants were reimbursed with course credit.

Sentences. Twenty-eight experimental sentence pairs and 28 filler sentence pairs were used in the experiment for each experiment block. The sentence pairs in one block involved an implied change in shape (change condition) and the sentences in the other block involved no change in shape (constant condition). See Figure 1 for an example of the sentence stimuli. Of the experimental sentence pairs, two versions were created for counterbalancing purposes to account for typicality effects. The first version ensured that the final sentence implied one of two possible object shapes (e.g., eagle – folded wings), while the second version ensured the final sentence implied the other possible shape (e.g., eagle – spread wings). To be able to counterbalance in this way without manipulating time aspect in the sentences, we could only use objects that were reversible in shape. For example, an eagle is able to fold its wings and then spread them again, and vice versa. An egg, however, could only ever start out as a whole egg and then become a fried egg, not the other way around.

		Picture	
Shape	**Sentence**	**Match**	**Mismatch**
Change	The eagle was moving through the air. That evening the eagle was resting in its nest. *The eagle continued perching in its nest. The eagle looked at ease in the nest.*		
Constant	The eagle was moving through the air. That evening the eagle was still moving through the air. *The eagle continued flying in the air. The eagle looked at ease in the air.*		

Figure 1. Example of stimulus material used in Experiment 1 and 2 (italics). Experiment 2 used the first sentences from Experiment 1 and had two sentences added at the end to continue emphasizing one shape. Pictures under "Match" refer to a picture matching the shape implied by the final sentence, where "Mismatch" refers to a picture mismatching the shape implied by the final sentence. The full list of stimuli can be viewed on: https://mfr.osf.io/render?url=https://osf.io/3hrq8/?action=download%26mode=render

Pictures. Eighty-four pictures were created for use in this experiment, of which 56 were used for the experimental sentences, and 28 for the filler sentences. During the experimental procedure, each sentence pair was followed by a picture that either matched or mismatched the implied shape, thus requiring the creation of two picture versions of a particular object. For example, if the sentence stated that the eagle was in the air, a picture of either an eagle with spread wings or a picture of an eagle with folded wings could be shown. The pictures were obtained from the internet (Google image search engine) and were edited with the Paint.NET software to be displayed as grayscale (to ensure that effects of color could not confound the results) to not exceed a 300 x 300 pixel resolution (approximately 7.9 x 7.9 cm on screen). The experiment was programmed using E-Prime 2.0 Professional and participants completed the experiments in isolated cubicles with computers equipped with 24.1” TFT-IPS screens with a resolution of 1920 x 1200 and a ratio of 16:10.

Design and Procedure. The experiment was a 2 (match: match vs. mismatch) x 2 (shape: change vs. constant) x 2 (block: block 1 shown first vs. block 2 shown first) x 2 (sentence: version 1 vs. version 2) mixed-subjects design, resulting in eight counterbalanced lists. ‘Match’ and ‘shape’ were tested within-subjects, meaning that participants viewed pictures that matched and mismatched implied shape, and viewed sentences that either implied a change in shape (change condition) or did not (constant condition). ‘Block’ and ‘sentence’ were between-subjects variables that ensured that, firstly, half of all participants were shown block 1 during session 1 (containing only the change condition items) and then block 2 during session 2, while the other half were shown block 2 during session 1, followed by block 1 during session 2. Participants completed each session one week apart to ensure that there were no carryover effects from one set of materials to another. Secondly, each experimental sentence pair could imply one of two object states (i.e., a sentence implying an eagle with spread wings or a sentence implying eagle with folded wings). This variable was also tested between-subjects so that half of all participants viewed the sentence implying one of the object states and that the other half viewed the sentence implying the other object state. Within the same block participants received each item once. We counterbalanced the experiment in this way to ensure that neither block order nor problems with typicality could be possible explanations for the results.

Participants were instructed that they would perform a self-paced reading task using the spacebar and that they would see a picture after the second sentence, and that they must answer whether the pictured object was mentioned in the previous sentence. They were instructed to press the “L” key for YES responses and the “A” key for NO responses. Participants first received 5 practice sentences. A trial proceeded as follows: the > sign was shown for 1000ms in the center of the screen (left-aligned) signifying that they would receive a new sentence sequence. Following this, they saw the first sentence (left-aligned), and had to press space to signify they understood the sentence, before immediately seeing the second sentence on the screen. Once they pressed the spacebar they saw a fixation cross in the center (center-aligned) of the screen for 500 ms to prepare them for a response. We chose 500 ms as the interstimulus interval as this is commonly used in sentence-picture verification studies (e.g., De Koning et al., 2017; Hoeben Mannaert et al., 2017; Zwaan & Pecher, 2012). Following the fixation cross they saw the picture in the center of the screen, which remained on screen until they had given a response. All of the experimental items (i.e., both the match and mismatch conditions) required a YES response while all of the filler items required a NO response. Half of all filler items also ended with a comprehension question, to ensure they properly read the sentences, where they had to give a YES/NO response using the “L”/ “A” keys, respectively.

Data Analysis. A repeated measures analysis of variance (rmANOVA) was conducted to test whether participants were faster in their response and whether they were more accurate when the picture matched than when it mismatched for both the change and constant conditions. Only the response times for accurate responses were used in the final analysis, and the median reaction times were analyzed instead of the mean response times. Many experiments using the sentence-picture verification task use the median for the analyses as it eliminates the necessity to make decisions on outliers (such as use of cutoffs based on standard deviations, absolute RTs, or other methods) (Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012). Subject analyses are shown with the subscript "1" (e.g., F_1) and item analyses[3] are shown with the subscript "2" (e.g., F_2).

Results

Accuracy. Table 1 shows a summary of the descriptives for the change and constant conditions. The rmANOVA for accuracy demonstrated a main effect for 'match', where participants in the match condition responded more accurately than in the mismatch condition ($F_1(1,83) = 17.63$, $p < .001$, $\eta^2_p = 0.175$; $F_2(1,27) = 15.79$, $p < .001$)[4]. A main effect for 'shape' was also found, where participants in the shape change condition responded more accurately than participants in the shape constant condition ($F_1(1,83) = 8.71$, $p = .004$, $\eta^2_p = 0.095$; $F_2(1,27) = 4.24$, $p = .049$). There was no significant interaction between 'match' and 'shape' ($F_1(1,83) = 0.003$, $p = .959$; $F_2(1,27) = 0.21$, $p = .648$). Paired-sample *t*-tests showed that there was a significant match effect in both the shape change ($t_1(83) = 3.59$, $p < .001$) and the shape constant condition ($t_1(83) = 2.18$, $p = .032$). The item analyses showed no significant effect of 'picture version' ($F_2(1,27) = 0.06$, $p = .809$).

---

[3] The item analyses were not pre-registered, but were included as part of the main analyses as this is conventional.

[4] These analyses excluded the counterbalancing measures such as 'sentence version' and 'block order' as these were not included in our pre-registered analysis plan. When included, the results for 'match' ($F(1,76) = 17.59$, $p < .001$) and 'shape' ($F(1,76) = 9.86$, $p = .002$) remained the same. The effect sizes of the counterbalancing measures are not reported due to lack of theoretical relevance (Madden-Lombardi, Dominey, & Ventre-Dominey, 2017; Pollatsek & Well, 1995; Raaijmakers, Schrijnemakers, & Gremmen, 1999).

Response time. The rmANOVA for response time (performed on correct responses only) found a main effect for 'match', where participants responded significantly faster when the picture matched than when it mismatched the shape implied by the final sentence ($F_1(1,83) = 10.80$, $p = .001$, $\eta^2_p = 0.115$; $F_2(1,27) = 9.95$, $p = .004$)[5]. There was no significant main effect of 'shape' ($F_1(1,83) = 0.46$, $p = .502$; $F_2(1,27) = 0.84$, $p = .366$), nor a significant interaction between 'shape' and 'match' ($F_1(1,83) = 0.189$, $p = .665$; $F_2(1,27) = 0.21$, $p = .654$). Paired-sample *t*-tests showed that participants were significantly faster in the match condition in both the shape change ($t_1(83) = -2.51$, $p = .014$) and the shape constant condition ($t_1(83) = -2.02$, $p = .047$). There was also a significant effect of 'picture version' in the item analyses ($F_2(1,27) = 4.72$, $p = .039$).

Exploratory analyses.

Comprehension accuracy. To examine whether participants did properly read both sentences in the experiment, the accuracy scores to the comprehension questions were analyzed. Sixteen comprehension questions focused on the final sentence in the pair, while 12 questions focused on the first sentence in the pair. Average comprehension accuracy was high at 89.07% (*SD* = 12.20%) for questions focused on the first sentence as well as for questions targeted at the second sentence (*M* = 88.69%, *SD* = 11.61%). A paired samples *t*-test found no significant difference between comprehension accuracy for questions targeting sentence one and sentence two ($t(83) = .31$, $p = .761$).

Discussion

As can be seen from Table 1, overall accuracy of participants was high in each condition. However, even though it was high, participants still responded less

---

[5] These analyses excluded the counterbalancing measures such as 'sentence version' and 'block order' as these were not included in our pre-registered analysis plan. When included, the results for 'match' ($F(1,76) = 10.36$, $p = .002$) and 'shape' ($F(1,76) = .65$, $p = .424$) remained the same.

accurately when the picture mismatched the final object shape in both the change condition (2% difference) and the constant condition (3% difference). The participants' task was to respond as quickly and accurately as possible whether the pictured object was mentioned in the previous sentence. Even though this was clearly the case, participants still responded in such a way that, when the picture mismatched the implied shape, they were slightly less likely to press the 'YES' response. This could suggest that participants are more inclined to match the picture they see to the image that is present in their mental simulation, rather than compare it to the textbase. Furthermore, participants also responded slightly more accurately in the shape change condition compared to the shape constant condition. The Event Horizon Model (Radvansky, 2012) provides a potential explanation for this effect. According to this model, memory is enhanced during the perception of an event boundary. As such, it could be that the change condition led to the perception of an event boundary, but that this was not the case when there was no implied change in shape. However, the fact that these percentage differences are so small limits the overall strength of these conclusions.

We also found a standard match effect for both the shape change and constant conditions, which went against our expectations. The results suggest that the most recently implied shape is more highly activated than the first-mentioned shape in situations that involve a change in shape. What this suggests is that mental simulations appear to update in a manner that replaces the initially simulated object, rather than activating both objects in unison. This would mean that, although an object can be reactivated when necessary (such as in Pecher et al., 2009), this is not required for the purpose of updating mental simulations.

Additionally, it looks as though the updating process does not seem to take additional cognitive effort during object changes compared to when there is no change occurring, as no significant differences in response times were found between the change and constant conditions. It is also possible, however, that no differences were found between the change and constant conditions because both required model creation, as the sentence items included a time shift, which has been related to slowdown of reading times in past studies (e.g., Speer & Zacks, 2005) and thus could have also influenced response times during object-verification.

Experiment 2

Experiment 1 illustrated that the final implied object shape is more activated in a mental simulation than the initial one. In Experiment 2 we wanted to see whether this result would replicate using four sentences instead of two. Most studies using the sentence-picture verification paradigm only examine the match effect using one sentence, therefore it is of interest to see whether this effect holds out under more natural conditions. To test this, the same stimuli were used as in the Experiment 1, except that two more sentences were added to increase the time between the activation of the initial shape and the response. Thus, in each item, the first sentence implied one shape of an object, whereas the final three sentences implied either another shape of that object (i.e., in the shape change condition) or again the same shape (i.e., in the shape constant condition). We expected to find a match advantage for both the shape constant and the shape change conditions in Experiment 2.

Method

Participants. We again aimed to have 84 participants in our sample, and therefore continued data collection until this goal had been met, replacing participants that had to be excluded due to having total accuracy scores below 80%. As a result of this measure, seven participants were excluded and replaced. The final sample consisted of 84 participants (aged 18-47 years, M_{age} = 21.08, SD_{age} = 4.32, 62 females) who were IBP students at the Erasmus

University Rotterdam in The Netherlands, and had a minimum level of 80 on the TOEFL, or a 6.0 on the IELTS. Participants were reimbursed with course credit.

Materials. The sentence items contained four sentences. The first two sentences of the item were identical to the ones from Experiment 1. The final two sentences continued implying the shape mentioned in the second sentence. Aside from this change in the stimuli, all materials were the same as that of Experiment 1. An example of the stimuli can be seen in Figure 1.

Design and procedure. The design and procedure of Experiment 2 were the same as that of Experiment 1.

Results

Accuracy. The data from Experiment 2 were analyzed using the same method as in Experiment 1. Table 1 shows a summary of the descriptives for the change and constant conditions. The rmANOVA for accuracy found a main effect for 'match', where participants in the match condition responded more accurately than in the mismatch condition ($F_1(1,83) = 13.81$, $p < .001$, $\eta^2_p = 0.14$; $F_2(1,27) = 11.90$, $p = .002$)[6]. A main effect for 'shape' was also found in the subject analyses (but not in the item analyses), where participants in the shape change condition responded more accurately than participants in the shape constant condition ($F_1(1,83) = 4.09$, $p = .046$, $\eta^2_p = 0.05$; $F_2(1,27) = 3.57$, $p = .069$). There was also a significant interaction between 'match' and 'shape' ($F_1(1,83) = 4.74$, $p = .032$, $\eta^2_p = 0.05$; $F_2(1,27) = 8.63$, $p = .007$). Paired-sample *t*-tests showed that the match effect was significant for both the shape change ($t_1(83) = 2.34$, $p = .022$) and the shape constant condition ($t_1(83)$

[6] These analyses excluded the counterbalancing measures such as 'sentence version' and 'block order' as these were not included in our pre-registered analysis plan. When included, the results for 'match' ($F(1,76) = 14.19$, $p < .001$), 'shape' ($F(1,76) = 4.06$, $p = .047$), and the interaction between 'match' and 'shape' ($F(1,76) = 4.78$, $p = .032$) remained the same.

= 3.36, p = .001). There was no significant effect of 'picture version' ($F_2(1,27) = 0.53$, p =.473).

Response time. The rmANOVA for response time (performed on correct responses only) found a main effect for 'match' in the subject analyses (but not in the item analyses), where participants responded significantly faster when the picture matched than when it mismatched the shape implied by the final sentence ($F_1(1,83) = 5.54$, p = .021, $\eta^2_p = 0.06$; $F_2(1,27) = 2.28$, p = .143)[7]. There was no significant main effect of 'shape' ($F_1(1,83) = 0.25$, p = .615; $F_2(1,27) = 0.08$, p = .776), nor a significant interaction between 'shape' and 'match' ($F_1(1,83) = 0.22$, p = .638; $F_2(1,27) = 1.10$, p = .304). Paired-sample t-tests illustrated no significant match effect for the shape change condition ($t_1(83) = -1.99$, p = .050) and the shape constant condition ($t_1(83) = -1.21$, p = .228). Figure 2 illustrates a boxplot comparison of the response times per condition for Experiment 1 and 2. There was also no significant effect of 'picture version' ($F_2(1,27) = 0.09$, p = .772).

[7] These analyses excluded the counterbalancing measures such as 'sentence version' and 'block order' as these were not included in our pre-registered analysis plan. When included, the results for 'match' ($F(1,76)$ = 4.75, p = .032), and 'shape' ($F(1,76)$ = 0.07 p = .797) remained the same.

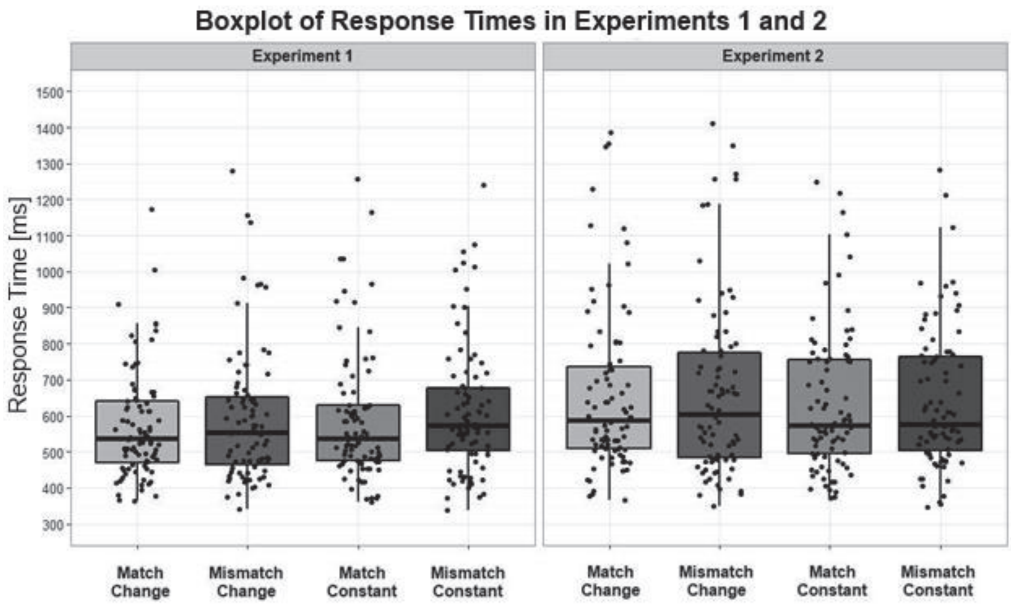


Figure 2. Box plot illustrating a comparison of response times and spread of data for Experiment 1 and 2.

Exploratory analyses.

Meta-analysis. We performed a meta-analysis on the data to examine the evidence for a match effect existing in the shape change and shape constant conditions (see Figure 3). The meta-analysis was performed using R 3.3.2 (R Core Team, 2016) using the package *metafor* (Viechtbauer, 2010). The code can be viewed in the Appendix. The results illustrated a 29.31 ms match advantage in the shape change condition, and a 20.66 ms match advantage in the shape constant condition, both significant (p = .0014 and p = .025, respectively) providing overall support for the notion that comprehenders do actively update their mental simulations.

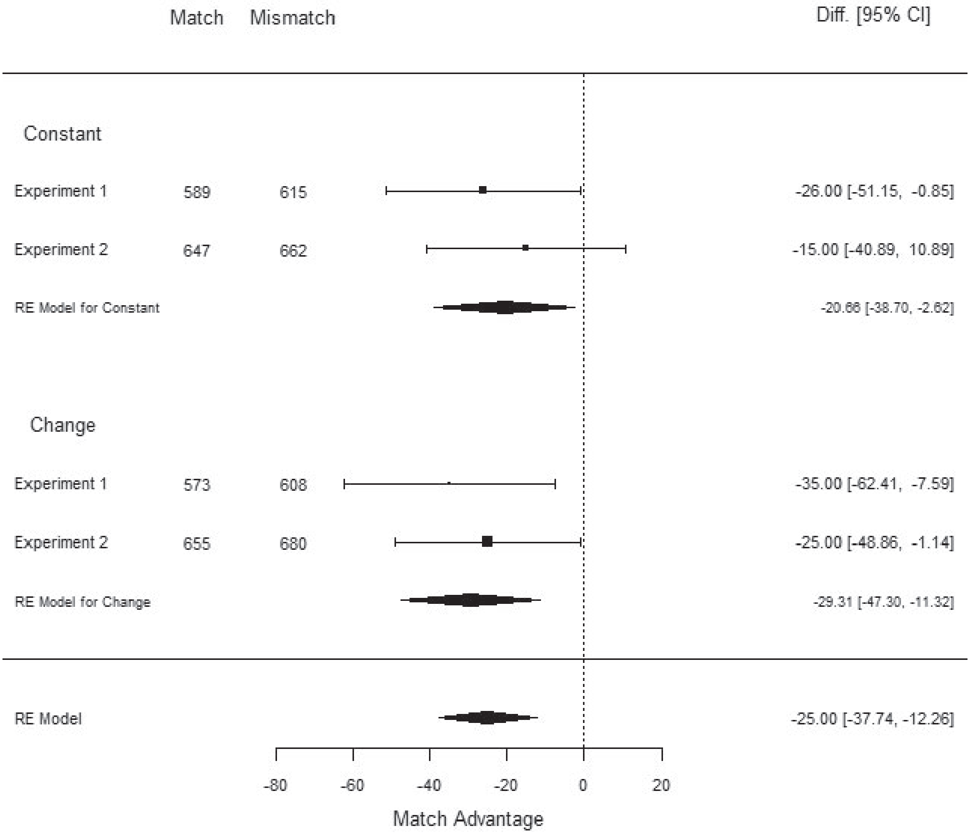


Figure 3. Meta-analysis of Experiments 1 and 2

Overview of Match Effects in Experiments 1 and 2

	Accuracy		**Response Time**			
	Match *M* (*SD*)	Mismatch *M* (*SD*)	Match *M* (*SD*)	Mismatch *M* (*SD*)	Effect size (Cohen's d_z)	BF_{10}
Experiment 1						
Change	0.97 (0.05)	0.95 (0.07)	572.90 (150.20)	607.50 (224.50)	-0.27	4.52
Constant	0.95 (0.09)	0.92 (0.08)	588.60 (181.30)	614.80 (181.50)	-0.22	1.62
Experiment 2						
Change	0.97 (0.05)	0.95 (0.06)	655.00 (233.60)	679.60 (268.40)	-0.22	1.53
Constant	0.97 (0.05)	0.91 (0.14)	646.60 (227.90)	662.40 (264.80)	-0.13	0.43

Table 1

Note. Response times are shown in ms. Reported effect sizes are for the comparison of response time. Bayes Factors were calculated using a Cauchy prior of 0.707 as a one-sided Bayesian paired samples *t*-test, using the JASP software (version 0.9.0.1) for the calculations.

Comprehension accuracy. To examine whether participants did properly read the sentences in the experiment, the accuracy scores to the comprehension questions were analyzed. Sixteen comprehension questions focused on the second sentence, while 12 questions focused on the first sentence. Average comprehension accuracy was high at 85.42% (*SD* = 10.97%) for questions focused on the first sentence as well as for questions targeted at the second sentence (*M* = 80.58%, *SD* = 10.72%). A paired samples *t*-test found that participants were significantly more accurate on questions targeting the first sentence compared to the second sentence ($t(83) = 3.96$, $p < .001$).

Mismatch accuracy. To further examine the interaction found in the accuracy scores, we performed a paired-samples *t*-test to see whether the difference in mismatch accuracy between the shape constant and shape change conditions were significant. The results showed that, when the picture mismatched the shape implied, participants were significantly more accurate in the shape change condition (*M* = .95, *SD* = .06) than in the shape constant condition (*M* = .91, *SD* = .14, $t(83) = 2.37$, $p = .020$).

Discussion

The aim of Experiment 2 was to find out whether the updating of mental simulations still takes place when participants read four sentences, where the first sentence implies one shape, and the final three imply another. The analysis of the accuracy scores revealed an interesting pattern, where participants were most accurate in the match condition, regardless of whether the sentences implied a change in shape or not, but were significantly less accurate in the mismatch condition, which interacted with the shape condition. Inspecting the data more closely, it appears that participants were more accurate in the mismatch condition when a change in shape was implied compared to when no change was implied. It is possible that this finding is due to enhanced memory during event boundaries as proposed by the Event Horizon Model, but then we would have expected to also find better accuracy in match condition when a change in shape was implied, which we did not. An alternative explanation for this finding is that participants in the shape constant condition experienced a continued reinforcement of a single shape (e.g., an eagle with spread wings) over the course of four sentences. If they are then confronted with a picture that displays a shape that does not overlap with what is present in their mental simulation, they are more likely to state that the pictured object was not mentioned in the previous sentences. In the shape change condition this would not be the case, as both shapes would have been mentioned in the sentences, resulting in slightly higher accuracy scores. Similar to Experiment 1, however, given that the differences in percentage accuracy is so small between conditions, these conclusions should be interpreted with necessary caution.

The analysis for response times in Experiment 2 did not show the interaction found for accuracy scores. Instead, only a main match effect was found for subjects, where participants were again faster at responding when the picture matched the shape implied by the final sentence than when it mismatched. The paired samples *t*-test, however, illustrated no significant match effect in the shape change and constant conditions. Similarly, the item analyses found no significant match effect. We expect that these findings may be explained by the less natural wording of the sentences in the shape constant condition. Although we had aimed to create more realistic stimuli by adding more sentences, it is possible that our manipulation caused the opposite effect. The continued emphasis of only one shape throughout all four sentences may have encouraged a surface-level representation, which may have resulted in a weaker activation of all relevant visual representations, leading to a lack of a match effect. Nevertheless, when we conducted a meta-analysis over the data from Experiment 1 and 2, we found a significant match effect of 29.31 ms for the shape change condition, and a significant match effect of 20.66 ms for the shape constant condition. This again supports the idea that mental simulations can be actively updated.

Finally, we found that participants were significantly more accurate on the comprehension questions targeting the first sentence compared to those targeting the second sentence. It is possible that this could be explained by the primacy effect, being that items coming first are remembered better than those that come directly afterward.

General Discussion

Over the past decade, much research has been conducted to find out which object properties are represented in mental simulations. Recently, however, more and more researchers have become interested in understanding what the underlying mechanisms of mental simulations are and how they unfold during language comprehension. We were specifically interested in how mental simulations update when a change in shape is implied over the course of two (Experiment 1) and four (Experiment 2) sentences, using a sentence-picture verification task to test this. We hypothesized that, if you imply one object shape in one sentence, and then imply the other in the sentence that follows, both shapes would remain active in a mental simulation, leading to no match advantage. We further hypothesized that, when more sentences are added that continue to emphasize the final object shape, only the final shape will remain active in a mental simulation, leading to a match advantage.

Our findings do not support our first hypothesis and provide tentative support for our second hypothesis. We found a significant match effect across the shape change and shape constant conditions in both Experiments 1 and 2. Upon closer investigation, we found that both the shape change and shape constant conditions in Experiment 1 had a significant match effect, but neither condition was significant in Experiment 2. In order to determine whether the most recent shape is more activated in the mental simulation when a change in shape is implied, a meta-analysis was conducted over the data from both experiments. The findings from the meta-analysis support the conclusion that participants have the final shape more activated in the mental simulation, with an overall match advantage of 29.31 ms.

During the process of language comprehension, a comprehensive situation model is built that results in the same (or similar) sensorimotor activation as when the described event is experienced. In a task where participants are asked to compare what they see in the picture to what they have read, it would be possible for them to either compare the picture to the surface structure of the text, or to compare it to the constructed situation model. In the first case, we would not expect a difference between any conditions as we tried to avoid making explicit references to object shapes. If the picture is compared at the level of the situation model, however, we would expect the matching picture to lead to faster responses, as the situation model includes all other information that is not only explicitly mentioned in the text (but is still relevant to the aspects of the situation). This is indeed what we found in the current study. If the sentence-picture verification task causes us to compare the viewed picture to our constructed situation model, it would make sense that, when there is overlap between the two, we are more likely to answer “yes” in the context of this task. This then also explains why accuracy was higher in the shape change condition, as the picture will always overlap with one of the implied shapes. In the shape constant condition, however, there are more constraints on what the object shape could be, so if there is less overlap between picture and model (i.e., in the mismatch condition), it makes sense that there are fewer people who would choose to answer in favor of the picture having been mentioned in the previous sentence.

An alternative explanation to this boosted accuracy effect in the shape change condition comes from the Event Horizon Model (Radvansky, 2012), which proposes that memory is enhanced at event boundaries. Although we cannot conclude that there were event boundaries in between the sentences of the shape change condition as this was not tested, if participants indeed perceive the two different object shapes described in the text as separate events (i.e., model creation, cf. Radvansky & Zacks, 2011), it would explain why there was enhanced accuracy in this condition.

In the introduction, we stated that it makes sense that mental representations can be reactivated at a later point in time if the context requires it (e.g., Pecher et al., 2009; Sundermeier et al., 2005). Indeed, if the comprehender cannot reactivate a previously encountered object state, then an explicit understanding of changes involving that state would become difficult to grasp. However, our results seem to suggest that for the purposes of updating mental simulations, no reactivation of prior object states are necessary.

The results from the current study support the idea that mental simulations are updated when there is an implied change in object shape. At this point we cannot determine whether our results are inconsistent with those reported by Hindy et al. (2015), who concluded in an fMRI study that both the initial object state (e.g., deflated balloon) and the final object state (e.g., inflated balloon) are encoded when an object changes shape in a sentence (e.g., “inflate the balloon”), as it is possible that there would still be some activation of the initial object state left. Our results do support the findings of Sato et al. (2013), who reported that mental simulations can be updated when participants receive information contradictory to the initially simulated shape.

Now that it has been established that mental simulations change the activation levels of the object traces during the process of updating, future research could focus on how much of the initial object trace remains activated when there is new incoming information, which could be investigated by increasing the ISI between the final sentence and the picture. Furthermore, it may prove beneficial to replicate the current study to improve the reliability of these findings.

Limitations

The current study has a few limitations that may limit the strength of our conclusions. Although the current study used more sentences than typically used in a sentence-picture verification paradigm with the intent to create more natural discourse, the stimuli used could still be considered impoverished textoids compared to that of common discourse seen in everyday life, which limits the generalizability of our findings (Graesser, Millis, & Zwaan, 1997). Additionally, due to the difficulty in the creation of items that both change and reverse shape it is possible that several items may refer to an object's shape more explicitly than others, though we did our best to ensure that this was not the case. As such, it would be prudent for future studies to continue trying to improve the quality of the texts used in experiments.

A second potential limitation to this study are the task demands of a sentence-picture verification task, as there is some debate as to whether this paradigm encourages explicit visual processing, and that by extension, outside of this experimental setting, the nature of these representations would vary. Indeed, sentence-picture verification tasks on their own are unlikely to be valid in terms of the conclusions that can be made regarding visual representations. However, there are many studies that have found match effects using various methods, such as electroencephalography (Coppens, Gootjes, & Zwaan, 2012), memory tasks (Pecher et al., 2009), and naming tasks (Zwaan et al., 2002), which all support the idea that the visual system is recruited during language comprehension.

A final limitation to our study is that the comprehension questions we asked the participants were not equally targeting sentences one and two, and that a significant difference was found in the comprehension accuracies between the two sentences in Experiment 2. Although comprehension accuracy was not a variable of interest to us during the design of this experiment, those results are still relevant, as the significant difference in comprehension accuracy could imply that participants read the first sentence more carefully than the remaining sentences in that item. However, as the number of questions was unbalanced with regard to the sentence's serial position (i.e., 12 questions targeting sentence one versus 16 questions targeting sentence two), strong conclusions cannot be drawn here. We can conclude that overall comprehension accuracy was high in both Experiments 1 and 2, suggesting that participants read the items. Furthermore, it is unlikely that we would have found the effects we did in the current study if participants had not read the other sentences in each item. Future studies, however, should ensure that comprehension questions equally target the various sentences in an item, so this can be explicitly tested.

We can conclude that mental simulations can be updated after changes in shape are implied across several sentences, and that this most likely occurs through the activation of the final shape and the deactivation of the initial shape. Future research could focus on how much of the initial object trace remains activated when there is a change occurring in a narrative.

Appendix

The meta-analysis was performed using R 3.3.2 (R Core Team, 2016) using the package *metafor* (Viechtbauer, 2010) with the following code:


```
library(metafor)
# Definition of vectors
version <- c("change", "change", "constant", "constant")
experiment <- c("Experiment 1", "Experiment 2", "Experiment 1",
"Experiment 2")
mmatch <- c(573, 655, 589, 647)
sdmatch <- c(150, 234, 181, 228)
mmismatch <- c(608, 680, 615, 662)
sdmismatch <- c(225, 268, 182, 265)
n <- c(84, 84, 84, 84)
corr <- c(.84, .91, .79, .89)
#Creation of data frame
data <- data.frame(version, experiment, mmatch, sdmatch, mmismatch,
sdmismatch, n, corr, stringsAsFactors = FALSE)
data
#Random-effects meta-analysis of all data
dat <- escalc(measure="MC", slab=experiment, m1i=mmatch, sd1i=sdmatch,
m2i=mmismatch, sd2i=sdmismatch, ni=n, ri=corr,
          data=data)
res <- rma(yi, vi, data=dat)
res
#Forest plot of all data
par(cex=.8, font=1)
forest(res, slab=experiment, ilab=cbind(data$mmatch,data$mmismatch),
```

```
      ilab.xpos=c(-100,-75), xlab="Match Advantage", ylim=c(-1,
12),rows=c(3:2, 8:7))
par(cex=.8, font=1)
text(-135, c(9,4), c("Constant", "Change"))
text(-100, 11, "Match")
text(-75, 11, "Mismatch")
text(75, 11, "Diff. [95% CI]")
par("usr")
#Random-effects meta-analysis of Change condition
dat <- escalc(measure="MC", m1i=mmatch, sd1i=sdmatch, m2i=mmismatch,
sd2i=sdmismatch, ni=n, ri=corr,
            data=data, subset=(version=="change"))
res.change <- rma(yi, vi, data=dat)
res.change
#Adding polygon for Change condition
addpoly(res.change, row=1, cex=.75, mlab="RE Model for Change")
#Random-effects meta-analysis of Constant condition
dat <- escalc(measure="MC", m1i=mmatch, sd1i=sdmatch, m2i=mmismatch,
sd2i=sdmismatch, ni=n, ri=corr,
            data=data, subset=(version=="constant"))
res.constant <- rma(yi, vi, data=dat)
res.constant
#Adding polygon for Constant condition
addpoly(res.constant, row=6, cex=.75, mlab="RE Model for Constant")
```

4

Situation Model Updating in Young and Older Adults

This chapter has been published as:

Hoeben Mannaert, L. N., Dijkstra, K. (2019). Situation model updating in young and older adults. *International Journal of Behavioral Development,* 1-8. doi:10.1177/0165025419874125

Abstract

Over the past decade or so, developments in language comprehension research in the domain of cognitive aging have converged on support for resilience in older adults with regard to situation model updating when reading texts. Several studies have shown that even though age-related declines in language comprehension appear at the level of the surface form and text base of the text, these age differences do not apply to the creation and updating of situation models. In fact, older adults seem more sensitive to certain manipulations of situation model updating. This article presents a review of theories on situation model updating as well how they match with research on situation model updating in younger and older adults. Factors that may be responsible for the resilience of language comprehension in older age will be discussed as well as avenues for future research.

Introduction

Cognitive aging research has been developing as a discipline in its own right over the past four decades. After an initial focus on age-related deficits in cognitive functions and processing speed (Salthouse, 1996), research has recently shifted toward the examination of preserved cognitive skills in language comprehension in older adults[8], specifically with regard to situation models (Radvansky & Dijkstra, 2007). There have been developments in approaches with regard to the representation of texts as well. Initially, three levels of representation were identified: the surface form, which refers to the exact words and syntax used; the propositional text base, which involves the abstract representation of the ideas in the text; and the situation model, which is the mental representation of the events described in a text (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). Successful comprehension of a text is considered to be the product of the creation and maintenance of an accurate situation model (Radvansky, 1999). Since the introduction to the concept of situation models (Van Dijk & Kintsch, 1983), many theories have been developed to explain not only how situation models are created but how they are updated as well. Readers create a new situation model whenever they encounter a change in the text and this is where updating of the situation model takes place.

Together, these developments provide better insight into the underlying mechanisms on situation model updating, either as a way to update changes gradually, taking one change at a time into account, or in a more global manner that keeps track of all of the available information in the changed situation. These developments also contribute to our understanding of how situation model updating occurs in young adult and older populations. The

[8] Older adults are defined here as healthy adults age 60 and up without diagnoses of memory impairments

following section describes several of these prominent theories on situation model updating to provide an explanation of how situation models are updated, before we turn to how aging affects situation model updating. We limit ourselves to those theories that focus on updating processes as a result of a change in dimension(s) and whose predictions also have been tested in empirical research that included younger[9] and older readers.[10]

Theories on Situation Model Updating

According to the event segmentation theory by Radvansky and Zacks (2011), to make sense of the complex and ever-changing world we live in, people segment ongoing perceptual activity into separate events (Kurby & Zacks, 2008). An event here is referring to a "segment of time at a given location that is conceived by an observer to have a beginning and an end," for example, brushing your hair (Kurby & Zacks, 2008, p. 72). Event segmentation is considered to be a cognitive process that creates these event models and is thought to result from the perceptual system trying to make predictions about the future. During an ongoing event, there exists a stable state that involves both perceptual predictions and the error monitoring of these predictions. As such, while an event is still ongoing (e.g., brushing your hair), it is reasonably easy to predict what will happen within that environment (e.g., looking at the mirror while brushing). Once these predictions are no longer accurate, an event boundary is perceived and the event model needs to be updated to accommodate the new information (e.g., leaving the house).

So when and how are these event boundaries perceived? According to the event-indexing model proposed by Zwaan, Langston, and Graesser (1995; see

[9] We define younger readers as adults, generally between the ages of 18-30 years of age.
[10] This means that certain theories, such as the memory-based text processing view (see McKoon, Gerrig, & Greene, 1996), the RI-Val model (Cook & O'Brien, 2014), and the Event Horizon Model (Radvansky, 2012) are not part of this review.

also Zwaan & Radvansky, 1998), readers track five dimensions of any situation: time, location, objects and characters, causal relationships, and intentions of protagonists. During the reading process, comprehenders monitor story events to see if the situation needs updating. Continuation of the event would be the default mode but whenever there is a discontinuation as the result of a change in any of the five dimensions, the current situation is deactivated and a new situation model is created (Zwaan, Langston, & Graesser, 1995). This process is associated with longer reading times as understanding the text becomes more time-consuming to accommodate the updating of the situation model.

There is much empirical support for the notion that updating occurs as a result of changes in the events in a narrative. A recent study by Hoeben Mannaert, Dijkstra, and Zwaan (2019) illustrated that when a text describes a change occurring to an object's shape, participants deactivate the initial object states in their mental representations, while the newer object state becomes active, suggesting that situation models do not require the activation of all associated information, but only that which is required for the active model. Additionally, studies have found that memory for an event is worse once a person crosses an event boundary, such as going through a door into another room (Radvansky & Copeland, 2006, 2010; Radvansky, Krawietz, & Tamplin, 2011), suggesting that the creation of a new situation model may interfere with a previous one. Further evidence for this interference between separate situation models comes from studies examining the fan effect (Anderson, 1974), which is the increase in the response times or error rates as a result of an increasing number of associations between concepts. Research has shown that when multiple situation models are created by referring to the presence of an object at separate locations, this fan effect occurs (Radvansky, 2005; Radvansky, O'Rear, & Fisher, 2017). However, when various objects are described as being in the same location, this fan effect does not occur. This shows that the integration of information into an existing situation model requires less cognitive effort than when new situation models need to be created.

Many studies have shown that changes on any of the five dimensions as proposed by the event-indexing model requires the situation model to update, evidenced by the longer reading times when spatial (Radvansky & Copeland, 2006, 2010; Radvansky et al., 2011), temporal (Radvansky & Copeland, 2010), or changes in causality, protagonists, objects, or motivations occur (Zwaan, Radvansky, Hilliard, & Curiel, 1998). Moreover, a study by McNerney, Goodwin, and Radvansky (2011) showed that situation model updating occurs not only in brief sentences created by experimenters but also when reading an entire novel, suggesting that situation model updating occurs during language comprehension in a naturalistic context and not simply in the context of an experiment.

Although both the event segmentation theory and the event-indexing model state that model updating occurs when changes are made to any of the five dimensions we discussed in the previous section, Kurby and Zacks (2012) argue that each theory proposes a distinct mechanism by which this updating occurs. According to them, the event-indexing model assumes incremental updating in situation models after a change occurs in any of the five dimensions. This incremental updating means that the model is continuously updated. Conversely, the event segmentation theory suggests that updating generally occurs globally (Kurby & Zacks, 2012), meaning that new models are created at event boundaries. These event models are kept in a stable state that is resistant to updating, as this would interfere with the prediction processes inherent to the model. These predictions are consistently compared to what is happening in a narrative (Kurby & Zacks, 2012). Once these predictions are no longer accurate and an increase in error is observed within the model, the model is abandoned and a new one is constructed. To summarize, the event-indexing model proposes an incremental updating mechanism, meaning that the model is continuously elaborated, while the event segmentation theory proposes a global updating mechanism, which argues that new models are created at event boundaries.

Are these two updating mechanisms mutually exclusive? Kurby and Zacks (2012) argue they might not be. Indeed, the mental representations of an ongoing narrative may be updated within one event (i.e., incremental updating) and may be updated entirely at event boundaries (i.e., global updating). Considering the fact that many studies have provided evidence for incremental updating or global updating, it seems natural to assume that both of these processes in fact exist (see Gernsbacher, 1997, for an overview). However, Kurby and Zacks (2012) argue that these processes have always been examined in isolation and that much of the evidence provided for incremental updating could in fact be interpreted as global updating occurring, and vice versa.

Kurby and Zacks (2012) found evidence of both incremental and global updating in an experiment in which participants performed think-aloud exercises where they typed their thoughts after finishing reading a clause and also had to segment the narrative into either short- or long-timescale events. Participants were more likely to mention characters, objects, space, and time when these changed in the narrative, illustrating the presence of incremental updating. Furthermore, event boundaries were significantly associated with the mention of characters, time, causation, and goal dimensions (but not for objects and space), providing evidence for global updating. Given that both global and incremental updating seem to occur in situation models, the authors conclude that neither the event segmentation theory nor the event-indexing model can independently explain how situation models are updated as they only consider one form of updating. Clearly, discovering whether the various dimensions that are tracked during language comprehension are updated using different updating mechanisms is an important next step for future research in this area.

Having concluded that the updating of situation models can occur both globally and incrementally, what can be said about where situation model construction takes place? Most models agree that situation models are built within working memory and that updating works via an interplay between working and long-term memory. Both the event segmentation theory and the event-indexing model state that working memory contains retrieval cues for long-term memory (Zacks, Speer, Swallow, Braver, & Reynolds, 2007; Zwaan & Radvansky, 1998). Situation models contain too much information to be stored and manipulated in short-term working memory alone, thus information in the narrative must be rapidly encoded into long-term memory (Zacks et al., 2007). This information can then easily be retrieved from long-term memory, as long as a part of the information is still available in working memory, with the help of retrieval cues. As such, in order for a situation model to be updated, a continuous interaction between working- and long-term memory processes is required.

To summarize, although there is evidence to suggest that updating occurs both incrementally and globally (e.g., Kurby & Zacks, 2012), more studies are required to establish exactly at which points in a narrative updating occurs, and whether updating differs for the five dimensions that are tracked during the reading of a narrative (i.e., time, space, characters and objects, goals, and causation). Given that (changes in) situation models contain too much information to be held temporarily in memory, working memory capacity and links with long-term memory are important for effective updating processes to occur. Here, the role of aging processes becomes particularly relevant as older adults may deal with situation model updating differently than young adults as a function of changes in their cognitive development.

Situation Model Updating in Younger and Older Adults

The discussion of theories on situation model updating focused on the mechanisms responsible for this process. These theories differ with respect to how and when these updating processes take place. Relevant for this review is how these theories may explain (differential) situation model updating processes in younger and older adults, possibly as a result of differences in cognitive developments across the life span (Zacks, Hasher, & Li, 2000). For example, older adults experience declines in some cognitive domains which may have repercussions for their ability to process and update information when reading. At the same time, other cognitive abilities are preserved or continue to develop in older age which may affect language comprehension processes in a more positive sense (Radvansky & Dijkstra, 2007). Below, areas of cognitive decline and preservation in language comprehension in older adults are discussed in the context of research illustrating how this has been demonstrated empirically in cohort comparisons of younger and older readers. After the areas of cognitive decline and preservation, as well as their impact on situation model updating have been discussed, the focus will turn to the issue of how these findings may or may not support the models on situation model updating discussed above.

One known area of cognitive decline in older adults is speed of processing (Salthouse, 1996). Processing of information occurs at a slower speed in older relative to younger adults and accounts for a substantial portion of age-related decline on various cognitive tasks (Salthouse, 1996). In accordance with the slowing hypothesis (Salthouse, 1996), research has shown that older adults need more time to process ideas in propositionally dense sentences in a text and at clause boundaries where information from the sentence is updated

(Payne & Stine-Morrow, 2014; Stine-Morrow & Hindman, 1994). The need to allocate more resources to the processing of more effort-demanding parts of a text could exhaust the available capacity to do so adequately (Kemper, Crow, & Kemtes, 2004). Indeed, older adults have shown marked declines in text processing ability, especially at the surface and textbase level (Radvansky, 1999; Radvansky, Zwaan, Curiel, & Copeland, 2001; Stine-Morrow, Loveless, & Soederberg, 1996).

Another area of cognitive decline in older age has to do with a decrease in working memory capacity with increasing age (Salthouse & Babcock, 1991). Limitations in working memory capacity increase chances that recently processed sentences are forgotten and that the construction of a text representation is hindered (De Beni, Borella, & Carretti, 2007; Norman, Kemper, & Kynette, 1992). This was demonstrated in older adults reading texts with higher syntactical complexity (Norman et al., 1992) and when they were processing cognitively demanding text components, such as clause and sentence boundaries (Payne & Stine-Morrow, 2014).

Inhibitory processes are a third area of cognitive ability that declines with age (e.g., Radvansky, Zacks, & Hasher, 1996). Specifically, the inhibition-deficit account (Hasher & Zacks, 1988) has been used to explain cognitive deficits in various tasks in older adults. According to this account, older adults are less able to prevent irrelevant information from entering working memory or to suppress information in working memory that is no longer relevant. Indeed, a study by Radvansky, Zacks, and Hasher (2005) found that older adults are less able to suppress information from competing situation models during a long-term memory retrieval task compared to young adults.

Given the impact of these age-related declines on textbase construction (Radvansky et al., 2001), text processing (Stine-Morrow & Hindman, 1994), and suppressing information, one would expect this to have a negative effect on situation model updating as well. As changes in events require more effort to incorporate these changes into an updated situation model according to the event-indexing model (Zwaan et al., 1995), situation model construction and updating should be more difficult for older adults than younger adults. A thorough review of the extant literature on this topic, however, reveals different results regarding age differences depending on the extent to which situation models are created and updated.

When situation models are created at a sentence level, age-related slowing in processing the information may occur, yet working memory may not be overly taxed in older adult because a limited amount of information has to be processed. Consequently, similar situation models may be constructed by younger and older adults. This issue was examined in a sentence–picture verification task in which participants read a sentence about an object (i.e., an eagle in the air) that was followed by a visual depiction of the object that either matched the implied shape of the object (e.g., an eagle with wings outstretched) in the sentence or mismatched (e.g., an eagle with the wings folded) with it (Dijkstra, Yaxley, Madden, & Zwaan, 2004). Generally, if readers create a situation model of the sentence, then the implied shape of the object matters and should result in faster response times for matching pictures than for mismatching pictures (Stanfield & Zwaan, 2001). The results indicated that older readers not only demonstrated a similar facilitation for the match effect as young adults, but even demonstrated a larger slowdown of responses when the picture mismatched, even when the variability in responses time was controlled for. Older adults had longer response latencies than younger adults overall, but similar performance on the comprehension questions as younger adults supporting the idea of a similar situation model construction.

Possibly, longer reading times for the sentences in older adults allowed them to build a more elaborate situation model that would be protected from overwriting by a mismatching picture. This coincides with a differential allocation of time in older adults in sentence comprehension studies where older adults spent more time at clause boundaries to comprehend the text (Kemper et al., 2004). Presumably, older adults have different strategies when creating situation models, compensating for declines in slowing and working memory capacity by allocating more resources to process the text where it is needed most (i.e., comparing the implied shape of the sentence with the picture presented after the sentence) to comprehend these texts effectively.

Would this lack of age difference regarding situation model construction hold for the same task but in a setting that taxes working memory to a greater extent? This issue was examined in another study using the sentence-verification task in younger and older adults but with participants listening to the sentences over headphones and naming the pictures that appeared after the sentence (Madden & Dijkstra, 2009). Here, the task could be considered more effortful because it required a more active response (i.e., naming the object) and the maintenance of the processed information from auditory information in working memory. The results again demonstrated a greater match effect in older adults despite the more taxing demands on working memory. Moreover, the match effect was larger in older adults with a higher working memory span than young adults with a higher working memory span. Again, older adults seemed to allocate more resources of their working memory capacity to maintain and update relevant information for the situation model. Why did higher task demands not have a negative effect on the performance of older adults? Possibly, the situation model that high span older adults created from their allocation of resources to a single sentence was even more difficult to override with a mismatching picture than the model created by low span older adults or young adults.

The results of these two studies suggest no age differences in situation model creation, not even when working memory capacity is taxed to a greater extent. In terms of the distinction between constructed and integrated situation models, these results imply that there is no age impairment for the construction of situation models. If anything, older adults are at an advantage for the construction of difficult-to-overwrite situation models, as they illustrated a larger slowdown in responses for mismatching than matching pictures. The question is whether this would also be true for integrated situation models as they not only require more working memory capacity to keep track of changes in the situation but also need to maintain links with the current situation model and long-term memory. We will discuss several studies below that looked into potential age differences regarding updating processes when one of the dimensions that are part of the event-indexing model (Zwaan et al., 1995), such as time, location, objects, or characters, changed and required updating of situation models.

One of the early studies on situation model updating focused on answers to probes about narratives that varied in distance from the protagonist. Older adults answered probes about objects in a room (e.g. shelves in the library) that were distant from a protagonist in narratives more slowly than objects in a room that were closer to the protagonist. Moreover, this distance effect was larger for older than for younger adults (Morrow, Leirer, Altieri, & Fitzsimmons, 1994). In a follow-up experiment, Morrow, Stine-Morrow, Leirer, Andrassy, and Kahn (1997) had younger and older participants again read narratives about a protagonist who moved through space, which resulted in varying distances of objects in those spaces. Reading times rather than probe times were assessed and again showed slowing in both age groups when the objects in a room were more distant from the protagonist, and again there was a larger slowdown in the older age group. Moreover, older adults with better reading comprehension of the narratives showed more slowing for sentences when updating required the integration of earlier information into the model. Both younger and older adults successfully managed to update the situation model, but this came at a cost for older adults. Only by slowing down in reading time were they able to update their situation model, yet their comprehension ability is the same as that of their younger counterparts.

Research on updating processes of other dimensions of the event-indexing model, such as time and goal completion, also showed longer reading times in older adults relative to younger adults when there were shifts in these dimensions. Radvansky, Copeland, Berish, and Dijkstra (2003) examined potential age differences in temporal updating in situation models. Young adults and older adults read narratives with either short (a moment later) or long time-shifts (a day later). Presumably, a short time shift requires little updating of the situation model, whereas a long time shift requires substantial updating, which should be noticeable in the response times. For example, a wall that is being painted does not look very different a moment later, but it will look entirely different a day later when one assumes the painting continued for some time. Indeed, this substantial updating was reflected in longer reading times in younger and older adults for the narratives that contained a longer time shift. The size of the effect was larger in older adults, again suggesting more sensitivity to the situation model updating manipulation among older adults.

Goal completion was examined among younger and older readers in a study by Radvansky and Curiel (1998). Younger and older participants read a narrative that contained a goal that the protagonist obtained, failed to obtain, or had a neutral outcome. Response times to probe questions about the goal revealed that the availability of the goal decreased, as indicated by longer response times, equally in younger and older readers for the completed goal relative to the failed goal. Both age groups showed similar differences in response times and were therefore equally sensitive to the manipulation. In contrast to the other studies discussed above, the slowdown in response time in Radvansky and Curiel's (1998) study was not greater for older than young adults. Apparently, it depends on the dimension of the event-indexing model that changed, location, time, or goal completion, whether or not updating occurs differently for younger versus older readers.

Based on the studies that tested predictions from the event-indexing model in a younger and older population, we can conclude that they hold equally well for both age groups. In general, younger and older adults demonstrate similar updating and integration ability of the situation model in narratives (Morrow et al., 1994; Morrow, Stine-Morrow, Leirer, Andrassy, & Kahn, 1997; Radvansky, Copeland, Berish, & Dijkstra, 2003; Radvansky & Curiel, 1998; Radvansky et al., 2001) despite stronger demands on available memory capacity to do so in older adults. Stronger age effects as reflected in longer reading or response times may be due to the establishment of more elaborate situation models based on more extensive reading experiences. Rather than having more difficulty establishing links between the current and integrated situation model and with long-term memory in older adults, their extensive reading experiences may actually help them to accomplish this. The ability to establish and maintain integrated situation models may also be due to a stronger emphasis on global and top-down text processing strategies relative to young adults who may focus more on surface-based and bottom-up processing. These successful strategies in older adults to deal with more effortful task demands at a global level may be a way to compensate for age-related declines at the lower surface or textbase level (Stine-Morrow, Morrow,

& Leno, 2002). For some changes in dimensions (location, time), this may be easier than for other changes (goal), hence the lack of stronger age effects there.

As stated earlier, updating processes along one dimension of the situation model is consistent with incremental updating. The event-indexing model supports the idea of continuous, incremental updating as demonstrated in longer reading and response times for manipulated dimensions in the narrative. As a whole, incremental updating appears intact in older adults, as seen by adequate situation model construction and updating ability when changes in location, time, and goals occur as long as sufficient processing time is allocated to allow this form of updating to occur. Stronger match effects in the picture-verification tasks discussed earlier (Dijkstra et al., 2004; Madden & Dijkstra, 2009) can also be considered as a form of incremental updating when a picture is compared with the mental representation of the implied shape of an object in the preceding sentence.

The question is how far this goes. Will older adults still be able to update their situation model when more extensive updating of the situation model is required? A study examining the fan effect (Radvansky, Zacks, & Hasher, 1996) demonstrated that, even though both older and younger adults could easily integrate information into a single situation model when reading sentences describing different objects in the same location (e.g., a potted palm and a bulletin board at an airport), age differences occurred when the same object was described to be in different locations (e.g., a potted palm in an airport, hotel, and restaurant). Because multiple situation models had to be constructed to represent objects in different locations, global updating was necessary and interference occurred during retrieval. Older adults suffered more from this interference as seen by the larger fan effect. It is therefore plausible that, when event boundaries are perceived and global updating has to occur (as predicted by the event segmentation theory), older adults may have more difficulty doing so.

This assumption is supported by the results of a study that required the construction of multiple situation models and global updating with regard to changes in characters in a narrative. Noh and Stine-Morrow (2009) demonstrated that older participants had more difficulty than younger adults in accessing previously mentioned protagonists in a narrative after a new one was introduced. This happened even though older adults over-allocated processing time to instantiate the first character in a narrative. As was the case in the previous study (Radvansky et al., 1996), multiple situation models had to be constructed to keep track of all protagonists in the narrative. This required global updating to create a new situation model that still contains all relevant information of the older model. Possibly, due to constraints to their working memory capacity, older readers had difficulty to allocate sufficient resources to characters introduced later in a narrative in an effort to maintain representations of characters that were introduced earlier. Their difficulty in doing so reflects a deficit in global updating among older adults.

This does not necessarily mean that older adults always have difficulty with global updating processes. Results of a study by Radvansky, Pettijohn, and Kim (2015) suggest that under certain circumstances, older adults are equally capable of updating their situation models globally (i.e., at event boundaries). In their study, both young and older adults had to move an object in a virtual environment, either within the same room or through a doorway. The results found a location updating effect as seen by participants' increased forgetting when an event boundary was crossed, suggesting global updating. Importantly, the effect sizes were similar for both age groups, suggesting that older adults do not have more difficulty updating their situation model at event boundaries compared to young adults.

The ability in older adults to update situation models globally in a similar manner as younger adults is further supported by a study on event segmentation by Magliano, Kopp, McNerney, Radvansky, and Zacks (2012), who had young and older participants segment either text-based or visually based narratives into separate events. They found that both older and young adults were similarly sensitive to situational changes resulting in good between-group segmentation agreement. However, older adults tended to create smaller segments in a narrative than younger adults. Possibly, older adults perceived event boundaries with fewer situational changes than younger adults. This is less taxing on their resources and helps older adults to update their situation model adequately. Although more research is needed to test this, it could suggest that older adults indeed tend update their situation models globally more often than young adults, possibly as a strategy to avoid placing heavy demands on their working memory capacity. Differential age differences with regard to global updating in different studies could be due to the extent to which how much updating is required to construct integrated situation models and how older adults may use strategies to do so. To assess whether this happens, changes within and beyond event boundaries have to be examined systematically in younger and older adults.

Bailey and Zacks (2015) did just that by controlling for changes in characters and locations in narratives. They found that, although older adults generally read more slowly than young adults, they had faster reading times for a probe that followed no change in the text than for a probe that followed a change in the text, which is indicative of global updating. Young adults, however, did not show such differences in changes for the probes. This suggests that, when changes occur in a narrative, for example, a character moving from the kitchen to the basement, older readers are not only slower when responding to probes about the basement but also to probes about the character (e.g., a change in hairdo), when compared to control probes. In other words, older adults do not only update the element that changes in the situation (location), which suggests incremental updating, but also to elements that did not change but are part of the new situation, which suggests global updating. No clear support for either incremental or global updating processes was found for young adults. It seems that older adults strategize their resources to allocate them where they are most needed and are thus able to successfully update their situation model.

Conclusion

Research on cognitive aging over the past several decades has often focused on declines in cognitive functioning (Zacks et al., 2000; Norman et al., 1992). Given that the updating mechanism of situation models appears to rely on the interaction between working memory and long-term memory, it is possible that this interaction is mediated by aging processes. Therefore, it would be reasonable to assume that the construction of situation models should be more difficult in older adults who have reduced working memory capacity relative to young adults, as the requisite linking between the current and integrated situation model would require more effort than if information was simply maintained in the current situation model.

Interestingly, several studies have shown that, even when controlling for the longer reading times, older adults still create mental representations during text comprehension both for shorter and longer texts (Dijkstra et al., 2004; Madden & Dijkstra, 2009; Morrow et al., 1994, 1997; Radvansky et al., 2001, 2003; Radvansky & Dijkstra, 2007). Furthermore, these findings support the notion that the constructed situation model can be more elaborate in older adults than in young adults, thus providing protection against contradictory (or mismatching) information. Perhaps most importantly, however, are the findings that younger and older adults appear to demonstrate similar capacities for updating and integrating information in the situation model, albeit that older adults do this more slowly.

We can draw the following conclusions from our discussion of situation model updating in younger and older adults. First, updating occurs when there is a change in the situation of the events described in a text. In the event-indexing model, this change can be along one dimension (Zwaan & Radvansky, 1998), such as location, time, or goals, or in the event segmentation theory along several dimensions (Kurby & Zacks, 2012). In both cases, updating occurs but the difference is whether only the new dimension is being updated (continuously in incremental updating) or all information is being updated (global updating). Older adults are able to update their situation model globally (Bailey & Zacks, 2015; Radvansky et al., 2015), but not always. When multiple situation models have to be created, for example, when objects are described in different locations, or when multiple characters are introduced in a narrative, updating processes may be too taxing on available working memory capacity in older adults to do so (Noh & Stine-Morrow, 2009; Radvansky et al., 1996). Future research could look into this matter more closely by examining the point at which updating in older adults no longer can be compensated for by allocating resources to the task. For example, what happens when older adults are not able or allowed to allocate more attentional resources to certain parts of the texts. Are they still able to construct and update situation models adequately then?

Secondly, older adults appear to be able to construct and integrate situation models along one or several dimensions as well as young adults but generally need more time for this, even if they have a higher working memory span. Situation model updating ability in older adults may be a way to compensate for needing to allocate more effort there (i.e., due to text complexity or task demand) where it is needed most. Only when they are able to allocate more resources to that task can they maintain successful updating performance. As described above, there may be a limit to the extent to which they are able to do that. Future studies could focus more specifically on how older adults utilize their extensive reading experiences to draw resources from long-term memory to construct and update a situation model. Older adults could be better at removing less relevant information from their situation models which would contradict an inhibitory deficit account (Bailey & Zacks, 2015). Alternatively, older adults could utilize different strategies relating to other goals when reading a narrative, or different self-regulatory activities, relative to young adults (Stine-Morrow, Miller, & Hertzog, 2006)

Apart from more research on incremental and global updating processes in younger and older readers, future research could incorporate insights from related domains. For example, language comprehension research from an embodied cognition perspective could be relevant to examine situation model updating from a different angle. Embodied cognition research has shown examples of how sensorimotor activation facilitates reading and updating processes (Dijkstra & Post, 2015). In a study on motor resonance, sensibility judgments about sentences by turning a knob clockwise or counterclockwise were faster when the manual response to the sentence was in the same rotation direction as the manual action described by the sentence (Zwaan & Taylor, 2006). Research on situation model updating could build on these findings by exploring different ways to examine updating processes in situation models in different age groups. Sensorimotor manipulations could be employed to see if older adults benefit differentially from such manipulations when they construct and update their situation model. For example, rotate a knob to move forward in a text that in which rotation is an important element in situation model construction could facilitate updating of a situation model, pressing a space bar faster when reading about a character speeding up in a narrative, and could facilitate updating processes as well. The interesting question here would be to assess whether sensorimotor facilitation would occur both for incremental and global updating processes.

To conclude, situation model updating is a process supporting language comprehension and appears to remain intact during aging, despite the declines in other cognitive processes. After a period in which research more heavily focused on the negative aspects of aging in relation to cognitive functions, evidence converges toward preserved abilities in aging and reading comprehension and how older adults successfully allocate their resources to maintain these skills.

5

Object Combination in Mental Simulations

This chapter has been submitted:

Hoeben Mannaert, L.N., Dijkstra, K., & Zwaan, R.A. (submitted). Object combination in mental simulation.

Abstract

Studies on the presence of mental simulations during language comprehension have typically focused only on single object properties. The current study investigates whether two objects are combined in mental simulations, and whether this is influenced by task instructions. In both experiments, participants read sentences describing animals using a tool in some way. After each sentence, they saw an image of a cartoon animal holding a tool, and they indicated whether the animal (Experiment 1) or the tool (Experiment 2) was mentioned in the previous sentence or not. The shown image either completely matched, partially matched, partially mismatched, or completely mismatched the preceding sentence. Ninety Dutch psychology students took part in Experiment 1, and 92 students took part in Experiment 2, both experiments were pre-registered. The results suggest that mental simulations indeed combine multiple objects during language comprehension, and that this is not influenced by task instructions. Regardless of the instruction type, participants always responded quickest in the complete match condition compared to the partial match condition, suggesting that language comprehension leads to the creation of a complete mental simulation.

Keywords: Mental simulation, conceptual combination, grounded cognition, language comprehension

Introduction

Mental simulations are defined as the "reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind" (Barsalou, 2008, p.618). Most studies on mental simulations have used the sentence-picture verification tasks in order to examine whether single object properties are simulated during language comprehension. For example, when participants read the sentence "The ranger saw an eagle in the sky" they are faster at responding to a picture of an eagle with spread wings than to an eagle with folded wings (Zwaan, Stanfield, & Yaxley, 2002). Thus far researchers have found evidence for the presence of several object properties in mental simulations, including color (Connell & Lynott, 2009; Hoeben Mannaert, Dijkstra, & Zwaan, 2017; Pecher & Zwaan, 2012), movement (Gentilucci, Benuzzi, Bertolani, Daprati, & Gangitano, 2000; Glenberg & Kaschak, 2002), orientation (Stanfield & Zwaan, 2001), and size (De Koning, Wassenburg, Bos, & Van der Schoot, 2016). Aside from studies using the sentence-picture paradigm, multiple neuroimaging studies have also found support for modality-specific sensorimotor and affective system activation during language comprehension (Binder & Desai, 2011, Hauk, Johnsrude, & Pulvermüller, 2004; Sakreida, Scorolli, Menz, Heim, Borghi, & Binkofski, 2013; Simmons, Ramjee, Beauchamp, McRae, Martin, & Barsalou, 2007).

Interestingly, no studies have yet explored how mental simulations are affected when multiple objects are included in a sentence, as most studies have included items referring to a single object only (e.g. an eagle with spread wings). Researchers tend to assume that for language comprehension to occur, a situation model is built which represents the meaning of the text (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998) However, for a situation model to be complete it should include a complete representation of the situation described by the text, but this has never been examined before. As past research has only used sentences with one object, we can only know for certain that a mental simulation includes one object. The question that remains then is: when we are reading a text, do we create separate mental simulations for each object that we encounter? Or do we combine multiple objects in a mental simulation in order to construct a comprehensive situation model? Although we believe this is likely to be the latter, this has never been tested empirically before.

The literature that has come closest to examining this is the literature on conceptual combination, defined as the ability to create new meaning out of already existing concepts (Lynott & Connell, 2010). The Embodied Conceptual Combination (ECCo) model proposed by Lynott and Connell (2010) builds on the idea from the Language and Situated Simulation (LASS) theory (Barsalou, Santos, Simmons, & Wilson, 2008) that both linguistic and perceptual information are required for concept representations, and propose the ECCo model as an explanation for how conceptual combination operates. The first principle of the theory states that, similar to the LASS theory, first the linguistic system is activated, which can be used for superficial processing tasks (e.g., when asked to verify the object properties “lemon-yellow”, this would only require the linguistic system to activate due to the strong word associations of those concepts, Solomon & Barsalou, 2004). This is followed by the activation of the simulation system, which is needed for more deep conceptual processing (e.g., when concepts are combined in novel ways, the linguistic system is insufficient to deal with this).

This principle was tested in a study by Louwerse and Connell (2011), who examined whether the response times in a property-verification task could be explained by linguistic associations or a simulation approach. This was done by testing whether switching between “linguistic modalities” (i.e., visual-haptic, auditory, and olfactory-gustatory) could predict faster responses than switches between perceptual modalities (i.e., visual, haptic, auditory, olfactory, and gustatory). The linguistic modalities were determined by using word co-occurrences to predict which modality an adjective belonged to. Indeed, they found that the linguistic factors best predicted short response times, while the perceptual factors best predicted longer response times, suggesting that the linguistic system peaks in activation before the perceptual simulation system. It should be noted that these findings were obtained in word-level tasks in which words are presented in isolation. It is not clear yet whether the temporal precedence of linguistic factors can be summarily extrapolated to larger stretches of text in which perceptual representation may have already been activated by prior text once a target word is processed.

The second principle of the ECCo model is that the head and modifier concepts reduce the number of affordances that can be integrated (or "meshed"; cf. Glenberg, 1997) in a simulation. In other words, the number of ways in which an object can interact with another is reduced during this "meshing" (Glenberg, 1997). This meshing can be completed in a destructive or a non-destructive way. If a conceptual combination is destructive, it means that either the modifier or head concept is significantly reduced, such as a *cactus beetle* being interpreted as a green and spiky beetle (Lynott & Connell, 2010). In this case, the *cactus* concept was destroyed and has modified the head concept. In a non-destructive conceptual combination, both concepts remain intact in the simulation. In this case a *cactus beetle* could be a beetle that feeds on cacti (Lynott & Connell, 2010). The linguistic system interacts with the simulation system by constraining which affordances are plausible (based on past experiences), and also helps to determine early on whether a destructive or non-destructive combination will take place. So when a novel compound noun is encountered, the linguistic information associated with the compound noun is activated first, which then activates the simulation system that can then provide further feedback to the linguistic system, continuing until the process of conceptual combination is complete (Connell & Lynott, 2011).

Several behavioral experiments have provided evidence for the perceptual nature of conceptual combinations. For example, Wu and Barsalou (2009) tested whether participants generated more properties of occluded or unoccluded features in both familiar and novel conceptual combinations. If participants use the simulation system to create conceptual combinations, they would generate more unoccluded properties than occluded properties. Indeed, they found that when participants read, for example, the noun "lawn", they produced more unoccluded features such as *soft* and *green* compared to occluded features such as *roots* or *dirt*. Conversely, when they read the noun phrase "rolled up lawn", *roots* and *dirt* became unoccluded, and were thus produced more frequently than *soft* and *green*. As such, it appears as though multiple object properties are simulated when a concept is activated, but that the contents of the simulation may be constrained by the linguistic system.

Connell and Lynott (2011) similarly examined the role of simulations when creating new concepts, where they gave participants a forced-choice interpretation task for novel noun-noun compound phrases (e.g., "octopus apartment"), where they had to answer as quickly as possible whether they could come up with an interpretation for the phrase or not, before providing the interpretation. The results showed that participants respond significantly slower when a destructive interpretation is used compared to a non-destructive interpretation. For instance, if *octopus apartment* is interpreted in a destructive manner, then it could be interpreted as "an apartment with eight rooms", while a non-destructive interpretation could be "a place where an octopus might live" (Connell & Lynott, 2011, p.4). It appears then that the integration of multiple objects in a situation model is easier than the manipulation of objects in the situation model, which appears to require more processing time.

As mentioned before, conceptual combination refers to the formation of a new concept from existing concepts. As such, most studies on conceptual combination have focused on how compound nouns are processed. In our study, we are interested in seeing whether multiple objects are combined in a mental simulation, but we believe that this combination works via the same mechanism as proposed by the ECCo theory. For example, if a person reads about a bear grabbing a broom from the corner in a room, we believe that the simulation system would activate and simulate a bear holding a broom. This would work similarly to the processes involved in non-destructive conceptual combination, as none of the concepts need to be reduced or "destroyed". In order to not confuse terminology, we refer to these type of sentences as involving object combinations, rather than conceptual combinations.

To our knowledge, no studies have yet been published on whether multiple objects are represented in mental simulations, and whether task instructions modulate these simulations. If unfamiliar object combinations are processed in a manner similar to what is proposed by the conceptual combination theories, then we would expect that, in our bear and broom example, both the bear and the broom would be present in the mental simulation. However, what if participants are told to either respond to the animal or to the tool, would they still simulate both objects, or just the one they were told to attend to? A study by Lebois, Wilson-Mendenhall and Barsalou (2015) suggests that the differences in the task instructions would affect the contents of the mental simulation. In their study, participants in one condition were instructed to pay attention to the verticality of words (e.g., "sky" or "basement") when responding with upward or downward responses based on word color. In the other condition participants were uninformed about the manipulation. Their study found only a congruency effect when participants paid attention to verticality, but found no significant effect when participants were unaware of the manipulation. These findings suggest that task instructions can modulate concept activation, and thus it is possible that influencing what comprehenders attend to while reading similarly modulates the contents of their mental simulations.

If a sentence is processed only superficially, and one only had to match a picture of an animal to a sentence, the presence of a matching or mismatching object should not significantly influence one's performance. However, if comprehenders routinely generate complete mental simulations during language comprehension, the presence of a mismatching object should create interference. If task instructions alone can influence what is included in a mental simulation, then this has significant consequences for the relevance of mental simulations in language comprehension, but if language comprehension requires mental simulations, then a complete simulation of all associated objects in a sentence should be performed automatically. If we are instructed, however, to only pay attention to one of the objects in the sentence, and subsequently no interference is created when a mismatching object is shown, this would imply that a complete simulation is not required for comprehension. These are the questions that we attempt to answer with the current study.

The Current Study

In Experiment 1, participants had to read sentences describing animals using a tool in some way. After they read a sentence, they saw a picture of a cartoon animal holding a tool, and they were instructed to answer whether the animal was mentioned in the previous sentence or not (see Figure 1 for an example). This picture either fully matched the preceding sentence (i.e., both the animal and the tool matched), partially matched (i.e., only the animal matched), partially mismatched (i.e., only the tool matched), or completely mismatched (i.e., neither tool nor animal matched). If participants only simulate the objects they are required to in order to complete the task, then we would not expect any differences in response times between the complete and partial match conditions. However, based on the fact that participants can generate non-destructive interpretations of novel concepts and are able to do this fairly quickly (Connell & Lynott, 2011), we predict that participants do simulate both objects and thus will respond faster to pictures that completely match the sentence, compared to when they partially match. Furthermore, if the complete scene is simulated, it should be easier to reject a picture when there is absolutely no overlap between the picture and the mental simulation (i.e., a complete mismatch) compared to when there is some overlap (i.e., a partial mismatch). Under the assumption of complete simulation, partial overlap should create interference, and thus give rise to longer response times than no overlap.

In summary, if participants respond fastest to the complete match condition compared to the partial match condition, this would be taken as evidence for object combinations in mental simulations and evidence for facilitation in the complete overlap condition. Moreover, if they respond fastest to the complete mismatch condition compared to the partial mismatch condition, this would be taken as evidence that partial overlap in a mental simulation generates interference.

In Experiment 2, participants received the exact same items as in Experiment 1, but were instructed instead to respond whether the tool in the picture was mentioned in the previous sentence or not. By having participants attend to different parts of the sentence, we can see whether instructions modulate the contents of mental simulations. It should be noted here that in the pictures the tool occupies much less of the visual space than the animal in order for the picture to maintain a semblance of realism. As such, it is possible that facilitation effects in Experiment 2 are smaller than in Experiment 1 as the target is smaller. Similarly, however, it would also be possible to have more interference effects because the larger object (i.e., the animal) mismatches the sentence. Our predictions for Experiment 1 were also our predictions for Experiment 2, namely that response times in the complete match condition would be significantly shorter than in the partial match conditions. If Experiment 2 similarly illustrates that participants respond faster in the complete match condition compared to the partial match condition, this would provide evidence that task instructions do not modulate the contents of mental simulations.

Ethics Statement

The participation in all experiments was voluntary. The participants subscribed to the experiments online via the university platform, and were told that by signing up for a study, they declare to voluntarily participate in this study. They were briefed with the content of each study and provided written consent. Participants were told they were free to terminate the experiment at any point in time without experiencing negative consequences. This study was approved by the Ethics Committee of Psychology at the Erasmus University Rotterdam, The Netherlands.

Preregistration

The predictions, exclusion criteria, design, methods, analyses, and materials of all the experiments reported in this article were pre-registered in advance of data collection and analysis on the Open Science Framework (OSF) to ensure confirmatory procedures were conducted according to a priori criteria, and can be viewed on https://osf.io/hqs7u. Analyses that were not pre-registered are referred to in this article under the heading "Exploratory Analyses".

Experiment 1

Method

Participants. Based on a power-analysis of the effect size reported in the study by Zwaan, Stanfield, and Yaxley (2002), who found a significant match effect for shape, we found that at least 84 participants would be required to find an effect if it exists (d = 0.31, α = .05, Power = .80). To ensure our study was not underpowered after exclusions, we recruited 100 Dutch participants (M_{age} = 20.58 years, SD_{age} = 3.02 years, 87 females, 13 males) from the Bachelor of Psychology at the Erasmus University Rotterdam. Participants were excluded if their average accuracy was below 80%. As a result of this exclusion criterion data from 10 participants were excluded in the analyses, resulting in a sample of 90 participants.

Materials. For the current study we wanted to ensure that the situations described in the sentences would be unfamiliar to participants, as this would lead to the increased likelihood that the sentence would be simulated rather than just superficially processed. As such, we decided to include sentences that described animals performing actions on tools similar to what could be experienced in a cartoon story. Readers are known to adapt quickly to such a cartoon world (Nieuwland & van Berkum, 2006). Ninety-six Dutch sentences were created that described an animal holding a certain object (e.g., *The dog walked with his umbrella across the street.*) and a total of 96 cartoon images were created for this study. There were two possible sentence versions (see Figure 1) and participants viewed only one of these versions (which were counterbalanced) and therefore saw in total 48 sentence items and 48 images. Per sentence, participants also answered a comprehension question, to which they responded after seeing the image. The cartoon images were created using images of cartoon animals and of tools found on the Google search engine, which were subsequently edited using the Paint.NET software (version 4.1.5) to look as though the cartoon animal was holding the object. The images were displayed in grayscale (to ensure that effects of color could not confound the results) and did not exceed a 300 x 300 pixel resolution (approximately 7.9 x 7.9 cm on screen). The experiment was programmed using E-Prime 2.0 Professional and participants completed the experiments in isolated cubicles with computers equipped with 24.1" TFT-IPS screens with a resolution of 1920 x 1200 and a ratio of 16:10.

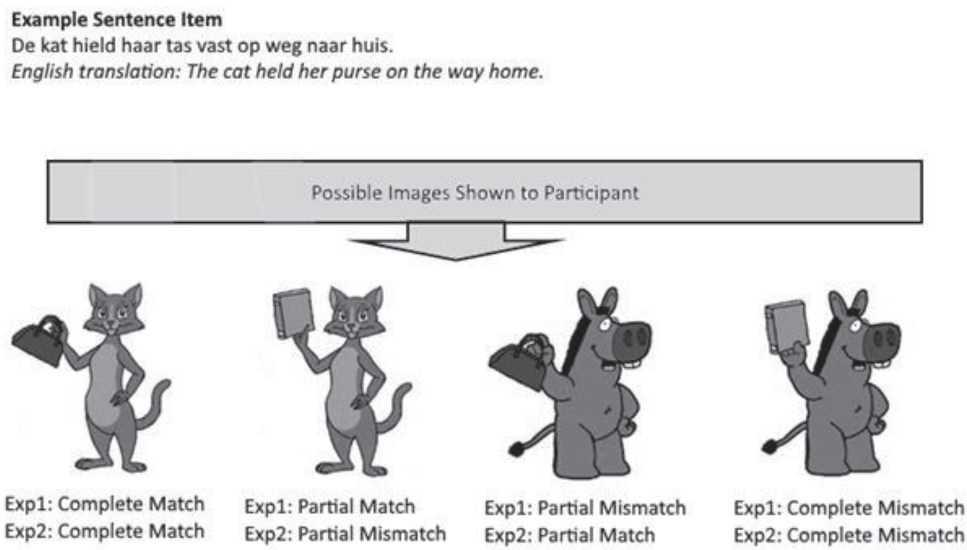


Figure 1. Example of an item in Experiments 1 and 2.

Design and Procedure. The study is a 2 (Type: partial vs. complete) x 2 (Match: match vs. mismatch) x 2 (Sentence: version 1 vs. version 2) x 4 (Image: version1 vs. version 2 vs. version 3 vs. version 4) mixed design. 'Type' and 'Match' are within-subjects variables: participants viewed 12 images per type-match condition. Per sentence, there were four possible images that could be shown to the participant:

1. **Complete match:** correct animal + correct object → Participants give "yes" response
2. **Partial match:** correct animal + wrong object → Participants give "yes" response
3. **Partial mismatch:** wrong animal + correct object → Participants give "no" response
4. **Complete mismatch:** wrong animal + wrong object → Participants give "no" response

'Sentence' and 'Image' are between-subjects variables which served to counterbalance this study, as for each sentence there were four possible images that could be shown to the participant, and there were two sentence versions. This led to a total of eight counterbalancing lists. An additional experiment from another study was performed by the participants in the same session, which was counterbalanced to be completed either before or after the current study.

Participants were instructed that they would perform a self-paced reading task using the spacebar and that they would see a cartoon image after each sentence and were instructed to answer YES (the "L" key) if the cartoon animal matched the preceding sentence, and answer NO (the "A" key) if the cartoon animal mismatched the preceding sentence. After responding to the image participants answered a comprehension question which asked what object the animal was holding, providing two answer options to the participants that they could respond to with the "L" and "A" keys. The purpose of the comprehension questions was to ensure that participants did not only read the name of the animal but that they properly read the entire sentence. Before starting the experiment they received 6 practice items.

A trial looked as follows: Participants saw the > symbol left-aligned in the center of the screen for 1000ms. Subsequently the sentence was shown left-aligned in the center of the screen and remained on-screen until participants pressed the spacebar. Subsequently a fixation cross appeared in the center of the screen (center-aligned) for 500ms, after which the image appeared in the center of the screen (center aligned) and remained on-screen until participants provided a response. After the response to the image they received the comprehension question: "What was the animal holding?" where two answer options were shown on opposite horizontal sides of the screen.

Results

Data analysis. Participants with an average accuracy below 80% were excluded from the analyses. If items in the "complete match" condition had an accuracy score of less than 60%, that item (across all conditions) would have been excluded as well, but as the accuracy scores were above this threshold, zero items were excluded from the analyses. All response-time analyses were performed on correct responses only. Per participant the median response time was taken per condition, as is common in sentence-picture verification studies (Hoeben Mannaert, et al., 2016; Hoeben Mannaert, Dijkstra, & Zwaan, 2019; Pecher & Zwaan, 2012; Zwaan et al., 2002) to prevent extreme values from influencing the data. To examine whether there is an interaction between the match and mismatch conditions, a repeated measures analysis of variance (rmANOVA) was performed with Type (partial vs. complete) and Match (match vs. mismatch) included as the repeated-measures factors, and List was included as the between-subjects factor.

Important to note here is that the mismatch condition always required a "no" response, while the match condition always required a "yes" response. Even though the pictures were counter-balanced across conditions, in general "yes" responses tend to be faster than "no" responses (e.g., Brouillet, Heurley, Martin, & Brouillet, 2010). This has no significant consequences for testing our hypotheses, as the main result of interest is the difference between the complete and partial match condition, which both required a "yes" response.

Subject analyses are denoted with the subscript "1" and item analyses are denoted with the subscript "2". Paired samples *t*-tests were performed to examine the difference between the complete and partial match conditions and between the complete and partial mismatch conditions.

Accuracy. The rmANOVA showed a significant effect of Type, illustrating that participants were significantly more accurate when the image completely matched or mismatched compared to when it partially matched or mismatched ($F_1(1,82) = 17.19$, $p < .001$, $F_2(1,47) = 14.26$, $p < .001$). There was also a significant match effect in the subject analyses (but not in the item analyses), meaning that, when the images showed the same animal as mentioned in the sentence, participants were significantly more accurate than when the animals differed ($F_1(1,82) = 4.41$, $p = .039$; $F_2(1,47) = 0.44$, $p = .513$). There was no significant interaction between Type and Match ($F(1,82) = 1.42$, $p = .238$; $F_2(1,47) = 1.71$, $p = .197$). There was no significant between-subjects effect of List ($F_1(7,82) = 0.40$, $p = .898$), but there was a significant three-way interaction with Type, List, and Match ($F_1(7,82) = 6.41$, $p < .001$).

A paired-samples *t*-test showed that participants were significantly more accurate in the complete match condition ($M = .99$, $SD = .04$) compared to the partial match condition ($M = .97$, $SD = .06$, $t_1(89) = 2.72$, $p = .008$, Cohen's $d = 0.29$), but this was not replicated in the item analysis ($t_2(47) = 1.93$, p

= .060). A second paired-samples *t*-test illustrated that participants were also significantly more accurate in the complete mismatch condition (*M* = .98, *SD* = .04) compared to the partial mismatch condition (*M* = .95, *SD* = .06, *t*(89) = 3.65, $p < .001$, Cohen's *d* = 0.38; $t_2(47) = 3.21$, *p* = .002).

Response times. Descriptive statistics can be seen in Table 1. The rmANOVA illustrated a significant effect of Type, showing that images that either completely matched or completely mismatched led to shorter response times compared to when they partially matched or mismatched ($F_1(1,82) = 10.22$, *p* = .002; $F_2(1,47) = 23.34$, $p < .001$). There was no significant effect of Match ($F_1(1,82) = 1.85$, *p* = .178; $F_2(1,47) = 0.16$, *p* = .687), nor a significant interaction between Type and Match ($F_1(1,82) = 0.01$, *p* = .925; $F_2(1,47)$ = 0.02, *p* = .878). There was no significant between-subjects effect of List ($F_1(7,82) = 1.25$, *p* = .288), but there was a significant three-way interaction with Type, List, and Match ($F_1(7,82) = 3.41$, *p* = .003).

Paired samples *t*-tests were performed to see whether there was a significant difference between the complete match and partial match conditions (which both required *yes* responses), and between the complete mismatch and partial mismatch conditions (which both required *no* responses). The analyses showed that participants responded significantly faster when there was a complete match (*M* = 1387ms, *SD* = 557ms) compared to when there was a partial match (*M* = 1461ms, *SD* = 574ms, $t_1(89) = -2.08$, *p* = .040, Cohen's *d* = -0.22; $t_2(47) = -3.17$, *p* = .003, see Figure 2). Moreover, participants also responded significantly faster when there was a complete mismatch (*M* = 1358ms, *SD* = 561ms) compared to when there was a partial mismatch (*M* = 1422, *SD* = 549, $t_1(89) = -2.59$, *p* = .011, Cohen's *d* = -0.27; $t_2(47) = -2.63$, *p* = .011).

Comprehension Accuracy. Participants on average had high comprehension accuracy ($M = .87$, $SD = .20$), suggesting that they did properly read the sentences in their entirety. An rmANOVA was performed on the comprehension accuracy data and found no significant main effect of Type ($F(1,82) = 0.41$, $p = .526$), nor a significant main effect of Match ($F(1,82) = 0.30$, $p = .583$), but found a significant interaction between Type and Match ($F(1,82) = 34.71$, $p < .001$). There was no between-subjects effect of List ($F(7,82) = 0.63$, $p = .733$). Paired-samples *t*-tests show that participants responded more accurately to the question "What was the animal holding?" in the partial mismatch condition ($M = .99$, $SD = .03$) compared to the complete mismatch condition ($M = .75$, $SD = .40$, $t(89) = -5.89$, $p < .001$), and responded more accurately in the complete match condition ($M = .99$, $SD = .03$) compared to the partial match condition ($M = .74$, $SD = .40$, $t(89) = 5.95$, $p < .001$).

5

Discussion

The aim of Experiment 1 was to see whether participants combined multiple objects in their mental simulations. The results showed that participants were significantly faster when there was a complete match compared to a partial match, suggesting that comprehenders simulate multiple objects during language comprehension. The findings further suggest that during this sentence-picture verification task, participants compare what is currently being simulated with what is shown in the picture. When there was a complete mismatch (i.e., no overlap) with the preceding sentence, participants were significantly more accurate and faster compared to when there was only a partial mismatch (i.e., only the object held by the animal matched). This provides support for the idea that when there is complete overlap between simulation and image, there is facilitation of the participant's response. Similarly, it is equally easy to identify when there is completely zero overlap between simulation and image. However, as soon as one of the objects in the picture overlaps with the simulation while the other does not, there appears to be interference, which can be seen in both the decreased accuracy scores and the increased response times. The interpretation of the accuracy scores, however, should be taken with caution, as the accuracy scores overall were nearly 100% and the percentage differences between conditions were between only 2% and 3%. Such small differences in accuracy may therefore not be very meaningful.

When examining the accuracy responses to the comprehension question “what was the animal holding?” participants on average had high accuracy scores, suggesting that they properly read the sentences. Interestingly, the participants performed best in the partial mismatch and complete match conditions, meaning that when the preceding image showed the correct tool it facilitated their response, and if it showed the incorrect tool it interfered with their response. It is possible that once they were presented with the comprehension question, they could no longer remember whether it was the sentence or the image that contained the correct answer to the comprehension question.

Experiment 2

Experiment 1 showed that participants respond significantly faster when the picture completely matches what was stated by the sentence compared to when it only partially matched. This suggests that we indeed combine objects in mental simulations during language comprehension. Experiment 2 was conducted to examine whether participants still show the same match effect as in Experiment 1 when participants are instructed to respond to the *tool* as opposed to the *animal*. Experiment 2, therefore, is a conceptual replication of Experiment 1. If the results from Experiment 2 do not replicate those of Experiment 1, this would suggest that task instructions can modulate the contents of mental simulations. Specifically, if asking participants to respond only to the *tool* leads to only the tool being mentally simulated, this would lead to no significant differences in response times between the complete match and partial match conditions. However, if changing the instructions does not lead to different results compared to Experiment 1, this would mean that comprehenders routinely generate complete mental simulations of texts.

Method

Participants. We recruited 100 Dutch participants (M_{age} = 20.42 years, SD_{age} = 3.72 years, 80 females, 20 males) from the Bachelor of Psychology at the Erasmus University Rotterdam. Participants were excluded if their average accuracy was below 80%, as a result of this exclusion criteria data from 8 participants were excluded in the analyses, resulting in a sample of 92 participants.

5

Materials. The same materials were used as in Experiment 1.

Design and Procedure. The design of Experiment 2 was identical to that of Experiment 1. The only difference in the procedure was that participants were instructed to respond "YES" (the "L" key) if the *tool* matched the one described in the sentence, and to respond "NO" (the "A" key) if it mismatched. As a result, the following four image types could be shown to the participants:

5. **Complete match:** correct object + correct animal → Participants give "yes" response
6. **Partial match:** correct object + wrong animal → Participants give "yes" response
7. **Partial mismatch:** wrong object + correct animal → Participants give "no" response
8. **Complete mismatch:** wrong object + wrong animal → Participants give "no" response

Results

Data analysis. The same analyses that were performed on the data of Experiment 1 were performed on the data from Experiment 2.

Table 1

Descriptives Experiments 1 and 2

Type	Match	*N*	Mean ACC (*SD*)	Mean RT (*SD*)
Experiment 1				
Partial	Match	90	.97 (*.06*)	1461 (*574*)
	Mismatch	90	.95 (*.06*)	1422 (*549*)
Complete	Match	90	.99 (*.04*)	1387 (*557*)
	Mismatch	90	.98 (*.04*)	1358 (*561*)
Experiment 2				
Partial	Match	92	.92 (*.16*)	1687 (*623*)
	Mismatch	92	.98 (*.07*)	1622 (*627*)
Complete	Match	92	.98 (*.05*)	1501 (*545*)
	Mismatch	92	.99 (*.04*)	1560 (*568*)

Accuracy. The rmANOVA showed a significant main effect of Type, ($F_1(1,91) = 20.78$, $p < .001$; $F_2(1,47) = 49.32$, $p < .001$), a main Match effect ($F_1(1,91) = 10.53$, $p = .002$; $F_2(1,47) = 53.84$, $p < .001$), and a significant interaction between Type and Match ($F_1(1,91) = 8.13$, $p = .005$; $F_2(1,47) = 30.46$, $p < .001$). List did not interact with any of the variables and therefore was excluded from the rmANOVA. A paired-samples *t*-test showed that participants were significantly more accurate in the complete match condition ($M = .98$, $SD = .05$) compared to the partial match condition ($M = .92$, *SD*

= .16, $t_1(91) = 3.93$, $p < .001$, Cohen's $d = 0.41$; $t_2(47) = 7.36$, $p < .001$). A second paired-samples *t*-test illustrated that participants were also significantly more accurate in the complete mismatch condition ($M = .99$, $SD = .04$) compared to the partial mismatch condition ($M = .98$, $SD = .07$, $t_1(91) = 2.61$, $p = .011$, Cohen's $d = 0.27$; $t_2(47) = 2.79$, $p = .008$).

Response Times. The rmANOVA illustrated a significant effect of Type, showing that images that either completely matched or completely mismatched led to shorter response times compared to when they partially matched or mismatched ($F_1(1,84) = 28.61$, $p < .001$; $F_2(1,47) = 14.09$, $p < .001$). There was no significant effect of Match ($F_1(1,84) = 0.05$, $p = .821$; $F_2(1,47) = 0.10$, $p = .750$), but there was a significant interaction between Type and Match ($F_1(1,84) = 4.91$, $p = .029$; $F_2(1,47) = 6.24$, $p = .016$). There was no significant between-subjects effect of List ($F_1(7,84) = 0.76$, $p = .620$), but there was a significant interaction between List and Match ($F_1(7,84) = 2.14$, $p = .048$).

In order to investigate the interaction found between Type and Match, a simple main effects analysis was performed using a Bonferroni adjustment[11]. The results of the analysis showed that participants responded significantly faster in the match condition when there was a complete match compared to when there was a partial match ($F(1,84) = 23.48$, $p < .001$). Conversely, in the mismatch condition participants showed no significant difference in response time when there was either a complete or partial mismatch ($F(1,84) = 3.84$, $p = .053$).

Paired samples *t*-tests were performed to see whether there was a significant difference between the complete match and partial match conditions, and

[11] Although the simple main effects analysis was not pre-registered, it is standard to include it when an interaction is found in an ANOVA. As such, it was included here and not under the heading "Exploratory Analyses".

between the complete mismatch and partial mismatch conditions. The analyses showed that participants responded significantly faster when there was a complete match (*M* = 1501ms, *SD* = 545ms) compared to when there was a partial match (*M* = 1687ms, *SD* = 623ms, $t_1(91) = -4.82$, $p < .001$, Cohen's $d = -0.50$; $t_2(47) = -3.96$, $p < .001$, see Figure 2). Interestingly, participants did not respond significantly faster when there was a complete mismatch (*M* = 1560ms, *SD* = 568ms) compared to when there was a partial mismatch (*M* = 1622ms, *SD* = 627ms, $t_1(91) = -1.86$, $p = .067$, Cohen's $d = -0.19$; $t_2(47) = -0.95$, $p = .347$).

5

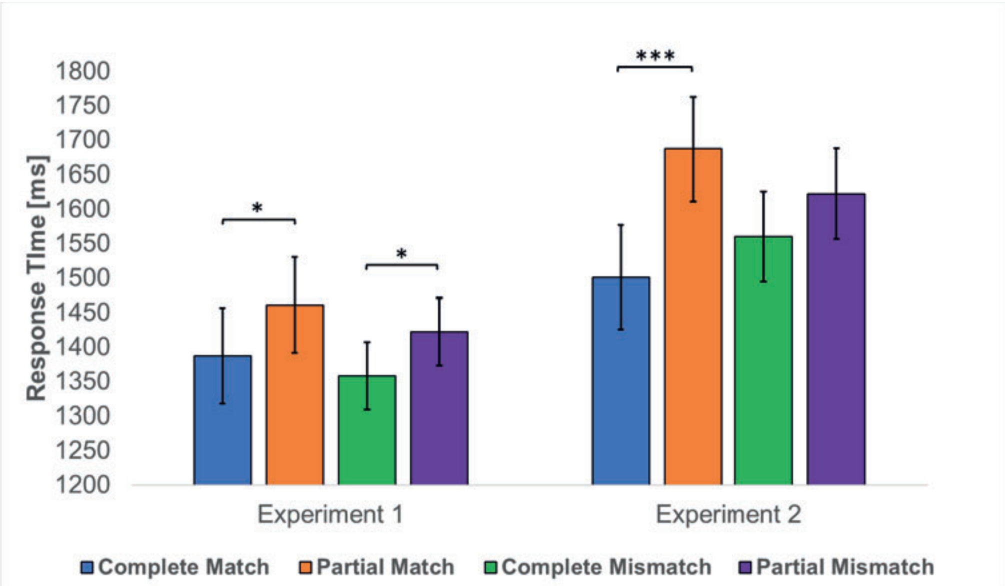


Figure 2. Average response times per condition of Experiment 1 and Experiment 2. * $p < .05$, *** $p < .001$

Comprehension Accuracy. Participants on average had high comprehension accuracy (*M* = .82, *SD* = .21), suggesting that they did properly read the sentences in their entirety. An rmANOVA was performed on the comprehension accuracy data and found a significant main effect of Type ($F_1(1,84) = 7.69$, $p = .007$), a significant main effect of Match ($F_1(1,84) =$

6.99, $p = .010$), and a significant interaction between Type and Match ($F_1(1,84) = 52.47$, $p < .001$). There was no between-subjects effect of List ($F_1(7,84) = 0.58$, $p = .769$). Paired-samples *t*-tests show that participants responded more accurately to the question "which animal was holding the tool?" in the partial mismatch condition ($M = .97$, $SD = .09$) compared to the complete mismatch condition ($M = .69$, $SD = .41$, $t(91) = -6.52$, $p < .001$), and responded more accurately in the complete match condition ($M = .98$, $SD = .06$) compared to the partial match condition ($M = .64$, $SD = .42$, $t(91) = 7.89$, $p < .001$).

5

Exploratory Analyses. The analyses in this section were not pre-registered online before data collection had started. A random-effects meta-analysis was performed using the *metafor* (Viechtbauer, 2010) package in R (version 3.6.0) to compare the facilitation effect (i.e., the difference between the complete and partial match conditions) and the interference effect (i.e., the difference between the complete and partial mismatch conditions) across Experiments 1 and 2. The code used in R for the analyses can be found in the Appendix. The meta-analysis for the facilitation effect showed that participants across Experiments 1 and 2 responded 129ms [95% CI: 19ms, 239ms] faster when there was a complete match compared to a partial match ($p = .021$). Heterogeneity was significant ($Q(1) = 4.56$, $p = .033$), indicating that the facilitation effect was larger in Experiment 2 than in Experiment 1. The meta-analysis for the interference effect illustrated that participants across Experiments 1 and 2 responded 63ms [95% CI: 24ms, 102ms] faster when there was a complete mismatch compared to when there was a partial mismatch ($p = .002$). Heterogeneity was not significant ($Q(1) = 0.002$, $p = .962$).

Discussion

The aim of Experiment 2 was to replicate the findings of Experiment 1 using the same method except for providing different instructions. In the current experiment participants were told to respond to the object that was in the cartoon animal's hand. The results of the analyses showed that participants were significantly faster when the picture completely matched what was stated in the sentence compared to when it partially matched. There was no significant difference between the complete mismatch and partial mismatch conditions. The effect in response times found in the match condition and the lack of effect in the mismatch condition is what drove the interaction in the rmANOVA. A simple main effects analysis confirmed that, indeed, participants responded significantly faster in the complete match condition compared to the partial match condition, while no significant differences are present within the mismatch condition.

Participants were also significantly more accurate when the image was either completely matching or completely mismatching the preceding sentence compared to when they were only partially matching or mismatching. As with Experiment 1, however, these small percentage differences are limited in their meaningfulness and should be interpreted with caution.

The meta-analysis supports the hypothesis that multiple objects are mentally simulated during language comprehension as when both objects were shown in the image there was clear facilitation of the response, as seen by the 129ms facilitation effect across both experiments. Moreover, when one of the two components of the image matched the sentence when the target item mismatched, there was a 63ms interference effect across both experiments.

Comprehension accuracy was highest in the partial mismatch condition compared to the complete mismatch condition, and in the complete match condition compared to the partial match condition. These findings illustrate that participants responded more accurately when the preceding image portrayed the correct animal. It is possible that once they were presented with the comprehension question, they could no longer remember whether it was the sentence or the image that contained the correct answer to the comprehension question.

General Discussion

Much research until now has focused on which object properties are present in mental simulations, but none have examined whether multiple objects can be combined in a mental simulation. The current study aimed to discover whether comprehenders combine multiple objects in their mental simulations, and whether this is dependent on task instructions. In Experiment 1 participants responded to images of cartoon animals holding a tool and were asked to response affirmatively if the pictured cartoon animal matched the animal described by the previous sentence. In Experiment 2 participants responded to the same stimuli but instead had to respond affirmatively if the pictured tool matched the tool described by the previous sentence. The images shown could either completely match, partially match, completely mismatch, or partially mismatch the preceding sentence.

The findings from both Experiments 1 and 2 provide support for our prediction that participants would respond faster when there was a complete match compared to a partial match. This can be taken as evidence that comprehenders combine multiple objects in their mental simulations during language comprehension. If they would not combine objects in a mental simulation there would be no difference between the complete match and partial match conditions, as participants were only instructed to respond to one of the objects in the image (either the cartoon animal or the tool), and thus would not have needed to simulate the other object for the sake of the task.

Therefore, task instructions also do not seem to modulate the contents of mental simulations.

These findings run counter to what was found by Lebois et al. (2015), who observed that task instructions can influence concept activation. However, it is similarly possible that the nature of our task caused both animal and tool to be simulated, as participants also had to respond to comprehension questions about the sentences, thus perhaps directing them to simulate both objects. We believe, however, that the comprehension questions only ensured that participants actually read each sentence, rather than stopping once they had read the name of the object which had to be verified in the subsequent picture. Although it may be of interest for future studies to explore the influence of comprehension questions on mental simulations, this was outside of the scope of the current study.

An additional finding from the current study was that in Experiment 1, participants responded significantly faster when there was a complete mismatch compared to a partial mismatch, but this was not replicated in Experiment 2, although the difference was in the predicted direction. A meta-analysis performed on the data from both experiments nonetheless found a significant interference effect, suggesting that when one of the objects matches what is mentally simulated, while the target mismatches, interference is generated. We hypothesized that comprehenders compare the contents of their mental simulation to the image during a sentence-picture verification task. Thus, when there is complete overlap between the pictured image and the mental simulation, facilitation occurs. Similarly, when there is no overlap whatsoever, it is very easy to respond that the image does not represent the previous sentence. However, if there is only a partial overlap, interference occurs. It is likely that the reason Experiment 2 did not replicate this finding is because in Experiment 2 participants were asked to react to the tool in the picture, which was smaller in size compared to the cartoon animal. If it were simply due to the fact that the agent in the text is more salient than the instrument, we should have seen no significant difference in the match condition either. As it is, it seems more likely that the interaction here is driven by the differences in object sizes in the picture. Perhaps if our experiment had used larger images this effect would also have been replicated, but this is something that could be investigated in future replications.

The ECCo theory proposed by Lynott and Connell (2010) proposes that during conceptual combinations, compound nouns can either be simulated in a destructive (where one of the concepts is reduced) or a non-destructive manner (where both concepts remain intact in the simulation) and that the linguistic system interacts with the simulation system to determine how these concepts need to be combined. The study by Connell and Lynott (2011) showed that it is easier for comprehenders to create non-destructive interpretations (i.e., for “octopus apartment”, the interpretation “a place where an octopus might live” is considered non-destructive) than it is to create destructive interpretations (i.e., “an apartment with eight rooms” requires the reduction of one of the properties in the compound noun) (p.4). The finding in our current study that objects can be combined in mental simulations fall in line with what is proposed to happen in a non-destructive conceptual combination. As conceptual combinations have not yet been tested using a sentence-picture verification paradigm, our study is the first to suggest that this link between visual representations and conceptual combinations actually exists in mental simulations.

Furthermore, in the Introduction we argued that, if participants read a sentence containing two objects (i.e., an animal and a tool), but only simulate one of them (i.e., the object requiring a response), it could be argued that a complete mental simulation is not required for language comprehension. If language comprehension occurs even without a complete mental simulation, this would imply that language comprehension does not require mental simulations. By showing that comprehenders do represent both objects in their mental simulation, regardless of task instructions, we provide preliminary evidence that a complete mental simulation becomes activated (and perhaps even is required) during language comprehension.

To conclude, studies on mental simulations have thus far only measured the activation of only a single object property in mental simulations during language comprehension. The current study examined whether multiple objects are combined in mental simulations and found support that this is indeed the case, regardless of task instructions. These results are in line with studies that have examined conceptual combinations in mental simulations. Future research should continue to focus on how mental simulations are constructed in novel and familiar contexts, and how they may be further altered by new incoming information.

Appendix

```
library(metafor)

# Definition of vectors

version <- c("match", "match", "mismatch", "mismatch")

experiment <- c("Experiment 1", "Experiment 2", "Experiment 1",
        "Experiment 2")

mpartial <- c(1461, 1687, 1422, 1622)

sdpartial <- c(574, 623, 549, 627)

mcomplete <- c(1387, 1501, 1358, 1560)

sdcomplete <- c(557, 545, 561, 568)

n <- c(90, 92, 90, 92)

corr <- c(.82, .81, .91, .86)


#Creation of data frame

data <- data.frame(version, experiment, mpartial, sdpartial, mcomplete,
        sdcomplete, n, corr, stringsAsFactors = FALSE)

data


#Random-effects meta-analysis of Match condition

dat <- escalc(measure="MC", m1i=mpartial, sd1i=sdpartial, m2i=mcomplete,
        sd2i=sdcomplete, ni=n, ri=corr,

          data=data, subset=(version=="match"))
```

5

```
res.match <- rma(yi, vi, data=dat)

res.match


#Random-effects meta-analysis of Mismatch condition

dat <- escalc(measure="MC", m1i=mpartial, sd1i=sdpartial, m2i=mcomplete,
        sd2i=sdcomplete, ni=n, ri=corr,

          data=data, subset=(version=="mismatch"))

res.mismatch <- rma(yi, vi, data=dat)

res.mismatch
```

6

Is Color Continuously Activated in Mental Simulations Across a Broader Discourse Context?

This chapter has been submitted as:

Hoeben Mannaert, L.N., Dijkstra, K., & Zwaan, R.A. (submitted). Is color continuously activated in mental simulations across a broader discourse context?

Abstract

Previous studies have provided contradictory information regarding the activation of perceptual information in a changing discourse context. The current study examines the continued activation of color in mental simulations across one (Experiment 1), two (Experiment 2), and five sentences (Experiment 3), using a sentence-picture verification paradigm. In Experiment 1, the sentence either contained a reference to a color (e.g., a red bicycle), or no reference to a color (e.g. bicycle). In Experiments 2 and 3, either the first or the final sentence contained a reference to a color. Participants responded to pictures either matching the color mentioned in the sentence, or shown in grayscale. The results illustrated that color was activated in mental simulations when the final sentence contained a reference to color. When the target object (e.g. bicycle) was mentioned in all sentences (i.e., in Experiment 2), color remained activated in the mental simulation, even when only the first sentence made a reference to a color. When the focus of the story was shifted elsewhere and the target object was not present across all sentences (i.e., in Experiment 3), color was no longer activated in the mental simulation. These findings suggest that color remains active in mental simulations so long as the target object is present in every sentence. As soon as the focus of the story shifts to another event, this perceptual information is deactivated in the mental simulation. As such, there is no continued activation of color across a broader discourse context.

Keywords: Mental simulation, situation models, color, language comprehension, grounded cognition

Introduction

"He wore a tall pointed grey hat, a long grey cloak, and a silver scarf. He had a long white beard and bushy eyebrows that stuck out beyond the brim of his hat." (Tolkien, 2005, p.25)

Whenever a new character is introduced in a story, certain details about the appearance of the character are mentioned. In the quote above, a clear description of the character Gandalf in the Lord of the Rings book series is given, and immediately the reader gains a good idea of how this character would look in real life. This mental representation of Gandalf is constructed at the beginning of the novel and is maintained throughout the series. When this character changes in appearance (Gandalf the Grey becomes Gandalf the White), this mental representation is presumably updated to accommodate these changes.

Mental representations do not merely describe the superficial text structure, but are thought to contain the meaning described by a text, also known as the situation model (Van Dijk & Kintsch, 1983). According to the Event Indexing Model, comprehenders integrate the characters and objects, goals, locations, events, and actions described in a text into a situation model (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995). Indeed, a plethora of studies have found evidence that multiple dimensions are tracked during language comprehension. For example, when discourse violates temporal, causal, protagonist-, and goal-related continuity, reading times increase (Zwaan, Radvansky, Hilliard, & Curiel, 1998). Moreover, spatial information is also tracked and incorporated into the situation model (Levine & Klin, 2001), especially when a narrative forces spatial relations to be causally relevant (Sundermeier, Van den Broek, & Zwaan, 2005). Even changes in neural activity have been associated with the tracking of the temporal dimension in short texts (Ditman, Holcomb, & Kuperberg, 2008), and it is thought that memory is worse for events that preceded a time-shift (Ditman et al., 2008; Speer & Zacks, 2005; Zwaan, 1996). As such, it is generally agreed upon that many dimensions are tracked during language comprehension and are incorporated into the updated situation model.

But how does the situation model update? Is all the information received in the narrative maintained and only select information updated when something changes? According to the Event Indexing Model (Zwaan et al., 1995), when changes occur in a narrative on any of the five dimensions, the situation model is updated. This form of updating is often referred to as *incremental updating,* as the situation model only updates the changed dimension while leaving the other dimensions intact. The Event Segmentation Theory (EST; Zacks, Speer, Swallow, Braver, & Reynolds, 2007), however, suggests that *global updating* occurs at event boundaries, meaning that the old situation model is replaced by a new situation model, and thus all dimensions are updated simultaneously. Findings in a study by Hoeben Mannaert, Dijkstra, and Zwaan (2019) could be taken as support for either theory, as changes to the implied shape of objects caused the initial shape to become deactivated as the more recent shape became activated. This could suggest that the altered shape completely replaced the initial shape in the situation model, potentially via global updating, but could also reflect incremental updating as only one entity needed to be altered. Interestingly, however, recent studies have shown that both incremental and global updating processes occur during language comprehension (Bailey & Zacks, 2015; Kurby & Zacks, 2012), suggesting that perhaps distinct mechanisms can play a role in updating situation models.

Given that comprehenders track events throughout a narrative, do they activate all of the associated information every time a particular dimension, for example entity, is mentioned? Or do they only activate when a change occurs on a particular dimension? For example, if one were to read the text: "*The boy* *rode on the red bicycle to the station. At the station he stepped off of his bicycle*." would the color *red* be reactivated when the word "bicycle" is mentioned the second time, or does color become irrelevant after the introduction of the object? Relating this back to the description of Gandalf, would readers create a mental simulation of Gandalf's appearance (including his grey or white cloak) each time the character is mentioned in the books, or are specific perceptual features irrelevant for these simulations?

Mental simulations are defined as the "reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind" (Barsalou, 2008, p.618). When the concept of the situation model was first introduced, the composition of the situation model was considered to be amodal in nature. More recently, however, many researchers are of the belief that the event representations that form the situation model are actually perceptual in nature (Barsalou, 1999; 2008; Zwaan, 2016). Indeed, much research has been published that provides support for sensorimotor activation during language comprehension (see Barsalou, 2008; Dove, 2016; or Kiefer & Pulvermüller, 2012, for extensive reviews on this topic). Specifically, many studies using the sentence-picture verification paradigm have found evidence that various object properties are included in mental simulations, such as object shape Zwaan, Stanfield, & Yaxley, 2002), orientation (Stanfield & Zwaan, 2001), motion Zwaan, Madden, Yaxley, & Aveyard, 2004), visibility (Yaxley & Zwaan, 2006), and color (Hoeben Mannaert, Dijkstra, & Zwaan, 2017; Pecher & Zwaan, 2012), but these have not examined the activation of these object properties over the course of more than one sentence.

So what happens to an object representation after its initial activation? A study by Pecher, Van Dantzig, Zwaan, and Zeelenberg (2009) showed that comprehenders can retain the implied shape and orientation of objects for 45 minutes, suggesting that mental simulations can be reactivated when a task

6

requires it. However, it has also been shown in several studies that when a time-shift occurs, or when a character changes location, that memory is worse for events that preceded those changes (Ditman et al., 2008; Morrow, Greenspan, & Bower, 1987; Radvansky & Copeland, 2006; Speer & Zacks, 2005; Zwaan, 1996). The lack of accessibility of previous information is thought to be due to the creation of a new situation model, which is thought to clear the information from previous events from active memory (Swallow, Zacks, & Abrams, 2009). Interestingly, it has also been found that memory for particular objects can be enhanced if those objects are present at event boundaries (Swallow et al., 2009). Therefore, if, for example, a protagonist were to cross event boundaries in a narrative, then the processing of this protagonist should be enhanced. But this tells us nothing about whether the perceptual information associated with that protagonist would also be maintained across event boundaries.

Interestingly, Swallow et al. (2009) found that perceptual information appears to be cleared from active memory at event boundaries, suggesting that mental representations at these boundaries are lacking perceptual detail. In their study, participants saw video clips containing objects that were either present at event boundaries or not, and after the clip participants had to choose which of the two shown objects was present in the video. The objects shown were tokens of the same category (e.g., two different chairs, of which one was shown in the video clip), enabling the authors to test the accessibility of perceptual information at event boundaries. Only when the target object was fixated and present at event boundaries was memory for perceptual information enhanced. These findings suggest that perceptual information is simulated when an event is ongoing (i.e., when only incremental updating is required), but when a new event occurs (i.e., when the situation model requires global updating) there is no purpose to maintaining that information.

Put simply, it appears that perceptual information is only carried across events when the objects are present at event boundaries. When this is not the case, the perceptual information is cleared from active memory. So what are the consequences of these findings in discourse processing? Although the findings from the studies on event perception provide indications for how situation models are created during discourse comprehension, because the studies were done on film clips and pertain to event perception, no strong conclusions can be drawn regarding discourse processing. Assuming for the moment that the same conclusions from event perception can be applied to discourse processing, this still leaves a lot of questions unanswered regarding what happens to perceptual information during a narrative. In a novel, objects and characters will not always be present at event boundaries, so does this mean that the perceptual information associated with them is lost? Or does it mean that this information is kept in long-term memory until it becomes relevant to the narrative and needs to be retrieved? Arguably, in our "red bicycle" example, the color of the bicycle would not *need* to be retrieved in subsequent mentions of the bicycle for readers to maintain a clear understanding of the situation described by the text.

On the other hand, if a comprehensive situation model is built at each section of a narrative, you would expect all relevant information to become reactivated at each mention of the object or character. Support for this assertion comes from the fact that comprehenders do retain perceptual information for long periods of time (Pecher et al., 2009). As such, there appear to be contradictory theories and studies regarding what happens to perceptual information during discourse processing, which is the focus of the current study.

The Current Study

To further our understanding of the underlying mechanisms of mental simulations in language comprehension, it is important that we know what the role of perceptual information is in mental simulations, and whether there is continued activation of this perceptual information. We conducted three experiments to examine this. Each experiment used a sentence-picture verification paradigm, where participants read sentences that described an object in combination with (or without) a color, followed by a picture that had to be verified. The picture shown either matched the color mentioned in the sentence, or was shown in grayscale. In Experiment 1, participants only viewed one sentence before they saw the picture, the sentence therefore either containing one reference to a colored object or only containing a reference to the object, without color (see Table 1). Based on the studies that have shown that participants respond significantly faster when pictures match the color implied in a text (Hoeben Mannaert et al., 2017; Pecher & Zwaan, 2012), we expect that participants will respond significantly faster to pictures shown in color compared to grayscale for sentences that make explicit reference to color. For sentences that contain no reference to color, we expected to find no difference in response times, as the items we used were low in color diagnosticity (Tanaka & Presnell, 1999).

Experiment 2 is an extension of Experiment 1 as here participants read two sentences, where either the first or the last sentence contains a reference to color. Based on the findings of Swallow et al. (2009) that suggested that perceptual information is erased from active memory once an event boundary is crossed, we expect to find no significant difference between the colored pictures and the grayscale pictures when only the first sentence contains a reference to color. Similar to Experiment 1, we expect to find a facilitation effect where participants respond significantly faster to the colored picture compared to the grayscale picture when the final sentence contains a reference to color.

For Experiment 3 we constructed stories in which the focus was shifted away from the target object for several sentences. Participants read five sentences before responding to the picture, where either the first or the final sentence contained a reference to color, while the middle three sentences were fillers that served to maintain coherence of the narrative, but were intended to shift attention away from the target object (see Table 1). Based on the findings from Experiments 1 and 2, we expect that, even when color is not mentioned in the final sentence, that participants will still show a facilitation effect, responding faster to colored pictures compared to grayscale pictures, in both sentence conditions (i.e. when the first or the final sentence contains a reference to color).

Ethics Statement

The participation in all experiments was voluntary. The participants subscribed to the experiments online via the university platform, and were told that by signing up for a study, they declare to voluntarily participate in this study. They were briefed with the content of each study and provided written consent. Participants were told they were free to terminate the experiment at any point in time without experiencing negative consequences. This study was approved by the Ethics Committee of Psychology at the Erasmus University Rotterdam, The Netherlands.

Preregistration

The predictions, exclusion criteria, design, methods, analyses, and materials of all the experiments reported in this article were pre-registered in advance of data collection and analysis on the Open Science Framework (OSF) to ensure confirmatory procedures were conducted according to a priori criteria. The preregistration for Experiments 1 and 2 can be viewed on https://osf.io/2nup7, the preregistration for Experiment 3 can be viewed on https://osf.io/bfm6p. Analyses that were not pre-registered are referred to in this article under the heading “Exploratory Analyses”.

Experiment 1

Method

Norming Study. As we were interested in testing the activation of color, it was important that the items we used were low in color diagnosticity (Tanaka & Presnell, 1999). For example, the word “pumpkin” is highly associated with the color orange, therefore even if the word “orange” is not included in the sentence, participants would still respond faster to a picture of an orange pumpkin compared to a grayscale pumpkin even without a color reference (Therriault, Yaxley, & Zwaan, 2009). To ensure that the findings from our study could not be confounded by effects of color diagnosticity, we performed a norming study to control for this. As such, we created a list of items that were partially taken from the low color diagnosticity items in the Tanaka and Presnell (1999) and Nagai and Yokosawa (2003) studies. As we needed more items than the ones used by those studies, we created the remainder of the stimuli ourselves.

Thirty-nine Dutch first-year Bachelor students at the Erasmus University Rotterdam (35 females, age range: 17-26 years old) took part in the norming study. Participants performed a word-picture verification task, where they first saw the word in the center of the screen, followed by a picture that was either shown in color or in grayscale. Forty-eight experimental items were shown in grayscale and in color (resulting in 96 experimental items shown in total), and 48 filler items were also shown in grayscale and in color (resulting in 96 filler items shown in total). Participants were instructed to respond “yes” (the “L” key) when the picture matched the preceding word, and were instructed to respond "no" (the "A" key) when the picture did not match. A paired-samples *t*-test found no significant color advantage in the response times for either experimental items ($t_1(38) = 0.06$, $p = .956$; $t_2(47) = 0.03$, $p = .980$) or for filler items ($t_1(38) = 1.80$, $p = .091$; $t_2 = 0.17$, $p = .864$). Accuracy scores also showed no significant color advantage for either experimental items ($t_1(38) = 0.89$, $p = .378$; $t_2(47) = 0.92$, $p = .361$) or filler items ($t_1(38) = 0.42$, $p = .680$; $t_2(47) = 0.54$, $p = .595$). As such, the items used in the current study show no evidence of having high color diagnosticity.

Participants. A power analysis was done using the results of Experiment 1 from Hoeben Mannaert et al. (2017), which used a similar paradigm to test whether color is represented in mental simulations. With an effect size of $f = 0.13$, it was calculated that a minimum of 82 participants would be required to find an effect if there is one ($\alpha = .05$, power = .80). To ensure our study had sufficient power after potential exclusions, 100 Dutch psychology students (77 females, $M_{age} = 20.79$ years, $SD_{age} = 3.07$ years) from the Erasmus University Rotterdam were recruited to take part in the current study. Participants were excluded if they had a total accuracy percentage of 80% or less, which led to the exclusion of five participants, resulting in a sample of 95.

Table 1.

Example of a sentence item and a picture item for each experiment.

<table>
<tr><th>Exp</th><th>Final Sentence</th><th>Sentence Example (in English)</th><th>Colored picture</th><th>Grayscale picture</th></tr>
<tr><td rowspan="2">1</td><td>Color</td><td>The boy rode on the red bicycle to the station.</td><td rowspan="6"></td><td rowspan="6"></td></tr>
<tr><td>Blank</td><td>The boy rode on the bicycle to the station.</td></tr>
<tr><td rowspan="2">2</td><td>Color</td><td>The boy rode on the bicycle to the station. At the station, he got off of his red bicycle.</td></tr>
<tr><td>Blank</td><td>The boy rode on the red bicycle to the station. At the station, he got off of his bicycle.</td></tr>
<tr><td rowspan="2">3</td><td>Color</td><td>The boy rode on the bicycle to the station. On the way he was passed by a bus. The bus suddenly cut him off. Luckily, he could evade in time and continue riding. At the station, he got off of his red bicycle.</td></tr>
<tr><td>Blank</td><td>The boy rode on the red bicycle to the station. On the way he was passed by a bus. The bus suddenly cut him off. Luckily, he could evade in time and continue riding. At the station, he got off of his bicycle.</td></tr>
</table>

Note. The examples provided here are in English, but the study used Dutch sentences. The translation therefore may not be exact.

Materials. One-hundred-ninety-two sentences were created that either included a reference to color (96 sentences) or omitted any reference to color (96 sentences). Of these sentences, half (96 sentences) were used as filler sentences, meaning that the picture shown afterwards did not match the object described in the sentence; the other half were experimental sentences. Given that each object was described by both a sentence containing a reference to color and a sentence containing no reference to color, each participant only received one version of these sentences, meaning that each participants read 48 experimental sentence items and 48 filler sentence items. Similarly, they saw 48 experimental pictures and 48 filler pictures, which were found using the Google search engine and edited using the Paint.NET software (version 4.1.5), were either depicted in one color or in grayscale, and did not exceed a 300 x 300 pixel resolution (approximately 7.9 x 7.9 cm on screen). In total they received 96 sentence items and 96 pictures. Additionally, participants received 24 comprehension questions to check whether they properly read the sentences. An example of the sentence items and pictures used in the current study can be seen in Table 1.

The experiment was programmed using E-Prime 2.0 Professional and participants completed the experiments in isolated cubicles with computers equipped with 24.1” TFT-IPS screens with a resolution of 1920 x 1200 and a ratio of 16:10.

Design. The experiment is a 2 (sentence: color vs. blank) x 2 (picture: color vs. grayscale) within-subjects design. Four lists were constructed to ensure sufficient counterbalancing, so that a sentence could either include a color referral or not, and that a picture could either be shown in color or in grayscale. An additional experiment from another study was performed by the participants in the same session, which was counterbalanced to be completed either before or after the current study, experiment order did not influence the results from the current study.

Procedure. Participants were instructed that they would perform a self-paced reading task using the spacebar and that they would see a picture after each sentence that either represented the object described in the sentence or did not. They were instructed to respond to the shape of the object and not to the color. If the picture matched the object in the sentence, they had to respond “yes” using the “L” key, and if it did not match then they had to respond “no” using the “A” key. Half of all filler items were followed by a comprehension question, which were closed questions requiring a “yes” or “no” response. The purpose of the comprehension question was to ensure participants properly read the sentences, rather than simply the object of the sentence. Before starting the experiment they received 6 practice items.

A trial looked as follows: Participants saw the > symbol left-aligned in the center of the screen for 1000ms. Subsequently the sentence was shown left-aligned in the center of the screen and remained on-screen until participants pressed the spacebar. Subsequently a fixation cross appeared in the center of the screen (center-aligned) for 500ms, after which the image appeared in the center of the screen (center aligned) and remained on-screen until participants provided a response.

Results

Data analysis. A repeated measures analysis of variance (rmANOVA) was run on the data, using ‘sentence version’ and ‘picture version’ as repeated measures variables. ‘List’ was used as a between-subjects variable, in order to improve the quality and power of our analyses (Pollatsek & Well, 1995). All response-time analyses were performed on correct responses only. Per participant the median response time was taken per condition, as is common in sentence-picture verification studies (Hoeben Mannaert, et al., 2016; Hoeben Mannaert et al., 2019; Pecher & Zwaan, 2012; Zwaan et al., 2002) to prevent extreme values from influencing the data. Subject analyses are denoted with the subscript "1" and item analyses are denoted with the subscript "2". As preregistered, we conducted rmANOVAs on accuracy scores and on response times.

Accuracy. The rmANOVA performed on the accuracy scores illustrated a significant effect of 'sentence' in both the subject and item analyses ($F_1(1,91) = 4.41$, $p = .039$; $F_2(1,47) = 4.78$, $p = .034$). There was a significant effect of 'picture' only in the item analyses ($F_1(1,91) = 2.52$, $p = .116$; $F_2(1,47) = 5.99$, $p = .018$). Similarly, there was only a significant interaction between 'sentence' and 'picture' in the item analysis ($F_1(1,91) = 3.40$, $p = .069$; $F_2(1,47) = 5.77$, $p = .020$). There was a significant interaction between 'list' and 'picture' ($F_1(3,91) = 2.87$, $p = .041$).

6

Exploratory analyses (accuracy). A paired samples *t*-test showed that participants responded significantly more accurately to the colored picture ($M = .98$, $SD = .04$) than to the grayscale picture ($M = .97$, $SD = 06$) when the sentence made reference to a color ($t_1(94) = 2.16$, $p = .033$, $d = 0.22$; $t_2(47) = 2.78$, $p = .008$). There was no significant difference in accuracy scores between the colored picture ($M = .99$, $SD = .04$) and the grayscale picture ($M = .98$, $SD = .04$) when the sentence contained no reference to color ($t_1(94) = 0.15$, $p = .880$, $d = 0.02$; $t_2(47) = 0.08$, $p = .936$).

Response Times. The rmANOVA performed on response times illustrated a significant effect of 'sentence', but only in the item analysis ($F_1(1,91) = 1.55$, $p = .216$; $F_2(1,47) = 5.00$, $p = .030$). Furthermore, both subject and item analyses showed a significant effect of 'picture' ($F_1(1,91) = 9.80$, $p = .002$;

$F_2(1,47) = 28.63$, $p < .001$) and a significant 'sentence'*'picture' interaction ($F_1(1,91) = 16.10$, $p < .001$; $F_2(1,47) = 11.72$, $p = .001$)

Exploratory analyses (response times). A paired samples *t*-test was conducted to examine the interaction between 'sentence' and 'picture', and found that participants responded significantly faster to the colored picture ($M = 821$ms, $SD = 324$ms) than to the grayscale picture ($M = 915$ms, $SD = 323$ms) when the sentence contained a reference to a color ($t_1(94) = -4.48$, $p < .001$, $d = -0.46$; $t_2(47) = -5.14$, $p < .001$). There was no significant difference between the colored picture ($M = 857$ms, $SD = 321$ms) and the grayscale picture ($M = 838$ms, $SD = 307$ms) when the sentence did not contain a reference to color ($t_1(94) = -1.17$, $p = .245$, $d = -0.12$; $t_2(47) = -0.56$, $p = .578$).

6

Discussion

As predicted, there was a significant color advantage when the sentence contained a reference to color, while no such advantage was present when the sentence did not contain a reference to a color. Although both the accuracy and the response time analyses support this conclusion, it should be noted that accuracy scores overall were very high (between 97% and 99% across conditions). Given that the significant difference in the color condition is only a difference of 1%, this is not very meaningful.

Experiment 1 has established that color is activated in mental simulations when it is mentioned for the first time, and thus support the findings of previous studies on color simulation (e.g., Hoeben Mannaert et al., 2017; Pecher & Zwaan, 2012). Experiment 2 serves to expand on this finding by examining whether this activation remains if participants read two sentences, where either the first or the final sentence contain a reference to color.

Experiment 2

The aim of Experiment 2 was to examine whether the activation of color in mental simulations would change across two sentences. In the current experiment, participants read sentences where either the first sentence contained a reference to a color, or the final sentence contained a reference to a color (see Table 1). We predicted that color would have deactivated if the second sentence made no reference to color, and in that condition expected to find no significant difference in response times between the colored picture and the grayscale picture. If color would not have deactivated by the second sentence, then we would expect to find a significant color advantage, similar to what was found in Experiment 1. For the condition where the final sentence contained a reference to color, we did expect to find a significant difference between the colored picture and the grayscale picture.

Method

Participants. One hundred Dutch psychology students (77 females, M_{age} = 20.47 years, SD_{age} = 3.34 years) from the Erasmus University Rotterdam were recruited to take part in the current study. Participants were excluded if they had a total accuracy percentage of 80% or less, as a result of this exclusion criteria three participants were excluded from the analysis. The final sample consisted of 97 participants.

Materials. The sentences from Experiment 1 were expanded to contain two sentences per item (see Table 1 for an example). The sentences either contained a reference to color in the first sentence or in the second sentence. The rest of the materials were identical to Experiment 1.

Design and Procedure. The design and procedure of Experiment 2 was identical to Experiment 1, except that participants were informed that they would see the picture after every two sentences.

Results

Data Analysis. The same analysis plan used for Experiment 1 was also used for Experiment 2.

Accuracy. The rmANOVA for accuracy scores found a significant effect of 'picture' ($F_1(1,93) = 5.97$, $p = .016$; $F_2(1,47) = 4.37$, $p = .042$), but no significant effect of 'sentence' ($F_1(1,93) = 0.0007$, $p = .980$; $F_2(1,47) = .001$, $p = .972$), nor a significant interaction between 'sentence' and 'picture' ($F_1(1,93) = 1.51$, $p = .223$; $F_2(1,47) = 0.53$, $p = .470$). 'List' significantly interacted with 'picture' ($F_1(3,93) = 14.61$, $p < .001$).

Exploratory analyses (accuracy). A paired samples *t*-test illustrated that participants responded significantly more accurately to the colored pictures ($M = .99$, $SD = .03$) compared to the grayscale pictures ($M = .97$, $SD = .04$) when the final sentence made a reference to color, but this was not significant in the item analysis ($t_1(96) = 2.48$, $p = .015$, $d = 0.25$; $t_2(47) = 1.96$, $p = .056$). There was no significant difference in accuracy scores between the colored picture ($M = .98$, $SD = .04$) and the grayscale picture ($M = .98$, $SD = .04$) when the first sentence made a reference to color ($t_1(96) = 0.87$, $p = .389$, $d = 0.09$; $t_2(47) = 0.68$, $p = .497$).

Response Times. The rmANOVA for response times found a significant main effect of 'picture' ($F_1(1,93) = 20.07$, $p < .001$; $F_2(1,47) = 39.40$, $p < .001$). There was no significant effect of 'sentence' ($F_1(1,93) = 2.90$, $p = .092$; $F_2(1,47) = 0.182$, $p = .672$), nor a significant interaction between 'sentence' and 'picture' ($F_1(1,93) = 0.31$, $p = .580$; $F_2(1,47) = 1.07$, $p = .307$). There was a significant interaction between 'sentence' and 'list' ($F_1(3,93) = 4.33$, $p = .007$).

Exploratory analyses (response times). A paired samples *t*-test on response times found that participants responded significantly faster to the colored picture (M = 842ms, SD = 309ms) compared to the grayscale picture (M = 891ms, SD = 276ms) when the final sentence made a reference to color ($t_1(96) = -3.27$, $p = .002$, $d = -0.33$; $t_2(47) = -4.23$, $p < .001$). Participants also responded significantly faster to the colored picture (M = 818ms, SD = 285ms) than to the grayscale picture (M = 878ms, SD = 295ms) when the first sentence made a reference to color ($t_1(96) = -3.54$, $p < .001$, $d = -0.36$; $t_2(47) = -2.82$, $p = .007$).

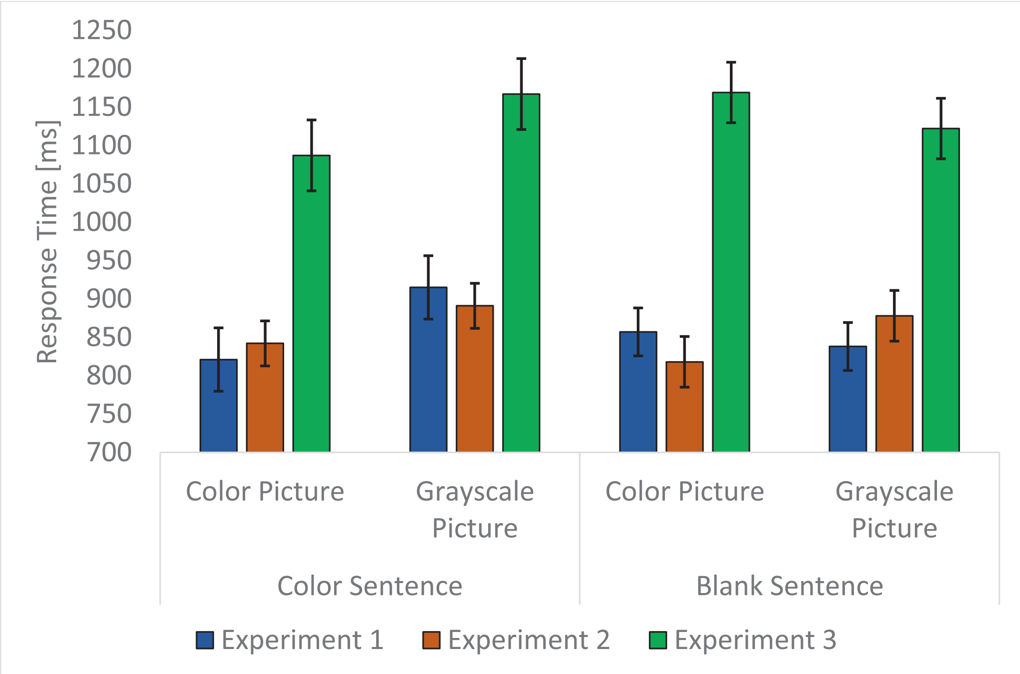


Figure 1. Bar graph displaying the average response times per condition for each experiment. Errors bars show 95% CI.

Discussion

We had predicted that color would have deactivated when the second sentence makes no reference to a color. Interestingly, the results showed that color had remained activated, regardless of whether color was mentioned in the first or the final sentence, as participants responded significantly faster to the colored picture compared to the grayscale picture. This suggests that, when a color is first mentioned, it becomes active in mental simulations, and remains active even in the following sentence.

Experiment 3

The aim of Experiment 3 was to examine how the activation of color would change in a wider discourse context. Participants in the current experiment read five sentences, where either the first or the final sentence contained a reference to color (see Table 1). The middle three sentences were filler sentences that maintained coherence within the story but referred to objects or events other than the target object in the first and final sentences. Examining the activation of color in mental simulations using several sentences allowed us to examine how this activation behaves in a more naturalistic context. Based on the findings from Experiment 2, we expected to continue to find a significant color advantage, regardless of whether color was mentioned in the first or final sentence in the texts.

Method

Participants. One hundred Dutch psychology Bachelor students were recruited from the Erasmus University Rotterdam (85 females, M_{age} = 19.93 years old, SD_{age} = 2.01 years). Four participants were excluded due to having an average accuracy below 80%, leaving us with a sample of 96 participants.

Design and Procedure. The design and procedure was the same as Experiment 1 and 2, except that participants read five sentences before seeing a picture. No other study was conducted before or after this Experiment.

Results

Analysis Plan. The same analysis plan was used as in Experiments 1 and 2.

Accuracy. The rmANOVA found a significant interaction between ‘sentence’ and ‘picture’, but only in the item analyses ($F_1(1,92) = 3.73$, $p = .057$; $F_2(1,47) = 4.77$, $p = .034$). There was no significant main effect of ‘sentence’ ($F_1(1,92) = 0.03$, $p = .867$; $F_2(1,47) = 0.02$, $p = .889$) or ‘picture’ ($F_1(1,92) = 2.68$, $p = .105$; $F_2(1,47) = 3.22$, $p = .079$). There was a significant interaction between ‘list’ and ‘picture’ ($F_1(3,92) = 3.58$, $p = .017$).

Exploratory analyses (accuracy). A paired samples *t*-test found that participants responded significantly more accurately when the picture shown was colored ($M = .99$, $SD = .03$) compared to when it was shown in grayscale ($M = .97$, $SD = .07$), when the final sentence made a reference to a color ($t_1(95) = 2.07$, $p = .041$, $d = 0.21$; $t_2(47) = 3.21$, $p = .002$). There was no significant difference between the colored picture ($M = .98$, $SD = .04$) and the grayscale picture ($M = .98$, $SD = .05$) when the first sentence referred to a color ($t_1(95) = 0.15$, $p = .880$, $d = 0.02$; $t_2(47) = 0.09$, $p = .929$).

Response Time. The rmANOVA found a significant interaction between ‘sentence’ and ‘picture’ in the subject analyses ($F_1(1,92) = 17.21$, $p < .001$), but not in the item analyses ($F_2(1,47) = 2.95$, $p = .092$). There was no significant main effect of ‘sentence’ ($F_1(1,92) = 1.77$, $p = .187$; $F_2(1,47) = 0.0001$, $p = .992$), nor a main effect of ‘picture’ ($F_1(1,92) = 0.96$, $p = .329$, $F_2(1,47) = 1.01$, $p = .321$). There was a significant interaction between ‘list’ and ‘sentence’ ($F_1(3,92)= 10.52$, $p < .001$).

Exploratory analyses (response time). A paired samples *t*-test was performed to examine the interaction between ‘sentence’ and ‘picture’. The results from the *t*-test showed that participants responded significantly faster to the colored picture ($M = 1087$ms, $SD = 368$ms) compared to the grayscale picture ($M = 1167$ms, $SD = 407$ms) when the final sentence made a reference to color ($t_1(95) = -3.38$, $p = .001$, $d = -0.35$; $t_2(47) = -1.92$, $p = .061$). When the first sentence contained a reference to color, the opposite pattern emerged. Participants responded significantly faster when the picture was shown in grayscale (M = 1122ms, SD = 382ms) compared to when it was shown in color (M = 1169ms, SD = 445ms, $t_1(95) = 2.33$, $p = .022$, $d = 0.24$), but this was not significant in the item analyses ($t_2(47) = 0.87$, $p = .388$).

Discussion

Consistent with Experiments 1 and 2, accuracy was very high across all conditions, and even though the participants were significantly more accurate when responding to the colored pictures compared to the grayscale pictures when the final sentence made a reference to a color, this difference was only 2%. Given that this difference is so small, it is questionable whether such a difference is meaningful.

Similar to the previous experiments (see Figure 1), when the final sentence contained a reference to color, color became activated in the mental simulations, as seen by the faster responses to the colored pictures compared to the grayscale pictures.

Contrary to our expectations, the analyses showed that participants do not respond faster to the colored pictures compared to the grayscale pictures when the first sentence referred to a color. In fact, the effect was reversed: participants responded significantly faster to the grayscale picture compared to the colored picture, suggesting that color now caused interference, rather than the facilitation observed in the previous experiments. However, this effect was found only in the subject analyses, not in the item analyses.

These findings suggest that, when participants read the final sentence that did not refer to the color mentioned in the first sentence, color was no longer activated in the mental simulation. The interference caused by seeing the colored picture suggests that the shown image does not match the mental simulation. As such, it appears that color is not needed in mental simulations to support language comprehension across a wider discourse context.

General Discussion

The aim of the current study was to examine the continued activation of color in mental simulations across a wider discourse context, as much contradictory information existed regarding the perceptual activation when changes occur to a situation model. Three experiments were conducted to test this using a sentence-picture verification paradigm. Experiment 1 examined the activation of color using a single sentence, Experiment 2 used two sentences, and Experiment 3 used five sentences.

Based on the findings of previous color simulation studies (Hoeben Mannaert et al., 2017; Pecher & Zwaan, 2012), we had expected to find a color advantage in Experiment 1 when sentences referred to a color (e.g. *The boy rode on the red bicycle to the station.*). Indeed, participants responded significantly faster to the colored picture compared to the grayscale picture when sentences contained a reference to color. When no color reference was given (*The boy rode on the bicycle to the station.*), there was no significant difference in response times between the colored and grayscale pictures. This experiment provided further support for color being activated in mental simulations when a reference is made to color.

Based on the findings by Swallow et al. (2009), we had expected to find that this color advantage would disappear when two sentences are provided, when the second sentence does not refer to a color (e.g., *The boy rode on the red bicycle to the station. At the station he stepped off of his bicycle*.”). Contrary to our expectations, Experiment 2 continued to show this color advantage. Participants responded significantly faster to the colored picture compared to the grayscale picture, regardless of whether the first or the final sentence contained a reference to color. The results from this experiment suggests that color continues to be active in mental simulations when only two sentences are provided. This result falls in line with the conclusions made by Swallow et al. (2009), whose findings suggested that perceptual information can remain activated in event models when objects are salient and present at event boundaries. In the current experiment, because the object (e.g., bicycle) was present in both sentences, even if a new event model had to be created, the perceptual information would not have been erased from active memory as it was still relevant to the situation model.

Linking this back to the updating mechanism proposed by the Event Indexing Model, it is possible that readers use incremental updating to update their situation model in this experiment, given that the event described by the two sentences can be considered to be ongoing, rather than two separate events. As such, the current situation model would only need elaborating to incorporate the incoming information, but a new situation model would not need to be built, thus making global updating unnecessary in this context.

In Experiment 3, we expected to continue to find this color advantage as seen in Experiment 2, when participants would read five sentences. Specifically, we expected that participants would respond faster to colored pictures compared to grayscale pictures when either the first or final sentence made a reference to a color. Importantly, in this experiment the middle three sentences did not explicitly refer to the target object (e.g., *The boy rode on the red bicycle to the station. On the way he was passed by a bus. The bus suddenly cut him off. Luckily, he could evade in time and continue riding. At the station he stepped off of his bicycle.*). In this way, we could examine whether the perceptual information would become deactivated in the mental simulation in a more naturalistic discourse context.

Interestingly, the findings from Experiment 3 were the exact opposite to our expectations, as participants responded faster to the grayscale picture compared to the colored picture, when the first sentence made a reference to a color. This might suggest that the colored picture caused interference by it not matching up with the object activated in the mental simulation. Moreover, it would mean that color does become deactivated over time as the focus of the narrative shifts to other events and therefore loses its relevance. As the color of the object is irrelevant to the goal of the reader (i.e., comprehension of the story), it appears that it is not needed to actively maintain perceptual information in the situation model. These results also fall in line with the findings by Swallow et al. (2009), who argued that perceptual information is erased from active memory if the target object is not present at event boundaries. As such, it is possible that in this experiment, a global updating mechanism (as proposed by the EST) is used by participants to update their situation model.

Two important conclusions can be drawn from this study. Firstly, perceptual information becomes active in mental simulations when they are referred to (Experiment 1), or when they remain relevant to the situation model (Experiment 2). When this perceptual information is no longer relevant to the narrative (Experiment 3), it no longer remains active in the mental simulation. Secondly, these findings suggest that a complete situation model, containing all related information, is not created during discourse processing. Only information which is required for language comprehension needs to be activated in the situation model.

The limitation of this study is as follows. Although we examined the activation of color in an arguably more naturalistic context than single-sentence studies, an experiment using five sentences can still be considered impoverished compared to texts occurring in real life (Graesser, Millis, &

Zwaan, 1997). As such, the generalizability of these findings to discourse processing as a whole is somewhat limited.

In conclusion, the current study has illustrated that color remains active in mental simulations so long as the target object is present in every sentence. As soon as the focus of the story shifts to another event, this perceptual information is deactivated in the mental simulation. As such, there is no continued activation of color across a broader discourse context.

7

General Discussion

The main research question of interest in this dissertation can be formulated as follows: how is the perceptual content of mental simulations affected by language comprehension? This question can be divided into several subquestions, which were discussed in the previous chapters.

What is the role of color in mental simulations?

Chapter 2 and Chapter 6 examined the role of color in mental simulations. In Chapter 2, participants read sentences that implied that an object had a particular color (e.g., *The driving instructor told Bob to stop/go at the traffic lights*), before responding to a picture of an object where the color either matched or mismatched the one implied by the previous sentence. Each experiment provided participants with the same sentences, but the pictures were shown in decreasing levels of saturation, in order to examine the activation of color in mental simulations. Given that previous studies had shown contradictory results regarding whether the match effect could be replicated for color as well, it was important to resolve this issue once and for all. The findings of our study provided evidence that color is activated in mental simulations, as participants responded significantly faster to the matching pictures compared to the mismatching pictures, when they were shown in full saturation. When the saturation of the pictures was reduced to the lowest level at which the hue could still be recognized, the match effect disappeared. Similarly, when the pictures were shown in grayscale, the match effect disappeared. These findings suggests that the match effect is actually created by the interference that occurs when there is a vivid difference between the pictured object and the mental simulation in the mismatch condition. Once this difference is removed (i.e., by reducing the saturation in pictures or by showing them in grayscale), there is no more interference, leading to the elimination of the match effect.

Chapter 6 examined the continued activation of color in a broader discourse context. In the first experiment, participants either read sentences that contained a reference to a color (e.g., *The boy rode on the red bicycle to the station*), or made no reference to a color (e.g., *The boy rode on the bicycle to the station*), before responding to pictures that were shown in a color matching the one described by the previous sentence, or shown in grayscale. Participants responded significantly faster to the pictures shown in color, when the sentence contained a reference to that color. When no reference was made to a color, there was no significant difference between the responses to the colored or grayscale pictures. The findings from this experiment support those from Chapter 2, as both studies provide evidence for the activation of color in mental simulations when this is relevant to understanding the events described in the text.

In the second experiment, participants read two-sentence items, where either the first sentence contained a reference to a color, or the final sentence contained a reference to a color. Here, participants responded significantly faster to the colored picture compared to the grayscale picture, regardless of whether the color was mentioned in the first or final sentence. This finding suggests that color remains activated in mental simulations when the target object is continuously being referred to. In the third experiment, the attention was shifted away from the target object using three filler sentences (out of five sentences in total), where either the first or final sentence made a reference to a color. Here, participants responded significantly faster to the picture shown in grayscale, compared to the colored picture, when the first sentence made a reference to a color. This suggests that color becomes deactivated in mental simulations when the focus is shifted away from the target object, thus making the activation of color unnecessary for understanding the events of the text.

The findings from Chapter 2 and Chapter 6 are in line with the studies that show that color is one of the object properties that is actively simulated during language comprehension (e.g. Zwaan & Pecher, 2012). Chapter 2 further highlights the importance of performing replication studies, as there was disagreement in the literature regarding the role of color in mental simulations. Furthermore, the findings from Chapter 6 fall in line with those from Swallow et al. (2009), who found that perceptual information can remain activated in event models when objects are salient and present at event boundaries. In Experiment 2 of our study, color remained activated as the target object was present in both sentences. When the focus was shifted away from the target object, color was no longer activated in mental simulations. This finding supports both weakly and strongly embodied theories of cognitions, as color is activated in mental simulations when it is required for the comprehension of the events in a narrative, but deactivates when it is no longer relevant to the narrative.

How are mental simulations updated?

Chapter 3 and Chapter 4 targeted the question of how mental simulations are updated, whereby Chapter 3 focused specifically on the activation of shape information in mental simulations in a changing discourse context, and Chapter 4 on the updating of situation models in older and young adults.

In Chapter 3, participants read sentences that either implied a change in shape over the course of two or four sentences, or implied no change in shape, before responding to a picture that either matched the initial shape (which was implied by the first sentence) or matched the final shape (which was implied by the final sentence). Although many studies that have been published have examined the updating of situation models, this has never been examined before at the level of the mental simulation. We were interested in examining how the activation of perceptual information changes across several sentences.

Given that comprehenders can reactivate perceptual information in a mental simulation if this is required, it is possible that this reactivation would be necessary in order to understand that changes are occurring to an object in a narrative. Another possibility is that, during the updating process, the contents of the mental simulations are replaced by the new incoming information. Our study found support for the second possibility, as participants responded significantly faster when the shape matched the one implied by the final sentence, suggesting that the initial shape becomes deactivated in a mental simulation as the more recent shape is activated. These findings support those of Sato, Schafer, and Bergen (2013), who reported that mental simulations are updated when participants receive information contradictory to the initially simulated shape.

As such, the findings from Chapter 3 suggest that, in order to comprehend changes occurring in a narrative, the mental simulation only activates the perceptual information that is required for understanding the current (i.e., most recent) event. Similar to the conclusions made from Chapter 6, this provides support for the embodied theories of cognition, as only the relevant perceptual information is activated in the mental simulation to support language comprehension.

Chapter 4 examined how situation models update in young and older adult populations by reviewing the current literature on this topic. Due to aging, many cognitive processes become impaired, which can lead to various problems in the everyday functioning of older individuals. Despite the large number of studies that report declines in cognitive functioning, many studies on situation model updating in older adults have focused on the preservation of cognitive skills. The review reports several studies that show that older adults are equally capable of updating their situation models compared to young adults, even if they take longer to do so. Interestingly, older adults also may create a more elaborate situation model compared to young adults, as their situation model appears to be resistant to contradictory information. As such, younger and older adults appear to have similar capacities for updating their situation models to represent the current events unfolding in a narrative.

Is perceptual information combined in mental simulations?

Chapter 5 examined whether perceptual information can be combined in mental simulations. Here, participants read sentences describing an animal performing an action with a tool, before responding to pictures of cartoon animals holding a tool. They were either instructed to judge whether the pictured animal was mentioned in the previous sentence, or whether the pictured tool was mentioned in the previous sentence. The shown picture could either be a complete match (both animal and tool matched the previous sentence), a partial match or mismatch (either the animal or the tool would match), or a complete mismatch (neither animal nor tool would match the previous sentence). If comprehenders create a complete mental simulation of an event to support language comprehension, then both the animal and the tool should be activated in a mental simulation, regardless of task instructions. However, if task instructions can modulate the contents of mental simulations, then this would have serious consequences for our current understanding of the role that mental simulations play in language comprehension. The results of our study illustrated that participants responded significantly faster to the complete match condition compared to the partial match condition, regardless of task instruction, suggesting that comprehenders do create a complete mental simulation to understand the events described in a sentence.

The findings from Chapter 5 are consistent with the findings of Chapter 6, as both studies illustrate that, when perceptual information is necessary for language comprehension, it is activated in a mental simulation. When this information is not necessary (as in the third experiment of Chapter 6), then it does not activate.

General conclusions

Several conclusions regarding the role of mental simulations in language comprehension can be drawn from the findings of this dissertation. In Chapter 1, several arguments were put forward regarding the requirement of sensorimotor information for language comprehension. The findings of the current dissertation are in line with the growing body of evidence that sensorimotor information is partially required for language comprehension. The word ‘partially’ is important here. We did not find support for a complete dependence on sensorimotor systems, as this would mean that perceptual information would have to be continuously activated in mental simulations. Instead, our findings illustrate that perceptual information is activated in mental simulations only when this is relevant for understanding the events described in the text. For instance, when color is referred to in a sentence, this leads to the activation of color in mental simulations. When attention is shifted away from the target object for a while before reintroducing that object – but without a reference to color – this leads to the deactivation of color in mental simulations. This means that, when an entity is introduced in a text, the perceptual information associated with that entity is stored, but is not reactivated each time that entity is reintroduced in the text.

Secondly, we found that, when changes to an object’s shape are implied in a text, the initially activated shape becomes deactivated as the more recent shape activates in a mental simulation. This suggests that, when changes are made to entities described in a narrative, that the initially simulated entity is replaced in the mental simulation by the new version. Putting this in the context of reading a novel, this would mean that if, for example, one reads about a character who cuts her hair, the mental simulation would update (or replace) the initial simulation of that character by deactivating that initial simulation first.

Thirdly, we found that, when multiple objects are mentioned in a sentence, both objects are activated in a mental simulation, irrespective of task instructions. This suggests that comprehenders routinely activate mental simulations to support language comprehension. If only one of the objects would have been simulated, then this would mean that mental simulations are not required for language comprehension. However, given that both objects were routinely activated during language comprehension, this points to a clear role of mental simulations in language comprehension.

Finally, we found that, despite the decline in cognitive functions as a result of aging, this does not impair the ability of comprehenders to update their situation models. If anything, due to their extensive reading experience, it appears as though older adults create a more elaborate situation model that is more resilient to contradictory information.

Together, these findings provide support for the embodied theories of cognition. When perceptual information is required to understand the events described in a text, this information is activated in a mental simulation. As soon as the events of a narrative change, this perceptual information is deactivated and replaced by more recent information in order to accurately reflect the current situation model. As such, mental simulations can be considered to be dynamic in the sense that they continuously update and change their perceptual contents to reflect the narrative.

Suggestions for future research

The main problem with theories of grounded (or embodied) cognition is that they lack specificity. As mentioned in Chapter 1, it is currently impossible to determine whether cognition occurs via interactions with the sensorimotor systems or via a partial dependence on sensorimotor systems. The research described in this dissertation suggests that sensorimotor information is required for language comprehension. However, because several aspects of these theories are vague (and at times contradictory) in how they describe the role of sensorimotor information in cognitive processes, it is essentially possible for both weak embodiment and secondary embodiment to serve as potential explanations for these findings. This means that in the future, theories need to be specified further so that it is possible to test empirically whether semantic content is grounded by interaction or by using the sensorimotor systems directly for the various cognitive functions.

Another problem with research on mental simulations is that they often use single sentences to study the activation of perceptual information during language comprehension. Although some of the research in the current dissertation has attempted to remedy this problem, it is important for future research to continue improving the quality of texts used in experiments, so that the findings of those studies can be generalized to how language comprehension functions in real life.

The current dissertation has examined the role of mental simulations in language comprehension by using empirical research to test specific hypotheses that were pre-registered before data collection started. One of the pillars of good research is the pre-registration of the hypotheses, method and analysis plan of a study before starting data collection. In a time where the reliability of psychological findings is questioned due to the occurrence of questionable research practices, it is crucial that researchers become accustomed to carrying out pre-registrations. By continuing the movement toward good research practices, the quality of research in psychology can be guaranteed, rather than speculated upon.

S

Summary

The underlying mechanisms of language comprehension have been a topic of debate for centuries. One of the most popular theories on language comprehension are theories of grounded cognition. These theories suggest that we use the same sensorimotor regions in the brain during activity as during cognitive processes, through the use of mental simulations. Many behavioral and neuroimaging studies have been conducted over the past two centuries to show that the contents of these simulations are modal in nature and that they are involved in language comprehension. The aim of the current dissertation is to further examine the mechanisms of mental simulations in language comprehension. This was accomplished by exploring the role of color in mental simulations **(Chapter 2)**, the process of updating in mental simulations **(Chapter 3)**, its use in language comprehension in younger versus older adults **(Chapter 4)**, the ability for multiple objects to be combined in mental simulations **(Chapter 5)**, and finally, the examination of the continuous activation of mental simulations **(Chapter 6)**.

Many behavioral studies using the sentence-picture verification paradigm have found that various object properties (e.g. shape, orientation, visibility, etc.) are simulated during language comprehension. Interestingly, research on the presence of color in mental simulations has led to mixed findings (Connell, 2007; Zwaan & Pecher, 2012). One of the aims of **Chapter 2** was to resolve this issue. This was accomplished by conceptually replicating the original study on color, using the same paradigm but a different stimulus set. The second goal of this study was to assess how much perceptual information is included in a mental simulation. If reduction of one aspect of color does not alter the match effect, it would suggest that not all perceptual information is relevant for a mental simulation. Our results did not support this: We found a match advantage when objects were shown at normal levels of saturation, but this match advantage disappeared when saturation was reduced. Taken

together, these results clearly show a strong match effect for color, and the perceptual richness of mental simulations during language comprehension.

Over the past several decades many studies have examined how situation models are updated, yet this had never been examined at the level of the mental simulation. **Chapter 3** examined how mental simulations are updated when there is an implied change of shape, over the course of several sentences. Given that comprehenders can reactivate perceptual information in a mental simulation if this is required, it is possible that this reactivation would be necessary in order to understand that changes are occurring to an object in a narrative. Another possibility is that, during the updating process, the contents of the mental simulations are replaced by the new incoming information. Our study found support for the second possibility, as participants responded significantly faster when the shape matched the one implied by the final sentence, suggesting that the initial shape becomes deactivated in a mental simulation as the more recent shape is activated.

Chapter 4 presents a review of the prominent theories on situation model updating, the evidence for preservation of this function in an aging population, and how this compares to young adults. The review reports several studies that show that older adults are equally capable of updating their situation models compared to young adults, even if they take longer to do so. Interestingly, older adults also may create a more elaborate situation model compared to young adults, as their situation model appears to be resistant to contradictory information. As such, younger and older adults appear to have similar capacities for updating their situation models to represent the current events unfolding in a narrative.

Studies on the presence of mental simulations during language comprehension have typically focused only on single object properties. **Chapter 5** investigates whether two objects are combined in mental simulations, and whether this is influenced by task instructions. In two experiments, participants read sentences describing animals using a tool in some way. After each sentence, they saw an image of a cartoon animal holding a tool, and they indicated whether the animal or the tool was mentioned in the previous sentence or not. If comprehenders create a complete mental simulation of an event to support language comprehension, then both the animal and the tool should be activated in a mental simulation, regardless of task instructions The results suggest that mental simulations indeed combine multiple objects during language comprehension, and that this is not influenced by task instructions, suggesting that language comprehension leads to the creation of a complete mental simulation.

Chapter 6 examined the continued activation of color in a broader discourse context. Participants read sentences that either contained a reference to a color, or no reference to a color, before responding to pictures that were shown in a color matching the one described by the previous sentence, or shown in grayscale. The current study examines the continued activation of color in mental simulations across one, two, and five sentences. The results illustrated that color was activated in mental simulations when the final sentence contained a reference to color. When the target object (e.g. bicycle) was mentioned in all sentences, color remained activated in the mental simulation, even when only the first sentence made a reference to a color. When the focus of the story was shifted elsewhere and the target object was not present across all sentences, color was no longer activated in the mental simulation. These findings suggest that color remains active in mental simulations so long as the target object is present in every sentence. As soon as the focus of the story shifts to another event, this perceptual information is deactivated in the mental simulation. As such, there is no continued activation of color across a broader discourse context.

Taken together, the chapters from this dissertation provide further support for the theories of embodied (or grounded) cognition. When perceptual information is required to understand the events described in a text, this information is activated in a mental simulation.

R

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1061-1070. doi:10.1037/0278-7393.19.5.1061

Anderson, J. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6*(4), 451–474. doi:10.1016/0010-0285(74)90021-8.

Aslan, A., Bäuml, K. H., & Pastötter, B. (2007). No inhibitory deficit in older adults' episodic memory. *Psychological Science, 18*(1), 72-78. doi:10.1111/j.1467-9280.2007.01851.x

Bailey, H. R., & Zacks, J. M. (2015). Situation model updating in young and older adults: Global versus incremental mechanisms. *Psychology and Aging, 30*(2), 232-244. doi:10.1037/a0039081

Barrós-Loscertales, A., González, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J., & Costumero, V., … Ávila, C. (2011). Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cerebral Cortex*, *22*(11), 2554-2563. doi:10.1093/cercor/bhr324

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(04), 577–660. doi: 10.1017/S0140525X99252144

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645. doi:10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). *Language and simulation in conceptual processing.* In M. De Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on*

R

meaning and cognition (pp.245-283). Oxford University Press. doi:10.1093/acprof:oso/9780199217274.001.0001

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536. doi:10.1016/j.tics.2011.10.001

Bramão, I., Reis, A., Petersson, K. M., & Faísca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica, 138*(1), 244–253. doi:10.1016/j.actpsy.2011.06.010

Brouillet, T., Heurley, L., Martin, S., & Brouillet, D. (2010). The embodied cognition theory and the motor component of “yes” and “no” verbal responses. *Acta Psychologica*, *134*(3), 310-317. doi:10.1016/j.actpsy.2010.03.003

Brunyé, T., Ditman, T., Mahoney, C., Walters, E., & Taylor, H. (2010). You heard it here first: Readers mentally simulate described sounds. *Acta Psychologica*, *135*(2), 209-215. doi:10.1016/j.actpsy.2010.06.008

Connell, L. (2007). Representing object colour in language comprehension. *Cognition, 102*(3), 476–485. doi:10.1016/j.cognition.2006.02.009

Connell, L., & Lynott, D. (2009). Is a bear white in the woods? Parallel representation of implied object color during language comprehension. *Psychonomic Bulletin & Review, 16*(3), 573–577. doi:10.3758/PBR.16.3.573

Connell, L., and Lynott, D. (2011). “Interpretation and representation: testing the Embodied Conceptual Combination (ECCo) theory,” in *European Perspectives on Cognitive Science*, eds B. Kokinov, A. Karmiloff-Smith, and N. J. Nersessian (Sofia: New Bulgarian University Press).

R

Available online at: https://www.lancaster.ac.uk/staff/connelll/papers/Connell-Lynott-2011-EuroCogSci.pdf

Cook, A. E., & O'Brien, E. J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes*, *51*(1-2), 26-49. doi:10.1080/0163853X.2013.855107

Copeland, D. E., & Radvansky, G. A. (2007). Aging and integrating spatial mental models. *Psychology and Aging, 22*(3), 569-579. doi:10.1037/0882-7974.22.3.569

Coppens, L. C., Gootjes, L., & Zwaan, R. A. (2012). Incidental picture exposure affects later reading: evidence from the N400. *Brain and Language*, *122*(1), 64-69. doi:10.1016/j.bandl.2012.04.006

De Beni, R., Borella, E., & Carretti, B. (2007). Reading comprehension in aging: The role of working memory and metacomprehension. *Aging, Neuropsychology, and Cognition, 14*, 189-212. doi:10.1080/13825580500229213

De Groot, A. D. (2014). The meaning of "significance" for different types of research [Translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han L. J. van der Maas]. *Acta Psychologica, 148,* 188–194. doi:10.1016/j.actpsy.2014.02.001 **(Original work published 1956)**.

De Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2017). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse Processes*, *54*(7), 493-503. doi:10.1080/0163853X.2015.1119604

R

Dijkstra, K., & Post, L. (2015). Mechanisms of embodiment. *Frontiers in Psychology, 6,* 1525. doi:10.3389/fpsyg.2015.01525

Dijkstra, K., Yaxley, R. H., Madden, C. J., & Zwaan, R. A. (2004). The role of age and perceptual symbols in language comprehension. *Psychology and Aging, 19*(2), 352. doi:10.1037/0882-7974.19.2.352

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2008). Time travel through language: Temporal shifts rapidly decrease information accessibility during reading. *Psychonomic Bulletin & Review*, *15*(4), 750-756. doi:10.3758/PBR.15.4.750

Dove, G. (2016). Three symbol ungrounding problems: Abstract concepts and the future of embodied cognition. *Psychonomic Bulletin & Review*, *23*(4), 1109-1121. doi:10.3758/s13423-015-0825-4

Engelen, J. A., Bouwmeester, S., de Bruin, A. B., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology, 110*(4), 659–675. doi:10.1016/j.jecp.2011.06.009

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Gentilucci, M., Benuzzi, F., Bertolani, L., Daprati, E., & Gangitano, M. (2000). Language and motor control. *Experimental Brain Research, 133*(4), 468-490. doi:10.1007/s002210000431

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes, 23*(3), 265-304. doi:10.1080/01638539709544994

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 430-445.

R

Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4301443/

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(1), 1-19.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558-565. doi:10.3758/BF03196313

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, *32*(2), 906-912. doi:10.1016/j.neuroimage.2006.03.037

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*(1), 163-189. doi: 10.1146/annurev.psych.48.1.163

Harris, I. M., & Dux, P. E. (2005). Orientation-invariant object recognition: Evidence from repetition blindness. *Cognition, 95*(1), 73–93. doi:10.1016/j.cognition.2004.02.006

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301-307. doi:10.1016/s0896-6273(03)00838-9

Hauk, O., Shtyrov, Y., & Pulvermüller, F. (2008). The time course of action and action-word comprehension in the human brain as revealed by neurophysiology. *Journal of Physiology-Paris*, *102*(1-3), 50-58. doi:10.1016/j.jphysparis.2008.03.013

Hindy, N. C., Solomon, S. H., Altmann, G. T., & Thompson-Schill, S. L. (2013). A cortical network for the encoding of object change. *Cerebral Cortex*, *25*(4), 884-894. doi:10.1093/cercor/bht275

Hoeben Mannaert, L. N., Dijkstra, K., & Zwaan, R. A. (2017). Is color an integral part of a rich mental simulation? *Memory & Cognition*, *45*(6), 1-9. doi:10.3758/s13421-017-0708-1

Hoeben Mannaert, L. N., Dijkstra, K., & Zwaan, R. A. (2019). How are mental simulations updated across sentences? *Memory & Cognition*, 1-14. doi:10.3758/s13421-019-00928-2

Holmes, K. J., & Wolff, P. (2011). Simulating realism in language comprehension. *Proceedings 33rd Annual Conference of the Cognitive Science Society, 1,* 2884–2889.

James, K.H., & Maouene, J. (2009). Auditory verb perception recruits motor systems in the developing brain: an fMRI investigation. *Developmental Science*, *12*(6), F26-F34. doi:10.1111/j.1467-7687.2009.00919.x

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics, 38*(7), 1033-1050. doi:10.1016/j.pragma.2005.09.012

Kemper, S., Crow, A., & Kemtes, K. (2004). Eye-fixation patterns of high-and low-span young and older adults: Down the garden path and back again. *Psychology and Aging, 19*(1), 157-170. doi:10.1037/0882-7974.19.1.157

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805-825. doi:10.1016/j.cortex.2011.04.006

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163-182. doi:10.1037//0033-295x.95.2.163

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*(2), 72-79. doi:10.1016/j.tics.2007.11.004

Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & Cognition, 40*(5), 812-826. doi:10.3758/s13421-011-0179-8

Lahiri, J. (2019). *The namesake.* Boston, Massachusetts: Houghton Mifflin Harcourt

Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*(8), 1764-1801. doi:10.1111/cogs.12174

Levine, W. H., & Klin, C. M. (2001). Tracking of spatial information in narratives. *Memory & Cognition*, *29*(2), 327-335. doi:10.3758/BF03194927

Long, S. A., Winograd, P. N., & Bridge, C. A. (1989). The effects of reader and text characteristics on imagery reported during and after reading. *Reading Research Quarterly*, 353-372. doi:10.2307 / 747774

Louwerse, M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35*(2), 381-398. doi:10.1111/j.1551-6709.2010.01157.x

Lupyan, G., & Bergen, B. (2015). How language programs the mind. *Topics in Cognitive Science, 8,* 408–424. doi:10.1111/tops.12155

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196-14201. doi:10.1073/pnas.1303312110

Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology*, *1*, 1-14. doi:10.3389/fpsyg.2010.00212

Madden, C. J., & Dijkstra, K. (2009). Contextual constraints in situation model construction: An investigation of age and reading span. *Aging, Neuropsychology, and Cognition, 17*(1), 19-34. doi:10.1080/13825580902927604

Madden-Lombardi, C., Dominey, P. F., & Ventre-Dominey, J. (2017). Grammatical verb aspect and event roles in sentence processing. *PloS ONE*, *12*(12), 1-13. doi:10.1371/journal.pone.0189919

Magliano, J., Kopp, K., McNerney, M. W., Radvansky, G. A., & Zacks, J. M. (2012). Aging and perceived event structure as a function of modality. *Aging, Neuropsychology, and Cognition, 19*(1-2), 264-282. doi:10.1080/13825585.2011.633159

Magliano, J.P. & Schleich, M.C. (2000) Verb aspect and situation models. *Discourse Processes, 29*(2), 83-112. doi:10.1207/S15326950dp2902_1

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, *102*(1), 59-70. doi:10.1016/j.jphysparis.2008.03.004

McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(4), 919-932. doi:10.1037/0278-7393.22.4.919

R

McNerney, M. W., Goodwin, K. A., & Radvansky, G. A. (2011). A novel study: A situation model analysis of reading times. *Discourse Processes, 48*(7), 453-474. doi:10.1080/0163853X.2011.582348

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*, *26*(2), 165-187. doi:10.1016/0749-596X(87)90122-7

Morrow, D. G., Leirer, V.O., Altieri, P., & Fitzsimmons, C. (1994). Age differences in creating spatial models from narratives. *Language and Cognitive Processes, 9*(2), 203-220. doi:10.1080/01690969408402116

Morrow, D. G., Stine-Morrow, E. A., Leirer, V. O., Andrassy, J. M., & Kahn, J. (1997). The role of reader age and focus of attention in creating situation models from narratives. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 52*(2), P73-P80. doi:10.1093/geronb/52B.2.P73

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*(2-3), 131-157. doi:10.1080/01638539809545042

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098-1111. doi:10.1162/jocn.2006.18.7.1098

Noh, S. R., & Stine-Morrow, E. A. (2009). Age differences in tracking characters during narrative comprehension. *Memory & Cognition, 37*(6), 769-778. doi:10.3758/MC.37.6.769

Norman, S., Kemper, S., & Kynette, D. (1992). Adults' reading comprehension: effects of syntactic complexity and working memory. *Journal of Gerontology, 47*(4), 258-265.

R

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry, 23*(3), 217–243. doi:10.1080/1047840X.2012.692215

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*(6), 615–631. doi:10.1177/1745691612459058

O'Brien, E. J., Cook, A. E., & Guéraud, S. (2010). Accessibility of outdated information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(4), 979-991. doi:10.1037/a0019763

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 499-508. doi:10.1037/xhp0000313

Palmer, S. D. (1999). *Vision science: Photons to phenomenology*. Cambridge: MIT Press.

Payne, B. R., & Stine-Morrow, E. A. L. (2014). Adult age differences in wrap-up during sentence comprehension: Evidence from ex-Gaussian distributional analyses of reading time. *Psychology & Aging, 29,* (2), 213-228. doi:10.1037/a0036282

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1108-1114. doi:10.1080/17470210802633255

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful

R

analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 785. doi:10.1037/0278-7393.21.3.785

Pulvermüller, F., Hauk, O., Nikulin, V., & Ilmoniemi, R. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793-797. doi:10.1111/j.1460-9568.2005.03900.x

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(20), 7865-7870. doi:10.1073/pnas.0509989103

Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT press.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Raaijmakers, J. G., Schrijnemakers, J. M., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, *41*(3), 416-426. doi:10.1006/jmla.1999.2650

Radvansky, G. A. (1999). Aging, memory, and comprehension. *Current Directions in Psychological Science, 8*(2), 49-53. doi:10.1111/1467-8721.00012

Radvansky, G. A. (2005). Situation models, propositions, and the fan effect. *Psychonomic Bulletin & Review, 12*(3), 478-483. doi: 10.3758/BF03193791

Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science, 21*(4), 269-272. doi:10.1177/0963721412451274

R

Radvansky, G. A., & Copeland, D. E. (2001). Working memory and situation model updating. *Memory & Cognition, 29*(8), 1073-1080. doi: 10.3758/BF03206375

Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *The American Journal of Psychology, 117*(2), 191-213. doi:10.2307/4149022

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition, 34*(5), 1150-1156. doi:10.3758/BF03193261

Radvansky, G. A., & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 210-216. doi:10.1037/a0017258

Radvansky, G. A., Copeland, D. E., Berish, D. E., & Dijkstra, K. (2003). Aging and situation model updating. *Aging, Neuropsychology, and Cognition, 10*, 158-166.

Radvansky, G. A., & Curiel, J. M. (1998). Narrative comprehension and aging: The fate of completed goal information. *Psychology & Aging, 13*(1), 69-79.

Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic Bulletin & Review, 14*(6), 1027-1042. doi:10.3758/BF03193088

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *The Quarterly Journal of Experimental Psychology, 64*(8), 1632-1645. doi:10.1080/17470218.2011.571267

Radvansky, G.A., O'Rear, A.E. & Fisher, J.S. (2017). Event models and the fan effect. *Memory & Cognition, 45*(6), 1028-1044. doi:10.3758/s13421-017-0713-4

Radvansky, G. A., Pettijohn, K. A., & Kim, J. (2015). Walking through doorways causes forgetting: Younger and older adults. *Psychology and Aging, 30*(2), 259-265. doi:10.1037/a0039259

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 608-620. doi:10.1002/wcs.133

Radvansky, G. A., Zacks, R. T., & Hasher, L. (1996). Fact retrieval in younger and older adults: The role of mental models. *Psychology and Aging, 11*(2), 258. doi:10.1037/0882-7974.11.2.258

Radvansky, G. A., Zacks, R. T., & Hasher, L. (2005). Age and inhibition: The retrieval of situation models. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60*(5), P276-P278. doi:10.1093/geronb/60.5.P276

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging, 16*(1), 145-160. doi: 10.1037/0882-7974.16.1.145

Richter, T., & Singer, M. (2017). *Discourse updating: Acquiring and revising knowledge through discourse.* In M.F. Schober, D.N. Rapp, & M.A. Britt (Eds.), *The Routledge Handbook of Discourse Processes* (2nd ed., pp. 167-190). New York, NY: Routledge. doi:10.4324/9781315687384-11

Sadoski, M. (1983). An exploratory study of the relationships between reported imagery and the comprehension and recall of a story. *Reading Research Quarterly, 19*(1), 110-123. doi:10.2307/747341

R

Sadoski, M. (1985). The natural use of imagery in story comprehension and recall: Replication and extension. *Reading Research Quarterly, 20*(5), 658-667. doi:10.2307/747949

Sadoski, M., Goetz, E. T., & Kangiser, S. (1988). Imagination in story response: Relationships between imagery, affect, and structural importance. *Reading Research Quarterly*, 320-336. doi:10.2307/748045

Sakreida, K., Scorolli, C., Menz, M. M., Heim, S., Borghi, A. M., & Binkofski, F. (2013). Are abstract action words embodied? An fMRI investigation at the interface between language and motor cognition. *Frontiers in Human Neuroscience*, *7*, 125. doi:10.3389/fnhum.2013.00125

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*(5), 763-776.

Santos, A., Chaigneau, S., Simmons, W., & Barsalou, L. (2011). Property generation reflects word association and situated simulation. *Language and Cognition*, *3*(1), 83-119. doi:10.1515/langcog.2011.004

Sato, M., Schafer, A. J., & Bergen, B. K. (2013). One word at a time: mental representations of object shape change incrementally during sentence processing. *Language and Cognition*, *5*, 345373. doi:10.1515/langcog-2013-0022

Simmons, W., Hamann, S., Harenski, C., Hu, X., & Barsalou, L. (2008). FMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology-Paris*, *102*(1-3), 106-119. doi:10.1016/j.jphysparis.2008.03.014

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802-2810. doi:10.1016/j.neuropsychologia.2007.05.002

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, *32*(2), 244-259. doi:10.3758/BF03196856

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language, 53*(1), 125-140. doi:10.1016/j.jml.2005.02.009

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*(1), 125-140. doi:10.1016/j.jml.2005.02.009

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153-156. doi:10.1111/1467-9280.00326

Stine-Morrow, E. A. L., & Hindman, J. (1994). Age differences in reading time allocation for propositionally dense sentences. *Aging, Neuropsychology, and Cognition, 1*(1), 2-16.

Stine-Morrow, E. A. L., Miller, L. M. S., & Hertzog, C. (2006). Aging and self-regulated language processing. *Psychological Bulletin, 132*(4), 582-606. doi:10.1037/0033-2909.132.4.582

Stine-Morrow, E. A. L., Morrow, D. G., & Leno, R. (2002). Aging and the representation of spatial situations in narrative understanding. *Journals of Gerontology, 57B*(4), P291-P297

Stine-Morrow, E.A.L., Loveless, M., & Soederberg, L. (1996). Resource allocation in on-line reading by younger and older adults. *Psychology and Aging, 11*(3), 475–486.

Sundermeier, B. A., Van der Broek, P., & Zwaan, R. A. (2005). Causal coherence and the availability of locations and objects during narrative comprehension. *Memory & Cognition*, *33*(3), 462-470. doi:10.3758/BF03193063

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236. doi:10.1037/a0015631

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences, 5,* 211–215. doi:10.1016/S1364-6613(00)01626-0

Tolkien, J. R. R. (2005). *The lord of the rings: 50th anniversary edition.* London: Harper Collins Publishers

Van Dam, W., Rueschemeyer, S., & Bekkering, H. (2010). How specifically are action verbs represented in the neural motor system: An fMRI study. *NeuroImage*, *53*(4), 1318-1325. doi:10.1016/j.neuroimage.2010.06.071

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York, NY: Academic Press.

R

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48. doi:10.18637/jss.v036.i03

Wilson, S., Saygin, A., Sereno, M., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701-702. doi:10.1038/nn1263

Yaxley, R. H., & Zwaan, R. A. (2006). Simulating visibility during language comprehension. *Cognition, 105*(1), 229–236. doi:10.1016/j.cognition.2006.09.003

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 293-357). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General, 138*(2), 307-327. doi:10.1037/a0015305

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin, 133*(2), 273-293. doi:10.1037/0033-2909.133.2.273

Zeng, T., Zheng, L., & Mo, L. (2016). Shape representation of word was automatically activated in the encoding phase. *PLoS ONE, 11*(10), e0165534. doi:10.1371/journal.pone.0165534

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1196-1207. doi:10.1037%2F0278-7393.22.5.1196

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review, 23*(4), 1028-1034. doi:10.3758/s13423-015-0864-x

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*(5), 292-297. doi:10.1111/j.1467-9280.1995.tb00513.x

Zwaan, R. A., & Madden, C.J. (2004). Updating situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 283-288. doi:10.1037/0278-7393.30.1.283

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science, 28*(4), 611–619. doi:10.1016/j.cogsci.2004.03.004

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *Plos One, 7*(12), e51382. doi:10.1371/journal.pone.0051382

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162-185. doi:10.1037/0033-2909.123.2.162

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading, 2*(3), 199-220. doi:10.1207/s1532799xssr0203_2

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168-171. doi:10.1111/1467-9280.00430

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology, 135*(1), 1–11. doi: 10.1037/0096-3445.135.1.1

R

S

Nederlandse Samenvatting

(Summary in Dutch)

Er wordt al eeuwenlang gediscussieerd over de onderliggende mechanismen van taalbegrip. Theorieën over *grounded cognition* zijn op het moment een van de meest populaire theorieën over taalbegrip. Deze theorieën suggereren dat we dezelfde sensomotorische gebieden in de hersenen gebruiken tijdens activiteiten als tijdens cognitieve processen, door het gebruik van mentale simulaties. Er zijn de afgelopen twee eeuwen veel gedrags- en neurobiologische onderzoeken uitgevoerd om aan te tonen dat de inhoud van deze simulaties modaal van aard is en dat deze simulaties betrokken zijn bij taalbegrip. Het doel van dit proefschrift is om de mechanismen van mentale simulaties in taalbegrip verder te onderzoeken. Daartoe zijn de volgende onderwerpen onderzocht: de rol van kleur in mentale simulaties (**hoofdstuk 2**), het updateproces in mentale simulaties (**hoofdstuk 3**), het gebruik van mentale simulaties in taalbegrip bij jongere versus oudere volwassenen (**hoofdstuk 4**), het vermogen om meerdere objecten te combineren in mentale simulaties (**hoofdstuk 5**) en de voortdurende activering van mentale simulaties (**hoofdstuk 6**).

Veel gedragsstudies hebben gebruik gemaakt van het *sentence-picture verification* paradigma om aan te tonen dat verschillende objecteigenschappen (bijv. vorm, oriëntatie, zichtbaarheid, etc.) worden gesimuleerd tijdens taalbegrip. Interessant genoeg is dat onderzoek naar de aanwezigheid van kleur in mentale simulaties heeft geleid tot gemengde bevindingen (Connell, 2007; Zwaan & Pecher, 2012). Eén van de doelstellingen van **hoofdstuk 2** was om duidelijkheid in deze situatie te brengen. Dit werd bereikt door de oorspronkelijke studie over kleur conceptueel te repliceren, met behulp van hetzelfde paradigma, maar met een andere stimulusset. Het tweede doel van deze studie was om te beoordelen hoeveel perceptuele informatie is opgenomen in een mentale simulatie. Als vermindering van één aspect van kleur het *match-effect* niet verandert, zou dit suggereren dat niet alle perceptuele informatie relevant is voor een mentale simulatie. Het *match effect* ontstaat wanneer proefpersonen sneller reageren op een plaatje dat overeenkomt met de zin (bijv. reageren op een plaatje van een groen stoplicht wanneer de een groen stoplicht impiceert) dan wanneer dit plaatje niet overeenkomt (bijv. reageren op een plaatje van een rood stoplicht wanneer de zin een groen stoplicht impliceert). Onze resultaten ondersteunden dit niet: we vonden een *match effect* wanneer objecten werden getoond op normale verzadigingsniveau's, maar dit effect verdween toen de verzadiging werd verminderd. Al met al laten deze resultaten duidelijk een sterk *match effect* zien voor kleur en ondersteunen het idee van de perceptuele rijkdom van mentale simulaties tijdens taalbegrip.

In de afgelopen decennia hebben veel studies onderzocht hoe situatiemodellen worden bijgewerkt, maar dit was nog nooit gedaan op het niveau van de mentale simulatie. **Hoofdstuk 3** onderzocht hoe mentale simulaties worden bijgewerkt wanneer er een impliciete verandering van vorm is, in een serie zinnen. Aangezien het mogelijk is om perceptuele informatie in een mentale simulatie te heractiveren als dit nodig is, is het plausibel dat deze heractivering nodig is om te begrijpen dat er veranderingen optreden bij een voorwerp in een verhaal. Een andere mogelijkheid is dat tijdens het updateproces de inhoud van de mentale simulaties wordt vervangen door de nieuwe informatie die binnenkomt. Onze studie vond ondersteuning voor de tweede mogelijkheid, omdat deelnemers sneller reageerden toen de vorm overeenkwam met de vorm die werd gesuggereerd in de laatste zin. Dit suggereert dat de oorspronkelijke vorm in een mentale simulatie wordt gedeactiveerd naarmate de meer recente vorm van het object wordt geactiveerd.

Hoofdstuk 4 presenteert een overzicht van de prominente theorieën over het bijwerken van het situatiemodel en het bewijs voor het behoud van deze functie bij een vergrijzende bevolking en hoe dit zich verhoudt tot jongvolwassenen. Het literatuuronderzoek rapporteert over verschillende studies die aantonen dat oudere volwassenen even goed in staat zijn om hun situatie modellen bij te werken in vergelijking met jongvolwassenen, ondanks het feit dat ze er langer over doen. Interessant is dat oudere volwassenen ook een uitgebreider situatiemodel kunnen maken in vergelijking met jongvolwassenen, omdat hun situatiemodel resistent lijkt te zijn tegen tegenstrijdige informatie. Jongere en oudere volwassenen lijken dus vergelijkbare capaciteiten te hebben om hun situatiemodellen bij te werken om de huidige gebeurtenissen in een verhaal weer te geven.

Onderzoek naar de aanwezigheid van mentale simulaties tijdens het begrijpen van taal richtte zich meestal alleen op eigenschappen van afzonderlijke objecten. **Hoofdstuk 5** onderzoekt of twee objecten kunnen worden gecombineerd in mentale simulaties en of dit wordt beïnvloed door taakinstructies. In twee experimenten lezen deelnemers zinnen waarin dieren worden beschreven die gereedschap gebruiken voor een doel. Na elke zin zagen de deelnemers een afbeelding van een cartoondier met gereedschap en ze gaven aan of het dier (of het gereedschap) in de vorige zin werd genoemd of niet. Als een volledige mentale simulatie van een gebeurtenis wordt gemaakt om taalbegrip te ondersteunen, dan moeten zowel het dier als het gereedschap worden geactiveerd in een mentale simulatie, ongeacht wat de taakinstructie is. Inderdaad suggereren de resultaten dat mentale simulaties meerdere objecten combineren tijdens taalbegrip, en dat dit niet wordt beïnvloed door taakinstructies. Dit impliceert dat taalbegrip leidt tot het creëren van een volledige mentale simulatie.

Hoofdstuk 6 onderzoekt de voortdurende activering van kleur in een bredere zinscontext. Deelnemers van de studie lazen zinnen die ofwel een verwijzing naar een kleur bevatten, ofwel geen verwijzing bevatten. Daarna reageerden ze op foto's die of werden getoond in een kleur die overeen kwam met de vorige zin, of werden getoond in grijswaarden. De huidige studie onderzocht de voortdurende activering van kleur in mentale simulaties in één, twee en vijf zinnen. De resultaten toonden aan dat kleur werd geactiveerd in mentale simulaties wanneer de laatste zin een verwijzing naar kleur bevatte. Wanneer het doelobject (bijv. fiets) in alle zinnen werd genoemd, bleef kleur geactiveerd in de mentale simulatie, zelfs wanneer alleen de eerste zin naar de kleur van het voorwerp verwees. Als de focus van het verhaal naar een ander gebeurtenis verschoof en het doelobject niet meer in alle zinnen aanwezig was, dan werd kleur niet meer geactiveerd in de mentale simulatie. Deze bevindingen suggereren dat kleur actief blijft in mentale simulaties zolang het doelobject in elke zin aanwezig is. Zodra de focus van het verhaal verschuift naar een andere gebeurtenis, wordt deze perceptuele informatie gedeactiveerd in de mentale simulatie. Er kan dus niet gesproken worden van een voortdurende activering van kleur in mentale simulaties in een bredere zinscontext.

Samengevat bieden de hoofdstukken uit dit proefschrift verdere ondersteuning voor de theorieën van embodied (of grounded) cognition. Wanneer perceptuele informatie vereist is om de in een tekst beschreven gebeurtenissen te begrijpen, wordt deze informatie geactiveerd in een mentale simulatie.

C

Curriculum Vitae

Curriculum Vitae

Lara Natasha Hoeben Mannaert was born in Auchenflower, Australia, on the 28th of August in 1992. In 2010 she obtained her International Baccalaureate Diploma from the International School of The Hague. In 2011 she started her Bachelor of Science (Honors) in Psychology at the University of Sheffield in England, and obtained her diploma in 2014. That same year she started a Psychology Master's degree at the Erasmus University Rotterdam (EUR) in The Netherlands, specializing in Brain and Cognition. In 2015 she graduated *cum laude,* and started working as a PhD student at the Department of Psychology, Education and Child Studies at the EUR. Whilst working as a PhD student, Lara presented her work at various national and international conferences, earned the award of Best Presentation at the Graduate Research Day in 2017, completed various courses, organized monthly meetings where several PhD students and professors came together to discuss recently published works in the field of language comprehension, worked as an academic teacher for first year Psychology students, and supervised third year Bachelor students in writing their thesis. After her work at the EUR, Lara started working as a Data Analyst in 2019 for the municipality of Capelle aan den IJssel, as part of her traineeship Business Analytics and Data Science with trainee.nl.

Publications

Hoeben Mannaert, L.N., Dijkstra, K., & Zwaan, R.A. (2017). Is color an integral part of a rich mental simulation? *Memory & Cognition, 45*(6), 974-982.

Hoeben Mannaert, L.N. & Dijkstra, K. (2019). Situation model updating in young and older adults. *The International Journal of Behavioral Development*, 1-8.

Hoeben Mannaert, L.N., Dijkstra, K., & Zwaan, R.A. (2019). How are mental simulations updated across sentences? *Memory & Cognition, 47*(6), 1201-1214.

Papers

Hoeben Mannaert, L.N., Dijkstra, K., & Zwaan, R.A. (submitted). Object combination in mental simulations.

Hoeben Mannaert, L.N., Dijkstra, K., & Zwaan, R.A. (submitted). Is color continuously activated in mental simulations across a broader discourse context?

Presentations

November 2017	Oral presentation on the updating of mental simulations: Graduate Research Day, Erasmus Universiteit Rotterdam. *Winner Best Presentation*
November 2016	Poster presentation on the updating of mental simulations: Psychonomic Society's 57^{th} Annual Meeting, Boston, United States.
July 2016	Oral presentation on the updating of mental simulations: Society for Text and Discourse Conference, Kassel, Germany
June 2016	Oral Presentation on the updating of mental simulations: Language and Perception International Conference, University of Trondheim, Norway.

C

A

Acknowledgements

"And there was Frodo, pale and worn, and yet himself again; and in his eyes there was peace now, neither strain of will, nor madness, nor any fear. His burden was taken away." (J.R.R. Tolkien, *The Lord of the Rings*). It is done! A task that at times felt like I myself was climbing Mount Doom has been accomplished, after four years of hard work. Similar to Frodo, who also had a Fellowship to support him, there are many members of my own fellowship that helped and supported me on this academic journey that I would like to thank.

First and foremost, I would like to thank my promotor. Rolf, when I stepped into your office five years ago and announced that I wanted to do a PhD in psycholinguistics and asked you what I had to do in order to accomplish that, I did not think that I would be lucky enough to be able to do it under your supervision. Thank you for the chance you gave me and the trust you placed in me to complete this project. Meetings with you were always interesting and entertaining, as we often ended up taking longer than planned because the discussions we had about the research ended up being about other things! Thank you for being my mentor. You helped me evaluate other studies critically, and embedded a strong mentality of ethical research practices in me. Without your guidance, this work would never have been possible.

Secondly, I would like to thank my copromotor. Katinka, thank you for your support and guidance these past years. Your kindness as a person made it very easy to step into your office and ask you questions. You always helped assuage any insecurities I had, and there were many moments where I needed that! Thanks to the way you gave me feedback on my work, I decided to incorporate your methods when giving feedback to my own students, by giving a lot of positive feedback first! I very much enjoyed the lunches we had together at the Paviljoen, and I enjoyed hearing your own thoughts on academia.

Aside from having spent much of my time behind a desk staring at a computer, much of my time was also spent at the ESB lab testing participants. Gerrit Jan and Christiaan, thank you so much for the support you gave me during this time and for answering the (what felt like) hundreds of questions I had. You (literally) gave me the tools I needed for the job!

The African proverb "it takes a village to raise a child" seems very appropriate to my life as a PhD student, as work on the 16th floor often felt like living in a PhD village where everyone was raising a PhD child. Thanks so much for the friendship and support from all the other PhD students on this floor. We had a lot of laughs together these past years! I'd like to give special thanks to the following (in no particular order): Julia, Ilse, Gloria, Lara, Rob, Donna, Işil, Loïs, Willemijn, Keri, Sabrina, Anniek, Jacqueline, Sander, and Tessa. Julia, thank you for being such a kind and caring person, your support has been sincerely appreciated the past years! Rob, thank you for always talking to me when I would get a coffee, even when I was in a bad mood you always managed to make me laugh! Sander and Tessa, thanks for geeking out with me and talking about games. Whenever I wanted to talk about something other than academia, you two provided some much-needed distraction! Lara II, thanks for all the laughs we had together. Being my namesake sometimes led to confusion but more often it led to lots of laughter!

As they say, there is no "I" in "team". As such, I would like to also sincerely thank the other PhD students from the Brain and Cognition section. Yiyun, Jianan, and Arnout: thank you for being great sparring partners. Thanks to your insights my research improved greatly!

Having the privilege to be part of the LanPercept project allowed me to meet a lot of great people. Harmen, thank you for all the conversations and discussions we had about PhD life. Many times while reading your messages I laughed out loud. Sobh, thank you for your guidance and the many laughs we had together. Staying at your place in Trondheim for two weeks was fantastic, and it was wonderful to meet your family. The way you would always smile and nod at me while I was doing a presentation helped me so much when I was nervous. This is something I also ended up doing for my students when they were presenting!

Most of my PhD time was spent in a room with four others: Marieke, Denise, Iris and Milou. In the beginning I was very intimidated by the prospect of starting a PhD track, but these ladies were so sweet and welcoming that I soon felt right at home there. They asked me hundreds of questions to get to know me, and it did not take long before I had been given the nickname ChocoLara, because of all the chocolate I kept in my drawer and ate continuously throughout the day. I cannot emphasize enough how much their friendship and support has meant to me these past four years. The many talks about dogs, television shows, movies, babies, and of course about our experiences during the PhD were often hilarious and sometimes tearful, but always heartwarming. Thank you girls for making my work feel like home.

My paranymphs, Denise and Iris, thank you for all the love and support you have given me and continue to give me. Denise, you are the kind of person that is always easy to talk to, and you are one of the kindest people I know, never stop being your sweet and bubbly self! One of the many things we had in common over the years is that we both moved to The Hague around the same time, we got dog(s), and we had a shared liking of the same television shows. All these things taken together meant that we never lacked material to talk about! I will also never forget the fun conversations we had about Stranger Things, all the shared pictures of our dogs, and the many creative projects you were always working on. Iris, you have been a great friend the past four years. Already from the beginning I had great respect and admiration for you as you were just a fountain of knowledge, you seemed to know everything! And if there was something you didn't know, it was not long before you found out. Whenever I needed someone to critically discuss my ideas and research, you were always the person I would come to. Aside from work, I loved to have our weekly Game of Throne recaps, or our hilarious discussions on topics ranging from the best type of chocolate to which yoga pose helps with a stiff neck to how to build furniture! And lastly, thank you for introducing me to the sport of bouldering! Leo and I have become complete fanatics thanks to you!

To my friends, Isabelle, Philipp, and Lavinia, thank you for being there for me in the past four years. The many times we spent together either watching movies, eating, bowling, bouldering, talking (and laughing!) always allowed for some much-needed relaxation during my off hours. Whenever things weren't going well or I needed a friend to talk to about my struggles you were there for me.

Finally, I want to mention the people to whom I am the most grateful: my family. To Soffy and Rodrigo: thank you for all the fun conversations, the movies, the games, the hugs, the food, and the time we spent together. During the times when I needed help, you were always the first to reach out. To Danny and Claudia: thank you for all the fun we've had talking about our shared experiences with our sometimes naughty but always hilarious dogs. Your support in my life has been invaluable. To Leonardo and Arelys: thank you for being such wonderful in-laws. Whenever I needed to talk you two were there, whenever I needed a hug you would provide one, and whenever I grew frustrated about something you listened and gave advice, helping me to put everything into context. Thank you for being my second parents. To Bas, my dear older brother: without having had someone in my life to constantly look up to I don't know where I would have ended up. Your continued

understanding and kindness has been a great support to me when the going got tough.

Dear mom and dad, I don't know where to begin when it comes to thanking you. Mom, whenever I came to you about anything you would always know what to say and list about a hundred solutions to the problem, so that I would always know what to do. And if there was no immediate answer, then the solution was simple: come to France for a bit and relax! Your diligence and care in raising me helped shape me into the person I am today. Without you helping me with my homework in high school I'm pretty sure I would never have gotten into the University of Sheffield, and would certainly have never managed to accomplish something as huge as a PhD! Dad, you may just be the smartest person I know, and you have been my role model my entire life. Having completed a PhD yourself, you recognized many of the struggles I faced these past few years, and therefore always knew what I needed to hear at those moments. Your constant unwavering assurance that I was capable of doing and completing a PhD has been incredible. And like mom, whenever things became too much, the advice was simple: just come to France, something I gladly did every year. Without your advice and guidance this PhD would have remained an impossibility.

And lastly, I want to thank the most important person in my life. Leito, what a crazy life we've had already! I was a mere fifteen years old when our adventure started together, and now, 12 years later that adventure still continues on. During the time of my PhD we moved houses twice, got two dogs, and went on epic holidays to Japan twice! You are the sweetest, kindest, funniest, most supportive and strongest person I know, and you continue to inspire me every day. I can't imagine my life without you, Akira, and Kami, and luckily, I don't have to. You make me want to do better and thanks to your example I continue to strive to be a better person. Let's hope the rest of

our lives are equally as filled with fun, joy, and love as the last 12 years have been.

A

STELLINGEN

behorende bij het proefschrift

Language Comprehension:

Examining the Dynamic Changes in Activation of Mental Simulations

Lara N. Hoeben Mannaert

1. One of the solutions to the replication crisis in psychology is for researchers to preregister. *(this dissertation)*

2. Mental simulations change the activation levels of the object traces during the process of updating. *(this dissertation)*

3. Younger and older adults appear to demonstrate similar capacities for updating and integrating information in the situation model, albeit that older adults do this more slowly. *(this dissertation)*

4. Comprehenders combine multiple objects in their mental simulations during language comprehension. *(this dissertation)*

5. A complete situation model, containing all related information, is not created during discourse processing. Only information which is required for language comprehension needs to be activated in the situation model. *(this dissertation)*

6. Bilinguals have two personalities.

7. The 'Lara phenomenon' shows that leading roles in video games are more and more frequently portrayed by female characters.

8. Companion animals improve mood and social interactions and reduce loneliness and problem behaviors in Alzheimer's patients.

9. Imagining a positive encounter with an outgroup member reduces prejudice.

10. Being sleep-deprived works great as a human repellant.

11. "Those who dare to fail miserably can achieve greatly." (J. F. Kennedy)