A Computational Theory of Visuo-Spatial Mental Imagery

Jan Frederik Sima

Dissertation
zur Erlangung des Grades eines Doktors der Naturwissenschaften

— Dr. rer. nat. —

Vorgelegt im Fachbereich 3 (Mathematik & Informatik)
Universität Bremen
Januar 2014

Dedicated to my parents

Acknowledgements

I want to sincerely thank

- Julia-Eva and Maja Mirabella.
- My first supervisor Christian Freksa for his valuable feedback, for sharing his ideas which inspired this work, and for thought-provoking discussions that strongly impacted my view on cognition. His feedback has greatly improved this thesis.
- My second supervisor Anna Borghi for being interested in my work and for taking the time to review it.
- The project R1 - Thomas Barkowsky, Sven Bertel, Maren Lindner, Ana-Maria Olteteanu, Holger Schultheis, and Rasmus Wienemann - for general support and inspiring conversations as well as whiskey and BBQ. Especially Holger Schultheis has spend a lot of time answering many of my questions.
- The CoSy working group for offering a great, liberal, and friendly work environment. Especially (in alphabetical order) Thomas Barkowsky, Sandra Budde, Lutz Frommberger, Julia Gantenberg, Gracia Kranz, Maren Lindner, Denise Peters, and Falko Schmid.

Abstract

The thesis develops a new theory of visuo-spatial mental imagery. The theory is concretized in a formal framework and implemented as a computational model. The theory and its model are evaluated against a set of empirical phenomena and compared to the contemporary theories of mental imagery. The new theory is shown to provide explanations for the considered phenomena that partly go beyond those of the contemporary theories.

The thesis is motivated by two main observations.

First, the observation that the lack of formalization of the current psychological and philosophical theories of mental imagery limits the progress of the imagery debate, i.e., the question about the nature of mental imagery. A formalized theory is able to provide more detailed explanations and predictions for the empirical data which can facilitate further empirical studies. Furthermore, sufficiently formalized theories become comparable with objective measures thus making similarities and differences between theories more transparent.

Second, some of the contemporary theories of mental imagery stress the involvement of rich mental representations in cognition and mental imagery. This approach has been considered problematic with respect to more recent results such as the functionality of eye movements during mental imagery as well as the neuropsychological findings on unilateral neglect. The enactive theory poses an exception and stresses the importance of sensorimotor interactions for mental imagery.

The new theory shares assumptions with the enactive theory with respect to direct and active vision and the relationship between vision and imagery. It combines this view with grounded mental concepts which function as hubs to low-level perceptional actions. The theory understands the process of mental imagery in the context of internal simulations of sensorimotor interactions. Mental images are based on grounded concepts whose semantics are made explicit by the overt and covert employment of the low-level perceptual actions they link to. This employment of perceptual actions makes low-level perceptual information available which represents an instance of the conceptually described mental image. Critically, this perceptual information is not made available by an activation of early visual areas but by mechanisms of proprioception and anticipation.

Zusammenfassung

Die vorliegende Dissertation entwickelt eine neue Theorie von räumlich-visueller mentaler Vorstellung. Diese Theorie wird formalisiert und als Computermodell implementiert. Die Theorie und das Modell werden anhand einer Menge von empirischen Phänomenen evaluiert und mit den anderen Theorien von mentaler Vorstellung verglichen. Es wird gezeigt, dass die neue Theorie Erklärungen für die Phänomene bietet, welche zum Teil über die Erklärungen der anderen Theorien hinausgehen. Die Arbeit ist durch zwei wesentliche Beobachtungen motiviert.

Dies ist erstens die Tatsache, dass die aktuellen psychologischen und philosophischen Theorien nicht formal beschrieben sind. Diese Tatsache limitiert den Fortschritt der sogenannten "imagery" Debatte. Diese Debatte dreht sich um die Frage, wie menschliche Kognition mentale Vorstellung realisiert. Eine formale Theorie ist in der Lage empirische Daten detaillierter zu erklären und Vorhersagen zu machen. Dies kann weitere empirische Untersuchungen theoretisch motivieren. Weiterhin sind formale Theorien objektiv vergleichbar, so dass Ähnlichkeiten und Unterschiede zwischen den Theorien transparenter werden. Dadurch wird der wissenschaftliche Fortschritt gefördert.

Zweitens stellen die meisten aktuellen Theorien die Rolle von mentalen Repräsentationen für die Realisierung von Kognition und mentaler Vorstellung in den Vordergrund. Die Erklärungmöglichkeiten dieses Ansatzes wurden vor allem hinsichtlich neuerer Ergebnisse kritisch bewertet, z.B. die Funktionalität von Augenbewegungen während mentaler Vorstellung sowie Ergebnisse aus der Neuropsychologie zu Aufmerksamkeitsstörungen. Die "enactive" Theorie von mentaler Vorstellung stellt hierzu eine Ausnahme dar, weil sie die Rolle von sensomotorischer Interaktion hervorhebt.

Die neue Theorie baut auf einigen der Annahmen der "enactive" Theorie hinsichtlich aktiver und direkter Wahrnehmung und dem Verhältnis von Wahrnehmung und Vorstellung auf. Die Theorie kombiniert dies mit geerdeten Symbolen (engl. grounded symbols) . Diese Symbole sind Assoziationen mit bestimmten Aktionen der visuellen Wahrnehmung, z.B. Augenbewegungen. Die Theorie sieht mentale Vorstellungen vor dem Hintergrund interner Simulationen von sensomotorischen Interaktionen. Mentale Bilder basieren auf abstrakten Symbolen. Die Semantik dieser Symbole ergibt sich

durch simulierte und tatsächliche Ausführung von Aktionen der visuellen Wahrnehmung. Diese Ausführung generiert eine konkrete perzeptuelle Instanz des, durch die Symbole konzeptuell beschriebenen, mentalen Bildes. Diese perzeptuelle Instanz wird nicht durch die Aktivierung von Arealen des visuellen Kortex generiert sondern durch Propriozeption und Antizipation.

Contents

Chapter 1

Introduction

1.1 Motivation – What is Mental Imagery?

What is mental imagery?

Mental imagery is one of those things that are easy to explain to a person, but incredibly hard to scientifically grasp. How many windows does your apartment or house have? Take your time to actually answer this question.

People usually report to solve this task by imagining themselves going from room to room adding all the windows together. If you also did it this way, then you just used mental imagery. Hearing a song play only in your head, imagining how to find your way from A to B, imagining what something or someone looks like, feels like, or tastes like; all that is also considered mental imagery.

How our capability to imagine such things can be understood or how it is realized has been a topic of philosophical and scientific discussion starting at least as early as ancient greek philosophy[1]. After the decline of behaviorism during which mental imagery naturally received little to no attention, it came back with a great impact on cognitive psychology with the first surprising experimental results on mental rotation (Shepard & Metzler, 1971) and later mental scanning (Kosslyn, 1973). Figure 1.1 depicts a set of stimuli from the mental rotation experiment. The task is to decide whether the left shape is the same as the right one or whether it is a mirrored version of it. The results showed that the response times are linearly proportional to the angle of rotation between the two shapes. That is, the finding is consistent with the assumption that one actually mentally rotates the shape to see if it fits. Such an interpretation suggests that mental images might have an uncanny structural similarity to the entities they represent, i.e., in this case that the mental representation of the figure is mentally rotated just like one would rotate an actual object. At the time of this study, these

[1] For a comprehensive overview on the history of the scientific and philosophical debate on mental imagery up until today, see (Thomas, 2013)

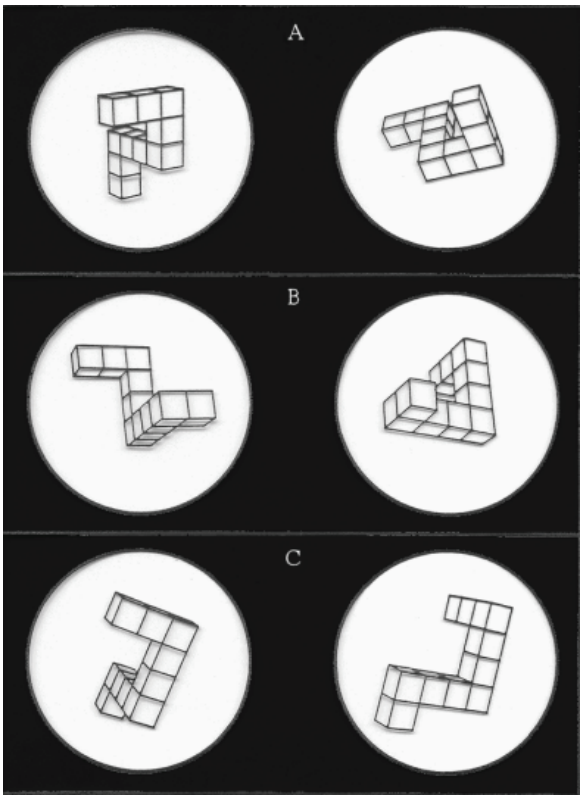


Figure 1.1: Mental rotation stimuli from (Shepard & Metzler, 1971).

results were surprising and seemed to challenge contemporary assumptions about cognition.

The prevalent view of cognition at that time is often referred to as computationalism or cognitivism (e.g., Fodor, 1983). That is, cognition is understood as information processing based on mental representations. Mental representations would specifically mean abstract and amodal symbols of entities in the real world which are used to build an internal model of the real world. The sensorimotor system had little to no relevance to the symbolic computation of cognition other than being input (perception) and output (action) to the central cognition module. If cognition is computation in this sense, then there seems to be no obvious reason why the angle of rotation should have an impact on computation time.

Results like this one inspired and motivated the pictorial theory of mental imagery (Kosslyn, 1980). Slightly simplified, the pictorial theory poses that mental imagery employs a specific mental representation in which the mental image is represented depictively. This mental depiction is located in the visual cortex in an area which during visual perception presumably holds the content of what one is seeing. During mental rotation this depic-

tion would then literally be rotated in order to solve the task. This new theory was opposed by an alternative theory – called the descriptive theory (Pylyshyn, 1973). The descriptive theory poses that no such specific mental representation is necessary, but that mental imagery just like all other cognitive processing can be explained with abstract and amodal symbols organized in propositional descriptions. The descriptive theory, however, faced the problem that mental rotation, and likewise mental scanning, do produce reaction time patterns seemingly inconsistent with the mere processing of abstract symbols. That is, the processing of abstract symbolic representations of the rotation stimuli should not show a dependance on the actual angle. This problem was tackled with the proposal that participants subconsciously emulate these reaction times using their knowledge about, for example, how long rotation around a certain angle usually takes. This is referred to as the tacit knowledge explanation.

The dispute between the proponents of these two theories became known as the imagery debate (Tye, 1991). The imagery debate was considered one of the hot topics in cognitive science and it has generated countless publications, studies, and empirical data up to this day. About 40 years since the onset of the debate, both sides have not changed their theoretical position much (Kosslyn, Thompson, & Ganis, 2006; Pylyshyn, 2007). But a new third position was established with the enactive theory of mental imagery (Thomas, 1999).

The enactive theory incorporates ideas of a paradigm shift in cognitive science. The aforementioned paradigm of computationalism has been followed-up by the paradigm of embodied cognition over the last years. At the core of embodied cognition is the assumption that the sensorimotor system, i.e., the processes of perception and action, constitute much more than just the input and output for internal mental representations which then realize the actual cognitive processing but rather are deeply involved in cognition themselves. The enactive theory differs critically from both the pictorial and the descriptive theory as it rejects the idea that mental images are realized through a mental representation which corresponds to the mental image. Instead, the experience of mentally "seeing" an image is proposed to result from a re-enactment of visual perception. That is, one goes through the motions of seeing something in order to mentally imagine "seeing" it. Critically, this includes the claim that these re-enacted processes are generally not directed at an internal mental representation but at the external world.

Given these three quite different theories of mental imagery, how can we decide which one describes the phenomenon of mental imagery most accurately? This is the fundamental question of the imagery debate. In the following, this question, its inherent problems, and a possible way of facilitating further progress of the imagery debate are discussed.

1.2 Problem and Method – State of the Imagery Debate

Relatively early in the imagery debate, it was argued that the problem of deciding whether mental imagery is realized by a depictive or a descriptive mental representation cannot be decided with behavioral data (Anderson, 1978; Palmer, 1978). The argument is based on the fact that different mental representations can always be made to fit arbitrary behavioral data equally well if the processes working on the respective representation are adjusted accordingly. Since the respective processes working on either a depictive or a descriptive mental representation during mental imagery are under-specified in both the pictorial and the descriptive theory, we cannot ultimately decide what type of mental representation better fits empirical data.

Generally, different theories can be ranked and evaluated by other measures than their ability to in principle explain empirical data. These measures include the efficiency of the proposed mechanisms, their plausibility, and how parsimonious the theory is (Anderson, 1978; Pylyshyn, 1979). Yet, none of these measures can be concretely applied to the imagery debate and the contemporary theories today. The reason is that all three theories are presented on a descriptive and often vague level. For example, the core conceptions of the three theories, i.e., the depictive representation of the pictorial theory, the descriptions or tacit knowledge of the descriptive theory, or the mechanisms behind the re-enactments of the enactive theory, are not formally defined. Instead, their concrete nature remains under-specified and in consequence the explanations and the predictions of the theories are necessarily often subject to the same under-specification.

In order to make the theories of the imagery debate comparable, they need to be formulated as explicitly and as formally as possible. The most thorough formalization is the implementation of theories as computational models. The computational implementation of psychological theories has several advantages which can facilitate progress of the research question at hand (e.g., Sun, 2009). For one, an implementation is essentially a detailed and formalized theory in itself (Sun, 2009). As such it is far less susceptible to ambiguity and misinterpretation – a problem that the imagery debate

currently displays[2]. Additional advantages of implemented theories include the ability to run simulations and to provide concrete explanations as well as concrete predictions for the empirical phenomena. The more concrete explanations and predictions are, the more directly can they motivate and facilitate new empirical studies. Resulting new and specifically inconsistent empirical results can in turn be integrated into the theory in a transparent manner by refinements and adjustments to the implementation. For under-specified theories, in contrast, descriptive ad-hoc extensions are often utilized to allow the explanation of specific (perhaps otherwise inconsistent) empirical data. This bears the danger that the consequences of such extensions for the overall framework of the theory and for specific explanations of other phenomena remain untested and unclear. Lastly, sufficiently formalized theories can be compared to each other with concrete measures of plausibility and efficiency that are otherwise not applicable. That is, beyond the fact that two theories are both generally able to account for the empirical data, two formalized theories can be compared and ranked according to time complexity (i.e., how complicated are the necessary calculations), space complexity (i.e., how much storage is necessary for the calculations to work), and ultimately Occam's razor (i.e., how parsimonious and simple is the (implementation of the) theory).

1.3 Aims and Theses

The aims of this thesis are

- the development of a theory of mental imagery which is able to provide explanations and predictions for a diverse set of empirical phenomena of mental imagery,
- the development of a formal framework of that theory which allows concrete implementations, and
- the development of a computational model based on the framework.

[2]Mutual misunderstandings are a prevalent problem of the imagery debate. For example, Kosslyn claims that the enactive theory would essentially be a form of the pictorial theory if it would be fleshed-out sufficiently (Kosslyn et al., 2006, p. 92) while Thomas (1999), in contrast, clearly states fundamental incompatibilities between the two theories. Another example are the diverging opinions on the concept of a functional space in which mental images are claimed to be represented in the pictorial theory. A discussion between Pylyshyn (2002) and Kosslyn, Thompson, and Ganis (2002) shows that the interpretations of such a functional space go so far apart, that Pylyshyn (2002, p. 218) even states that the assumption of a functional space is either incorrect (and a literal space is actually meant) or that it would follow that the pictorial theory does not differ from his descriptive theory. These examples underscore the current inability to fully understand, compare, or evaluate the contemporary theories as a result of their ambiguous description, i.e., their lack of formalization.

The thesis is based on the assumption that a more formal theory of mental imagery is able to facilitate the imagery debate by providing more detailed explanations and predictions for empirical phenomena of mental imagery than the contemporary theories currently do.

1.4 Structure of the Thesis

Chapter 2 "Phenomena and Theories of Visuo-Spatial Mental Imagery" summarizes important empirical results of visuo-spatial mental imagery and presents the three main contemporary theories of visuo-spatial mental imagery. Furthermore, the explanations and potential problems of the three theories with respect to the discussed phenomena are reviewed.

Chapter 3 "The Perceptual Instantiation Theory" presents and explains a new theory of visuo-spatial mental imagery. The chapter discusses visual perception and how mental imagery builds upon the mechanisms of visual perception.

Chapter 4 "A Formal Framework of PIT" summarizes the core commitments of the perceptual instantiation theory (PIT) and presents a formal framework of it. Lastly, it compares PIT to the three contemporary theories.

Chapter 5 "The Computational Model" presents a computational implementation of PIT based on the formal framework developed in the previous chapter.

Chapter 6 "Evaluation" presents the evaluation of the presented theory and the computational model. The explanations and predictions for those empirical phenomena discussed in Chapter 2 are elaborated.

Chapter 7 "Conclusion and Outlook" discusses the contributions of the thesis and provides an outlook on future work.

Chapter 2

Phenomena and Theories of Visuo-Spatial Mental Imagery

This chapter summarizes important empirical results on visuo-spatial mental imagery. Visuo-spatial mental imagery is the imagination of information that is usually conveyed via visual perception, i.e., visual and spatial information. This thesis is concerned with visuo-spatial mental imagery in contrast to mental imagery of haptics, acoustics, etc. The chapter furthermore reviews the three major contemporary theories of mental imagery and discusses their explanatory power with respect to the empirical results.

2.1 Empirical Results of Mental Imagery

There is a vast amount of empirical data on visuo-spatial mental imagery in the literature. This chapter can only report on a subset of these studies. This subset of phenomena has been selected considering the following factors:

- phenomena that are well established, relatively well researched, and have been reproduced;
- phenomena for which the contemporary theories differ in their explanation or lack a satisfactory explanation;
- phenomena which cover different aspects of mental imagery.

The considered areas of empirical data cover the general findings that visuo-spatial mental imagery shows similarities to visual perception (e.g., mental scanning), yet, also shows striking differences to visual perception (e.g., mental reinterpretation). The apparent embodied nature of mental imagery (e.g., eye movements) is considered as well as the complex role of

Figure 2.1: Island stimulus for mental scanning used in (Kosslyn, Ball, & Reiser, 1978). The island contains different locations that differ in their distance to each other. In the lower left corner a hut, a well, a lake, and a tree are visible. On the top is a rock and further locations include grass and a beach.

attentional processes in both mental imagery and visual perception (e.g., unilateral neglect).

2.1.1 Mental Scanning

In studies on mental scanning participants learn a stimulus, for example, the map of the island in Figure 2.1, which they later mentally imagine. Using their mental image, participants are asked to shift their attention from one entity in the image to another entity. It turned out that participants take significantly longer for attention shift between, for example, the hut and the rock, than they do for a shift between the hut and the well. This is called the mental scanning effect. The mental scanning effect is a strong linear correlation between the time it takes to scan between two entities in the mental image and the distance between these two entities in the original stimulus. Several studies have reproduced and shown the robustness of this effect (for an overview, see Denis & Kosslyn, 1999). In particular, the effect was shown to persist the following variations:

- Whether participants were explicitly instructed to use mental imagery (e.g., Kosslyn et al., 1978) or not (e.g., Finke & Pinker, 1982, 1983;

Pinker, Choate, & Finke, 1984);

- Presentation of additional distance information inconsistent with a stimulus (Richman, Mitchell, & Reznick, 1979), e.g., indicating that some routes on a map have a certain distance while the actual distance in the stimulus is inconsistent with that information;

- Variation of the experimenters' expectancy of the experimental results, i.e., the experimenters have a certain (partially false) belief about, a) the time it generally takes to mentally scan mental images (Intons-Peterson, 1983; Jolicoeur & Kosslyn, 1985), b) how scanning time depends on the to-be-scanned distance (Jolicoeur & Kosslyn, 1985), and c) how scanning time depends on the type of stimulus (Jolicoeur & Kosslyn, 1985);

- Variation of the participants' expectancy of the experimental results about a) the time mental scanning takes in general (Goldston, Hinrichs, & Richman, 1985), b) how scanning time depends on the to-be-scanned distance (Goldston et al., 1985);

- Whether participants are instructed to imagine movement (e.g., Kosslyn et al., 1978; Richman et al., 1979; Jolicoeur & Kosslyn, 1985), i.e., participants are instructed to imagine a little black speck moving between entities, or not (e.g., Finke & Pinker, 1982, 1983);

- Whether the mental image is generated based on a previously presented visual stimulus (e.g., Kosslyn et al., 1978), generated from information in long-term memory (Pinker et al., 1984), or generated from verbal descriptions (e.g., Denis & Cocude, 1992);

- Variation of the salience of the entities used in the mental image (Denis & Cocude, 1997).

All the above studies report a linear correlation between scanning time and distance. However, intercept and slope of the linear function describing the relationship between time and distance have been shown to vary significantly based on certain variations, e.g., instructions of the task, the expectations of the experimenters (Intons-Peterson, 1983), as well as the belief of the participants (e.g., Richman et al., 1979). In particular, it has been shown that the slope of the function, i.e., the rate or pace of scanning, can be significantly altered by the participant's belief about the time-distance relationship for mental scanning (Goldston et al., 1985). Furthermore, pseudo-experiments[1] have shown that participants generally expect a linear relationship between time and distance for mental scanning

[1] Pseudo-experiments are experiments in which participants are verbally described an experimental setting and asked how they think they would behave as participants of that experiment.

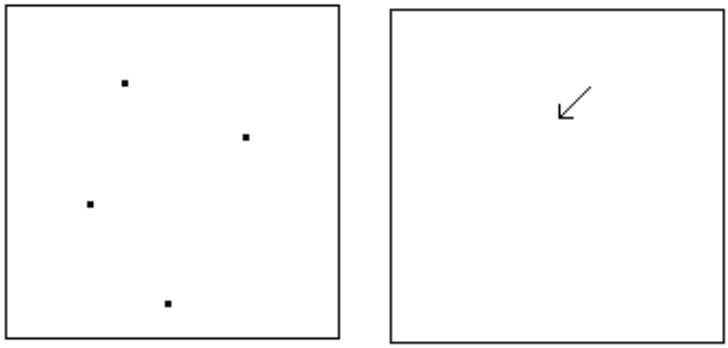


Figure 2.2: Mental scanning stimuli of (Finke, Pinker, & Farah, 1989). Participants were first presented with a dot pattern as displayed on the left side. The dot pattern was removed and an arrow was presented. The participants were to judge whether that arrow would point at one of the previously visible dots or not.

experiments (e.g., Richman et al., 1979; Mitchell & Richman, 1980). This allows the possibility that participants are not actually employing a "genuine" mental scanning process but emulating reaction times subconsciously (as it has been suggested in, e.g., Pylyshyn, 1981). However, Reed, Hock, and Lockhead (1983) report a comparison between participants' estimations of scanning time and the time actually taken to mentally scan the same stimulus. It was found that participants' estimations differ from the actual scanning time for at least some stimuli.

Considering the influence of the above discussed factors on the mental scanning process, Denis and Kosslyn (1999) recognized the need to minimize any suggestive context in mental scanning experiments and argued for "no explicit imagery instructions" and that "if participants form and scan images, they [should] do so spontaneously" (p. 427). Finke and Pinker (1982) report a mental scanning study that realized such control conditions. Figure 2.2 shows the stimuli of their experiments. Participants were presented with a pattern of four black dots on a white background. The pattern was removed and an arrow appeared for 4s. Participants had to judge whether the arrow would point at one of the previously visible dots. The distance between the arrow and the location of the dots was varied. The instructions did not mention mental images nor (mental) scanning, yet the results showed the mental scanning effect. This study allowed to visually perceive the arrow during the decision making, which allows the possibility that participants make eye movements between the visible arrow and the previous location of the respective dot. This might be problematic as the observed reaction times could potentially be the result of eye movements instead of a mental scanning process. For this reason, Pinker et al. (1984) altered the scanning

task so that the arrow was not visually shown but its location and orientation were verbally described. The variation of reaction time with distance was again significant. This result is important as it gives strong support to the mental scanning effect being a result of an actual functional mental process and not a non-functional emulation due to demand characteristics.

Summarizing, the literature reports the following findings: 1) a robust mental scanning effect, and 2) different factors that influence the mental scanning effect.

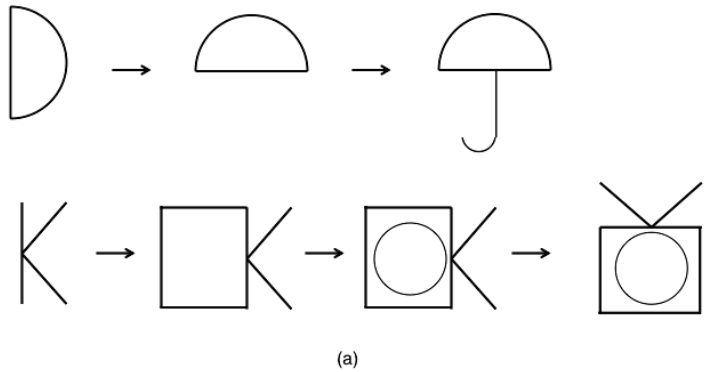
2.1.2 Mental Reinterpretation

Mental reinterpretation is the discovery of a second meaning of an ambiguous stimulus by only inspecting one's mental image of that stimulus. The studies on mental reinterpretation are particularly important for the question of the nature of mental images as they make the differences between literal pictures and mental images evident.

One can identify two general classes of stimuli in the literature on mental reinterpretation. Stimuli from these two classes apparently differ in the difficulty of their mental reinterpretation. In the following, the studies reporting on rather easily reinterpretable stimuli are summarized first before the studies using stimuli that are difficult to mentally reinterpret are summarized.

Finke et al. (1989) have shown that simple geometric shapes mostly resembling alphanumeric characters can be successfully transformed mentally, i.e., by rotation, superimposition, and juxtaposition, so that new geometric shapes emerge which can be mentally inspected and recognized. Figure 2.3 shows some of these stimuli and their transformation. The success rate for the reinterpretation, i.e., the recognition of the emerging new shape, ranged from roughly 50% to clearly above 50% for those participants able to correctly follow the transformations. In these experiments the starting stimuli were described verbally, which is in contrast to nearly all other mental reinterpretation studies which present the initial stimuli visually. Slezak (1995) reports an experiment with similar stimuli, i.e., mirrored numbers, and found that about 65% of the participants were able to reinterpret the stimuli mentally by identifying the number hidden in the shape. In this experiment the stimuli were presented visually. These stimuli are also shown in Figure 2.3.

Contrasting these comparably simple stimuli, there are studies using more complex ones, i.e., more complex in the sense that they do not consist of alphanumeric characters nor only simple geometric shapes. The by far most used stimuli are ambiguous drawings such as those shown in Figure 2.4. Among these ambiguous drawings the duck-rabbit is used in nearly all of the considered studies. Several experiments have shown that it is very hard to find the second meaning of the duck-rabbit only by inspection of the mental



(a)

(b)

Figure 2.3: Reproduction of stimuli used in (Finke et al., 1989) (a) and (Slezak, 1995) (b). For the upper stimuli, labeled (a), the first figure in each line is described to the participants verbally who then mentally transform their mental images according to verbal instructions so that the depicted intermediate figures should result. The final figure is to be interpreted as a new object. For the lower two stimuli, labeled (b), the respectively left one is briefly shown to the participants who then have to find an alternative interpretation of just the right side of the stimulus using their mental image. The alternative interpretations are the numbers depicted on the respectively right side.

image of it (e.g., Chambers & Reisberg, 1985; Peterson, Kihlstrom, Rose, & Glisky, 1992; Brandimonte & Gerbino, 1993). The success rate of its mental reinterpretation without any hints ranges from 0% to 5% in the different studies. Additionally, most studies also report that the large majority (if not all) of the participants were able to successfully find the second interpretation afterwards using their own drawing of the stimulus from memory. Slezak (1995) has conducted a series of reinterpretation experiments with a range of different types of other ambiguous stimuli. Examples of these stimuli are depicted in Figure 2.5. In this study no participant was able to find the second interpretation except for some participants for some of the rotated stimuli, which were, however, attributed to guessing.

It has been shown that different factors can significantly increase successful mental reinterpretation. These factors are: 1) explicit reference frame hints (e.g., Reisberg & Chambers, 1991), 2) proper selection of training

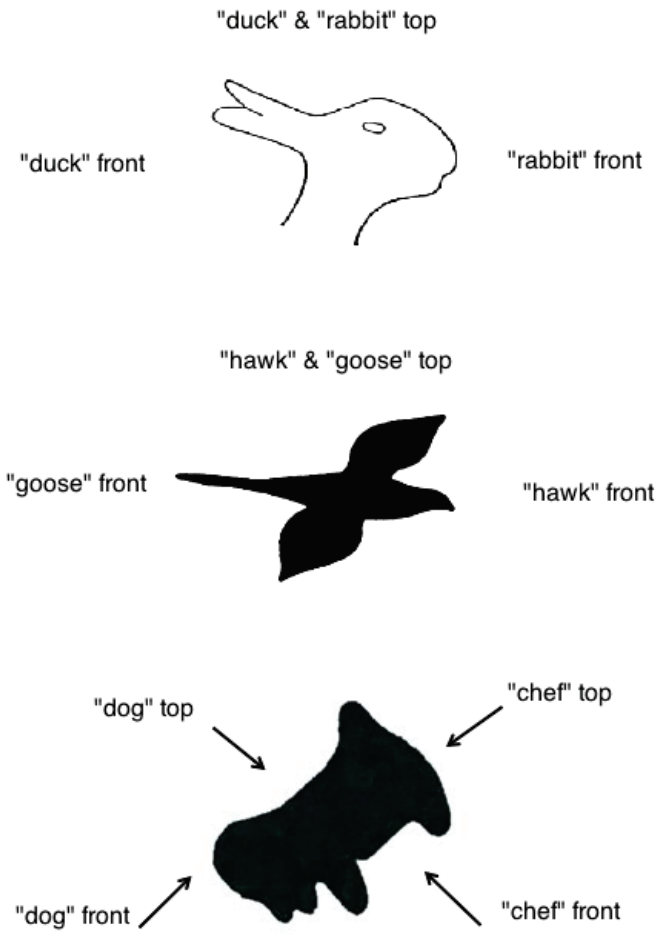


Figure 2.4: Reproduction of stimuli used by (Peterson et al., 1992). From top to bottom, the stimuli show a duck-rabbit, a goose-hawk, and a chef-dog.

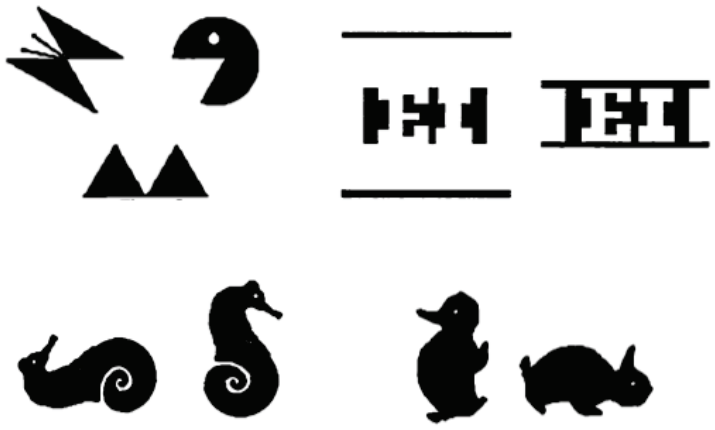


Figure 2.5: Stimuli used in (Slezak, 1995). The top left shows a Kanizsa illusion (Kanizsa, 1955) in which the emergent triangle is to-be-discovered mentally. The top right shows a stimulus in which the bars are to be mentally attached to the middle in order to then find the emergent letters as depicted to the right. The lower two stimuli show ambiguous drawings which depict different animals depending on the orientation. These stimuli were mentally rotated in order to discover the alternative meaning.

stimuli (e.g., Peterson et al., 1992), 3) partitioning of the stimulus (e.g., Peterson et al., 1992), and 4) articulatory suppression during the presentation of the stimulus (e.g., Brandimonte & Gerbino, 1993). In the following, these factors are elaborated.

Hyman and Neisser (1991) gave participants information about the orientation and category of the to-be-discovered second meaning of the duck-rabbit during the inspection of the mental image. These conditions significantly increased the participants' ability to successfully reinterpret the mental image of the duck-rabbit. Brandimonte and Gerbino (1993) replicated these experiments and report success rates between 5% and 20% when the different hints were provided. Reisberg and Chambers (1991) used similar explicit reference frame hints. They found that the mental reinterpretation of a seemingly arbitrary shape that was a rotated version of the shape of Texas is significantly influenced by whether participants are told to rotate and reinterpret their mental image or participants are told that the left side is to be considered the new top of the shape and then reinterpret their mental image. In the first case no participant successfully reinterpreted the stimulus while in the second case more than half of the participants were successful. The same effect was found for another rotated stimulus as well as for a figure-ground reversal stimulus (in this case participants either had no information or were told to reverse figure and ground).

Successful mental reinterpretation was also increased when participants were given implicit reference frame hints in the form of training examples of ambiguous stimuli which require the exact same reference frame transformation as the later presented experimental stimulus. Peterson et al. (1992) found that using the goose-hawk (see Figure 2.4) as a training example resulted in significantly higher success of the mental reinterpretation of the duck-rabbit than other ambiguous training stimuli that require different reference frame transformations, e.g., the chef-dog (see Figure 2.4).

Peterson et al. (1992), furthermore, tested the effect of different partitioning of ambiguous stimuli during the presentation. They partitioned stimuli into what they called "good" and "bad" parts. The parts were shown to participants one after the other and participants had to mentally "glue" them together. After they discovered a first interpretation of the stimulus, they were asked to find the second one. A partition of the initial shape was defined as "good" when the parts were "cut" at the minima of curvature. This method aimed at making the parts more familiar in the sense that the resulting shapes seem more natural than those of a "bad" partition. Without giving any reference frame hints, it was found that both "good" and "bad" partitioning improved mental reinterpretation compared to the normal presentation of the stimulus. Furthermore, "good" parts led to a significant increase in successful mental reinterpretation compared to "bad" parts.

Brandimonte and Gerbino (1993) used articulatory suppression during the presentation of the stimuli. This is achieved by participants saying "lalala" aloud during the initial presentation of the stimuli, e.g., the duck-rabbit. This suppression consistently led to a significant increase in successful mental reinterpretation.

Summarizing, the literature on mental reinterpretation reports: 1) a distinction between "easy" and "hard" stimuli with differing difficulty of mental reinterpretation, and 2) a set of different factors that can significantly increase successful mental reinterpretation.

2.1.3 Eye Movements

Several studies report the occurrence of spontaneous eye movements during mental imagery. These studies usually present participants with a stimulus which they later mentally imagine to describe or answer questions about. It is generally found that eye movements during such imagery tasks reflect the content of the mental image (e.g., Brandt & Stark, 1997; Spivey & Geng, 2001; Laeng & Teodorescu, 2002; Demarais & Cohen, 1998; Johansson, Holsanova, & Holmqvist, 2006; Johansson, Holsanova, Dewhurst, & Holmqvist, 2011).

In the experiments reported by Johansson et al. (2006) and Johansson, Holsanova, Dewhurst, and Holmqvist (2011), a distinction between local

and global correspondence of eye movements to the processed content of the mental image is defined. Global correspondence requires that an eye movement is not only directed towards the expected direction, e.g., to the left when processing the spatial relation *left of*, but also to a location consistent with the participant's gaze pattern over the whole experiment, i.e., the gaze is directed to the same location every time the same entity is referred to. Local correspondence requires the eye movement to only match the expected direction. Johansson et al. (2006) and Johansson, Holsanova, Dewhurst, and Holmqvist (2011) report experiments in which participants were either shown a complex and detailed picture or were presented with the verbal description of a complex and detailed scene. After this *perception phase* an *imagery phase* followed in which participants had to describe the picture/scene from memory while their eye movements were tracked. During this phase participants were facing a blank white screen. It was varied whether participants are allowed to freely move their eyes during the *perception phase* and during the *imagery phase*. For participants allowed to freely move their eyes during both phases, there is a significant local and global correspondence of their eye movements to the mental image. These results were reproduced in total darkness. The correspondence remained significant even when participants were forced to keep a fixed gaze during the *perception phase*. When participants had to keep a fixed gaze during the *imagery phase* after freely moving their eyes during the *perception phase*, it was found that recall is inhibited. Participants reported significantly less detail, objects, and locations compared to a control group. Furthermore, an analysis of the given verbal description showed that participants reported more abstract properties of the stimulus, e.g., global gestalt properties, whereas a control group reported more concrete details. These results provide evidence that eye movements during mental imagery are 1) functional for the recall of information from a mental image; 2) occur independently of the input modality of the stimulus; and 3) are not exact re-enactments of the eye movements of the visual perception of the stimulus.

Furthermore, it has been found that the spatial dispersion of eye movements during mental imagery depends on individual differences (Johansson, Holsanova, & Holmqvist, 2011). The spatial mental imagery score of the "Object-Spatial Imagery and Verbal Questionnaire" (Blazhenkova & Kozhevnikov, 2009) was found to be negatively correlated to the spatial dispersion of the gaze pattern produced during mental imagination of a complex scene. This spatial mental imagery score reflects a person's preference and ability to use spatial mental imagery compared to, for example, visual mental imagery or language-like thought. Concretely, the spatial distribution of the eye movements, that is, the area participants looked at during imagery shrinks with higher scores in the ability to use spatial mental imagery.

Summarizing, the literature reports 1) the robust occurrence of spontaneous eye movements during mental imagery; 2) that these eye movements

reflect the content of the mental image; 3) that forcing a fixed gaze affects mental imagery performance; and 4) that individual differences affect eye movements, in particular, their spatial dispersion.

2.1.4 Unilateral Neglect

Unilateral neglect is a neuropsychological condition defined by a deficit in attention or awareness of one side of space. Left unilateral neglect is much more common than right side unilateral neglect and can be the result of brain injury to the right cerebral hemisphere usually affecting the parietal lobe. Right-sided damage causes a neglect for the left side because information from the left side of the visual field is processed in the brain's right hemisphere. Patients suffering from visual neglect cannot properly attend to one side of the visual field or one side of objects during visual perception. For example, they might not eat food on the left side of their plate or produce drawings in which one side is missing or distorted as shown in Figure 2.6. Sometimes patients with visual neglect also show imaginal neglect, also referred to as representational neglect. Imaginal neglect is the inability to correctly attend to or process one side of one's mental images. Visual and imaginal neglect are highly complex neuropsychological conditions which are not properly understood theoretically or neurally. Due to the complexity of the topic and the limited scope of this thesis, only the core findings on unilateral neglect will be considered. The reason to include unilateral neglect is the fact that it poses critical constraints on the theories of mental imagery and is in general hard to concretely reconcile with all current theories.

Several studies have shown that patients with imaginal neglect fail to properly process or access the information on one side of their mental images. For example, patients with imaginal neglect show a great asymmetry when naming french towns based on an imagined map of France: they mention mostly towns on the non-neglected side (Rode, Rossetti, Perenin, & Boisson, 2004). This effect was not significant when they were asked to just name french towns without using a mental image. Bartolomeo, Bachoud-Levi, Azouvi, and Chokron (2005) report that imaginal neglect patients take longer to judge whether a french town is left or right of Paris when it is on the (neglected) left side. In other experiments, patients with imaginal neglect were impaired in their description of the left side of a familiar place, but were able to report formerly left-sided details once they imagined standing at the other side of that same place while then neglecting details of the former right side (Bisiach, Luzzatti, & Peranid, 1979; Bisiach, Capitani, Luzzatti, & Perani, 1981).

It is generally assumed that the symptoms of unilateral neglect are the result of several deficits with different severity playing together in the individual patient. Yet, the role of attention, specifically exogenous attention

Figure 2.6: Copies of drawings made by patients with left (visual) unilateral neglect (taken from (Thomas, 2013))

(i.e., attending to cues bottom-up), is accepted to be fundamentally involved in neglect (Bartolomeo & Chokron, 2002, 2001; Bourlon, Oliviero, Wattiez, Pouget, & Bartolomeo, 2010). For example, patients with visual neglect show abnormal eye, head, and hand movements, i.e., not attending the neglected side (Behrmann, Watt, Black, & Barton, 1997; Husain et al., 2001), despite their limbs and eyes being functionally normal. The critical role of attention is further supported by the fact that the effects of neglect can sometimes be alleviated by directly guiding a patient's attention towards the neglected side, for example, by presenting a very salient stimulus while the non-neglected side contains little to no cues (e.g., Bartolomeo, 2007).

It is an important fact, that there is a double dissociation between imaginal neglect and visual neglect (Coslett, 1997). This means, there are patients who display healthy vision while showing neglect in mental imagery and, similarly, there are patients that display healthy mental imagery but show visual neglect. This dissociation means that (partially) different processes must underlie the two types of neglect.

Summarizing, there are three major findings on unilateral neglect which are relevant for theories of mental imagery: 1) the impairment of accessing information on one side of a mental image in imaginal neglect; 2) the dissociation of visual and imaginal neglect; and 3) the apparent role of attentional processes in visual and imaginal neglect.

2.2 Theories of Mental Imagery

This section reviews the three main theories of visuo-spatial mental imagery: the pictorial theory (Kosslyn, 1994; Kosslyn et al., 2006), the descriptive theory (Pylyshyn, 2002, 2007), and the enactive theory (Thomas, 1999). There are other theories that deal with mental imagery which are not explicitly discussed here. The reason for omitting them is that these theories (currently) do not aim at explaining a broad range of phenomena of mental imagery but either focus on the explanation of specific results or on providing a general framework of human cognition and do not elaborate in depth on the framework's application to different mental imagery phenomena (e.g., perceptual symbol systems and the simulation theory of cognition, see Barsalou, 2008; Hesslow, 2012). The theory of visuo-spatial working memory of Logie (2003) is formulated more broadly as a theory of working memory and focusses on aspects often not directly addressed by the other theories of mental imagery, such as the relation between different components of working memory and the selective interference between them. The mental model theory is largely concerned with reasoning on mental models which are generally considered amodal with mental imagery being a specific case of a mental model (Johnson-Laird, 2001). Both the visuo-spatial working memory model and the mental model theory are addressed in Section 7.2.2 where the theory

presented in this thesis is related to them.

2.2.1 The Pictorial Theory

The pictorial theory has been developed and shaped mostly by Stephen Kosslyn. I will focus on the theory in its current form (Kosslyn, 1994; Kosslyn et al., 2006) which replaced a previous version (Kosslyn, 1980). The pictorial theory comprises several components that interact during mental imagery. Most components and processes are proposed to also be employed in visual perception, in particular, for object recognition.

The theory distinguishes between two types of mental images: spatial mental images and visual mental images. A spatial mental image is described as an object map that is held in the spatial-properties-processing subsystem (SPP) and generated from information from associative memory (AM). The object map comprises information such as location, size, and orientation of entities. It does not hold any visual information, e.g., color and shape. New spatial relations can be inferred by the SPP.

A visual mental image also relies on an object map in the SPP, but it further depicts visual information in the visual buffer (VB). Visual information, e.g., shape and color, are stored in an encoded form in the object-properties-processing subsystem (OPP). The information is decoded into a depictive pattern of activation that is evoked in the VB during visual mental imagery. The SPP determines properties such as size and location of the shape that the OPP maps into the VB. If a visual mental image contains multiple parts, i.e., shapes, this process is successively repeated for each part. The resulting mental image is not necessarily complete as parts might be missing due to fading or not being generated yet. Parts can be "refreshed" by mapping them from the OPP to the VB again.

The VB is described as a hybrid depictive representation. It is a depictive representation in the sense that space is used to represent space. That is, a shape is represented by an activation in the VB that resembles that shape. But, each "point [...] [in the VB] represents more than the presence or absence in space. Rather, properties such as color, intensity, depth, and motion are also specified at each location, using a symbolic (propositional) code" (Kosslyn et al., 2006, p. 136). It is stressed that the visual buffer and visual mental images are functionally depictive. This means that they do not have to be literally depictive in the sense of a picture in which two adjacent points are physically adjacent, but that these two points are accessed (in generation or inspection of the mental image) as two adjacent points even though they might be physically further apart in the neural substrate of the VB. For the rest of the thesis I will use the term "depictive" in this sense.

In order to generate and inspect visual mental images, a part of the VB is accessed by the attention window (AW). That is, only a part of the VB can be processed at one time. Inspection of visual mental images uses matching

with stored (encoded) shapes in the OPP to recognize the content of the VB. Accessing a given part of the mental image in the VB is either realized by scanning, i.e., successively moving the attention window, or parts can "pop out" without mentally scanning to the respective location in the VB. The theory assumes that the processing of mental images in the VB is the same as the processing of actual visual input, i.e., "[o]nce a configuration of activity exists in the visual buffer, input is sent to the ventral and dorsal systems and is processed in the usual ways – regardless of whether the activity arose from immediate input from the eyes or from information stored in memory" (Kosslyn, 1994, p. 336).

The core of this theory is the assumption of a depictive mental representation that holds the mental image so that it resembles what it represents. This resemblance is what conveys the meaning of the content of the mental image. The content of the mental image is accessed by the same inspection processes used in visual perception.

2.2.2 The Descriptive Theory

The descriptive theory, also known as the propositional theory, is proposed as a null hypothesis to the pictorial theory (Pylyshyn, 2002). Its essential claim is that there is not sufficient evidence and no need for a specific mental representation, i.e., a depictive representation, to account for mental imagery. Instead it is proposed that the empirical results of mental imagery can be explained by mental representations in the form of symbolic descriptions (Pylyshyn, 2002, p. 163), e.g., mentalese (Fodor, 1975), which are furthermore assumed to underlie (high-level) cognition in general. Those empirical results which indicate a spatio-analogical nature of mental images, e.g., mental rotation and mental scanning, could prima facie not be explained by description-like mental representations. Therefore, the employment of *tacit knowledge* was proposed in the context of the descriptive theory (Pylyshyn, 1981) to account for these results.

Pylyshyn (2002) stresses that in order to understand the mental mechanisms underlying mental imagery it is necessary to distinguish between observable behavior due to the intrinsic nature of mental imagery, i.e., the fixed mental representations and their processes, and behavior due to the tacit knowledge of a participant. Tacit knowledge refers to the mental state of a participant that can be directly or indirectly altered, e.g., a participant's goals or beliefs. The type of tacit knowledge that is mostly relevant to explain the common phenomena of mental imagery is a participant's tacit knowledge about what it would be like to visually perceive something. The intrinsic nature of mental imagery would, on the other hand, be defined by what is called the *cognitive architecture*. The cognitive architecture comprises the mental representations and processes which cannot be altered by a participant's tacit knowledge.

The descriptive theory claims that many of the observed properties of mental imagery, e.g., the mental scanning effect, are not a result of a spatio-analogical format of mental images, i.e., mental images being depictive, but that they result from the application of tacit knowledge of what it would be like to perceive what is to-be-imagined. To be clearer, the task to mentally imagine an entity X leads participants to simulate as many of the properties as possible of what visual perception of X would be like using their tacit knowledge about seeing X as well as psychophysical skills such as estimating the time it would take to see X (Pylyshyn, 1981). The underlying cognitive architecture of mental images could still have the form of symbolic descriptions which are not spatio-analogical. From this claim it follows that altering a participant's tacit knowledge will also alter the analogical properties of mental imagery. If the observed behavior for a given task can be altered by altering the participant's tacit knowledge, this task is said to be *cognitively penetrable*.

For some mental imagery tasks, other mechanisms than the application of tacit knowledge are proposed to explain the empirical data. For example, visual indexing is the process of subconsciously binding parts of a mental image to different locations in the visual field, e.g., a chair or a stain on the wall. When processing a part of the mental image eye movements are made to the respective location thus possibly creating the mental scanning effect and spontaneous eye movements.

Summarizing, the descriptive theory is to be understood not as a fleshed-out theory but as a null hypothesis to the pictorial theory. As such it serves the purpose of providing "[...] a test for the irrelevance of assumptions about the image format." (Pylyshyn, 2002, reply to comments, p. 227). More precisely, it provides an alternative explanation for the empirical data on mental imagery. And this explanation does not rely on a special format of the mental representation, i.e., one that is different from symbolic descriptions.

2.2.3 The Enactive Theory

The enactive theory, or perceptual-activity theory, of mental imagery contrasts the two traditional theories in some fundamental aspects. The enactive theory of mental imagery is described by Thomas (1999). It builds upon ideas of the ecological approach to vision (Gibson, 1986) and the work of Neisser (1976) on schemata.

The enactive theory generally applies to all modalities of perception and the respective type of mental imagery, but its focus has been on visual perception and visuo-spatial mental imagery. I identify four fundamental assumptions made by the enactive theory: 1) non-existence of explicit mental representations, 2) perception as an on-going process of active interrogation of the environment, 3) the existence of specialized perceptual instruments

used to retrieve specific information, and 4) schemata as subconscious data structures guiding the employment of the perceptual instruments. In the following, these assumptions are elaborated.

The enactive theory rejects the existence of explicit mental representations. An explicit mental representation is to be understood as a mental state which directly corresponds and thus stands in for an entity. Examples include specifically depictive and descriptive mental representations of the content of mental images. In the enactive theory there is "no thing or state in the mind or brain [that] corresponds to the percept or image" (Thomas, 1999, p. 223). From the perspective of the traditional theories of mental imagery explicit representations, image-like or description-like, are created as (end) products of perception that comprise information of what is seen. These representations are processed/inspected when one imagines the entity they represent. In the enactive theory such end products of perception are never created, instead perception is an on-going process.

This directly leads to the second assumption of the enactive theory: the understanding of perception as an active and on-going process and not as an input mechanism for mental representations. For example, the enactive theory states that the experience of visually perceiving a cat is constituted by the employment of those perceptual processes that "fit" the stimulus of a cat. In particular, visually perceiving a cat is not creating or activating a certain mental representation that symbolizes a cat, but perceiving a cat corresponds to the activity of successfully applying the respective perceptual processes.

The third assumption states that these perceptual processes are made up by specific employments of different perceptual instruments. The perceptual instruments of visual perception contain, for example, different types of eye movements, head movements, and also, in principle, querying neural states. These instruments are actively employed in order to retrieve specific types of information[2]. The recognition of a cat can thus be imagined as an interrogation of the environment about the necessary properties of a cat. This process is highly dynamic as the choice of the next interrogation step directly depends on the feedback of the previous interrogation steps.

The fourth assumption is that this interrogation process is controlled by data structures termed *schemata*. Schemata can be imagined as acquired procedural knowledge of how and when to use which perceptual instrument given the current feedback of the perceptual instruments. Concretely, a *schema* is defined as "a data structure, implemented in the brain, that functions to govern perceptual exploration of the world so that appropriate perceptual tests are applied at appropriate times and places, and that is continuously modified or updated by the results returned by those tests so

[2]What different perceptual instruments exist and how they are used to retrieve specific types of information is, however, an open question.

as to be able to govern perceptual exploration more efficiently in the future" (Thomas, 2002). We can thus think of a set of schemata that instruct the perception and recognition of a cat by specifying which visual features and spatial relations between them have to be successfully tested by the respective perceptual instruments. The successful testing of these aspects corresponds to the experience of seeing something.

Mental imagery comes about when, for example, the schemata for cat are granted (at least partial) control of the respective perceptual instruments and try to recognize a cat while there is actually no cat to be perceived. Compared to visual perception, the employment of the perceptual instruments is either not fully executed or the bottom-up input of the perceptual instruments is ignored during mental imagery. These differences account for the distinct experiences of perception and imagination.

Summarizing, the enactive theory explains (the experience of) mental imagery with the employment of several different perceptual processes guided by respective schemata, which implicitly represent how one visually recognizes a given object.

2.2.4 Summary and Comparison of the Theories

The following provides a brief summary and comparison of the three contemporary theories with respect to 1) the representation of the mental image; 2) the spatio-analogical character of mental imagery, and 3) what constitutes mental imagery.

The representation of the mental image:

- Pictorial theory: the mental image is depictively represented in the visual buffer
- Descriptive theory: the mental image is propositionally represented by amodal descriptions
- Enactive theory: the mental image is not represented directly. Instead the processes that lead to the experience of mental imagery are encoded in the respective schemata

The spatio-analogical character of mental imagery refers to the fact that behavior in mental imagery is often analogical to behavior expected for an actual picture. The mental scanning effect is an example that shows this spatio-analogical character of mental imagery. There are several more examples, e.g., inspecting "smaller" parts of a mental images takes longer than inspecting "bigger" parts (for an overview of similar studies, see Kosslyn, 1980). The three theories explain this spatio-analogical character of mental imagery as follows:

- Pictorial theory: The spatio-analogical character of mental imagery results from the spatio-analogical structure of the visual buffer which holds the depictive mental image. That is, the processing of the mental image is determined by the structure of the mental representation.
- Descriptive theory: The spatio-analogical character results from the non-functional application of one's tacit knowledge. That is, applying the knowledge of what perceiving the to-be-imagined entity would be like and subconsciously emulating of these properties, e.g., expected reaction time patterns.
- Enactive theory: The employment of the processes of visual perception including non-mental processes such as eye movements give mental imagery the same spatio-analogical properties that the visual system has, e.g., longer attention shifts (such as saccades) take more time.

The three theories, furthermore, differ in their assumption of what mental imagery is:

- Pictorial theory: mental imagery is the processing of the mental image in the visual buffer using processes of visual perception. This understanding is based on the assumption that the visual buffer is similarly used during visual perception to provide a mental representation of what is currently perceived.
- Descriptive theory: mental imagery is the processing of the respective amodal descriptions which represent the mental image. These descriptions are not processed by modality-specific mechanisms such as processes of visual perception. Mental imagery is further defined by the concurrent (non-functional) application of one's tacit knowledge about how the content of the current mental image would be perceived in visual perception. Tacit knowledge causes the characteristic behavior, e.g., reaction time patterns, of mental imagery. If descriptions are processed without the application of tacit knowledge, this would be considered general cognitive processing and not mental imagery.
- Enactive theory: mental imagery arises through the employment of those schemata which are otherwise used to perceive real-world entities. It is those entities which are mentally imagined when these schemata are employed without fitting real-world stimuli. That is, the re-enactment of the perception of an entity corresponds to the mental imagination of that entity.

2.3 Evaluation of the Theories

The following gives a brief overview of the explanations and problems of the three contemporary theories with respect to the above reviewed phenomena

of visuo-spatial mental imagery.

2.3.1 Mental Scanning and Cognitive Penetration

The mental scanning effect in its general form is fundamental to the study of mental imagery and accordingly all contemporary theories provide plausible accounts of it.

The pictorial theory provides a **structural explanation** as the mental image is assumed to be represented in the visual buffer which has the property that the metrics of the stimulus are kept in its mental representation. The inspection processes working on the visual buffer are constrained so that they process the mental image successively, i.e., scanning from one point to another on the mental image shifts attention through all the points in between. The linear relation of reaction time and distance is therefore the result of the metrical representation and the respective inspection processes.

The descriptive theory provides a **tacit knowledge explanation** which states that participants use their tacit knowledge of what the scanning task would be like in visual perception and subconsciously emulate reaction times accordingly.

The enactive theory provides what I term an **equivalence explanation** for the general mental scanning effect. The enactive theory proposes the employment of perceptual processes of visual perception during mental imagery so that the process of visually perceiving is re-enacted. Because the mental scanning effect exists in visual perception, e.g., a saccade over a longer distance takes longer, the mental scanning effect is also evident in mental imagery.

The findings that the mental scanning effect can be varied in specifically its slope, i.e., the speed of scanning, by a variety of different factors (as reviewed in Section 2.1.1) poses a more difficult challenge than the general mental scanning effect itself for the **structural explanation** of the pictorial theory. Because the speed of scanning varies, for example, with the expectation of the participants about how long mentally scanning a certain distance takes, the observed reaction times can at most partially result from the structure of the visual buffer. This problem is an instance of the **cognitive penetration argument** against the pictorial theory made by Pylyshyn (e.g., Pylyshyn, 2002). The argument is that if a participant's belief or knowledge can alter his behavior, e.g., reaction time, during mental imagery, then the measured behavior cannot be due to the properties of a fixed representational structure such as the visual buffer. Figure 2.7 explains an experiment of Richman et al. (1979) which is an example of how the mental scanning effect can be manipulated. These findings suggest that if the structure of the mental representation contributes to the mental scanning effect then it does so as one out of several factors. That is, because participants reliably show different scanning speeds due to individual differences,

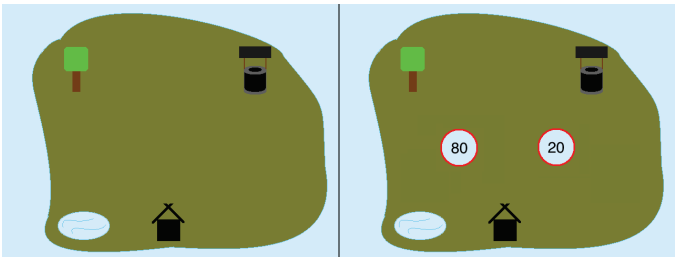


Figure 2.7: Two stimuli depicting an island similar to those used in the mental scanning experiments of (Richman et al., 1979). The island on the right differs from the left one in that it additionally contains sign posts indicating inconsistent distances between the hut and the tree (80 miles), and the hut and the well (20 miles). Richman et al. (1979) found that these sign posts had a significant effect on the mental scanning time along those routes so that mentally scanning along the "80" route took longer than mentally scanning along the "20" route.

e.g., current beliefs, or different demand characteristics (e.g., Goldston et al., 1985) scanning speed cannot be constrained only by the structure of the mental image. However, the pictorial theory does not offer concrete mechanisms how both structure and additionally task demands and individual differences contribute to the variations of the mental scanning effect.

Neither the descriptive nor the enactive theory have been concretely applied to such variations of mental scanning. The descriptive theory seems consistent with the results of the experiment in Figure 2.7, because it assumes mental imagery to be based on interpreted descriptions which could in principle account for the integration of actual and suggested distances. However, the descriptive theory does not make any claims with respect to the concrete mechanisms or the role of tacit knowledge for these cases. The enactive theory is currently not fleshed out enough to be applied to these results. Specifically, perceptual re-enactments alone seem insufficient to explain how the semantics of the sign posts influence the scanning times.

2.3.2 Difficulty of Mental Reinterpretation

The pictorial theory would at first predict that mental reinterpretation of stimuli such as the duck-rabbit is possible and likely according to the theory's assumptions that "[o]nce a configuration of activity exists in the visual buffer, input is sent to the ventral and dorsal systems and is processed in the usual ways – regardless of whether the activity arose from immediate input from the eyes or from information stored in memory" (Kosslyn, 1994, p. 336). The pictorial theory explains the fact that the lack of successful mental reinterpretation of the duck-rabbit, the Necker cube, and the Schröder staircase (Chambers & Reisberg, 1985) clearly does not fit this assumption

by the complexity of these stimuli. Their complexity does not allow one to properly maintain enough of the mental image in the visual buffer at the same time due to the parts of the image constantly fading (Kosslyn, 1994). This **"complexity argument"** has been criticized (e.g., Thomas, 1999; Slezak, 1995) based on the grounds that other stimuli which are assumed to be processed using mental imagery by the pictorial theory are of similar if not higher complexity, e.g., the figures of the mental rotation experiments (Shepard & Metzler, 1971), the island map used in mental scanning experiments (e.g., Kosslyn et al., 1978) or the assumed imagination of two animals in a comparison task (Kosslyn, 1975). If fading plays a role in the inspection of the duck-rabbit, then it is surprising that fading is not considered or necessary to consider in these other imagery tasks. Stimuli which are easier to mentally reinterpret (e.g., Finke et al., 1989) are consequently assigned a lower complexity explaining the difference to stimuli such as the duck-rabbit. Kosslyn (1994) acknowledges the empirical data showing that different hints facilitate mental reinterpretation. This is assumed to affect the current organization of one's mental image into perceptual units. A reorganization of these perceptual units into which a mental image is partitioned is necessary in order to successfully reinterpret a mental image. Such perceptual units can be understood as the composition of the mental image. For example, a mental image of the star of David, can be seen both as two overlapping triangles and as a hexagon with six attached triangles. The theory does, however, not elaborate on the relationship between what is depicted in the visual buffer and the organization of that depiction into perceptual units with respect to the generation and inspection as well as the fading of mental images.

The difficulty of mental interpretation in general lends support to the descriptive theory's assumption that mental imagery relies entirely on abstract and interpreted descriptions and does not utilize modality-specific processes or representations of visual perception. The mental reinterpretation of easier stimuli can be attributed to general (symbolic) reasoning processes that do not require any depictive representations. Given that these stimuli consist of simple geometrical shapes is it plausible that symbolic descriptions alone can represent them with sufficient detail. The descriptive theory does, however, not go into any detail about the possible mechanisms of mental reinterpretation. Also it remains unclear how different hints can generally facilitate mental reinterpretation of stimuli such as the duck-rabbit which are assumed to generally not be mentally reinterpretable because they are solely represented by interpreted and abstracted symbolic descriptions.

The enactive theory only briefly discusses its account of mental reinterpretation. The difficulty of stimuli such as the duck-rabbit is attributed to the fact that different schemata essentially represent how one looks at something in order to recognize it. That is, if one first interpreted the picture of the duck-rabbit as a rabbit, one will employ the respective schemata

that correspond to seeing something as a rabbit during mental imagery. Thus one is inhibited in seeing the mental image as something different, i.e., a duck. The enactive theory further suggests that the easier stimuli of Finke et al. (1989) can be mentally reinterpreted because of a specific (acquired) familiarity with simple shapes and letters. This familiarity with such shapes in many different orientations and circumstances supports the ability to imagine them in many different ways, e.g., after they have been rotated or otherwise manipulated as it is the case for the tasks employed by Finke et al. (1989). It is not discussed how and why different types of hints can facilitate mental reinterpretation.

2.3.3 Functionality of Eye Movements

Mast and Kosslyn (2002) have argued that spontaneous eye movements during mental imagery are consistent with the pictorial theory. They suggest that eye movements might be stored and recalled during mental imagery in order to trigger sequences of memories and specifically help to correctly position parts of a mental image in the visual buffer. A more detailed explanation on the relationship between eye movements and the content of mental images has not been provided so far. Note that spontaneous eye movements cannot be directly integrated into the pictorial theory as they lead to new bottom-up information being sent from the eyes to the early visual areas likely "overwriting" the current mental image in the visual buffer. Furthermore, it has been speculated that the saccadic suppression[3] of eye movements being executed during mental imagery might be responsible for the fast fading of mental images in the visual buffer (Kosslyn, 1994, p. 101).

The concept of visual indexing has been proposed within the framework of the descriptive theory (e.g., Pylyshyn, 2007). Visual indexing is the process of binding entities of one's mental image to visual cues in the environment in order to facilitate processing and saving working memory resources by outsourcing information to the external world. Visual indexing explains spontaneous eye movements during mental imagery as such eye movements would be made towards those visual cues in the environment to which mental entities have been bound. This explanation is, however, at odds with the findings that such spontaneous eye movements have been found even when participants were facing a blank white wall and also when in total darkness (Johansson et al., 2006). In these cases the environment likely could not have provided any visual cues for indexing. Furthermore, spontaneous eye movements during mental imagery would generally not be predicted by the descriptive theory because of its assumption that mental imagery does not require the employment of modality-specific processes or representations.

[3]Saccadic suppression is the fact that "seeing" is suppressed during the execution of a saccade.

Eye movements during mental imagery link naturally to the enactive theory because of its assumptions that exploratory perceptual processes of visual perception are (partially) executed during mental imagery. Furthermore, it follows from the functional role of these perceptual processes that a restriction of eye movements negatively affects the performance of recall and availability of information of the mental images. The enactive theory does currently not provide further details on the link between spontaneous eye movements and mental imagery such as, for example, what types of information are conveyed through the employment of eye movements.

2.3.4 The Constraints of Unilateral Neglect on Theories of Mental Imagery

There are three major findings on unilateral neglect which are relevant for theories of mental imagery: 1) the impairment of accessing information on one side of a mental image that is otherwise available (i.e., imaginal neglect), 2) the dissociation of visual and imaginal neglect, and 3) the apparent role of attentional processes in visual and imaginal neglect.

The pictorial theory could in principle account for imaginal and visual neglect by assuming damage to the respective side of the visual buffer. However, the fact that visual and imaginal neglect can be dissociated from each other (Coslett, 1997) voids this explanation, because the visual buffer is assumed to be employed alike in both vision and imagery. Thus such a damaged visual buffer would always show neglect in both vision and imagery. Bartolomeo (2002) concluded that the pictorial theory requires visual and imaginal neglect to be caused by damage to anatomically and functionally different components outside of the visual buffer. Locating such a component for the case of preserved mental imagery but impaired visual perception, i.e., visual neglect without imaginal neglect, is particularly difficult as "[...] one wonders where the anatomical locus of impairment should be located" (Bartolomeo, 2002, p. 361) given that the visual buffer already comprises the earliest areas of the occipital lobe and that the pictorial theory explicitly assumes that content of the visual buffer is processed alike independently of whether the content of the buffer came from the eyes during vision or memory during imagery (Kosslyn, 1994, p. 336). Accordingly, Bartolomeo (2002) draws the conclusion that the existence of such a visual buffer for both bottom-up visual perception and top-down imagery is not supported by the sum of the results for visual and imaginal neglect. In defense of the pictorial theory in this respect, I see the possibility that visual neglect is caused by damage to processes of attention which control what information of the world is projected into the visual buffer by controlling where the eyes are looking at. At the same time imagery would remain healthy in case of such damage because it presumably does not rely on external attention nor input from the eyes in the pictorial theory. Imaginal neglect with healthy

vision could be caused by damage to those processes that project content from memory to the visual buffer. It is, however, questionable to which extent this function, i.e., projecting information from memory to the buffer, is necessary for vision as it has been proposed that mental imagery plays a role in pattern matching during visual perception (Kosslyn, 1994, p. 259). Either way, the pictorial theory still faces the related problem that the neuropsychological literature reports that damage to the occipital lobe (i.e., the presumed locus of the visual buffer) is neither sufficient nor necessary for deficits in mental imagery (Bartolomeo, 2002; Goldenberg, 1998; Trojano & Grossi, 1994). For example, a recent study reports of a patient experiencing vivid visual mental imagery despite near-complete cortical blindness caused by damage to area V1 (Bridge, Harrold, Holmes, Stokes, & Kennard, 2012). These neuropsychological results show that early visual areas are not functional for mental imagery. That is, they show the implausibility of the visual buffer as assumed by the pictorial theory.

The fact that processing of mental images is disrupted with respect to the information on only one side in imaginal neglect seems hard to account for by the descriptive theory. The descriptive theory assumes that mental images are represented and processed in an amodal format that is not analogically structured. If the underlying representation of the mental image is not structured (spatio-)analogically, then there is no plausible reason why accessing exactly those parts that refer to entities on one side should be disrupted. This problem is acknowledged by Pylyshyn who states that "it would be odd for a symbolic encoding system by itself to have directional preferences, such as found in neglect, and I also agree that most cases of imaginal neglect are unlikely to be due to tacit knowledge" (Pylyshyn, 2002, in reply to comments, R5.3).

Altogether, the combined empirical results of unilateral neglect seem incompatible with both the descriptive and the pictorial theory. Furthermore, Bartolomeo (2002) interprets the combined results of visual and imaginal neglect to provide evidence that theories of mental imagery that assume attentional processes to underlie mental imagery, instead of an internal representation, are most plausible. In particular, he refers to the enactive theory as it assumes schemata as attentional procedures employed in visual perception as well as mental imagery. He states that "[i]f the application of these procedures can be constrained either by the external environment or by memory processes, with distinct neural correlates subserving these occurrences, then double dissociations between perceptual and imagery abilities are expected to arise in brain-damaged patients" (Bartolomeo, 2002, p.374). It is, however, currently not possible to further evaluate the applicability of the enactive theory to unilateral neglect in more depth due to its currently under-specified state.

2.3.5 Summary

The above review of the contemporary theories of mental imagery and their applicability to the considered empirical phenomena has shown that the three theories' descriptive level of formulation does often not allow in-depth explanations for some aspects of the phenomena. Furthermore, the phenomena of functional eye movements and the constraints of the findings on unilateral neglect are hard to reconcile with both the descriptive and the pictorial theory. The enactive theory promises the possibility to overcome some of the problems of the other two theories; yet, it is also the theory that is described most vaguely and currently considered more of a sketch than a fleshed-out theory (Thomas, 1999). In the next chapter, a new theory of mental imagery will be proposed which builds upon some of the assumptions of the enactive theory, but is aimed at providing a comparatively concrete description of the involved mechanisms of mental imagery.

Chapter 3

The Perceptual Instantiation Theory

This chapter introduces the perceptual instantiation theory (PIT) of visuo-spatial mental imagery. It discusses how PIT understands visual perception and visuo-spatial mental imagery. Chapter 4 will then present a formal summary of PIT and Chapter 5 will present a computational model of PIT.

3.1 Visual Perception

Visuo-spatial mental imagery is intimately related to visual perception. Therefore, in order to explain mental imagery, it is necessary to do so based on a clear conception of visual perception.

3.1.1 Visual Perception in the Enactive Theory

The summaries of the three contemporary theories of mental imagery and their comparison in Section 2.2 made it evident that the enactive theory stands out against both the pictorial and the descriptive theory in a few important points. In the following, two such points regarding how the enactive theory understands visual perception and mental imagery are explicitly elaborated:

- The enactive theory emphasizes the role of the attentional and perceptual processes directed at external stimuli, e.g., the role of eye movements, in perception and mental imagery. In contrast, the pictorial and the descriptive theory are generally not concerned with these processes and do, instead, assume mental imagery to be realized on a "higher" level. That is, the processing of mental representations by mechanisms aimed not at external entities but at the mental representation of entities.

- Accordingly, the understanding of visual perception of the enactive theory differs from that of the other two theories. Perception in the enactive theory consists of several different specialized perceptual instruments which are selectively used to retrieve specific information from the environment. In the other theories, in contrast, perception seems to be assumed as a much more generic process whose major task is providing input to the visual buffer (in the pictorial theory) or to the propositional descriptions (in the descriptive theory). The relevant processing of visual perception and mental imagery is accordingly based on these resulting mental representations.

The enactive theory's view of visual perception is partly based on and partly consistent with ideas of active vision (Ballard, 1991), sensorimotor contingencies (O'Regan & Noe, 2001), and ecological vision (Gibson, 1986).

Also because of its different understanding of visual perception, the enactive theory seems to be the most promising candidate to plausibly incorporate the finding of spontaneous functional eye movements during mental imagery as well as being potentially consistent with the constraints posed by the findings on unilateral neglect (see Section 2.3). Yet, the applicability of the enactive theory to all of the considered phenomena is limited as its functional details have not been worked out.

PIT adopts the above mentioned assumptions about visual perception from the enactive theory as well as the functional "re-use" of perceptual processes in mental imagery. These assumptions will be fleshed-out and combined with multi-modal grounded symbols – referred to as mental concepts.

3.1.2 An Example of Visual Perception

The recognition of an object, for example, a square, in visual perception is the result of the application of several different low-level processes of the human visual system. The information necessary to recognize a square includes the edges, their intersection, the relative location and orientation of these features, the separation of the features from the background, whether some of the features go into depth, and much more. The human visual system actively and in a largely top-down fashion applies specific **perceptual actions** to actively find out about these different types of information with the aim to identify the most likely alternative of all those things that the currently perceived object could be.

This view of visual perception means that recognition is not a comparison or pattern-matching process working on a mental representation of what is currently being perceived. But recognition is the successful application of specific perceptual actions to the external stimulus, e.g., the eye movements and their respective feedback lead to the recognition of a square. After

the perceived object has been interpreted as a square, much information is discarded and an abstract conceptual description of the object remains in long-term memory. That is, if one remembers the object in question after some time, the fact that it was, for example, a red small square, remains, but many details, such as the exact size, location, orientation, or color, are often missing or have been replaced by generic information. The information that has been lost by this abstraction comprises the low-level information that was made available by the perceptual actions. This includes, specifically, the coordinates of the object in an ego-centric reference frame, through which information about concrete size, orientation, depth, location, and visual features of the object can be determined. Such information is available during and shortly after the perceptual process on that level of granularity that the visual system is capable of perceiving and distinguishing. This information is referred to as **perceptual information**. In contrast, the abstracted conceptual memory of that object – *red, small, square* – could be seen as qualitative information, but will be referred to as conceptual information, or simply **mental concepts**.

In this section, three important aspects were introduced: 1) perceptual actions, 2) perceptual information, and 3) mental concepts. A fourth important aspect is a modality-specific long-term memory of visual perception referred to as **visuo-spatial long-term memory** which controls the interplay between perceptual actions, perceptual information, and mental concepts. The next sections will elaborate on these aspects.

3.1.3 Visuo-Spatial Long-Term Memory

The visuo-spatial long-term memory (abbreviated VS-LTM) constitutes the procedural knowledge of how to look at the world in order to recognize entities, properties, and relations. For this purpose the VS-LTM provides two mappings: 1) a mapping of perceptual information onto mental concepts, and 2) a mapping of mental concepts onto perceptual actions. These mappings are acquired procedural knowledge and are continuously adapted.

Let's looks at a simple example of how those two types of mappings of the VS-LTM work during visual perception.

Mapping of perceptual information onto mental concepts: From the current fixation of an object O_1 a saccade is triggered bottom-up towards the salient object O_2. Saccades are a type of perceptual actions. The execution of the saccade yields a type of perceptual information. The type of perceptual information is the change of the coordinates of the gaze from the previous position (on O_1) to the new position (on O_2). This information is mapped onto a set of mental concepts, which are thereby identified. In this case these are spatial relations that hold between O_1 and O_2.

Figure 3.1 shows how this example of a mapping of perceptual information onto mental concepts can be described more formally. The figure also

shows how the mapping of perceptual information onto mental concepts is a many-to-many mapping. That is, different perceptual information can be mapped onto one same mental concept and the same perceptual information can be mapped onto different mental concepts. For example, one specific saccade can be mapped onto both the mental concepts *close* and *left-of*. Also, two different saccades can both be mapped onto the same concept, for example, *left-of*.

Mapping of mental concepts onto perceptual actions: Mental concepts can be fully or partially identified. Imagine, for example, that *square* is partially identified when three out of four edges with respective relative locations and orientations are already known.

The currently identified or partially identified mental concepts and the temporarily available corresponding perceptual information determine the perceptual action that will be executed next. This step is realized by the other mapping provided by the VS-LTM: the mapping of mental concepts onto perceptual actions. It is assumed that the strategy of choosing a perceptual action in a given situation follows the principle of maximum information gain. That is, the strategy is to choose that perceptual action which is expected to give the maximum gain of information about what the perceived object or scene is. Such strategies are considered in vision research for scene and object recognition (e.g., Schill, Umkehrer, Beinlich, Krieger, & Zetzsche, 2001). For the example of the partially identified mental concept *square*, i.e., the location and orientation of three edges have already been perceived, the next perceptual action would be an attention shift towards an anticipated fourth edge to gather support for the hypothesis that the object in question is indeed a square. This attention shift might be another saccade. The planning of that saccade takes into account the available perceptual information of the already known edges. That is, the to-be-attended-to fourth edge is anticipated to have a location and orientation fitting with the already known edges. The saccade is then executed towards this specific location. Given that the edge is found at the respective location, the next chosen perceptual action would retrieve the orientation and other visual features. The perceptual feedback, i.e., the location provided by the saccade and information about the orientation, lead to the full identification of the mental concept *square*.

The mapping of mental concepts onto perceptual actions is also a many-to-many mapping. That is, one mental concept can be identified by several different (sets of) perceptual actions and different mental concepts can be identified by employing the same perceptual actions. For example, to check for the fourth edge of a square, both an appropriate saccade and an appropriate head movement are possible. Furthermore, information about distance as well as information about direction between two given objects can be retrieved by the same perceptual action such as a saccade.

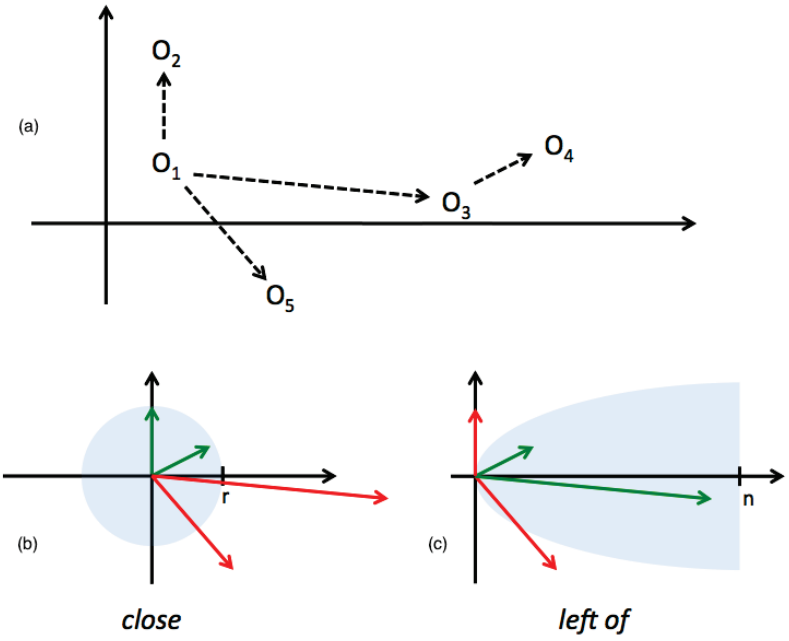


Figure 3.1: Example of the mapping of perceptual information onto mental concepts by the VS-LTM. The top (labeled (a)) shows a set of objects $(O_1, \ldots, O_5)$ in a coordinate system representing the visual field. The dashed arrows represent attention shifts such as saccades made between the objects during visual perception. Each of these attention shifts yields information about the starting and ending coordinates. The relative shift between these coordinates can be represented by a vector. For example, a saccade from the coordinates (2,3) to the coordinates (3,4) can be represented by the vector $v = (1, 1) = (3, 4) - (2, 3)$. These resulting vectors are mapped onto mental concepts such as spatial relations. Two examples of spatial relations are shown in the lower part of the figure (labeled (b) and (c)): *close* and *left-of*. These mappings can be imagined as a check whether a vector, that represents the respective attention shift, falls into a set of points that corresponds to the respective mental concept. For example, the concept *close* is defined by the region of points in blue. An attention shift falls into this region when it does not exceed a certain length, that is, when the distance between two objects is not larger than that certain length. A red vector indicates the lack of recognition whereas a green vector indicates the recognition of the mental concept. For example, it can be seen that an attention shift from O_3 to O_4 would trigger the recognition that O_3 is left of O_4 as well as that O_3 is close to O_4.

3.1.4 Perceptual Actions

It is an assumption of PIT that almost all perception is mediated by and thereby connected to respective perceptual actions. Examples of perceptual actions of visual perception include saccades, micro-saccades, head and body movements, adjusting the focal length of the lens as well as covert actions such as covert attention shifts. Different types of information are retrieved using different perceptual actions. For example, information about locations and spatial relations can be retrieved by saccades, while smooth pursuit is used to track the movement of an object, and adjusting the focal length of our lenses gives information about depth.

Covert and Overt Attention Shifts

Many perceptual actions fall into the category of attention shifts, for example, all different kinds of eye movement can be understood as attention shifts. Two general types of attention shifts are distinguished: covert and overt attention shifts. The latter include all sorts of observable eye movements, such as saccades, micro-saccades, head and body movements (which indirectly move the gaze). Covert attention shifts are those attention shifts which are not directly observable. For example, while keeping our gaze fixated at a certain point, we can still shift our attention to another point in the periphery of our gaze. Such covert attention shifts have been interpreted as part of the planning of overt attention shifts as they usually precede a respective saccade (Theeuwes, Belopolsky, & Olivers, 2009). Covert attention shifts have been shown to be sufficient on their own instead of overt attention shifts as an aid in problem solving (Thomas & Lleras, 2009). This makes sense because the information of a saccade, i.e., specifically the start and ending coordinates, are necessarily already provided by the planning of it. In order to execute a saccade the respective muscles must be controlled so that the gaze actually ends up near the goal position. For this control the information of the goal position in space must be available on some level. The planning of an overt attention shift can thus similarly be used to provide feedback that can then be mapped onto respective mental concepts.

3.1.5 Mental Concepts

The memory of a previously perceived scene corresponds to a conceptual description of that scene in conceptual long-term memory (C-LTM). Conceptual descriptions consist of mental concepts. Mental concepts include spatial relations (e.g., *left-of, close*), objects (e.g., *square, house*), and properties (e.g., *red, big*). The C-LTM comprises all mental concepts and associative links between them. The C-LTM can be understood as what is often referred to as declarative or associative memory (Anderson, 2005). The mental concepts of PIT have two important properties: 1) they are grounded symbols

and 2) they incorporate input from all modalities. These two properties will be elaborated in the following.

The mental concepts of PIT are grounded symbols in that they function as hubs linking to perceptual actions. The linked perceptual actions are those which are used for the perception of the entity that the respective mental concept represents. That is, for example, the mental concept of the relation *left of* comprises the different ways of perceiving the relation *left of* such as certain eye movements, hand movements, attending to certain sound patterns, and hearing the words "left of" in a sentence. The so-defined mental concepts of PIT differ from symbols as often used in cognitive science and artificial intelligence (for example, ACT-R (Anderson et al., 2004), physical symbol system (Newell, 1990), or mentalese (Fodor, 1975)), because 1) they do not contain the semantics of the entity they represent, and 2) they do not directly reflect properties of the entity they represent. The semantics of a mental concept, that is, what the mental concept means to the organism, corresponds not to the processing or the activation of the mental concept, but the semantics are manifested in the process of executing the linked perception actions. That is, the semantics of an entity are the perception of that entity, i.e., what seeing, touching, or otherwise perceiving the entity is like. The mental concepts also do not directly reflect the properties of the represented entity. Consider, for example, that a depictive mental representation of an entity does preserve and thus reflect properties of the represented entity (e.g., Kosslyn, 1994). The properties of an entity instead become available by the perceptual feedback of the perceptual actions in visual perception. In mental imagery, as it will be discussed later, the employment of the linked perceptual action generates a perceptual instance of the represented entity which makes some of the properties of the entity available.

A set of mental concepts describing a scene incorporates the input of all modalities. That is, the perception of, for example, a cheese contains not only the visual and spatial information of it conveyed via visual perception but also the smell that was perceived and its texture and feel when it was touched. The resulting mental concepts of the perception through the different modalities are combined in one final conceptual description of that cheese. Importantly, also subtle and fully subconscious information such as that communicated via different demand or task characteristics (Orne, 1962) is assumed to be included in the final conceptual description. The different modalities can also give conflicting input as in, for example, the McGurk effect (McGurk & MacDonald, 1976). The McGurk effect is an example of sensory integration. When seeing a video of someone saying "ga" without sound but at the same time hearing the sound "ba", we perceive the person in the video actually saying "da" which is a mixture of those two sounds. Conflicting mental concepts can be part of the conceptual description of a scene. When this conceptual description is processed during the

mental imagination of the scene, these conflicting mental concepts might be integrated to make the mental image of the scene consistent.

From the above definition of mental concepts, it follows that mental concepts are not independent of the organism's body as the linked perceptual actions are specific to the perception and action capabilities of the body.

3.1.6 Additional Aspects of Visual Perception

Top-Down and Bottom-Up Control

The employment of perceptual actions can both be triggered top-down and bottom-up. A bottom-up triggered perceptual action is, for example, a saccade that is automatically made towards a moving or otherwise salient object. The execution of that saccade yields perceptual information which is mapped onto mental concepts and influences the selection of the next (top-down triggered) perceptual action. Another way that perceptual actions are guided bottom-up is when they adjust and correct a top-down triggered perceptual action. For example, if three edges of what is presumably a square are already known, an attention shift is executed in order to find the fourth edge. The attention shift is made towards the anticipated location of the fourth edge according to the available perceptual information and the general knowledge of what a square looks like. This anticipated location might not be accurate enough or actually fit with the stimulus at hand, so that the attention shift is corrected bottom-up via, for example, detecting the edge in the periphery of one's gaze.

A top-down triggered perceptual action is one that is selected based on the currently identified and partially identified mental concepts and their corresponding perceptual information.

Interpretation

The iterative process of selecting perceptual actions, executing them to retrieve perceptual information, and identifying mental concepts from the perceptual information results in a set of identified mental concepts and corresponding perceptual information. From this set of identified mental concepts a subset is drawn as the interpretation of what is "seen". This interpretation is usually one out of several possible descriptions of the perceived scene. The selected interpretation is the most plausible and coherent alternative out of all possible ones. This decision is based on the individual background knowledge and the current situation. The interpretation process, furthermore, considers not only the modality of visual perception but the currently available information of all modalities. Figure 3.2 depicts an example of an ambiguous stimulus and its interpretation.

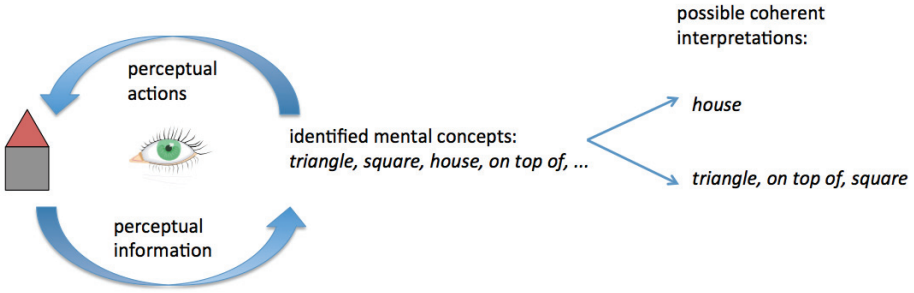


Figure 3.2: Interpretation of an ambiguous stimulus. The stimulus on the left can be seen as a schematized house or as a triangle on top of a square. An interpretation is drawn based on the identified mental concepts. Which interpretation is chosen depends on the context. For example, on a map for wayfinding, the house is more plausible, whereas the alternative interpretation is more plausible for a drawing about geometry.

Short-Term Memory

The short-term memory keeps the currently (partially and fully) identified mental concepts, the gathered perceptual information and the current interpretation of what is perceived. The perceptual information includes specifically the coordinates of the recognized entities and the spatial relations in an ego-centric reference frame. The short-term memory as proposed here describes a mental representation that is in parts functionally similar to other short-term memory structures such as ego-centric perceptual maps (Wang & Spelke, 2002; Sholl, 2001) or the visual cache as an assumed passive visual store in visuo-spatial working memory (Logie, 2003).

3.2 Mental Imagery

This section will present how PIT understands the underlying mechanisms of mental imagery. First, the relationship between mental imagery and visual perception is elaborated.

3.2.1 How Mental Imagery Relates to Visual Perception

This section will give an overview of how mental imagery relates to and builds upon the processes and representations of visual perception. First, I will summarize the process of visual perception.

Figure 3.3 gives a schematic overview of the visual perception of a scene. The recognition of a scene is realized through the dynamic interplay of choosing a perceptual action and retrieving the respective perceptual informa-

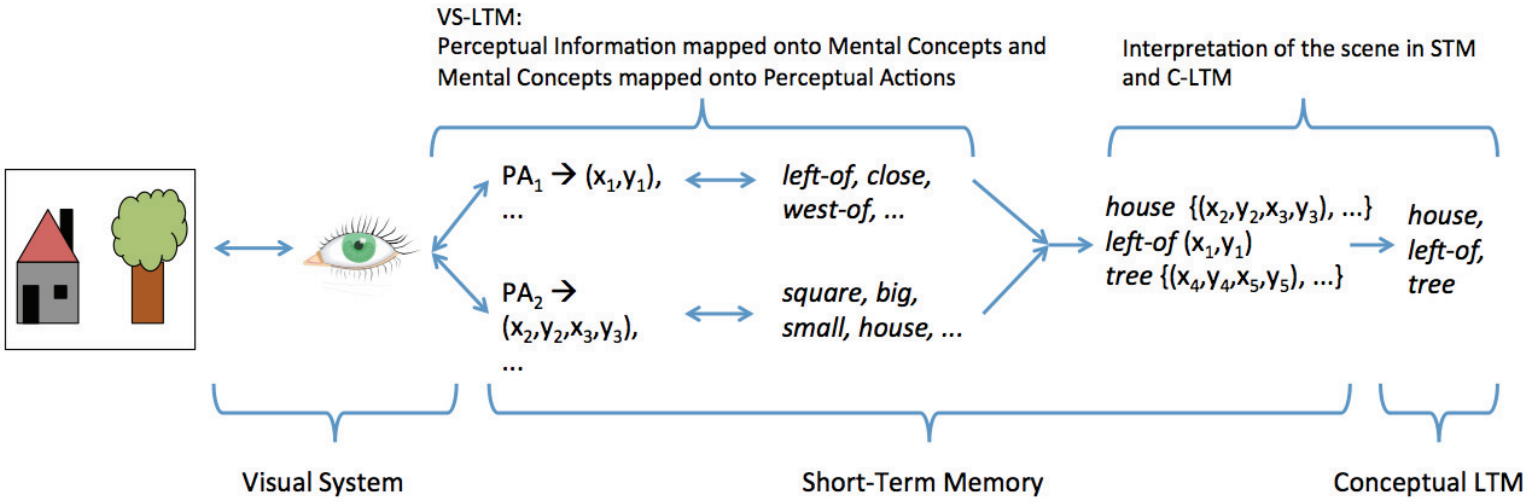


Figure 3.3: Example of visual perception as proposed by PIT. The employment of perceptual actions (PA_i) provides perceptual information which is mapped onto mental concepts. The selection of further perceptual actions is based on the currently identified or partially identified mental concepts and the perceptual information. From the set of identified concepts the most plausible subset is drawn as the interpretation of what is seen. This interpretation is available in short-term memory and represents the current perception. Only the conceptual description without perceptual information is stored in conceptual long-term memory.

tion it yields. The retrieved perceptual information is then used to identify mental concepts based on the gathered perceptual information. The identified mental concepts and the available perceptual information determine the choice of the next perceptual action. From the set of all identified mental concepts, one consistent and most plausible subset is drawn. This subset is the interpretation of what is perceived. Finally, this interpretation is stored in long-term memory. However, in long-term memory the interpretation is abstracted from the perceptual information and only the set of mental concepts is stored. That is, the (long-term) memory of a perceived scene includes a conceptual description such as "red small square" but not the concrete details of the perception such as the actual size, shade of red, or location of the square.

The process of mental imagery can be thought of as the task to (re-)create a set of perceptual information for the set of mental concepts that describe the scene or object that is to be imagined. In other words, mental imagery is the process of (re-)creating one instance of perceptual information for a given conceptual description. For every conceptual description there are several possible instances of perceptual information, because the conceptual description is an abstraction. For example, squares of marginally different sizes, orientations, and colors might all be abstracted to the same conceptual description "small red square".

Figure 3.4 depicts the process of generating a mental image. The process of mental imagery starts with a conceptual description of what is to be imagined. That is, mental imagery starts with the end product of visual perception. A conceptual description of a scene consists of a set of mental concepts such as *house*, *left of*, *tree*. As discussed in Section 3.1.5 mental concepts consist of links to those perceptual actions which are used to identify the mental concepts, i.e., they are used to look at that thing which the respective mental concept represents. In mental imagery these links from mental concepts to perceptual actions are used to create an instance of perceptual information that corresponds to the respective set of mental concepts. The mental concepts are successively mapped onto perceptual actions which are then executed either overtly, e.g., spontaneous eye movements, or covertly. The execution yields perceptual information, for example, information about the change of gaze position. This process of picking and employing perceptual actions for a given mental concept in order to retrieve perceptual information is referred to as **instantiation**. The term instantiation is used because the perceptual information which is made available by the employment of perceptual actions represents one (perceptual) instance of the mental concepts that conceptually describe the mental image. The perceptual information made available through instantiation is mapped onto mental concepts by the VS-LTM just as it is the case in visual perception. The perceptual information and the mental concepts that have been identified based on it, influence the instantiation of further

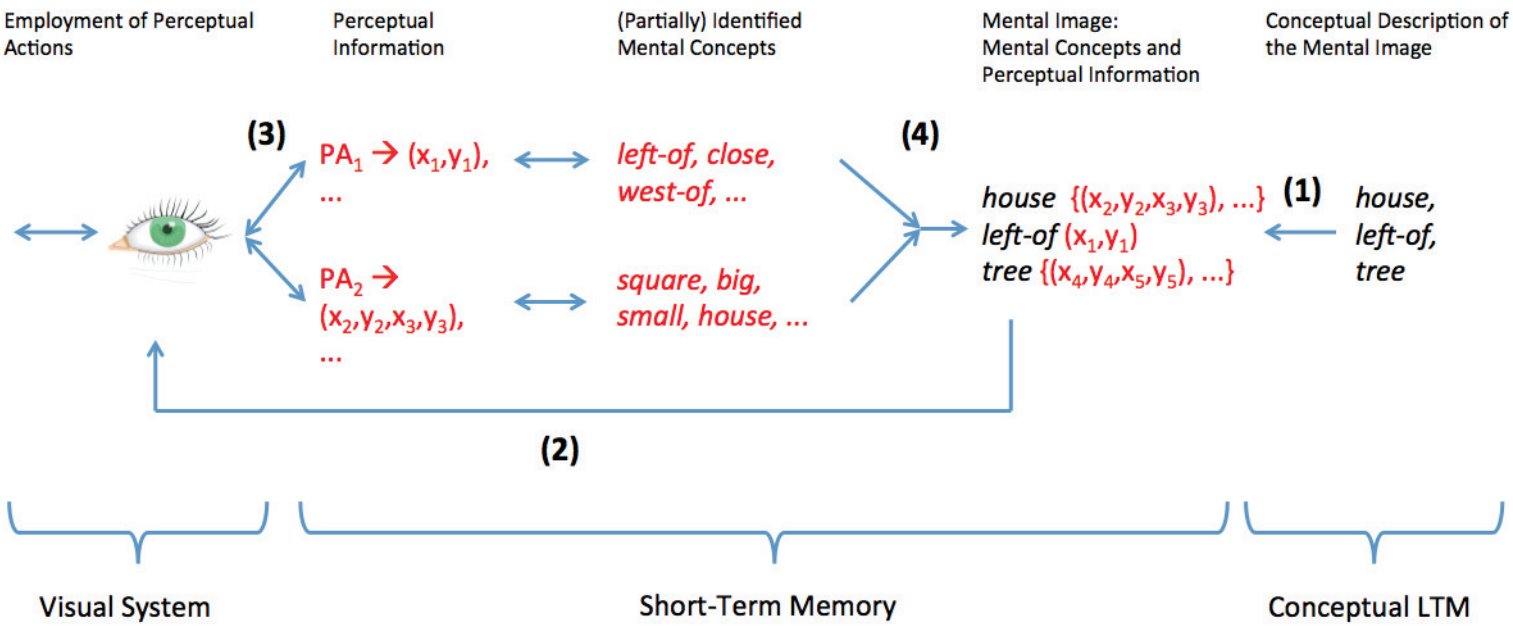


Figure 3.4: Example of how a mental image is generated. (1): the conceptual description of a scene is retrieved from conceptual LTM and held in short-term memory.
(2): the mental concepts describing the scene are successively mapped onto perceptual actions (PA_i). This step is realized by the mapping of mental concepts onto perceptual actions of the VS-LTM.
(3): the perceptual actions are executed by the visual system and yield perceptual information as feedback. This information is mapped onto mental concepts using the mapping from perceptual information onto mental concepts that is provided by the VS-LTM.
(4): from all identified mental concepts a subset is selected as the current interpretation and in case of mental imagery as the current mental image.
Information in red print is information that has been made available by mental imagery.

mental concepts. Again, similar to visual perception, from all identified mental concepts and their respective perceptual information, an interpretation is drawn and held in short-term memory. The interpretation is the most plausible subset of all identified mental concepts with their perceptual information. This interpretation in short-term memory constitutes the mental image. The perceptual information of the mental image held in short-term memory can be used for further processing, e.g., inferring new information such as previously not identified spatial relations.

Mental imagery understood as described above is very similar to what is often described as simulation in the context of embodied and grounded cognition. For example, it is proposed that word and sentence comprehension is realized by respective simulations of the described objects and situations in terms of how they are perceived and which affordances they offer (e.g., Glenberg & Kaschak, 2002; Kaup, Yaxley, Madden, Zwaan, & Luedtke, 2007) and sometimes it is even assumed that all cognition comprises such simulations (e.g. Hesslow, 2012).

3.2.2 Instantiation: Parsimony and Context-Sensitivity

During mental imagery mental concepts are successively mapped onto perceptual actions using the mappings of the VS-LTM. Given that these mappings are many-to-many mappings, a single mental concept can be mapped onto several different perceptual actions. Consider, for example, the spatial relation *left-of* which can be mapped onto several different overt and covert attention shifts. Which perceptual action is chosen depends on several factors.

One factor is parsimony: mental imagery is assumed to be an economic process. That is, to instantiate a mental concept, perceptual actions are chosen so that they require a minimum of (motor) effort to perform. For example, in order to mentally imagine a scene described as "A left of B" the spatial relation *left-of* can be instantiated using a covert attention shift rather than a longer and more expensive saccade. Using shorter attention shifts also means that less time is taken (the timing of perceptual actions is discussed later in Section 3.2.4). Mental imagery is, furthermore, economic with respect to which mental concepts are instantiated. That is, a mental concept is instantiated on demand when required. It is therefore possible that some mental concepts of the description of a given scene are not instantiated at all. For example, given the task "The house is left of the tree and the person is left of the house; what is the relation between the person and the tree?", it is not necessary to instantiate the shapes of any of the entities to solve the task and therefore visual details will not be instantiated to the same degree as for tasks in which shape is relevant such as "Is the tree higher than the house?".

Another factor that determines the choice of perceptual actions for a

given mental concept is the context in which a mental concept is to be instantiated. The context consists of the other mental concepts of the conceptual description of the to-be-imagined scene and the perceptual information of any already instantiated mental concepts. If the conceptual description of a scene contains not only *left-of* but other mental concepts further describing the spatial relation such as *close* or *far*, these are considered for the instantiation of the spatial relation. That is, *left-of, far* is generally mapped onto a perceptual action such as an attention shift that is longer than for *left-of, close*. Already available perceptual information such as the extent of a shape can also influence which perceptual action a mental concept is mapped onto. If the reference object of the relation *left-of* has some extent (i.e., perceptual information about its shape has been instantiated), it influences what *left-of* concretely means in this context, i.e., when we imagine something being left of a tiny object the relation *left-of* has a different concrete distance than when we imagine something being left of a huge object.

Apart from context-sensitivity and parsimony, mental concepts are mapped onto perceptual actions so that prototypical[1] perceptual actions are chosen. That is, without any further information *left-of* would not be mapped onto any of the extreme cases, i.e., very short of very long distance attention shifts, but onto an attention shift of prototypical length and direction.

3.2.3 Perceptual Information and Bodily Feedback

The employment of perceptual actions yields perceptual information which is mapped onto mental concepts. This section discusses what such perceptual information is. I will first focus on visual perception and afterwards discuss the consequences for mental imagery.

I distinguish between three types of perceptual information that result from the employment of perceptual actions during visual perception. The first type is information that directly depends on the stimulus. That is, for example, if one focusses on an object, the photoreceptor cells on the retina fire according to the different wavelengths of the light particles emitted by that object. Much of this activation is transferred to the early visual areas of the occipital lobe and causes the activation of neural representations selective to, among other information, the existence of edges, their orientation, and location (Gazzaniga, Ivry, & Mangun, 2009). This information is directly caused by the perceived object because the activation is caused by the light emitted by the object.

The second type of perceptual information is bodily feedback, specifically, proprioception. This type of perceptual information only indirectly depends on the stimulus as it is derived from the state one's body is in when perceiving the object. Some of the ways we perceive depth are good exam-

[1] Prototypical with respect to the experience of the individual

ples for this kind of feedback. Figure 3.5 shows and explains an example of proprioceptive feedback which yields information about depth, i.e., how far an object is away from the observer. Another example of depth perception is the focal length of the lenses in our eyes. In order to put an object in focus the lenses have to be adapted to the distance of the object just like the lens of a camera. The change in focal length of the lenses is realized by a set of respective muscles. It follows that the state of these muscles corresponds to the distance between the object and the observer. Similar to these examples, attention shifts such as all types of eye movements, head and body movements provide bodily feedback through the resulting positional change of the gaze. The position of one's gaze depends on the positions of one's eyes, head, and body which are derived through appropriate proprioceptive feedback.

The third type of perceptual information made available by the employment of perceptual actions is information that is neurally encoded during the preparation or planning of a perceptual action. This information also does not directly depend on the stimulus. An eye movement such as a saccade is not made randomly but planned. Generally, the planning of attention shifts requires knowledge how to engage one's muscles in order to move one's attention, e.g., one's gaze, to the desired position. That is, the desired position has to be known on some level. Information about the next planned gaze position has been found to be encoded in different neural representations, for example, the frontal eye fields (Schall & Hanes, 1993). Information such as the current and next position of one's gaze can yield information about the (anticipated) metrical properties of the object that is currently inspected as well as spatial relations that hold between inspected objects. Section 3.1.3 already discussed how spatial relations can be identified based on attention shifts. Covert perceptual actions might generally correspond to the planning of respective overt perceptual actions as it is theorized for covert attention shifts (Section 3.1.4 also discussed this aspect of covert attention shifts). Further evidence from different neuroimaging studies indicates that such covert actions comprise the planning of overt actions with the actual execution of the action being cancelled (for a summary, see Hesslow, 2012).

For mental imagery it is obvious that the employment of perceptual actions can lead to at least the second type of perceptual information (for overtly executed perceptual actions) and the third type of perceptual information (for covertly executed perceptual actions), because these types of perceptual information do not necessarily depend on the presence of an actual stimulus. It is, however, possible that also perceptual information of the first type is made available by the employment of perceptual actions if these perceptual actions trigger a recurrent top-down activation of low-level neural representations in early visual cortex. Note that the pictorial theory assumes that exactly the first type of perceptual information is recurrently activated as a pattern of neural activation in the early visual areas of the

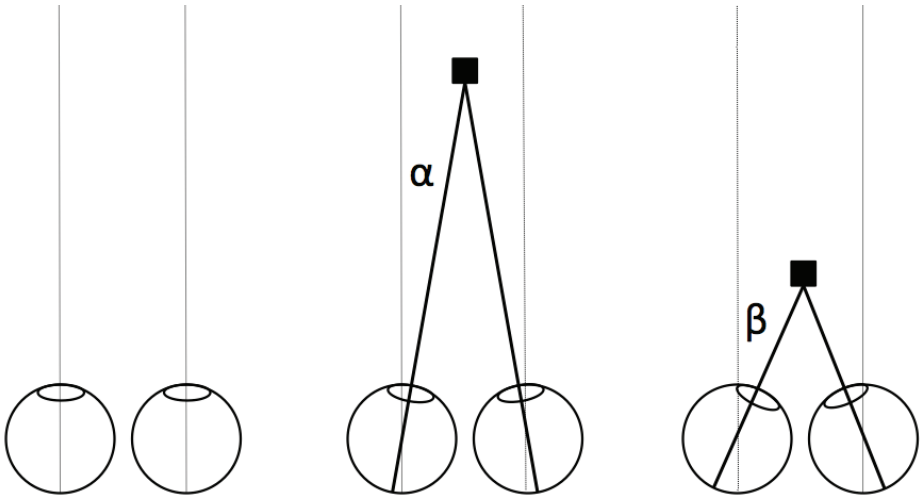


Figure 3.5: The figure depicts the convergence depth cue in binocular vision. The leftmost part shows the position of the eyeballs when the horizon is fixated. The middle part shows the fixation on an object at a medium distance and the rightmost part shows the fixation on an object at close distance. The angle between the gaze direction when an object is fixated and the direction when the eyes look straight ahead depends on the distance of the object. This angle is controlled by a set of respective muscles so that their current state corresponds to the distance between the fixated object and the observer.

occipital lobe (see Section 2.2.1 for more details on the pictorial theory). Neuroimaging studies have found contradicting results with respect to activation of these early visual areas during mental imagery (e.g., Kosslyn & Thompson, 2003; Mellet, Petit, Mazoyer, Denis, & Tzourio, 1998). Neuropsychological data, however, clearly shows that damage to the occipital lobe (which includes the early visual areas in question) is neither sufficient nor necessary for deficits in mental imagery (Bartolomeo, 2002; Bridge et al., 2012).

PIT accordingly assumes that the perceptual information of the second and third type, i.e., bodily feedback and anticipation, is used for the instantiation of mental concepts in mental imagery. Given this understanding, mental imagery critically relies also on non-neural states and feedback of the body.

3.2.4 The Spatio-Analogical Character of Mental Imagery

It is a well-supported finding that mental imagery generally exhibits similar reaction time patterns as visual perception. For example, in both visual perception and mental imagery it is generally the case that the time to scan over a distance increases roughly linearly with the length of that distance, the time to inspect a stimulus increases for small stimuli, and the perception or inspection of a set of objects takes longer the more objects are present (Kosslyn, 1980); also see Section 2.1.1 on mental scanning. It was these findings that first motivated the proposal of spatio-analogical mental representations for mental imagery (for a summary, see Kosslyn, 1980).

Because mental imagery is assumed to employ the same perceptual actions as visual perception, PIT accounts for these similarities in a trivial way and without appeal to specifically structured mental representations. The timing patterns of different perceptual actions determine the timing of both visual perception and mental imagery.

It is worth noting that these timing patterns persist even when perceptual actions are not executed overtly, such as spontaneous eye movements during mental imagery, but covertly, such as shifting attention within the periphery of one's gaze. Covert execution of perceptual actions is assumed to at least include the planning phase of the respective overt perceptual action (see Section 3.1.4), while the final overt behavior is not executed. The timing patterns of the planning phase of an overt perceptual action thus persists for covert perceptual actions. Thomas (1999) and Hesslow (2012)[2] both generally make the assumption that the timing patterns between overt and covert actions show comparable temporal properties. Summarizing, this means that the spatio-analogical character of mental imagery does not result from the properties of an internal mental representation but from the

[2]In (Hesslow, 2012) this assumption refers to the case of conscious simulations, i.e., mental imagery.

properties of the employed perceptual actions which are generally directed at external stimuli. Consequently, it is the physical properties of the visual apparatus or more generally the body that constrain the temporal properties of not only vision but also mental imagery.

3.2.5 Reasoning with Mental Images

Reasoning with mental images is realized by the top-down employment of perceptual actions based on the perceptual information of the mental image. As stated before a saccade between two objects yields perceptual information about the spatial relations that hold between these objects. In a similar way, new information can be inferred not based on actual stimuli but the perceptual information of a mental image. For example, to infer the spatial relation between two entities in a mental image, the perceptual information of their (imagined) locations is used to trigger a top-down attention shift from one location to the other. This attention shift yields new perceptual information which is mapped onto mental concepts just as if that attention shift was executed during visual perception. The so identified mental concept could be a spatial relation which was not explicit before.

3.2.6 Differences between Mental Imagery and Visual Perception

Mental imagery and visual perception both employ the VS-LTM and the same perceptual actions, yet, there are some important differences between the two processes. In the following, two important differences are elaborated. Note that, Section 3.2.3 already discussed a third important difference between mental imagery and visual perception with respect to the types of perceptual information that are used in perception and imagery.

Interpretation

The interpretation process of visual perception selects a subset of the set of all identified mental concepts with their respective perceptual information (see Section 3.1.6). This subset is selected so that it forms the currently most plausible and coherent description of what is presumably perceived. This interpretation is abstracted as only the mental concepts without their respective perceptual information are stored in conceptual long-term memory. This end product of visual perception is also the starting point of the mental imagery process: the conceptual description of a scene is retrieved from conceptual long-term memory and held in short-term memory. The mental concepts of that description are then instantiated as described in Section 3.2.2. In mental imagery an interpretation from the set of all identified mental concepts with the respective perceptual information is drawn

similarly to visual perception. This interpretation corresponds to the mental image. That is, if the task is to form a mental image of a certain scene, this includes the interpretation to be the conceptual description of the given scene. Accordingly, there can be no mental image without an interpretation fitting with the imagery task. If a second alternative interpretation of a mental image is to be found (e.g., in mental reinterpretation tasks), then this alternative interpretation will have to "overwrite" the initial interpretation. In contrast, an ambiguous stimulus can be inspected in visual perception "from scratch", i.e., without the necessity of a previous interpretation.

Attention

In contrast to visual perception, attention is not influenced bottom-up in mental imagery. In visual perception, a salient object in the periphery can draw attention and trigger a bottom-up attention shift towards it. In mental imagery, these types of bottom-up triggered attention shifts do not happen because there is no unexpected information, such as a suddenly appearing object. All perceptual information created during mental imagery is the result of the application of perceptual actions which have been chosen based on the mental concepts that describe the mental image. That is, all attention shifts that are executed as a result of the instantiation of mental concepts during mental imagery are top-down triggered attention shifts. Also attention shifts executed to infer new information from a mental image are based on the perceptual information that was made available through the instantiation of the mental concepts. That is, the starting and ending coordinates of the attention shifts are taken from the perceptual information and the attention shifts are therefore triggered top-down.

Chapter 4

A Formal Framework of PIT

This section presents a formalization of the perceptual instantiation theory (PIT). First, the core commitments of PIT are summarized. Based on these, the processes involved in visuo-spatial mental imagery are described formally. Lastly, PIT is compared to the three contemporary theories so that differences and similarities between PIT and the other theories become clear.

4.1 Core Commitments of PIT

The following summarizes the core commitments of PIT based on its description in Chapter 3:

- Mental concepts: mental images are based on mental concepts. Mental concepts are abstracted conceptual descriptions. A mental concept is grounded in those perceptual actions that are used to recognize the entity the mental concept represents (see Section 3.1.5).
- Visuo-spatial long-term memory (VS-LTM): visual perception is controlled by the procedural knowledge of how perceptual information maps onto mental concepts and how mental concepts map onto perceptual actions (see Section 3.1.3).
- Instantiation: during mental imagery the mental concepts describing a scene are instantiated with one concrete instance of perceptual information (see Section 3.2.2).
- Perceptual information: perceptual information is generated during mental imagery by the (overt or covert) employment of perceptual actions. Perceptual information is conveyed via bodily feedback, i.e., proprioception, and anticipation, i.e., neural encodings of (planned) perceptual actions (see Section 3.2.3).

- Identification: perceptual information is mapped onto mental concepts using the mapping of the VS-LTM, i.e., mental concepts are identified based on the available perceptual information (see Section 3.2.2).

- Interpretation: in visual perception and mental imagery an interpretation is drawn from the set of all identified mental concepts and their respective instances of perceptual information. This interpretation represent the most plausible alternative of all subsets. In visual perception the interpretation corresponds to what is consciously perceived (see Section 3.1.6).

- Mental image: in mental imagery this interpretation, i.e., the subset drawn from all identified mental concepts together with their instance of perceptual information, constitutes the mental image (see Section 3.2.1).

- Spatio-analogical character: the temporal properties of visuo-spatial mental imagery are similar to those of visual perception because the same perceptual actions are employed in both perception and imagery. The temporal properties of these perceptual actions are determined by the physical properties of the human visual system. (see Section 3.2.4).

4.2 Formal Framework of PIT

In this section the process of generating a mental image, that is, the imagination of a scene from long-term memory, is described within a formal framework. The framework is intended to clarify the principles underlying the process of mental imagery. This means that the process is broken down into functions. These functions and the operands they process will be described regarding their functionality, i.e., what purpose they serve and what they represent. Chapter 3 discussed these aspects on a theoretical level with respect to the theoretical and empirical psychological literature. Here, the discussion will focus on making clear how these aspects can be understood from a formal perspective.

4.2.1 Functions

I will start with the operands on which the functions of mental imagery operate. These are 1) perceptual information, 2) perceptual actions, and 3) mental concepts.

Perceptual information are the low-level features the human visual system can perceive. For example, edges, orientation and location of edges, brightness, and color.

Perceptual actions are the basic actions of the human visual system. For example, eye movements, head movements, adjusting the lens, and covert attention shifts.

Perceptual information and perceptual actions are intimately connected in two ways. First, perceptual actions are used to retrieve perceptual information from the environment. Second, they are connected through mental concepts. **Mental concepts** are associative hubs linking perceptual information and perceptual actions. They are identified when a certain combination of perceptual information has been perceived. For example, the mental concept *square* is identified when the features of a square have been perceived, e.g., four edges in a certain arrangement. The recognition of a square in visual perception corresponds to the full identification of the mental concept *square*. Furthermore, a mental concept links to those perceptual actions which are used to perceive these defining features.

Through this mapping of perceptual information to mental concepts and the mapping of mental concepts to perceptual actions, a top-down guided active perception is realized. For example, the perception of three edges triggers the (partial) identification of the mental concept *square*. The mental concept then provides the respective perceptual action to test for the existence of a further fitting edge to confirm the hypothesis of a square being present.

Figure 4.1 depicts these relationships between perceptual information, perceptual actions, and mental concepts. The figure contains the three functions that realize these relationships: 1) **execute**, 2) **identify**, and 3) **select**. These functions are elaborated in the following.

The function **execute** represents the execution of a perceptual action by the visual/motor system. The execution of a perceptual action yields perceptual information. The time of executing a perceptual action depends on the physical constraints of the visual/motor system. Both overt and covert perceptual actions can be executed. Overt perceptual actions include specifically spontaneous eye movements. The function **execute** has to include a mechanism to decide whether a perceptual action is to be executed overtly or covertly.

The function **identify** identifies or partially identifies mental concepts based on the newly retrieved and previously retrieved perceptual information.

The function **select** then selects the next to-be-executed perceptual action based on the identified mental concepts as well as the available perceptual information. For example, consider a saccade towards the location of an anticipated fourth edge of a presumed square. The mental concept square provides the perceptual action to check for a fourth edge while the current perceptual information, e.g., the location and distance between the already perceived edges, is used to adjust the perceptual action so that the edge is checked for in the "correct" location. The selection of a perceptual

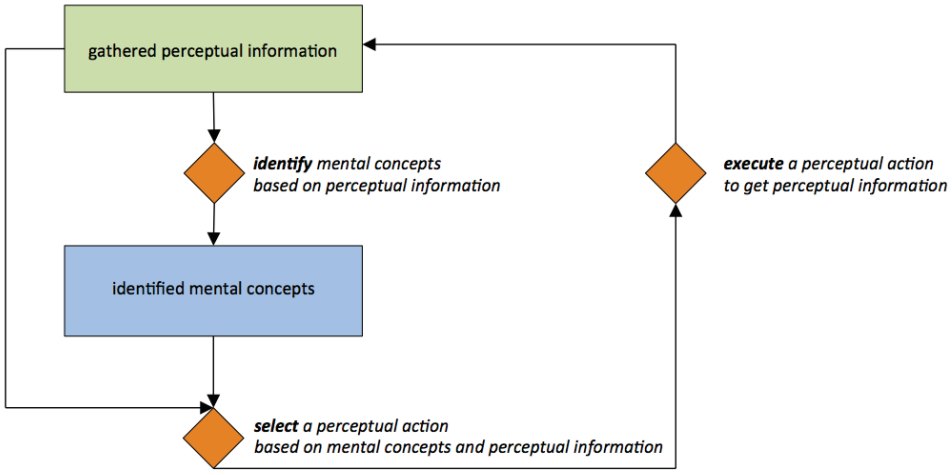


Figure 4.1: The cyclic process of select-execute-identify used during visual perception for object and scene recognition and during mental imagery for the instantiation of mental concepts. The cyclic process comprises 1) the selection of a perceptual action based on the identified mental concepts and available perceptual information; 2) the execution of the perceptual action to retrieve further perceptual information; and 3) the identification of mental concepts based on the available perceptual information.

action is subject to the principle of parsimony (see Section 3.2.2).

Now that the functions that govern the relationships between perceptual information, perceptual actions, and mental concepts have been clarified, they can be embedded into a large picture including the components of the framework of PIT. Figure 4.2 depicts this framework which also incorporates the functions and operands of Figure 4.1. The figure contains two additional functions: 1) **retrieve**, and 2) **interpret**. These are discussed in the following.

The function **retrieve** retrieves a set of mental concepts from conceptual long-term memory. This set of mental concepts is the conceptual description of the to-be-imagined scene. These mental concepts are used to start the cyclic process of **select-execute-identify** until at some point the function **interpret** selects a subset of mental concepts with their instantiation of perceptual information, i.e., the perceptual information generated by the perceptual actions that were selected based on these mental concepts. This subset is the interpretation of what is perceived in the case of visual perception and of what is imagined in the case of mental imagery. That is, the **mental image** is constituted by the drawn subset of mental concepts and their instances of perceptual information.

The function **interpret** selects what is considered the most plausible

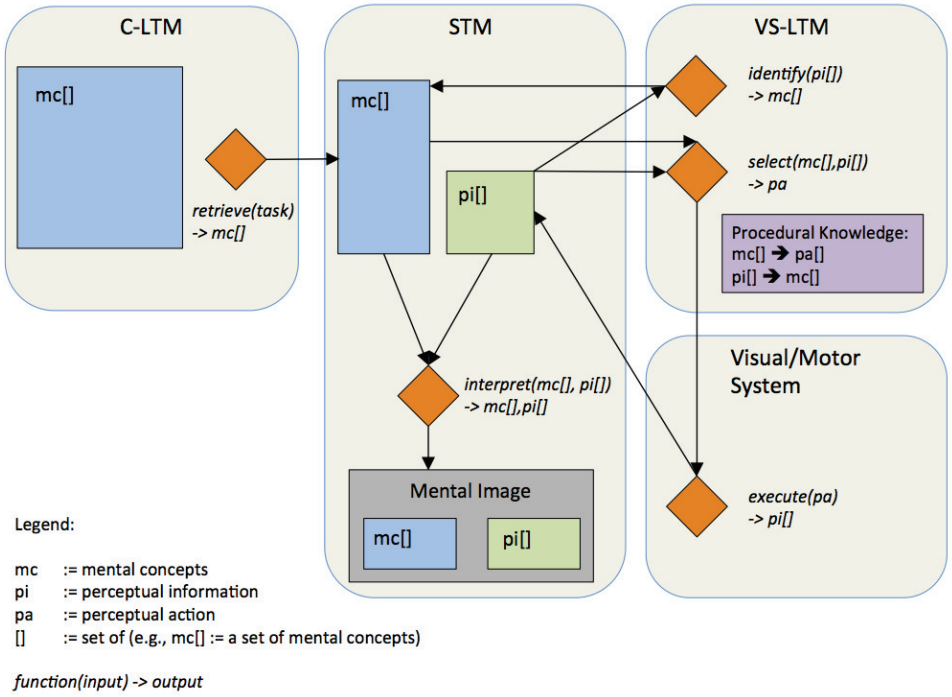


Figure 4.2: The formal framework of PIT. The mental imagination of a scene starts with 1) the retrieval of a set of mental concepts from C-LTM which conceptually describe the scene; 2) these mental concepts are successively instantiated with perceptual information by the cyclic process of select-execute-identify; 3) an interpretation is drawn from all identified mental concepts with their instances of perceptual information; 4) this interpretation constitutes the mental image of the scene.

subset. For example, a set of perceptual information might correspond to both the mental concepts *square* and *triangle* as well as to a (schematic) *house*. Then the interpretation whether one "sees" a house or two geometric shapes depends on what is most plausible given the current situation and the background knowledge.

The components of the framework are 1) conceptual long-term memory (C-LTM), 2) short-term memory (STM), 3) visuo-spatial long-term memory (VS-LTM), and 4) visual/motor system. The **C-LTM** is the long-term memory of conceptual information. It consists of the set of all mental concepts. The **STM** temporarily holds perceptual information and mental concepts relevant to the current process of perception or imagery as well as the mental image. The **VS-LTM** holds the procedural knowledge of how to interact with the environment in order to recognize familiar objects and relations. This procedural knowledge is represented by 1) a mapping of mental concepts onto perceptual actions realized by the function **select** and 2) a mapping of perceptual information onto mental concepts realized by the function **identify**. The **visual/motor system** represents the human visual system which executes perceptual actions to retrieve perceptual information.

4.3 Comparison to the Contemporary Theories

In the following PIT is compared to the three contemporary theories of mental imagery (see Section 2.2 for a summary of the three theories).

4.3.1 The Pictorial Theory

The comparison of PIT to the pictorial theory is presented along three main questions for which the two theories give different answers. These questions are: 1) what type of information is stored in long-term memory that is retrieved and used to generate a mental image?; 2) what type of low-level perceptual information does a mental image comprise?; and 3) how does the spatio-analogical character of mental imagery come about?

What Information is Stored?

PIT is similar to the pictorial theory in that it also assumes that mental imagery makes previously implicit information explicit. It does, however, differ with respect to what this implicit information is.

In the pictorial theory the initially implicit information are the encoded mental percepts which are made explicit through a representation of them in the visual buffer. Once the depictive mental image is generated in the visual buffer, it is inspected by the same processes that process content of the visual buffer during visual perception. This inspection can yield information formerly not accessible. For example, in order to recall the shape of Snoopy's

ears, the encoded percept of Snoopy is retrieved and decoded by placing it in the visual buffer to then "read off" the required information.

In contrast, no percept of Snoopy would be stored in PIT, but instead Snoopy's shape is indirectly described by the procedural knowledge of how the mental concepts describing Snoopy are linked to perceptual actions. The execution of those perceptual actions then creates an instance of perceptual information corresponding to the shape of Snoopy.

In a way, PIT takes a shortcut by leaving out the depictive mental percept which is mentally inspected and, instead, directly utilizing the processes of inspection themselves without a mental percept that is to-be-inspected. In this sense, PIT can be seen as a more parsimonious version of the pictorial theory. However, the inspection mechanisms posed by PIT are quite different to those of the pictorial theory. This difference is rooted in the different assumptions of the two theories about visual perception. The pictorial theory assumes that both visual perception and mental imagery rely on the same processes which are employed to process content of the visual buffer. These processes are different than the perceptual actions proposed by PIT, because perceptual actions are employed not to process an inner mental representation but are directed at external stimuli. The pictorial theory further assumes an inspection mechanism which relies on matching the content of the visual buffer against the stored mental percepts. Whereas PIT assumes that perception and recognition are the result of the successful application of different specific perceptual actions to the external stimulus.

What Low-Level Perceptual Information Does a Mental Image Consist of?

Another difference between PIT and the pictorial theory is the type of information that is assumed to make up the mental images. In the pictorial theory it is low-level information of early visual areas which is recurrently activated during mental imagery. This information corresponds to the (partly) depictive information that is transferred from the retina into the visual cortex during visual perception. PIT does not rely on this type of low-level information although it does not claim that such information might not somehow be recurrently activated, but it specifically poses that proprioceptive feedback of executed perceptual actions as well as information made available through covert perceptual actions makes up mental images (see Section 3.2.3 for more details on this distinction of low-level information). Accordingly, the pictorial theory emphasizes the functional involvement of modality-specific mental representations in mental imagery whereas PIT emphasizes the functional involvement of processes of sensorimotor interaction with the environment as constitutive of mental imagery.

The pictorial theory, furthermore, poses that the depictive mental percepts that are stored in long-term memory constitute the mental image once

they are encoded and placed in the visual buffer. That is, these percept-like memories are the low-level perceptual information that the mental image consists of. In contrast, in PIT the low-level perceptual information that makes up the mental image is not stored at all, but has to be actively "created" by the employment of respective perceptual actions.

Where Does the Spatio-Analogical Character of Mental Imagery Come From?

In the pictorial theory, it is (to a large extent) the spatial properties of the mental image in the spatio-analogical visual buffer that give rise to the fact that mental imagery shares many properties with visual perception. To be clearer, because the mental image is similar to an actual image, i.e., the mental representation preserves metrical properties, mental imagery shows behavioral similarities to visual perception. In contrast, PIT poses that the spatio-analogical character of mental imagery is determined by the characteristics of the perceptual actions. The perceptual actions are physically constrained by our visual system, e.g., to which extent and with which speed we can execute a saccade is limited by our body's ability to move our eyeballs in both visual perception and mental imagery.

4.3.2 The Descriptive Theory

PIT is compared to the descriptive theory with respect to two aspects: the symbolic descriptions and the tacit knowledge.

Mental Concepts vs. Amodal Symbols

The descriptive theory poses that mental imagery does not require a specific mental representation, but that it can be realized using the same type of abstract descriptions that it assumes for other cognitive functions. This means that mental images are fully represented in an amodal and non-analogical format using symbolic descriptions. There are two critical differences between the symbolic descriptions of the descriptive theory and the mental concepts of PIT. First, in contrast to the descriptive theory, the mental concepts of PIT are not amodal. As described in Section 3.1.5, they consist of links to perceptual actions of several modalities. Second, in PIT a mental image is not fully represented by its conceptual description, i.e., the mental concepts. Instead, the mental concepts describing a given mental image only form the basis for the selection of respective perceptual actions whose execution then creates the perceptual content of the mental image, i.e., instances of perceptual information for the given mental concepts. That is, in order to imagine a square the execution of perceptual actions which the mental concept *square* links to is necessary.

Procedural Knowledge vs. Tacit Knowledge

The tacit knowledge proposed by the descriptive theory serves the purpose of explaining why we show similar behavior in mental imagery and visual perception despite the fact that the mere processing of symbolic descriptions should not be restricted in such a way[1]. Tacit knowledge specifically includes the subconscious knowledge of what visual perception is like. It is assumed that we use tacit knowledge to mimic, for example, reaction times during mental imagery to fit those expected during visual perception of the same stimulus (Pylyshyn, 1981). A crucial aspect of tacit knowledge is that its application in mental imagery is not functional, that is, the simulation that it is used for is not necessary to access or process the information of the mental image. To make this point clearer, using tacit knowledge during mental imagery is optional, yet, it is often induced by different demand characteristics (Orne, 1962). The VS-LTM and its mappings could at first glance be seen as an instance of tacit knowledge as it essentially contains the same information that is assigned to tacit knowledge, i.e., the implicit knowledge of how visual perception works. However, tacit knowledge contains information what visual perception is like and the skills to simulate it, whereas the VS-LTM contains the knowledge and skills of how to visually perceive. That is, without the VS-LTM there is neither vision nor imagery. Critically, this means that the VS-LTM is functional for mental imagery and in that respect distinct from tacit knowledge.

4.3.3 The Enactive Theory

PIT is most similar to the enactive theory. Many of the assumptions of PIT were adopted from the enactive theory. Given that the enactive theory has only been described on a relatively coarse level compared to PIT, PIT could be understood as a fleshed-out version of the enactive theory. This role is, however, limited as some of the core commitments of the two theories are incompatible. In the following, I will point out to which extent the two theories are compatible and at which point they part in their assumptions.

Schemata and the VS-LTM

The employment of several different and specialized perceptual processes both in (visual) perception and in (visuo-spatial) mental imagery is assumed by the enactive theory and adopted by PIT. From this, it follows that explanations for phenomena of mental imagery that generally show similar behavioral properties as visual perception, e.g., mental scanning, are explained by both theories in principle in the same way. Furthermore, the

[1]For example, when shifting attention across a mental image, we reliably take more time the longer the distance is that we shift across. In this sense the processing of the mental image is restricted. Section 2.1.1 discusses mental scanning in detail.

role of the VS-LTM in PIT can be treated as functionally equivalent to the schemata of the enactive theory. Both schemata and the VS-LTM work in a circular manner by activating and receiving input from several specialized perceptual processes. The difference between them is, that the VS-LTM maps what is perceived onto a set of mental concepts. Also the interpretation of what is perceived in visual perception and what is "perceived" as the mental image is represented by mental concepts (with respective perceptual information). In contrast, the enactive theory explicitly states that it assumes no such thing as an end product of visual perception and, in fact, it assumes no explicit mental representations at all in the brain or mind (Thomas, 1999, p. 218). Accordingly, the two theories further differ as there are no corresponding components in the enactive theory to PIT's short-term memory, conceptual long-term memory, and mental concepts, which can all be considered mental representations.

As the enactive theory assumes no explicit end products of visual perception, it instead poses that all experience and knowledge is implicitly stored through adjustments to the schemata themselves. That is, changes to the dynamic processes of visual perception. PIT similarly assumes that the mappings of the VS-LTM are being adjusted to implicitly store new knowledge such as new or finer distinctions between mental concepts.

Open Issues in the Enactive Theory

Two open issues of the enactive theory are 1) the question how schemata can be concretely imagined, and 2) how schemata can be embedded in a general cognitive framework. PIT can be understood as a further development of the enactive theory that addresses these two issues. However, the answers offered by PIT are in contradiction to the core assumptions of the enactive theory as elaborated above. Nevertheless, because of the overlap between the two theories' assumptions about (visual) perception and the "re-use" of perceptual processes in mental imagery, a successful application of PIT to phenomena of mental imagery should also be considered as (at least partly) supporting the enactive theory.

Chapter 5

The Computational Model

This chapter describes a computational implementation of the framework of the perceptual instantiation theory (PIT). I describe the architecture of the model and present two examples showing how the model works. For a superficial understanding of the computational model it is sufficient to read Section 5.2 describing these two examples. Lastly, some general aspects about implementations of the framework of PIT are discussed.

5.1 The Architecture of the Model

In the following sections I first discuss the different components of the model and what representations they include. Afterwards, a list of all the data types used in the model is given. Lastly, I describe how the functions are implemented in the model.

5.1.1 The Components and Representations of the Model

As outlined in the formal framework of PIT, the computational model comprises the four components: 1) C-LTM, 2) STM, 3) VS-LTM, and 4) the visual/motor system.

The C-LTM holds the set of all mental concepts. This set is represented in a directed graph. The nodes of the graph correspond to scenes, objects, and properties of objects while the edges correspond to spatial relations. This graph holds not only different remembered scenes but also the general background knowledge. Figure 5.1 depicts a subgraph of the graph in C-LTM.

The STM holds a set of mental concepts and a corresponding set of perceptual information. The mental concepts are represented in the same way as in C-LTM, i.e., as a directed graph. Perceptual information that is made available via instantiation extends the respective mental concepts. The perceptual information is subject to decay. This decay is implemented by the

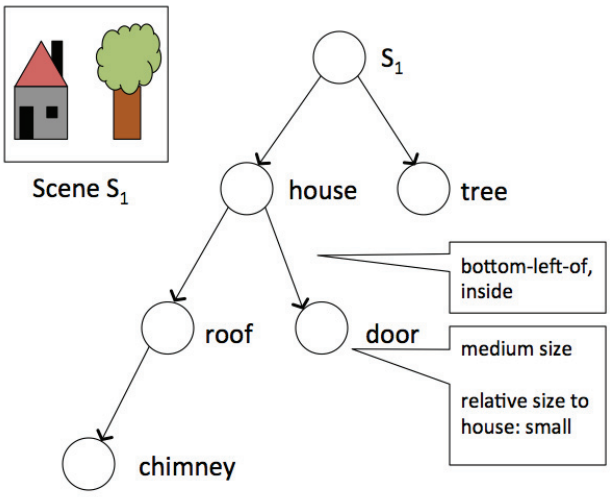


Figure 5.1: A subgraph of the graph of all mental concepts in C-LTM. The depicted subgraph shows parts of the conceptual description of the also depicted scene S_1. The nodes represent scenes and objects and further include properties of objects. The edges represent the spatial relations between objects and parts of objects. The same graph structure is also used in STM where the edges (spatial relations) and nodes (objects) are extended with instances of perceptual information.

removal of perceptual information after a set time interval. Decayed perceptual information will have to be re-instantiated if needed. Furthermore, there is an upper bound to the number of mental concepts that can be held in STM. If this bound is exceeded by retrieving or inferring further mental concepts, previously retrieved mental concepts are removed. This removal first affects those mental concepts that have not been used recently.

The VS-LTM holds the procedural knowledge of what perceptual actions to employ and how to identify mental concepts based on perceptual information. The representations in VS-LTM that are used by the two functions **identify** and **select** to realize this procedural knowledge are: a normalized vector for each direction, a scaling factor for each distance and each size, and a set of normalized vectors for each object. How these representations are used by the two functions is elaborated in Section 5.1.3.

The implementation of the visual/motor system comprises one variable, one constant, and the function **execute**. The variable represents the current focus of attention and is implemented as a coordinate with an initial value of (0,0). It is referred to as **focus**. The constant is a fixed value representing the distance from the current focus within which attention shifts are executed covertly. This value is referred to as **radius**.

5.1.2 The Data Types Used in the Model

In the following, a complete list of all data types used in the model is given. The extended Backus-Naur-Form syntax is used[1].
scene := id, object^n, spatial_relation$^{[(n-1)...n(n-1)]}$
object := id, [size], [part^n], [coordinate], [imagined_size],
[shape_information], instantiated
part := object, relative_size, spatial_relation
imagined_size := size
size := large_size | medium_size | small_size
spatial_relation := direction, [distance], [topology]
direction := left | right | top | bottom | top_left | top_right |
bottom_left | bottom_right | center
topology := inside_outside
distance := far_distance | medium_distance | close_distance
shape_information := line^n, extent^2
line := coordinate^2
location := coordinate
attention_shift := id, vector^n
focus := coordinate
coordinate := x, y

[1] http://standards.iso.org/ittf/PubliclyAvailableStandards/s026153_ISO_IEC_14977_1996(E).zip; raised numbers indicate repetitions, e.g., A := B^x indicates that A comprises x elements of B.

vector := x, y

relative_size := -2 | -1 | 0 | 1 | 2
id $\in \mathbb{S}$ (the set of strings)
x, y, extent, radius $\in \mathbb{R}$ (the set of real numbers)
inside_outside, instantiated $\in \mathbb{B}$ (the set of boolean values)
far_distance := 4
medium_distance := 2
close_distance := 1
large_size := 4
medium_size := 2
small_size :=1
left := 1
right := 2
top := 3
bottom := 4
top_left := 5
top_right := 6
bottom_left := 7
bottom_right := 8
center := 9

Perceptual Information, Perceptual Actions, And Mental Concepts

Perceptual information is implemented by the following data types of the above list: shape information, location, and vectors (as instances of spatial relations).

The model generalizes perceptual actions to attention shifts which are implemented by vectors.

Mental concepts correspond to all the qualitative data types that are used to describe scenes or objects in the model; that is, scenes, objects, the sizes of objects, and spatial relations, i.e., directions, orientations, and topology.

5.1.3 The Functions of the Model

The Function Retrieve

The function **retrieve** provides a set of mental concepts matching a to-be-imagined scene from C-LTM to STM. Each remembered scene corresponds to a node in the graph of the C-LTM. This is the scene-node. The nodes directly linked to from that scene-node correspond to the objects in the scene. These are retrieved together with the spatial relations of the scene (which

are the edges from the scene-node to the object-nodes). These initially retrieved mental concepts do not fully describe the scene, because the parts of the objects are not retrieved. The parts of the objects are the nodes that the object-nodes link to. If these details are required for a mental imagery task, they will be retrieved with an additional call of the **retrieve** function asking for the respective object-node. This way further and further details of the to-be-imagined scene, i.e., parts of parts of parts, are retrieved on demand.

The Function Interpret

The function **interpret** selects a subset of mental concepts and perceptual information from the set of mental concepts and perceptual information in STM. This selection represents the mental image. The function **interpret** is implemented in a trivial way. Section 5.3.2 will discuss why a plausible and non-trivial implementation of this function is an unsolved and hard problem in itself and thus outside the scope of this thesis.

The function **interpret** is implemented so that those mental concepts which are retrieved from C-LTM as the conceptual description of the mental image will always be part of the interpretation together with the perceptual information that is made available via the instantiation of those mental concepts. It is possible that perceptual information leads to the identification of other mental concepts that are not part of the initial conceptual description (see the function **identify** described later in this section), but these mental concepts will not be part of the final interpretation.

Functions of the User Interface

The computational model can run different simulations of the generation of mental images and the inference of information from mental images. These functions are implemented as part of the STM component and are listed in the following:

- **Visual imagination of a scene**: instantiation of the shapes of the objects and the spatial relations of the given scene. This does not include the instantiation of the parts of the objects.

- **Spatial imagination of a scene**: instantiation of the spatial relations and assignment of locations to the objects of the given scene. The shape information of the objects is not instantiated; the objects are abstracted to points.

- **Elaboration of an object**: instantiation of the parts of the given object and of the spatial relations between the parts and the object.

- **Detailed imagination of a scene**: equals the visual imagination of the scene with a subsequent elaboration of each object in that scene.
- **Inferring spatial relations between two objects**: the location and optionally further instantiated perceptual information (e.g., shape information) are used to identify a spatial relation.
- **Identification of objects based on an instance of shape information**: the instantiated shape information is identified, i.e., it is mapped onto a set of mental concepts that (fully or partially) fit with the shape information.

The Function Select

The function **select** realizes the procedural knowledge of the VS-LTM which maps mental concepts onto sets of perceptual actions. This function is context-sensitive so that already available perceptual information is taken into account when mapping mental concepts onto perceptual actions. In the following, it is distinguished between the selection of perceptual actions for mental concepts describing spatial relations and for mental concepts describing shapes.

Mapping Spatial Relations Onto Attention Shifts Spatial relations comprise of a direction and optionally a distance and topological information. The implemented directions are *top-left*, *top*, *top-right*, *right*, *bottom-right*, *bottom*, *bottom-left*, *left*, *center*). *Center* represents the lack of a direction and is used to describe how parts relate to their father-object. For example, to describe the relative location of the nose in a face: the nose is located in the middle of the face, i.e., in the *center*. Directions are implemented by normalized vectors, e.g., the prototypical *left of* is implemented by (-1,0). Distance is represented by three mental concepts *close*, *medium*, and *far* which are used as the factors 1, 2, and 3. These factors scale the normalized vectors that represent the direction. Topological information is restricted to the distinction of whether a part is inside or outside of its father-object.

The following describes how a given spatial relation is mapped onto an attention shift, i.e., a vector. Three cases are distinguished; each new case is an extension of the former: 1) only direction and distance are provided, 2) additionally shape information of the reference object is provided, 3) additionally topological information is provided.

Case 1: a spatial relation describing direction and distance. The normalized vector corresponding to the given direction is multiplied with the factor for the given distance. If no distance is provided the default distance of *close* is used.

Case 2: a spatial relation within context. Same as case 1 but the vector is multiplied with the extent of the shape along the respective axis, e.g., for *left-of* the horizontal extent is used. Additionally the extent is multiplied with a scaling factor with the default value of 1[2]. The results of the above calculations are depicted in Figure 5.2.

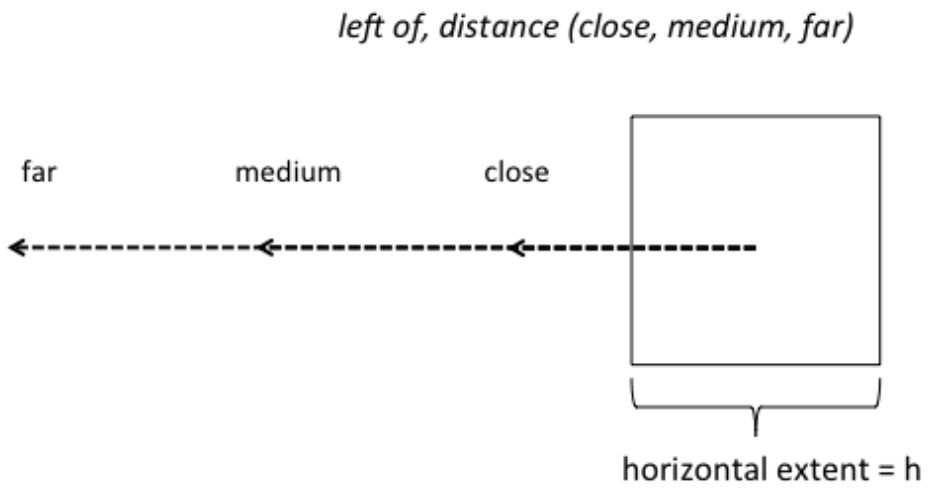


Figure 5.2: The mapping of spatial relations between objects onto attention shifts. The figure depicts how the selection of attention shifts for spatial relations depends on the distance and the extent of the imagined shape of the reference object. How the resulting vectors are computed is explained in Section 5.1.3.

Case 3: a spatial relation between a part and its father-object. This case is restricted to mental concepts describing a spatial relation between a part and its father-object. If the part is inside the object, the normalized direction vector is multiplied by the factor for the provided distance divided by 6 and multiplied by the respective extent of the shape of the father-object. If the part is outside the object, the vector is calculated as in the inside-case and then added to the shape's extent (along the direction of the spatial relation) multiplied by a scaling factor that is set to 0.5 by default[3]. Figure 5.3 depicts the results of these calculations.

Mapping Shapes Onto Attention Shifts The attention shifts associated with a shape are so that the vectors implementing the attention shifts describe the respective shape as a polygon. For example, the set of attention

[2]The purpose of this factor is to allow adjustments to how much the imagined shape of the reference object influences the selection of an attention shift for spatial relations.

[3]As above, this factor serves the purpose of adjustment.

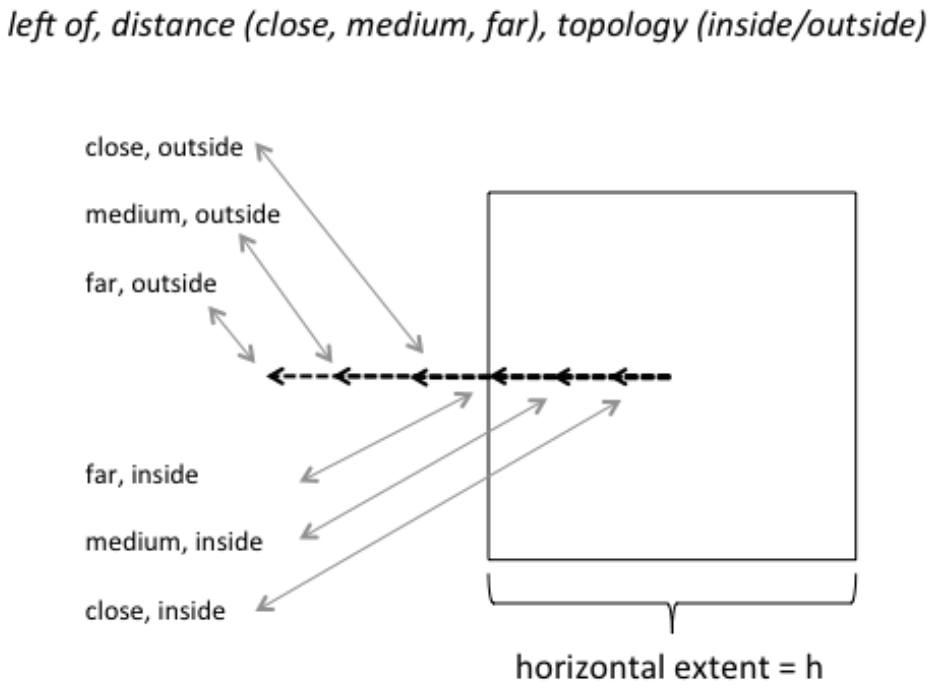


Figure 5.3: The mapping of spatial relations between objects and their parts onto attention shifts. The figure depicts how the selection of attention shifts for a spatial relation that describes the location of a part relative to its father-object depends on distance, topology, and the extent of the imagined shape of the reference object. How the resulting vectors are computed is explained in Section 5.1.3.

shifts for *square* would be vectors that describe the four orthogonal edges that form a square. The actual perceptual actions that the human visual system uses to perceive different shapes are not fully known and thus this simplification is used for practical purposes.

The vectors are normalized so that the recognition of shapes is size-invariant. By using an appropriate scaling factor the vectors can be matched to a shape of different sizes. Three different sizes in which shapes can be imagined are distinguished in the model: *small*, *medium*, and *large*. The sizes are implemented as factors: 1, 2, and 3. In order to retrieve a set of vectors corresponding to the shape of an object, the object and optionally the to-be-imagined size are provided to the function **select**. If a size is provided, its respective factor is multiplied with each of the vectors. If no size is provided, the default size of *small* is used. The resulting vectors are provided to the function **execute** of the visual/motor system.

The Function Identify

The function **identify** realizes the procedural knowledge of the VS-LTM which maps perceptual information onto mental concepts. This way spatial relations and shapes are identified based on the perceptual information made available by executing perceptual actions. The following discusses this function for spatial relations and shapes separately.

Mapping Perceptual Information Onto Spatial Relations An attention shift, for example, made between the locations of entities in a mental image, yields perceptual information in the form of a vector. The mapping of that vector onto a fitting spatial relation is done as follows. First, that direction vector that is closest to the given vector is identified. This is realized by comparing the angle between each direction vector and the given vector. The direction vector closest in angle is chosen and determines the direction of the returned spatial relation. If additional information, such as the extent of the reference object or the fact that the to-be-inferred spatial relation holds between a part and its father-object, is available, it is used for the calculation of the distance of the spatial relation. For this, the given vector is projected onto the already identified direction vector and afterwards the calculations as already described above for the function **select** are simply reversed. The resulting distance is then projected onto one of the three mental concepts of distances (*small*, *medium*, *large*) based on which it is closest to. The resulting spatial relation containing direction, distance, and optionally topology is returned.

Mapping Perceptual Information Onto Shapes The mapping of perceptual information of a shape onto a fitting set of mental concepts is done as follows. The perceptual information of a shape is represented by a set of line segments, i.e., pairs of coordinates. These pairs of coordinates are transformed to vectors by simply subtracting the coordinates of each pair. The resulting vectors now correspond to the representation of attention shifts which the mental concepts describing shapes are linked to. The given vectors are compared to these sets of attention shifts to find a full or partial match. The mental concepts representing objects whose shape is a full or partial match are returned together with the sizes of the matching shapes.

Full match: A full match is found when the given vectors correspond to a set of vectors in VS-LTM with some constant scaling factor that is applied to all vectors. This scaling factor is then projected onto one of the mental concepts of size based on which it is closest to. The mental concepts of shape and size are returned.

Partial match: A partial match is found when the given vectors correspond to a subset of a set of vectors in VS-LTM. A constant scaling

factor is again projected onto one of the sizes. The mental concepts of shape and size are returned.

There can be several matches so that one set of vectors correspond to several shapes at once. For a trivial example, consider a square, which is simultaneously a square, a rectangle, a parallelogram, and a schematized "U", i.e., "⊔". The "⊔" is an example for a partial match.

The Function Execute

The function **execute** takes as input a set of perceptual actions. The model distinguishes between attention shifts that instantiate a spatial relation and those that instantiate shape information. Spatial relations are instantiated by a single attention shift while shape information is instantiated by a set of more than one attention shift. The set of attention shifts which are the input to the function **execute** are either provided by the VS-LTM, i.e., the function **select** selects a set of attention shifts for a mental concept, or attention shifts are derived from the already instantiated perceptual information when inferring new information from a mental image. In the following, these different cases are discussed separately.

Instantiating a spatial relation: An attention shift is retrieved from the function **select** of the VS-LTM. The attention shift is executed by simply adding the vector (which represent the attention shift) to the coordinate of the current **focus**. The resulting perceptual information is the new coordinate of **focus** and the relative change of **focus**. This perceptual information constitutes a location (the new focus) and a spatial relation (the change of focus).

Instantiating a shape: A set of attention shifts is retrieved from the function **select** of the VS-LTM. These attention shifts are executed by successively adding the vectors to the coordinates of **focus**. The perceptual information that results are line segments derived from the different coordinates that **focus** was changed to, the location, and the extent of the imagined shape (which is derived from the set of line segments). This perceptual information constitutes an instance of a shape.

Inferring new information: The difference between the above cases and the case of inference is that the input for the inference is not provided by the VS-LTM but directly from the perceptual information that is already available in STM. For inference no mental concept needs to be mapped onto attention shifts, but perceptual information such as two already instantiated locations are used to directly execute an attention shift between them. By subtracting the coordinates of two locations a vector representing an attention shift results. This attention shift is executed and yields perceptual information, i.e., the change of the coordinate of **focus**. This perceptual information is used by the function **identify** of the VS-LTM and mapped onto a set of mental concepts. For the case of a single attention shift be-

tween two locations, the identified mental concepts will be spatial relations containing direction, distance, and topology.

Similarly, instead of inferring a spatial relation, this process can be based on already instantiated shape information, i.e., a set of line segments represented as a set of pairs of coordinates. Pairs of coordinates are transformed to vectors by subtraction. These vectors can be executed as attention shifts and the resulting perceptual information can be identified as a different type of shape.

Time of execution: The time of executing perceptual actions is implemented to be linear to the length of the vectors that represent the perceptual actions. This approximates the time constraints of the human visual/motor system as discussed in Section 3.2.4.

Overt vs covert attention shifts: Attention shifts will be executed overtly if their respective vector would shift the **focus** beyond the distance given by the **radius**. Otherwise it will be executed covertly. This behavior is in analogy to our ability to shift attention within the periphery of our gaze during a fixation in visual perception.

5.2 Examples

The following two examples serve the purpose of giving an easy and slightly simplified overview of how the model realizes mental imagery. In order to fully understand how the model works it is recommended to read Section 5.1. However, for a superficial understanding of the model the following two examples are sufficient.

5.2.1 Generating a Mental Image

Figure 5.4 depicts how the model generates a mental image of a scene from memory. After the task has been given, the conceptual description of the to-be-imagined scene is retrieved from C-LTM. The C-LTM contains scenes which each consist of a set of objects and a set of spatial relations. As shown in the figure, an object can further have properties, e.g., qualitative size, and link to other objects.

In the depicted case, the STM retrieves a scene consisting of two objects and one spatial relation. The object *triangle* has the property *small*. The mental concepts *triangle* and *small* are instantiated with a set of perceptual information. The first step is the selection of perceptual actions that correspond to the perception of a *small triangle*. This step is realized by the function **select** of the visuo-spatial long-term memory (VS-LTM). *Triangle* is mapped onto a set of vectors that represent attention shifts. The vectors are adjusted with the factor α_1 which represents the property *small*. The resulting vectors are used by the function **execute** of the visual/motor system. The model realizes the execution of attention shifts by successive

addition of the vectors to a coordinate (which represents the focus of attention). The function **execute** makes a set of line segments (i.e., pairs of coordinates) available. These represent the perceptual information that forms an instance of the shape of *small triangle*. From this shape information further perceptual information is derived such as the location and the extent of the imagined shape.

As a next step, the spatial relation *left of* is instantiated. For this step, the already available perceptual information of the *small triangle* is considered so that the concrete instantiation of *left of* is realized within this context. Concretely, the imagined extent of *triangle* is taken into account by the function **select** of the VS-LTM when it selects one of many attention shifts that correspond to the perception of *left of*. In the depicted case, the attention shift is determined by a prototypical attention shift and the horizontal extent of *triangle* as well as a scaling factor β_1. The attention shift is executed by the visual/motor system and yields as perceptual information a location of *square* and an instance of *left of*.

The next step, which is not depicted in the figure anymore, would be the instantiation of the shape of *square*.

5.2.2 Inferring Information in a Mental Image

Figure 5.5 depicts how the model simulates the task of inferring information from a mental image. The model was given the two premises "A left of B" and "C right of B". These two premises have already been imagined so that the locations of the three objects A, B, and C are instantiated. The objects do not contain a conceptual description of their shape and accordingly no shape information was instantiated; instead the objects are abstracted to points which equal their location. The model is then queried for the spatial relation between A and C. This spatial relation was not given and is therefore inferred from the already available perceptual information, i.e., in this case the locations of A and C. An attention shift is executed by the visual/motor system based on the two available locations. This attention shift yields perceptual information which is identified by the VS-LTM. That is, the function **identify** maps it onto a set of mental concepts, in this case the spatial relation consisting of the two mental concepts *left of* and *far*. These inferred mental concepts are used to answer the question.

5.3 Notes on Implementations of PIT

5.3.1 Modeling Approaches

This chapter has presented one possible implementation of PIT. However, the framework of PIT can be implemented using a wide range of different modeling approaches as well as combinations of such approaches.

For example, qualitative spatial calculi (e.g., Wallgrün, Frommberger, Wolter, Dylla, & Freksa, 2007; Freksa, 1991) could plausibly be used to model the different types of mental concepts and their representation in more detail. The calculi of qualitative spatial reasoning aim to imitate the qualitative representations used by humans by employing qualitative descriptions of spatial situations including direction, orientation, topology, and mereology.

Furthermore, probabilistic approaches such as bayesian modeling could plausibly be used to model, for example, the background knowledge used by the VS-LTM. That is, modeling the acquired information of how likely it is to perceive one mental concept compared to another in a given situation. Such statistical information would also allow to determine which perceptual action will maximize information gain.

5.3.2 Problematic Aspects

There are two problematic aspects to consider when implementing PIT. These are discussed in the following.

Visual Perception

Research on visual perception is limited in its knowledge about what low-level actions the human organism employs during visual perception, how theses actions are concretely realized, and what purposes they serve exactly. Similarly, it is an open question what exactly the low-level perceptual features are that the human visual system can process and how exactly it processes such features. For any computational implementation of the framework of PIT, it is thus necessary to make several assumptions and simplifications with respect to the perceptual information, perceptual actions, and mental concepts used for visual perception and mental imagery.

Background Knowledge

It is without doubt that human cognition and major aspects of human cognition such as (visual) perception and mental imagery can only be understood with respect to both the current context that an organism is in and the history of that organism (i.e., the acquired procedural and conceptual knowledge). This means, that any model of mental imagery has to either be able to acquire such background knowledge from autonomous interaction with the environment or this knowledge has to be modeled and put into the model by the designer. The latter option practically means that the background knowledge will be highly simplified. The option of letting a model of human cognition learn such knowledge autonomously in a non-trivial environment poses a hard problem in itself.

In particular, the function **interpret** of the framework of PIT strongly depends on such background knowledge. The interpretation of ambiguous perceptual information corresponds to the problem of categorical perception. Categorial perception, i.e., the ability to see and recognize entities as what they are, is considered an AI-complete problem in the strong-AI community (Shapiro, 1992). This means that the difficulty of this problem is considered equivalent to that of achieving strong AI.

An approach to investigate these issues in a simple artificial domain is discussed in Section 7.2.1.

5.4 Summary

This chapter has presented a computational implementation of PIT. This implementation together with the framework of Chapter 4 serve the purpose of making the concepts and ideas described in Chapter 3 more clear. The computational model also serves as a proof-of-concept showing that PIT is sufficiently complete and consistent as a theory so that implemented instances of it are possible. Lastly, the computational model can run simulations of mental imagery tasks which offer concrete mechanistic explanations of phenomena as well as specific predictions. In Chapter 6 both the theory and the computational model will be applied to show how PIT can account for the different phenomena of mental imagery that were previously presented in Chapter 2.

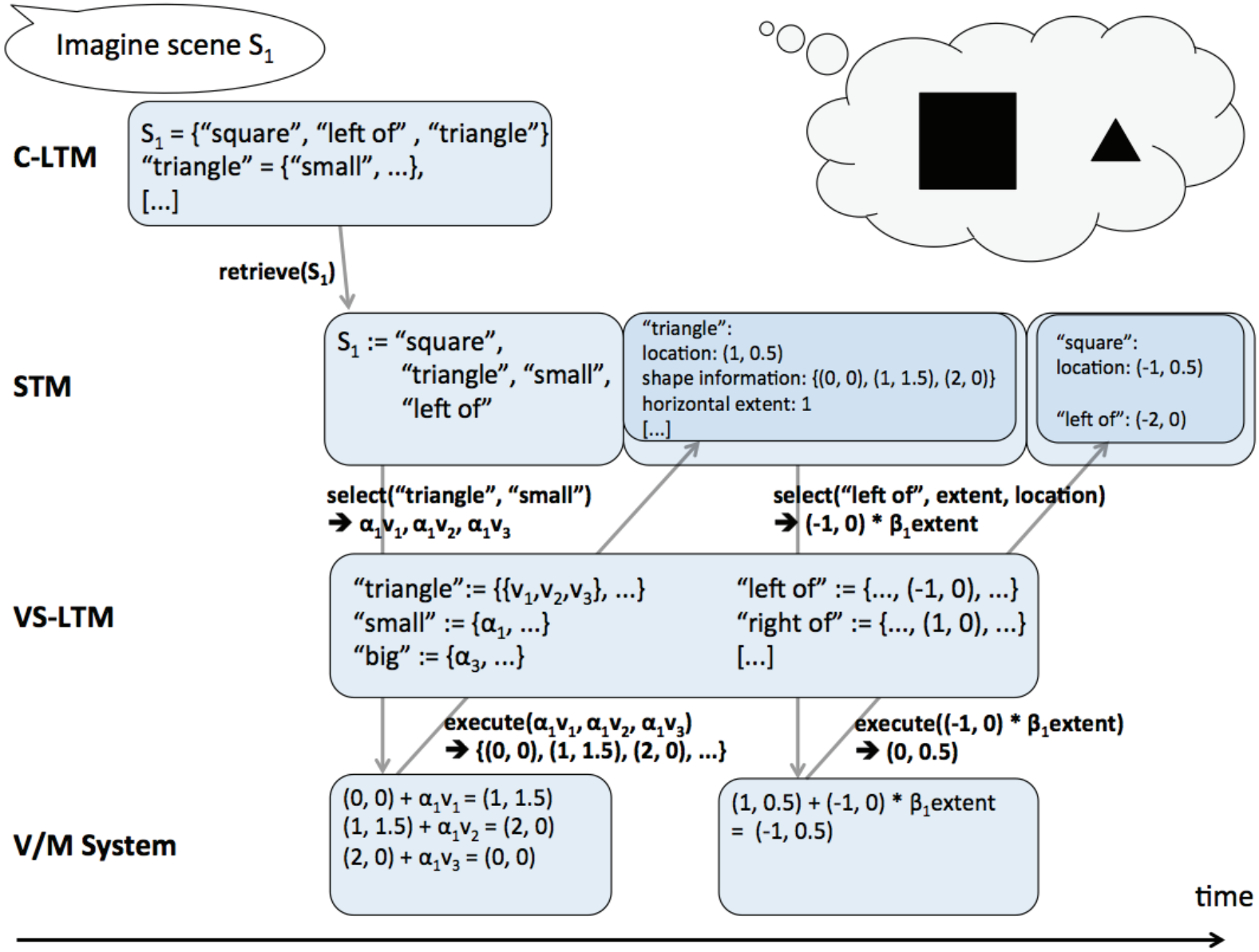


Figure 5.4: Generation of a mental image. The figure depicts how the model simulates the generation of a mental image of a scene. The to-be-imagined scene is shown in the upper right corner of the figure. The workflow follows the arrows from left to right. Details are given in the text.

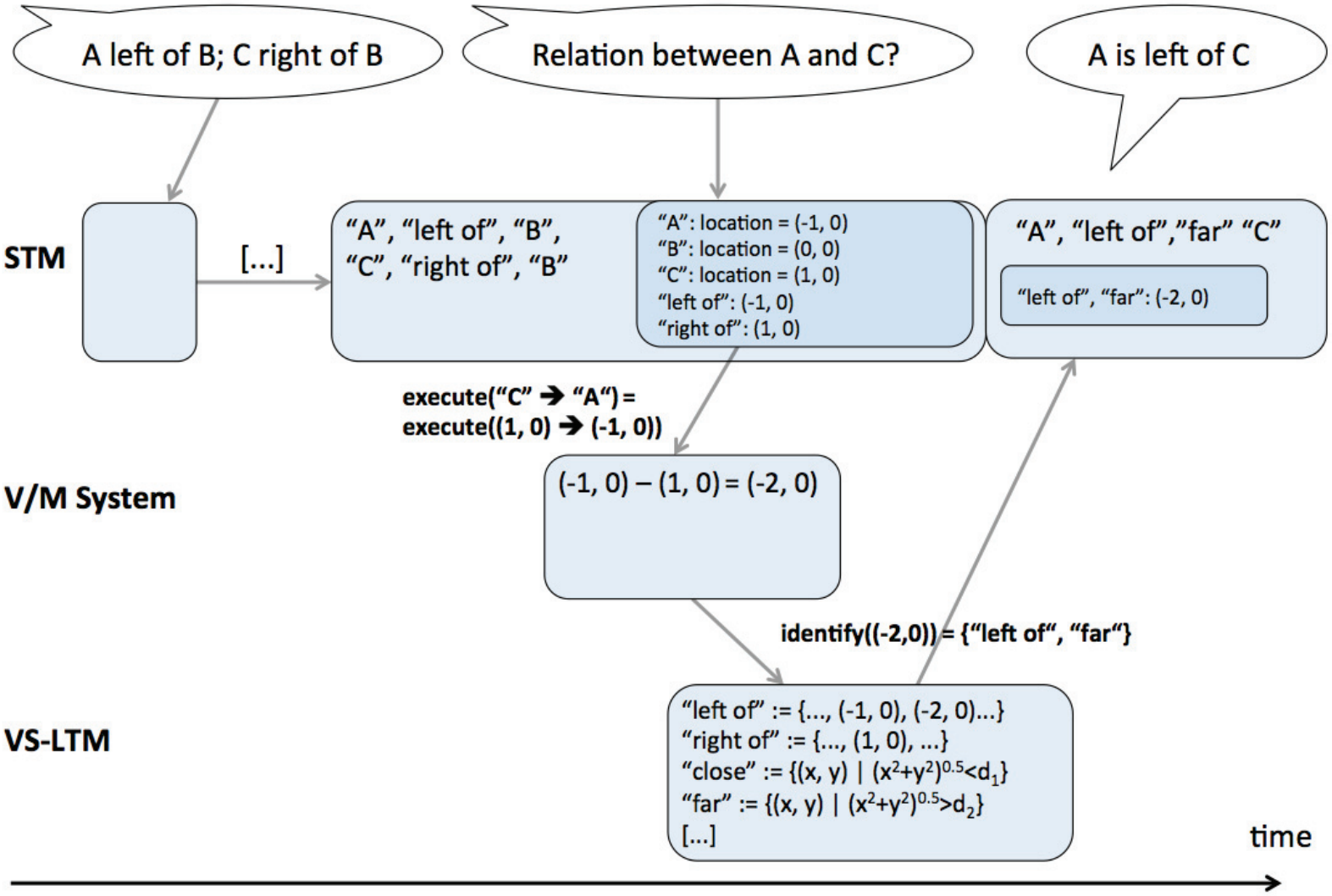


Figure 5.5: Inference in a mental image. The figure depicts how the model simulates the inference of new information from a mental image. The workflow follows the arrows from left to right. Details are given in the text.

Chapter 6

Evaluation

This chapter presents the evaluation of the perceptual instantiation theory (PIT) and its computational model. The previously identified and discussed empirical phenomena of Chapter 2 are picked up in this chapter and the respective explanations and predictions of PIT are elaborated. The phenomena of mental scanning and eye movements are explained with support of simulations of the computational model while the explanations for mental reinterpretation and unilateral neglect are based on the theoretical description of PIT.

6.1 Mental Scanning

The relevant empirical findings on mental scanning can be summarized as follows: 1) there is a robust mental scanning effect and 2) the specific parameters of that effect, specifically the slope of the linear relation between reaction time and distance, are significantly affected by a number of different factors. I will first discuss the general mental scanning effect and then how it can be influenced.

6.1.1 The General Mental Scanning Effect

The mental scanning effect is the finding that humans show an approximately linear relationship between the time of shifting their attention from one point to another in a mental image and the distance between these two points. The compared distance is the distance in the original stimulus, because distance in a mental image cannot itself be measured. PIT and its model explain this effect with what I will term the **equivalence explanation**[1]. PIT poses that mental imagery employs the same perceptual processes used during visual perception. Therefore, similar reaction time

[1]The enactive theory uses the same equivalence explanation to explain the general mental scanning effect (see Section 2.3.1).

Figure 6.1: Two stimuli depicting islands as used in mental scanning experiments. The island on the right additionally contains sign posts indicating inconsistent distances (Richman et al., 1979). Compare Figure 2.7.

patterns as observed during visual perception will also be observed during visuo-spatial mental imagery. During visual perception a shift in attention - whether it is realized as, for example, a saccade or a head movement - shows the property that shifting over a longer distance generally takes longer than shifting over a shorter distance. This property results simply from the physical structure of the human body, e.g., one's gaze cannot go from A to B without going through the intermediate space (also see Section 3.2.4 for a discussion on this constraint for perception and imagery). Therefore, PIT explains the general mental scanning effect trivially due to the common employment of the same perceptual processes as in visual perception.

The model of PIT reflects this explanation in more detail. A remembered stimulus is encoded in a scene which consists of the conceptual description of the objects and the spatial relations between them. The spatial relations of the scene reflect the remembered distances. For the generation and processing of the mental image of that scene, the respective mental concepts are instantiated with perceptual information. For this instantiation the mental concepts describing, for example, the spatial relation *left-of* and *close* are mapped onto a fitting attention shift implemented in the model as a vector. The length of the vector corresponds to the distance qualitatively described by the spatial relation. The time for executing an attention shift is linear to the length of the vector that represents that attention shift. It follows that a spatial relation describing a longer distance will be instantiated using an attention shift over a longer distance. And that the execution of that attention shift will take proportionally longer the longer the conceptually described distance is. Table 6.1 shows the model's output for a mental scanning task using the left stimulus of Figure 6.1.

6.1.2 Variations of Mental Scanning

There are several different variations of the mental scanning paradigm which all show the general mental scanning effect. Yet, many variations significantly affect the slope of the linear relation between distance and time. That

Table 6.1: The model's reaction times (RT) are averaged over 10 trials and include noise. Correlation: $r = 0.94$

Scan path	RT Model	Actual Relative Distance
house → tree	29.61	4.47
house → well	31.85	4.47
house → lake	12.14	2
lake → tree	27.67	4
lake → well	42.34	5.67
tree → well	35.49	4

Table 6.2: Mental scanning (Richman et al., 1979). Reaction times (RT) of the model are averages over ten trials and include noise.

Condition	RT Experiment [s]	RT Model
20 route	3.118	25.36
80 route	3.496	35.03

is, the speed of mental scanning is affected. I will only focus on one of those variations here, because the explanation that PIT gives for this specific case similarly applies to all other variations as well. The right side of Figure 6.1 shows a stimulus similar to the one used by Richman et al. (1979) for their mental scanning experiment. In contrast to the original mental scanning paradigm, the stimulus additionally contains two sign posts indicating certain distances between entities. These distances are obviously inconsistent with the actual distances in the stimulus. Specifically, the distance between the hut and the tree is of the same length as that between the hut and the well. The sign posts, however, indicate that the distance between the hut and the tree is much longer, i.e., 80 miles, than the distance between the hut and the well, i.e., 20 miles. Participants were asked to mentally scan several routes using their mental image of the stimulus. A reliable effect of the sign posts on the reaction times of scanning was found so that participants took longer to scan along the "80 miles" route than along the "20 miles" route. This experiment is an example of how cognitive penetration affects mental imagery. That is, the reaction times of the mental scanning task were significantly affected by additional knowledge or belief of the participant; in this case the suggested, albeit incorrect, distances. Other variations of mental scanning can be seen as similar, as they also vary the participants's belief or knowledge. For example, participants are made belief that mentally scanning over certain distances takes a certain time (Goldston et al., 1985) or

more indirectly, the experimenters are made belief that certain mental scanning outcomes are to be expected (Intons-Peterson, 1983). In both these cases the induced belief affects the general mental scanning effect in the expected way.

PIT's explanation for the experiment of Richman et al. (1979) is based on the theory's assumption that the mental concepts underlying mental imagery are the result of the integration of multi-modal input. That is, they combine the sensory input of several senses including rather subtle information such as the suggested distances in the stimulus or different demand characteristics. For this reason, the conceptual description of the stimulus does not only reflect the metrical properties of the stimulus but also integrates the semantics of the sign posts suggesting different distances. That is, the spatial relation between, for example, the hut and the tree is not just *top-left-of* as it might be for the same mental scanning stimulus without sign posts, but rather *top-left-of, far* with *far* corresponding to the suggested "80 miles" distance. As elaborated above for the general mental scanning effect, such changes in the conceptual description lead to respective changes in reaction time. In this case the scanning time increases for the "80 miles" route and decreases for the "20 miles" route. Table 6.2 shows the resulting reaction times of the model for a simulation of the mental scanning variation of Richman et al. (1979).

6.1.3 Predictions

PIT makes two clear and testable predictions with respect to mental scanning. The first prediction results from the **equivalence explanation** for mental scanning posed by both the enactive theory and PIT. This prediction has, however, not been mentioned by Thomas (1999) or in other publications on the enactive theory. The second prediction results from the fact that the mental concepts of PIT are the result of the integration of different inputs. These two predictions are discussed in the following.

Both the enactive theory and PIT assume that (visual) perception comprises of perceptual actions which are also employed during mental imagery. The linear relation between distance and time in mental scanning is explained by the fact that shifting attention in perception also has this property. Actually, shifting attention using, for example, saccades does not show a strictly linear relationship between the time and the to-be-shifted-over distances. This is because other effects such as saccades reaching a higher acceleration and velocity over longer distances and, additionally, varying corrective movements for "overshooting" of saccadic eye movements also play a role. The prediction, made by both the enactive theory and PIT, is that the observed relationship between time and distance in mental scanning should upon closer observation show a stronger correlation to the (non-linear) one of attention shifts as employed in visual perception rather than to a strictly

linear one.

The second prediction that follows from PIT's explanation of the variations of mental scanning is that the scanning time is determined not only by the properties of the stimulus but, furthermore, by any other sort of related and even conflicting information obtained through other modalities. That is, if additional information is systematically varied in both content and mode of communication, PIT predicts that the scanning time will be affected according to that additional information. PIT poses that the mental concepts on which mental images are based are multi-modal and integrative so that they combine related input and subsequently the outcome of the instantiation of these mental concepts will change. It should be expected that conflict resolution, i.e., the process of mapping conflicting spatial relations onto one set of perceptual actions during mental imagery, requires additional time.

A prediction about eye movements during mental scanning is described in Section 6.3.4 which discusses predictions of PIT with respect to eye movements during mental imagery.

6.2 Mental Reinterpretation

The empirical findings on mental reinterpretation that were identified in Chapter 2 can be summarized as follows. First, the literature reports that there are two classes of stimuli that differ in their difficulty of mental reinterpretation. On the one hand, ambiguous drawings like the duck-rabbit have been shown to be very hard to mentally reinterpret while, on the other hand, simpler stimuli based on elementary geometrical shapes including alphanumeric characters have been shown to be comparatively easy to mentally reinterpret. The second major finding is that there are several factors that significantly improve performance in the mental reinterpretation of stimuli that are otherwise hard to mentally reinterpret.

6.2.1 Differences Between Stimuli of Mental Reinterpretation

I will first elaborate why stimuli like the duck-rabbit are very hard to mentally reinterpret as a mental image and then apply this explanation to the finding that other types of stimuli are easier to mentally reinterpret.

Stimuli That are Difficult to Mentally Reinterpret

Mental reinterpretation generally poses the question why the same ambiguous stimuli are very easy to reinterpret in visual perception while they are hard to mentally reinterpret using mental images. I will use the duck-rabbit as a representative example for those stimuli whose mental reinterpretation

has been shown to be hard. The duck-rabbit (depicted in Figure 6.2) has been used in almost all considered studies on mental reinterpretation and is, furthermore, very similar to other stimuli such as the goose-hawk or the chef-dog (see Section 2.1.2 for an overview of different ambiguous stimuli). The hardness of mental reinterpretation strongly indicates that there must be a critical difference between visual perception and mental imagery with respect to (re-)interpretation. Section 3.2.6 outlined differences between visual perception and mental imagery based on the assumptions of PIT. PIT poses that the key difference with respect to mental reinterpretation is the fact that in visual perception we are able to draw an interpretation basically "from scratch", i.e., with little bias towards a (previous) interpretation. In mental imagery, in contrast, the process of reinterpreting an imagined stimulus requires a mental image of that stimulus. That is, before a mental image is inspected, it needs to be generated. The representation of a mental image, however, corresponds to an interpretation drawn from the set of all mental concepts and their instantiation of perceptual information (see Section 3.2.6). This interpretation will include the mental concepts of the conceptual description retrieved from conceptual long-term memory. The reason these mental concepts are included in the initial interpretation is simply that they are the conceptual description of what is to be imagined. To put it simply, generating a mental image of the duck-rabbit stimulus that was recognized as a duck, will lead to a mental image with the interpretation "duck" and the respective mental concepts describing the parts of a duck. That is, before the mental image could be potentially reinterpreted as "rabbit", it is necessarily imagined as "duck".

In order to find an alternative interpretation, a set of mental concepts has to be identified from the perceptual information so that these mental concepts could form a coherent alternative interpretation. However, in mental imagery, the perceptual information is not taken from the (ambiguous) real-life stimulus but it is generated as instances of those mental concepts which have been retrieved from conceptual long-term memory. Therefore, the perceptual information available "fits" the initial interpretation. Because mental images are based on abstracted mental concepts, the generated perceptual information will not exactly resemble that of the original stimulus but will rather be prototypical for the given mental concepts. This point is supported by an experiment reported in (Chambers & Reisberg, 1992). They briefly showed participants the duck-rabbit so that only one interpretation of it was recognized. Participants then compared their mental image to pictures of slightly modified duck-rabbits. The modifications were made to parts of the duck-rabbit which are only relevant for one of the two interpretations, e.g., removal of the mouth of the rabbit and changes to the beak of the duck. Participants were less likely to notice changes to the original stimulus that are irrelevant to their interpretation and more likely to notice changes relevant to their interpretation. These results support the

assumption that a mental image is formed specific to one's initial interpretation and might even lack some of the details of the original stimulus that would allow a successful mental reinterpretation.

Summarizing, in mental imagery, a mental (re-)interpretation has a strong bias towards the initial interpretation of the mental image. This bias has two reasons. The first reason is the fact that there already is an initial interpretation which would have to be "overwritten" by an alternative interpretation. And the second reason is that the perceptual information from which (alternative) mental concepts can be identified has been generated to specifically fit the mental concepts of the initial representation.

This bias can explain why mental reinterpretation is generally hard. It is worth repeating that mental reinterpretation without any sort of hints and with visually presented stimuli has been shown to be very hard. For example, no participant managed to find the second interpretation of either the duck-rabbit, the Necker cube, or the Schröder staircase (all are depicted in Figure 6.2) in the original experiments of Chambers and Reisberg (1985). Slezak (1995) reports similar results for a variety of different ambiguous stimuli, e.g., requiring rotation, figure/ground reversal, the Kanizsa Illusion[2], which almost none of the participants could mentally reinterpret. Additionally, Reisberg and Chambers (1991) report a series of experiments using different types of ambiguous stimuli that again were not successfully reinterpreted by almost all participants except when hints were given.

Stimuli That are Easy to Mentally Reinterpret

But there are also stimuli for which successful mental reinterpretation even without hints has been reported. Slezak (1995) used "mirrored number" stimuli and Finke et al. (1989) used a variety of stimuli made up of simple geometric shapes and alphanumeric characters. Examples of both of these stimuli are depicted in Figure 2.3 of Chapter 2. For these types of stimuli often a majority of participants was able to successfully reinterpret them mentally. As discussed for the difficult stimuli, in PIT a mental image always has an initial interpretation. Therefore, in order to mentally reinterpret a mental image this initial interpretation has to be replaced with a new interpretation. That is, the new interpretation will have to provide a more plausible description of the stimulus than the initial interpretation does. I will now discuss to which extent this explanation is consistent with these reinterpretable stimuli.

Observing the stimuli in Figure 2.3, it can be plausibly claimed that these stimuli either do not really represent anything meaningful beyond their very shapes or they do so in a highly schematized way. For example, the "mirrored 2" could at best be described as a "heart on a plate"; an image that

[2]These stimuli require the combination of the contours of parts of the image to form a new emerging shape.

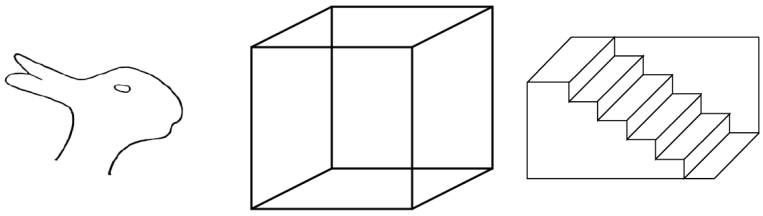


Figure 6.2: The duck-rabbit, the Necker cube and the Schroeder staircase. The duck-rabbit can be both seen as a duck or a rabbit. The Necker cube and the Schroeder staircase are ambiguous with respect to which part is interpreted to be in front and which part to be in the back. For the Necker cube either the lower left side or the upper right side of the cube can be seen as being in the front. The Schroeder staircase can be interpreted so that the lower left part or the upper right part is extending towards the observer.

seems rather odd and unfamiliar. The point to make here is that these stimuli, in contrast to even a simple drawing like the duck-rabbit, seem much less realistic and meaningful, or simply less plausible to depict something familiar. There are studies which support this point that concrete "meaning" of stimuli plays a critical role in mental reinterpretation: Brandimonte, Hitch, and Bishop (1992a, 1992b) have shown that figures that are easy to name are more difficult to mentally reinterpret than figures that are difficult to name.

Another difference that can be assumed between these stimuli and the duck-rabbit is that the latter is much more likely to be conceptually represented as a composition of parts (in this case natural to an animal) such as *ears*, *head*, *eyes*, and *nose* with the respective spatial relations between them. Whereas, the "heart on a plate" offers perhaps two parts and alphanumeric characters or simple geometric shapes might be represented holistically as consisting of just a single part[3]. The same observation applies also to the to-be-discovered second meaning of those stimuli. The new interpretations would similarly be conceptually represented by rather few mental concepts. For the mirrored numbers, the new interpretation is in fact simpler than the initial interpretation as only one half of the stimulus is considered and the new interpretation can be conceptually described trivially as "2". The complexity of a previous and new interpretation is likely to affect mental reinterpretation. Concretely, it requires less effort to replace a trivial interpretation consisting of only very few mental concepts and at the same time it is of less effort to form a new interpretation that is of very low complexity, because it requires only very few "new" mental concepts to be identified.

[3]Such an assumed holistic representation of letters, numbers, and simple shapes might very well be due to our strong familiarity and daily exposure to them as suggested by Thomas (1999). Such a specific representation of letters and numbers is further supported by the neuropsychological findings of selective neglect of letters (Goldenberg, 1993).

The fact that the stimuli of Finke et al. (1989) were presented verbally and not as usual in such studies visually likely adds to the success of their mental reinterpretation. PIT assumes that the mental concepts underlying mental imagery are the product of the integration of all modalities, which means that in this case the mental concepts are derived from verbal input only and such a verbal description is naturally much less restricting in terms of concreteness and details than a visual presentation[4]. Consequently, the plausibility of or converging evidence for a verbally given interpretation is less strong than that derived from a more detailed visual presentation given that everything else remains equal.

Summarizing, the stimuli (including both interpretations) that have been shown to be comparatively easy to mentally reinterpret would be 1) conceptually represented very simply, i.e., one to very few parts and spatial relations, 2) their resemblance to real objects is weak or non-existent, and 3) in case of a verbal presentation much less detailed and settled than for a visual presentation. All these aspects decrease the plausibility of the initial interpretation of these stimuli. The less plausible a current interpretation is, the more likely it becomes to find a more plausible alternative interpretation, i.e., successfully mentally reinterpret the mental image.

6.2.2 Why Mental Reinterpretation can be Improved

Section 2.1.1 reported the different factors that have been shown to significantly improve mental reinterpretation of stimuli that are otherwise hard to mentally reinterpret such as the duck-rabbit. These factors can be divided into four groups: 1) explicit hints about reference frame manipulation and identity of the to-be-discovered meaning, 2) training stimuli with the same reference frame reversals, 3) partitioning of the stimulus during presentation, and 4) articulatory suppression during presentation of the stimulus. Each of these four types of factors are discussed in the following.

Several studies reported that mental reinterpretation of ambiguous stimuli such as the duck-rabbit improve significantly when hints are provided during reinterpretation (e.g., Reisberg & Chambers, 1991; Hyman & Neisser, 1991). Such hints include: 1) hints about what to "see", e.g., telling participants that they are looking for an animal, and 2) hints about the alternative reference-frame, e.g., "the front of the rabbit could be the back of another animal" or "the left side is the new top". According to PIT's explanation of why the duck-rabbit is hard to mentally reinterpret, these hints should help mental reinterpretation because they specify that (and to some extent how) the current interpretation should be discarded or altered. Reisberg and Chambers (1991) report experiments in which participants were presented drawings which were rotated versions of meaningful stimuli such as

[4]The saying that "a picture is worth a 1000 words" seems to apply here.

the shape of Texas. They discovered that the instruction to mentally rotate the stimulus did not lead any of the participants to discover the shape of Texas in their mental image. But the explicit hint to understand the left side of their mental image as the new top, in fact, led to successful mental reinterpretation for more than half of the participants. This striking result supports PIT's explanation that hints are successful because they explicitly induce a re-structuring of the conceptual description of the current interpretation of the mental image. Whereas only mental rotation does not induce such re-structuring but only changes the orientation while keeping the current interpretation.

It has also been shown that participants that have been provided with training examples of ambiguous drawings which include the same reference-frame reversals as the later presented duck-rabbit show a significant increase in successful mental reinterpretation (Peterson et al., 1992). Again, such training likely increases the propensity of participants to discard the conceptual description of the current interpretation thus aiding the ability to draw an alternative interpretation.

Peterson et al. (1992), furthermore, showed that a partitioning of the ambiguous stimuli significantly increases the success of mental reinterpretation. In this study the ambiguous stimulus was presented in a piecemeal fashion so that participants had to mentally "glue" the presented parts together to get the complete stimulus. The fact that participants never perceived the full stimulus could likely lead to a lower plausibility of their interpretation than had they seen the full picture. This argument is similar to the one made previously about the study of Finke et al. (1989) who presented their stimuli verbally and not visually. In both cases the resulting initial interpretation of the stimulus should be less fleshed-out and should contain less fixed visual details than if one visually perceives the stimulus as a whole. This aspect should negatively affect the current interpretation's plausibility and thus increase the likelihood of finding a new more plausible interpretation.

Brandimonte and Gerbino (1993) showed that the mental reinterpretation of the duck-rabbit significantly improves when participants are instructed to loudly say "lalala" during the initial presentation of the duck-rabbit. This procedure is termed articulatory suppression. This finding can be explained because the mental concepts which underlie mental images integrate multi-modal input. In the case of this study, the participants essentially link nonsense verbal input to the visual input of the duck-rabbit. Given that the conceptual description of the duck-rabbit contains information from both these sources, the overall converging evidence for the found interpretation is decreased as the verbal part simply does not fit with the interpretation "duck" or "rabbit". Given a therefore lower plausibility of this interpretation the likelihood of replacing it with a more plausible interpretation is again increased.

6.2.3 Summary and Predictions

The phenomenon of mental reinterpretation is complex and includes many different aspects such as the different types of stimuli and the different types of hints. The interpretation process (as described in Section 3.1.6 and Section 4.2) is fundamentally involved in the explanation of this phenomenon. Also the interpretation process is at the heart of (categorical) perception and thus a hard problem for which no formal implementation exists (see Section 5.3.2). Accordingly, PIT's explanations for the different aspects of mental reinterpretation have been made on a descriptive level. It is therefore not possible to make predictions as concretely as for mental scanning (Section 6.1) or eye movements (Section 6.3) for which the computational model can be applied directly. This section will thus provide a summary of the identified factors relevant for the success of mental reinterpretation and more general predictions that follow from that.

The explanations of PIT showed why mental reinterpretation is generally very hard unless either specific simple stimuli are used or additional hints and help is provided. The successful mental reinterpretation of a stimulus without additional hints depends mainly on the overall plausibility of the initial interpretation. The plausibility of the initial interpretation determines the success of mental reinterpretation in so far as that discarding the current interpretation (and replacing it with a new interpretation) becomes more likely the less plausible the initial interpretation is. Factors that contribute to the plausibility of an interpretation are: 1) how realistic the stimulus is and 2) how much converging evidence for the current interpretation exists.

The first point is illustrated in Figure 6.3. The duck-rabbit on the left side of the figure is predicted to be harder to mentally reinterpret than the classic duck-rabbit[5]. This is because the duck-rabbit on the left side simply looks much more like an actual rabbit or duck (whatever the initial interpretation might be) and thus the initial interpretation will be more plausible and therefore harder to "overwrite". This example also shows the second point, i.e., converging evidence for an interpretation, as the fur/feathers texture present in the left duck-rabbit supports the initial interpretation. Another way of varying converging evidence for ambiguous stimuli like the duck-rabbit is the presentation of additional information such as presenting sounds or verbal labels that would either support or provide evidence against one of the interpretations. For example, presenting the duck-rabbit stimuli together with a depiction of a pond, sounds made by ducks, or simply the

[5]As already mentioned the classic duck-rabbit has been shown to be very hard to mentally reinterpret. Therefore, a comparison between the mental reinterpretation of the two types of duck-rabbits would need to involve appropriate hints to induce successful mental reinterpretation (see Section 2.1.2 for an overview of such hints).

Figure 6.3: Two version of the duck-rabbit stimulus with different levels of realism.

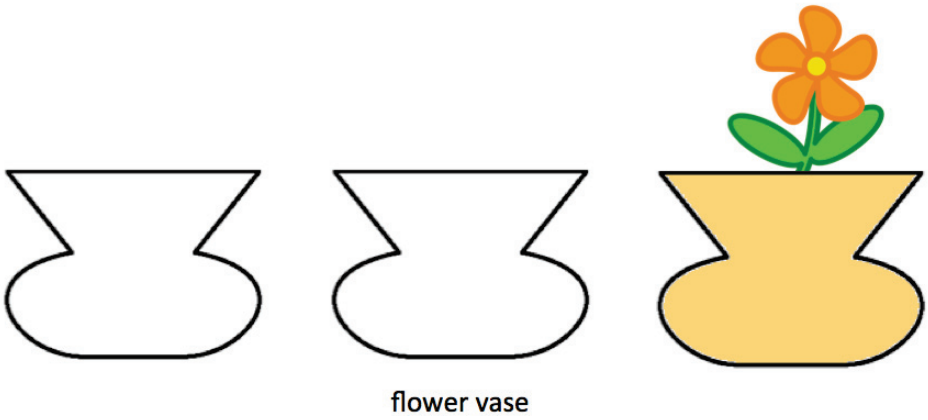


Figure 6.4: Variations of the "mirrored 3" stimulus of (Slezak, 1995). Depicted are two possible variations of the "mirrored 3" stimulus which PIT predicts to be harder to successfully mentally reinterpret, because the initially rather meaningless figure is assigned a concrete meaning (flower vase) or made more realistic by adding details.

subtitle "duck" should lead to decreased rates of mental reinterpretation[6]. Another way of testing the above factors is to vary them for stimuli which have been shown to generally be mentally reinterpretable such as the "mirrored number" stimuli of Slezak (1995). These stimuli could be varied so that their parts resemble actual objects more clearly. Figure 6.4 shows some possibilities to make these stimuli more realistic and less abstract.

[6]Assuming that the initial interpretation is "duck" and the to-be-found interpretation is "rabbit".

6.3 Eye Movements

As reviewed in Chapter 2, the literature reports the robust occurrence of spontaneous eye movements during mental imagery. These eye movements reflect the content of the mental image and have been shown to be functional in mental imagery. In particular, the recall of memories using mental imagery is negatively affected qualitatively and quantitatively when participants have to maintain a fixed gaze. Furthermore, individual differences in the spatial dispersion of such spontaneous eye movements have been reported.

6.3.1 Eye Movements in PIT

The computational model of PIT directly incorporates spontaneous eye movements during mental imagery because saccades are part of the perceptual actions used in visual perception. During mental imagery these same perceptual actions are employed to instantiate the conceptual description a mental image is based on. The model implements a distinction between overt and covert attention shifts. Overt attention shifts are assumed to correspond to, in particular, spontaneous eye movements, whereas covert attention shifts correspond to non-observable attention shifts such as within the periphery of one's gaze. The distinction between overt and covert attention shifts is made based on the length of the vector that represents the attention shift. That is, vectors with a length larger than the a-priori set threshold, will be executed as overt attention shifts, i.e., eye movements. This means that if attention is shifted beyond a certain distance from the current focus of attention, the attention shift will be observable as a spontaneous eye movement.

6.3.2 Functionality of Eye Movements

In PIT, attention shifts are functional for mental imagery; they reflect the currently processed content and their suppression will restrict instantiation and thereby the generation and inspection of the mental image. These properties follow straight-forwardly from the fact that mental imagery is realized by the instantiation of mental concepts and that the process of instantiation is realized by employing perceptual actions such as (overt) attention shifts. If the spatial relation *left-of* is instantiated during mental imagery this could be realized by a respective eye movement which would then also directly reflect the currently imagined spatial relation. If eye movements are suppressed then consequently instantiation is inhibited. This means that the processing of the mental image is inhibited in so far as overt attention shifts cannot be executed. This will restrict the generation and inspection of the mental image. It has been shown that keeping a fixed

gaze during mental imagery produces such inhibitions in recalling content of the mental image independent of how the to-be-imagined stimulus has been presented previously, i.e., verbally or visually. This finding is in line with PIT's assumption that the mental concepts underlying mental images are the result of the integration of all modalities. That is, the instantiation is not directly related to the mode of perception of the to-be-instantiated mental concept.

Another aspect of the inhibition of mental imagery due to keeping a fixed gaze is the fact that not only the amount of information, e.g., the number of recalled entities of the stimulus, is decreased, but, additionally, also the quality of what is recalled, i.e., the type of information, changes when eye movements are inhibited. Johansson, Holsanova, Dewhurst, and Holmqvist (2011) reported that participants would rather recall global and more abstract information about the stimulus such as "it was a living room" or "the walls were colored in blue" when gaze was kept fixed during imagery. In contrast, the descriptions given in the condition in which eyes could move freely rather referred to referents, states and events of the stimulus, e.g., "the man was digging". It is pointed out that the former more global information would also be expected to be perceived during visual perception with a fixed gaze, because it refers to the type of information that can be gathered through a single fixation and the surrounding peripheral information. This exact analogy between (fixed gaze) vision and (fixed gaze) imagery is also found in the computational model. An eye movement (i.e., an overt attention shift) is employed exactly when attention is to be shifted beyond what would be accessible by covert attention shifts. This means, the model could also only instantiate that information that requires no such overt attention shifts in a simulation of a fixed gaze mental imagery task. That information would then naturally be of the kind reported for fixed gaze vision, i.e., rather global and abstract information.

6.3.3 Individual Differences in Eye Movements

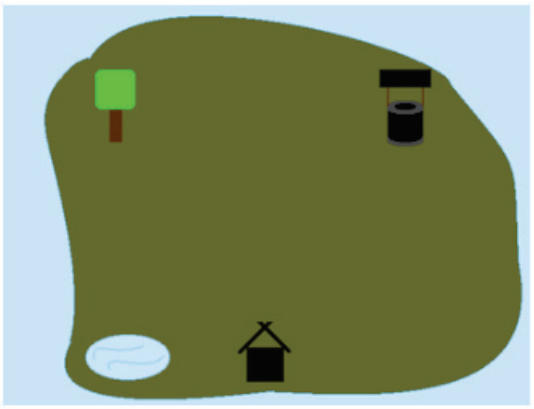
The dispersion of spontaneous eye movements during mental imagery is subject to individual differences and has been linked to the participants' score in the "Object Spatial Imagery and Verbal Questionnaire" (OSIVQ) of Blazhenkova and Kozhevnikov (2009). This questionnaire assesses individual differences in cognitive style with respect to one's ability and preference to use object imagery (i.e., visual mental imagery) and spatial mental imagery. The two scores for object and spatial mental imagery are negatively correlated to each other which indicates a trade-off between the two types of mental imagery (Kozhevnikov, Blazhenkova, & Becker, 2010). Johansson, Holsanova, and Holmqvist (2010) report a negative correlation between the spatial dispersion of eye movements during mental imagery and the spatial mental imagery score of the OSIVQ. That is, the stronger the

Table 6.3: Types of attention shifts for the generation of a mental image with and without shape information. The island depicted in Figure 6.5 is imagined as a mental image without shape information and including shape information.

Attention Shifts (AS)	Shapes	No Shapes
Total Overt AS	9	0
Total Covert AS	8	3
Overt/Covert for Spatial Relations	3/0	0/3

preference/ability of a person to use spatial mental imagery, the lower the dispersion of spontaneous eye movements will be. There are two ways to account for this finding that can be derived from the model of PIT. The first possibility is that people with a preference to use spatial mental imagery have the skill of using spatial mental imagery very efficiently. Such efficiency could be understood as being able to instantiate spatial mental concepts, such as spatial relations, with particularly short attention shifts. That is, the concept *left-of* would be instantiated by a shorter vector by a participant with a high spatial mental imagery score than for a participant with a lower spatial mental imagery score. The shorter the vectors used in the instantiation process, the faster one can imagine spatial configurations, because reaction times depend on the length of the attention shift. This aids one's ability (and thereby likely also one's preference) to use spatial mental imagery. Shorter attention shifts naturally lead to a lower dispersion of the overall pattern of (overt) attention shifts.

The second possibility offered by the model is that people with a high spatial mental imagery score will mentally imagine much less visual information, e.g., shapes, textures, than a person with a low spatial imagery score. The reason is that the spatial mental imagery score is negatively correlated with the object (i.e., visual) imagery score which indicates the preference/ability to imagine visual information. When less shape information is instantiated, the instantiation of the spatial relations will in consequence utilize shorter attention shifts. The reason is that the instantiation of, for example, *left-of* is context-sensitive so that available perceptual information of the shape of a referenced entity will affect the length of the vector of *left-of* proportional to the extent of the entity's (imagined) shape. Section 3.2.2 and Section 5.1.3 elaborate on the mechanisms of this context-sensitivity. Concretely, when the shape of an entity is not instantiated, its shape is abstracted to a point with no extent. For such a shape-less entity the length of the spatial relations is not affected so that the default short length is used. This property of the model can be observed in Figure 6.5. Table 6.3 shows a comparison of the employed overt and covert attention shifts for the two conditions.



```
> imagine_v
concept? > island_map
shape quantification:
Covert from (0.0, 0.0) to (3.0, 0.0)
Covert from (3.0, 0.0) to (3.0, -3.0)
Covert from (3.0, -3.0) to (0.0, -3.0)
Covert from (0.0, -3.0) to (0.0, 0.0)
done
Overt from (0.0, 0.0) to (-6.0, 6.0)
shape quantification:
Overt from (-6.0, 6.0) to (-1.7573593128807143, 10.242640687119286)
Overt from (-1.7573593128807143, 10.242640687119286) to (-6.0, 6.0)
Overt from (-6.0, 6.0) to (-12.0, 6.0)
done
Overt from (-12.0, 6.0) to (1.2426406871192857, 6.0)
shape quantification:£
Covert from (1.2426406871192857, 6.0) to (4.242640687119286, 6.0)
Covert from (4.242640687119286, 6.0) to (4.242640687119286, 3.0)
Covert from (4.242640687119286, 3.0) to (1.2426406871192857, 3.0)
Covert from (1.2426406871192857, 3.0) to (1.2426406871192857, 6.0)
done
Overt from (1.2426406871192857, 6.0) to (-4.757359312880714, 0.0)
shape quantification:
Overt from (-4.757359312880714, 0.0) to (-0.5147186257614287, 4.242640687119286)
Overt from (-0.5147186257614287, 4.242640687119286) to (-4.757359312880714, 0.0)
Overt from (-4.757359312880714, 0.0) to (-10.757359312880714, 0.0)
done

> imagine_s
concept? > island_map
Covert from (-10.757359312880714, 0.0) to (-13.757359312880714, 3.0)
Covert from (-13.757359312880714, 3.0) to (-10.757359312880714, 3.0)
Covert from (-10.757359312880714, 3.0) to (-13.757359312880714, 0.0)
```

Figure 6.5: Output of two simulations of the model. The scene of the island depicted on the top is imagined by the model in two conditions: 1) including the visual information of the entities' shapes, and 2) without shape information of the entities. Table 6.3 summarizes the resulting overt and covert attention shifts.

6.3.4 Predictions

The model makes predictions about the occurrence of spontaneous eye movements during mental imagery.

The size of an attention shift, i.e., the length of the vector, that is employed during instantiation is the crucial factor that determines whether the attention shift will be executed covertly or overtly. The distance of an attention shift is determined by two main factors: 1) the distance of the spatial relation in the conceptual description of the to-be-imagined stimulus, and 2) the concrete vector upon which a spatial relation such as the mental concept *left-of* will be mapped onto by the **select** function of the VS-LTM.

The first point is trivial, as the properties of the given stimulus are encoded on a conceptual level, i.e., as sets of mental concepts such as *left-of* or *left-of, close* or *left-of, far*. In mental imagery, these mental concepts will be mapped onto attention shifts that reflect the conceptually described distance of the spatial relation. The second point can be subdivided into two further aspects: 1) individual differences, and 2) content of the to-be-imagined stimulus. Individual differences can determine how, for example, a prototypical *left-of* is instantiated. Such differences are the result of how the mappings of the **select** function of the VS-LTM have been learned and in which context they are commonly used. The second aspect is the content that is imagined and this aspect can be easily tested empirically. The model of PIT instantiates spatial relations so that the available context of the spatial relation is considered. This aspect has already been discussed in Section 6.3.3. Essentially, the availability of perceptual information about shapes will affect the instantiation of the related spatial relations so that the vectors of the spatial relation are adjusted in their length. They will get longer proportional to the extent of the imagined shapes of the entities which the spatial relations refer to.

From these two main points, the following predictions regarding the occurrence of spontaneous eye movements that reflect the content of the mental image are inferred.

- The more visual information, i.e., specifically shapes, is contained in a mental image, the more eye movements are expected;
- The more complex a mental image is, i.e., the more entities and thus necessarily spatial relations it contains, the more eye movements are expected;
- The former two points can be summarized by the prediction that the more realistic and rich in detail the mental image is, the more eye movements are expected;
- The longer the distances in a mental image are, the more likely eye movements become;

- The larger the shapes in a mental image are, the more likely eye movements become.

A concrete experiment to both test PIT's explanations and predictions about eye movements and mental scanning (see Section 6.1.2) would be the reproduction of the mental scanning experiment of Richman et al. (1979) with the addition of eye tracking so that the participants' gaze is recorded during mental imagery. PIT would predict that spontaneous eye movements occur and reflect the different imagined distances of specifically the "20 miles" and "80 miles" routes. That is, these eye movements would reflect the integration of the given metrics of the island stimulus and the additional suggested (inconsistent) distances of the sign posts. A verification of this prediction would strongly support the assumption that mental images rely on mental concepts which can include potentially conflicting information and that attention shifts such as eye movements are used to instantiate these mental concepts so that the integrated direction and distance is reflected in the concrete attention shifts.

Note that PIT does not automatically predict the non-employment of mental imagery if no relevant eye movements occur during a given task. It is possible that the task at hand does not elicit such eye movements because the required conditions are not given. For example, many spatial reasoning tasks used in the literature on mental model reasoning (e.g., Jahn, Knauff, & Johnson-Laird, 2007) are very abstract and contain no or very little (relevant) visual information, e.g., "A is left of B; B is above C; what is the relation between A and C". The relationship between mental model theory and PIT with respect to the role of eye movements in such tasks is discussed in Section 7.2.2.

6.4 Unilateral Neglect

Unilateral neglect in visual perception (visual neglect) and in mental imagery (imaginal neglect) are complex neuropsychological phenomena whose underlying mechanisms are not fully understood. Any attempt to explain unilateral neglect and its properties requires a much broader scope than that offered by a theory of mental imagery. Therefore, an explanation of unilateral neglect is not the goal of this section. Instead, it will be discussed if and how PIT is compatible with the empirical findings of unilateral neglect. The reason unilateral neglect is considered in this thesis is that it poses strong constraints on theories of mental imagery and it has been argued that both the descriptive and the pictorial theory are inconsistent with the findings on unilateral neglect (see Section 2.3.4).

6.4.1 Unilateral Neglect and PIT

As reported in Chapter 2, the underlying causes of unilateral neglect are assumed to be an interplay of several different deficits with varying severity. Yet, specifically the role of exogenous attention, i.e., attention triggered by external cues such as appearing objects, is accepted to play a major role in visual neglect (Bartolomeo & Chokron, 2002, 2001; Bourlon et al., 2010). A deficit in exogenous attention causes the patients' attention to be overly strongly drawn towards cues on their non-neglected side of their visual field. This effectively leads to the failure to properly attend to objects on their neglected side[7]. According to PIT both visual perception and mental imagery generally employ the same perceptual mechanisms and thus also the same attentional processes. There is, however, a difference with respect to attention between visual perception and mental imagery in PIT that has been discussed in Section 3.2.6: bottom-up attention plays no role in mental imagery where attention is controlled by top-down guidance only. The instantiation of mental concepts is exactly such a top-down guidance as attention shifts are perceptual actions which are selected based on the available mental concepts and instantiated perceptual information. In visual perception, in contrast, it is fundamental that attention is also drawn to salient cues in our visual field in a bottom-up fashion. The distinction between bottom-up and top-down attention is the same as the distinction between exogenous and endogenous attention in the literature on unilateral neglect. Given that bottom-up attention shifts are relevant for visual perception but not mental imagery, damage to respective neural areas involved in bottom-up but not top-down attention, could explain the occurrence of visual neglect without imaginal neglect (Bartolomeo, D'Erme, & Gainotti, 1994; Rode et al., 2010; Bourlon et al., 2010). This would also explain why only few patients with visual neglect also show signs of imaginal neglect (Bartolomeo, 2007) as imaginal neglect would additionally require damage to brain areas involved in top-down attention processes. Under the assumption that deficits in bottom-up, or exogenous, attention cause visual neglect but not imaginal neglect, PIT is consequently consistent with the occurrence of visual neglect without imaginal neglect.

The case of imaginal neglect without visual neglect seems slightly more difficult. Deficits in top-down attention with healthy bottom-up attention should intuitively lead to deficits in both vision and imagery as both obviously depend on top-down guidance of attention. It has, however, been argued that patients showing only imaginal neglect but not visual neglect might have initially shown symptoms of both neglects, but over time adapted

[7] Other deficits are likely relevant in combination with such a deficit in exogenous attention. For example, it has been proposed that an inability to properly disengage attention from those objects on the non-neglected side additionally plays a role in visual neglect (e.g., Posner, Walker, Friedrich, & Rafal, 1984).

and learned to compensate their visual neglect to the point at which symptoms are hardly noticeable (Bartolomeo, 2007). Such a compensation would have to then rely on aspects that are only given in vision but not imagery, perhaps a stronger reliance on bottom-up attention.

Additionally, Rode et al. (2010) report on two patients, one suffering from both visual and imaginal neglect and one suffering from imaginal neglect only. They compared the specific damaged brain areas of the two patients and proposed that damage disconnecting the posterior callosal in the patient suffering only imaginal neglect might prevent a symmetrical processing of spatial information from long-term memory. Such an explanation of imaginal neglect could easily be incorporated into the model of PIT as an inhibition of (top-down) attention shifts during mental imagery. Those top-down attention shifts that would be directed into the neglected half of the underlying coordinate system could be inhibited. Such a manipulation would work to simulate imaginal neglect for scenes (in which objects located on one side could consequently not be instantiated), objects (for which parts on one of their sides could not be instantiated), and single shapes of objects (which would only be partially instantiated so that the resulting shape information would be incomplete, i.e., lacking those contours from the neglected side). The reason why such a simulation of imaginal neglect would work for all these cases is that the employed coordinate system is not absolute but the origin of the coordinate system of PIT's model is relative to the currently imagined entity or scene.

Summarizing, given the above assumptions from the literature on unilateral neglect, PIT is in principle able to account for visual and imaginal neglect and their dissociation. As PIT is more concretely described, its consistency with the results of unilateral neglect also serves the purpose of clarifying more concretely how the enactive theory could be consistent with these results (as it was already suspected by Bartolomeo (2002)).

6.5 Summary

In this chapter PIT was applied to all considered phenomena of mental imagery including specifically the more detailed aspects. PIT and its implementation were shown to provide explanations and predictions for mental scanning, mental reinterpretation, and eye movements during mental imagery. PIT was shown to be in principle consistent with the constraints posed by the findings on unilateral neglect. For the phenomena of mental scanning and eye movements the implemented mechanisms of the computational model were directly applicable and simulations supported the proposed explanations. Overall, the explanations and predictions of PIT and its computational model go considerably beyond the explanations of the three contemporary theories (see Section 2.3).

Chapter 7

Conclusion and Outlook

> *"But if it is asked whether the devils could have deluded the onlookers by the above-mentioned method of working upon the mental images, and not by assuming aerial bodies like flying birds, the answer is that they could have done so."*
> (Malleus Maleficarum Part 2, Chapter VIII)

This chapter discusses the contributions of the thesis and gives an outlook on future work.

7.1 Contributions

7.1.1 Contributions to the Imagery Debate

PIT and its computational model make contributions to some of the most fundamental questions of the imagery debate. These questions are 1) does mental imagery rely on a depictive mental representation?; 2) how can the spatio-analogical character of mental imagery be explained?; and 3) are modality-specific representations and processes functionally involved in mental imagery?

The computational model of PIT constitutes a proof of concept that mental imagery does not require a depictive mental representation. That is, the computational model provides a concrete example how the spatio-analogical character of mental imagery, for example, evident in mental scanning, need not result from a specifically structured mental representation (as assumed by the pictorial theory) or from the non-functional application of tacit knowledge (as assumed by the descriptive theory) but can result from the employment of perceptual actions. The temporal properties of covertly or overtly employing these perceptual actions determine the temporal properties of mental imagery. That is, the physical structure of the human visual system gives mental imagery its spatio-analogical character. In this respect, PIT provides support to the enactive theory by providing

a more formal framework and a concrete model that build upon the enactive theory's assumption that attentional processes of visual perception are re-used in mental imagery.

PIT proposes the functional involvement of modality-specific processes of visual perception in mental imagery. However, these processes are generally aimed at the inspection of external stimuli and not at internal mental representations. Specifically, the involvement of modality-specific representations of early visual areas is not necessary, because perceptual information of a mental image can be retrieved by processes of proprioception and anticipation without the need for recurrent activation of early visual representations. This assumption of PIT, furthermore, allows the otherwise conflicting neuroimaging results on the activation of early visual areas during mental imagery to be resolved, as such activation can now be argued to be non-functional.

Additionally, the computational model of PIT constitutes a first step towards a model-based investigation of the questions of the imagery debate. The computational model can in the future be compared to implementations of other theories of mental imagery. The development and comparison of theories using computational models should allow for a more efficient progress on the questions of the imagery debate. A comparison of implemented theories will reduce misunderstandings between researchers of the different theories as implemented concepts and processes are less ambiguous and more transparent.

The application of the computation model to the phenomena of mental imagery resulted in several concrete predictions. Ways of empirically testing these predictions have been proposed. Future empirical work of testing these prediction can provide stronger evidence for PIT as well as suggest corrections and refinements for the framework and model. This way the computational model can facilitate a tighter coupling of theory and experiments which can further support efficient progress of our knowledge on the nature of mental imagery.

7.1.2 Contributions to the Understanding of the Empirical Phenomena of Mental Imagery

The application of PIT and its computational model to the different phenomena of mental imagery led to explanations that covered aspects which were previously only discussed on a vague level. The explanations provided by PIT aid a deeper understanding of these aspects.

In the context of mental scanning, the important concept of cognitive penetration was covered. The computational model provides the first mechanistic account of how and why cognitive penetration affects mental imagery. What is cognitively penetrated, or altered, are the mental concepts that conceptually describe the mental image. Based on these mental con-

cepts perceptual information is generated which will then also be altered as it constitutes an instance of the mental concepts. This way the perceptual information of the mental image will reflect a participant's belief and knowledge.

The recent finding that spontaneous eye movements during mental imagery not only reflect the content of the mental image but are in fact functional for mental imagery is incorporated in PIT and implemented in the computational model. Eye movements in mental imagery have been the topic of recent discussion and their concrete role remained unclear. The model gives a mechanistic account of how eye movements are functionally involved for the generation and inspection of mental imagery as they are employed for the instantiation of mental concepts. That is, spontaneous eye movements during mental imagery can be understood as a means to make an abstract mental concept concrete through the "replay" of its perception. This understanding automatically gives an explanation for the fact that these eye movements have been found to correspond to the content of mental images.

The explanations of PIT for the findings on mental reinterpretation show how and why a mental image is necessarily always interpreted. Several researchers (e.g., Fodor, 1975; Chambers & Reisberg, 1992; Cornoldi, Logie, Brandimonte, Kaufmann, & Reisberg, 1996) have proposed that mental images always come with an interpretation or caption and PIT's model provides concrete support for these assumptions. In PIT a mental image corresponds to an interpretation drawn from a set of mental concepts with instantiated perceptual information.

Although not aiding the understanding of unilateral neglect, it was shown that PIT and its model can be consistent with the findings on unilateral neglect. A consistency of these findings with theories of mental imagery was not available before. Rather these findings have been interpreted to be inconsistent with the pictorial and the descriptive theory.

7.1.3 Contributions to the Enactive Theory

PIT contributes to the development of the enactive theory in that the assumption of the enactive theory that perception is active vision with an emphasis on the interaction with the environment instead of the inspection of internal mental representations is adopted and fleshed out in the formal framework and the computational model. PIT's implementation offers a concrete way of understanding the previously only abstractly described schemata proposed by the enactive theory. The VS-LTM and its functions **select** and **identify** correspond to the schemata and furthermore link them to the concept of grounded symbols. This linkage allows the ideas of the enactive theory to be embedded into common frameworks of cognitive systems, e.g., systems including long-term memory and working memory based

on (grounded) symbols.

Furthermore, the similarity of PIT with the enactive theory should allow a transfer of (some of) the explanations given by PIT for the phenomena of mental imagery to the enactive theory.

7.1.4 Contributions to Embodied Cognition

The paradigm of embodied cognition is the currently prevalent paradigm in cognitive psychology and cognitive science for understanding the nature of cognition. The previously dominant paradigm of computationalism (or cognitivism) emphasized the role and importance of rich internal mental representations of the world for cognition. Embodied cognition, in contrast, emphasizes the sensorimotor interactions of an organism with its environment, i.e., action and perception capabilities, as constitutive of cognition. Cognition is understood as bootstrapped from acquired sensorimotor interactions and their internal simulation (e.g., Hesslow, 2012) and/or symbols grounded in these interactions (e.g., Barsalou, 1999). The contemporary theories of mental imagery with the exception of the enactive theory are rooted in the previous paradigm with a focus on the role of mental representations.

PIT is an example of a theory that is rooted in the concepts of embodied cognition. Mental imagery in PIT is realized by the simulation of sensorimotor interactions and their (anticipated) feedback. Furthermore, the non-brain body of an organism plays a critical role in PIT as proprioception of the state of muscles yields perceptual information during mental imagery. Lastly, PIT's mental concepts are an example of grounded symbols which implement associations between actions and perceptions. Accordingly, the computational model can be interpreted as a concrete instance of embodied cognition for the domain of visuo-spatial mental imagery. PIT's applicability to a wide range of diverse phenomena of mental imagery provides support to the validity of these implemented assumptions of embodied cognition.

7.2 Outlook

The section is divided into two parts: 1) extending the model of PIT and 2) comparing and/or complementing PIT with other theories of visuo-spatial information processing.

7.2.1 Extending the Model of PIT

PIT assumes that mental imagery relies on acquired procedural knowledge of visual perception (see Section 3.1.3). Accordingly, the computational model of PIT also requires such procedural knowledge. In particular, the following information is necessary:

- knowledge about what entities/relations exist (to inform mental concepts);
- knowledge of the likelihood of perceiving a certain entity/relation in a certain situation (to inform the functions **interpret** and **select**);
- knowledge how to perceive a given entity/relation (to inform the functions **select** and **identify**).

Currently, this knowledge has been designed and put into the model. If this knowledge could instead be autonomously acquired, that would make a stronger case for the plausibility and validity of the theory and model. However, the domain of human perception is so complex and still insufficiently understood, that the acquisition of such knowledge is highly difficult (see Section 5.3.2 for more details on this point). There is, however, the possibility of implementing the model of PIT on an artificial system such as a robot so that the system is able to learn this knowledge itself over time. The reason why such an approach is possible for an artificial system but hardly so for a model of human cognition, is that the internals of a robot are fully known and common robotic systems are much simpler than humans with respect to their perception and action capabilities.

It is a fundamental assumption of PIT as well as of theories of embodied cognition, in particular, perceptual symbol systems (Barsalou, 1999) and the simulation theory of cognition (Hesslow, 2012), that such knowledge is acquired and used for mental imagery and simulations. The internal simulations assumed by the two theories of embodied cognition can be understood as the "offline" employment of processes of perception and action. Mental imagery as proposed by PIT is exactly such simulations with two exceptions: 1) mental imagery is conscious whereas simulations are largely subconscious and 2) mental imagery employs perceptual actions not only covertly but also overtly, e.g., eye movements, whereas simulations are generally assumed to rely on the covert employment of actions.

Mental imagery can thus be seen as a special (conscious) case of the simulations proposed by the above mentioned theories. The theories of embodied cognition propose that these simulations constitute cognition so that thought corresponds to the simulation of acting and mentally perceiving the consequences of the actions. However, it remains an unsolved problem how exactly such simulations are acquired and realized and how they can give rise to further cognitive abilities.

There is previous computational work (e.g., Ziemke, Jirenhed, & Hesslow, 2005; Moeller & Schenk, 2008) that showed how tasks such as collision-free navigation and simple object recognition can be realized by acquired sensorimotor interactions and the simulations of such interactions in simulated robots. Computational investigations on the question how simulations are acquired and how aspects of high-level cognition can emerge from these

simulations is regarded a necessary next step in research on embodied cognition (e.g., Pezzulo et al., 2011; Barsalou, 2010). Therefore, it is a promising future endeavor to investigate how the computational model of PIT can be bootstrapped from acquired sensorimotor interactions. This would not only provide stronger support for PIT and its model, but potentially provide insight on the fundamental assumptions of embodied cognition.

The following outlines one way of bootstrapping the computational model of PIT from sensorimotor interactions.

Bootstrapping PIT From Sensorimotor Interactions

Similarly, to previous studies on bootstrapping behavior from sensorimotor interactions (Ziemke et al., 2005; Moeller & Schenk, 2008), a simple simulated domain and a simulated robot would be used. Let's concretely assume a simple agent in a domain containing free spaces and obstacles. The agent is able to move in different directions. In the following, I will show how the processes and representations of PIT can be transferred and acquired from an autonomous exploration of the agent in the domain.

Mental concepts in PIT are abstracted descriptions of situations. They are grounded in the actions that are available and meaningful for the described situation. This concept can be directly applied to the assumed domain. The agent can learn which actions, i.e., movements, are useful in different situations. That is, it learns to link a perception to a set of actions. It does so by abstracting the perception to a category. These categories are defined by which actions are afforded by the perceptions of that category (for the concept of affordances, see, e.g., Gibson, 1986). Categories are learned by feedback on the agent's actions in different situations. Concretely, moving into a wall leads to negative feedback and moving into free space leads to positive feedback. What is learned from this feedback is the function **identify** which maps the robot's actual or imagined perceptions onto mental concepts. Figure 7.1 gives examples of mental concepts in the domain.

Internal models will be used to learn the functions **select** and **execute**. Internal models are the common approach to learn and engage in goal-directed sensorimotor interactions, that is, to learn how to act in a given situation given a certain goal (Kawato, 1999). There is evidence that the human brain implements such internal models for this purpose (Wolpert, Miall, & Kawato, 1998). Internal models usually comprise of two types of models: forward models and inverse models. A forward model receives an action as input and outputs the predicted change in perception. An inverse model receives the current perception as well as a goal state or a motivation and outputs that action which is predicted to change the current perception towards the goal state.

In PIT the function **select** selects an action given a set of mental con-

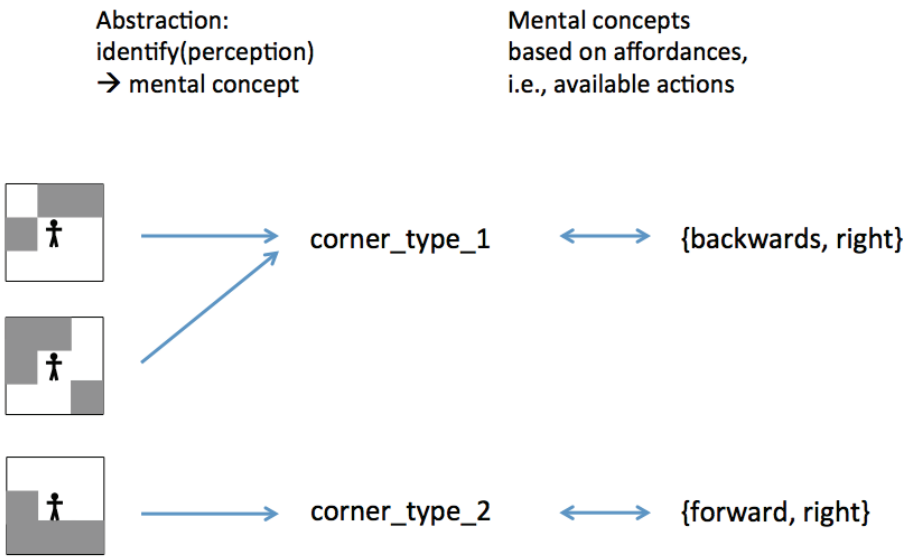


Figure 7.1: The figure gives an example how perceptions can be abstracted to mental concepts based on affordances. The perception of the agent is his surrounding. The grey fields represent obstacles. The function **identify** has been trained to identify viable actions for a given perception. Perceptions are categorized based on the actions that they afford. Using this abstraction concepts such as different corners, corridors, and dead-ends emerge.

cepts. In the domain, **select** is learned as an inverse model that suggests an action given the current mental concepts. As mental concepts are categories that contain the viable actions, **select** has to learn which of the available actions will meet the agent's goals. For example, such an inverse model could be learned based on the goal to not employ an action inverse to the previous one, e.g., going right after going left. That would mean, that even though the mental concept "corner" offers two viable actions, **select** will be trained to take the option that meets this goal.

In PIT the function **execute** internally simulates an action to provide perceptual information. **Execute** can be learned as a forward model. That is, it will simply learn how perceptions change given an action, e.g., moving to the left leads obstacles in the perception to "move" to the right. Such a forward model can then predict the next perception given an action. As a perception is internally abstracted to mental concepts, **execute** will have to fill in unknown parts of the perception, i.e., create a concrete instance of the abstract mental concept. This filling-in process can be based on learned statistics of the domain, e.g., which concrete "corner" is most likely.

The function **interpret** interprets an ambiguous set of mental concepts in PIT. The domain described here is too simple to provide such ambiguous situations. However, for a more complex domain **interpret** can be learned by using a statistical measure of how likely different interpretations are for a given domain and a given situation. That is, the likelihood of different perceptions in the domain.

The above showed how PIT can in principle be bootstrapped from sensorimotor interactions. However, in the above assumed domain the capability of mental imagery would not be of much use. It is promising to further extend the simple domain with more complicated tasks for the agent such as finding resources and exploring the world. Mental imagery as internal simulation could then be used as a planning tool for these tasks. Such an extension could be of high relevance for embodied cognition as it might provide a concrete proof-of-concept that abilities such as planning and reasoning can be bootstrapped from sensorimotor interactions via mental imagery.

7.2.2 PIT and Other Theories of Visuo-Spatial Information Processing

In the following, potential future work of comparing, combining, or complementing other theories of visuo-spatial information processing, i.e., visuo-spatial working memory and the mental model theory, with PIT is briefly discussed.

Visuo-Spatial Working Memory

The visuo-spatial working memory theory (VSWM) (Logie, 2003) is a well-established and well-supported theory of working memory. It proposes that working memory consists of multiple components which are each specialized mental systems that deal with particular types of information and particular types of manipulations of information. One such component is the visual cache with the inner scribe. The visual cache operates as a passive visual temporary store while the inner scribe is associated with attentional control and is involved in planning and executing movements. In contrast to the visual buffer of the pictorial theory of mental imagery (Kosslyn et al., 2006), the visual cache is not directly linked to visual perception but instead the visual cache holds information that has been processed and interpreted by visual perception and respective background knowledge. The visual buffer is critically different than the visual cache as it directly processes sensory input and mediates it to long-term memory.

This difference between the visual buffer on the one hand and the visual cache on the other hand, has led to a recent discussion about whether the proposed structures overlap to some degree or whether they are distinct (e.g., Borst, Niven, & Logie, 2012; Meulen, Logie, & Sala, 2009). These investigations might form the basis of a unification of the two theories. In the following, I will briefly discuss how PIT (as an alternative to the pictorial theory and as potentially overlapping with structures of the VSWM) is consistent with the results of the study of Borst et al. (2012), who compare properties of the visual buffer (of the pictorial theory) with the visual cache (of the VSWM).

Borst et al. (2012) report three experiments that investigated whether the cognitive processes underlying mental image generation and short-term retention of mental images are the same or different. They employed two interference conditions: spatial tapping and irrelevant visual input (IVI). Their results provide support for the following conclusions. Retention of mental images is realized in a representation different than the one used for the generation of mental images. The retention would be realized by the visual cache which is not disrupted by IVI as it is not directly connected to visual perception but holds the already processed and interpreted content of mental imagery. The generation of mental images would be realized by the visual buffer which is disrupted by IVI because it is located in the early areas of the visual cortex. Spatial tapping interfered only to a lesser degree with the generation of mental images in the visual buffer which is assumed to not be involved in the (blind) spatial tapping task. Spatial tapping did interfere with the retention of information in the visual cache, which is to be expected as the visual cache is also involved in executing movement (via the inner scribe).

Summarizing, the study proposes that the visual buffer of the pictorial

theory is used to construct mental images which are then immediately stored in the visual cache of the VSWM.

In PIT we can find corresponding structures and processes for the visual buffer and the visual cache. The generation of mental images is realized by the process of instantiation which employs different perceptual processes including overt and covert attention shifts. Instantiation would therefore be expected to be disrupted by IVI as it would cause bottom-up triggered attention shifts which would interfere with those attention shifts executed for instantiation. Perceptual information generated through instantiation is stored in short-term memory and extends the corresponding mental concepts at which point it would not be expected to interfere with IVI. The short-term memory holding the instantiated perceptual information would accordingly correspond to the visual cache. Spatial tapping involves top-down guided movements and can thus be assumed to rely on perceptual feedback and thus employ the same (multi-modal) short-term memory as perceptual feedback from mental imagery. Therefore, the interference between spatial tapping and the mental image in PIT's short-term memory would be expected.

It would be an interesting topic of investigation to compare PIT to the the VSWM framework in more depth. This could go along two different directions. One direction would be the investigation how PIT is generally consistent or inconsistent with the vast literature on interference studies and neuropsychological results that provide support for the structure and functions of the components of the VSWM (e.g., Logie, 1995). The other direction would be to investigate to which extent the explanations of PIT for the phenomena considered in this thesis can be transferred to the framework of the VSWM. Given the prima facie compatibility of PIT with the VSWM, the successful transfer of the relatively detailed explanations of PIT and its model could provide one concrete instance of the VSWM framework for the phenomena considered in this thesis. This would additionally provide an easy way of linking concepts of embodied cognition as realized in PIT to the established VSWM framework.

Mental Model Theory and Preferences in Reasoning

Mental model theory (e.g., Johnson-Laird, 2001) is an established and well-supported theory on human reasoning including specifically reasoning with visuo-spatial information. Mental model theory postulates that there are three representational levels involved in human reasoning: propositional representations, mental models, and mental images (Johnson-Laird, 1998). The relationships between these three levels are hierarchical in the sense that the more specific representation depends on the information of the more general representation. The example in Figure 7.2 helps to illustrate this point. It has been shown that there is a considerable overlap between the mental model theory and the pictorial theory of mental imagery (Sima, Schultheis, &

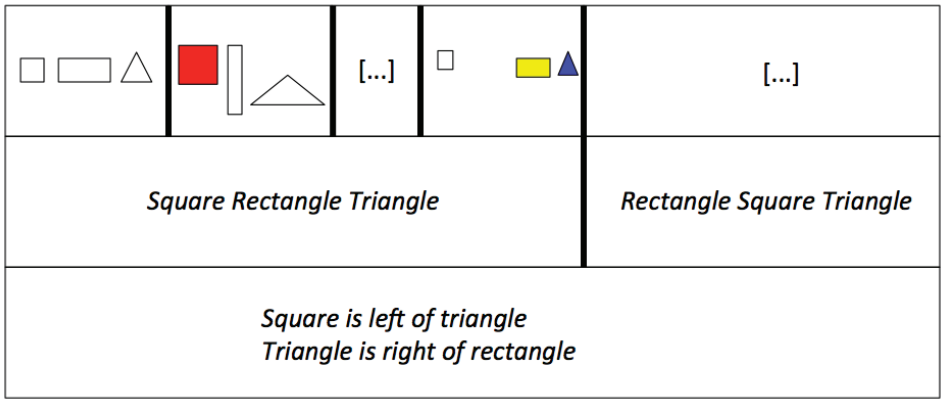


Figure 7.2: The three levels of the mental model theory. The bottom represents the propositional level, i.e., language-like descriptions. The middle level is the mental model level which is a specification of the more general propositional level, because a mental model might only represent one of many valid configurations described by the propositions. In the depicted example there are two valid configurations given the propositional premises. The upper level is the mental image level which again is a specification of the more general mental model level, because it additionally specifies properties that the underlying mental model representation might be invariant to, such as color, distance, and shape. As depicted a variety of valid specifications are possible.

Barkowsky, 2013). The two theories assume the same or at least very similar representational levels, structures of mental representations, and anatomical localizations. This overlap makes it likely that the two theories describe the same reasoning apparatus while, however, focussing on different aspects of it. It would be a promising future endeavor to investigate how and to which extent PIT is comparable to mental model theory and which extensions would be necessary for it to account for the specific reasoning phenomena, such as preferred mental models, that the mental model theory has been successfully applied to (e.g. Jahn et al., 2007). In the following, a first application of PIT to preferred mental models is described.

A recent study (Sima et al., 2013) has tested the explicit claim of mental model theory that human reasoning is realized on the level of mental models and that the employment of visual mental images can even impede this reasoning process when visual information is irrelevant to the reasoning task at hand (Knauff & Johnson-Laird, 2002). The study used three-term series spatial reasoning problems of the form: "X is west of Y" (premise 1), "Z is north-east of X" (premise 2), "What is the relation between Z and Y?" (conclusion). These problems are under-specified, that is, there are different answers which are valid given the premises. Figure 7.3 depicts the different

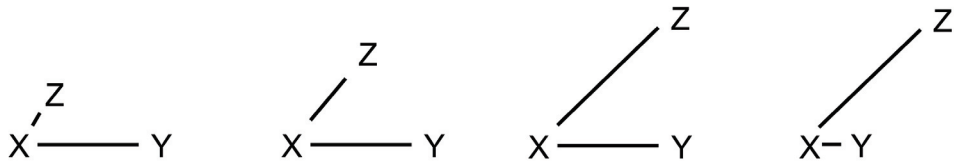


Figure 7.3: Different valid solutions for a spatial reasoning task. The figure shows the four valid solutions for the spatial reasoning task: "X is west of Y" and "Z is north-east of X"; "What is the relation between Z and Y?". The solutions are from left to right: "west", "north-west", "north", and "north-east".

valid answers for the above example. Problems of this type are commonly used to study preferred mental models. A preferred mental model is a robust within-subject and between-subject preference for one of many valid answers (e.g., Jahn et al., 2007).

The study used the above reasoning tasks for two experiments which only differed in their instructions.

The first experiment (the mental model experiment) used no instructions other than just asking participants to solve the tasks as it is common for studies investigating reasoning with mental models. In the second experiment (the imagery experiment) the instructions were designed to induce the employment of visual mental images, i.e., "imagine the letters as cities on a map". Note that the induced visual information of imagining the letters as cities on a map is irrelevant to the actual reasoning task which is the same in both experiments. The study found two main results: 1) only in the imagery experiment a majority of participants showed significant spontaneous eye movements along the given spatial relations of the tasks[1], and 2) there were significant preferences for one of the different solutions, but they did differ between the two experiments. The critical finding is the second one, that is, the fact that the employment of visual mental imagery even with irrelevant visual information has led to different reasoning outcomes for the same spatial reasoning task. This finding is not predicted by mental model theory and requires additional hypotheses on the relationship between the imagery-level and the mental-model-level of the mental model theory.

Without additional adjustments the model of PIT explains both the findings on eye movements and preferences, i.e., the occurrence of eye movements in the imagery experiment, the lack of eye movements in the mental model experiment, and the emergence of different reasoning preferences depending on the addition of (irrelevant) visual details. In PIT, a mental model can be understood as simply a mental image for which no shape information

[1]Eye movements were recorded using an eye-tracker.

is instantiated. Because the instantiation of spatial relations depends on additional instantiated information, such as shape information, spatial relations will be instantiated using attention shifts of greater length when shape information is available (see Section 3.2.2 and Section 5.1.3 for an in-depth explanation). The employment of attention shifts of greater length leads to 1) more spontaneous eye movements, because attention shifts are executed overtly when they exceed a certain length, and 2) the generated concrete instance of the imagined situation, i.e., the mental image, is different because the different attention shifts lead to entities having different locations. The inference of new spatial relations depends on the locations of the entities. Therefore, different spatial relations will be inferred, that is, different reasoning outcomes will result. Different reasoning outcomes accordingly lead to different preferences.

Summarizing, without any additional adjustments PIT is able to account for the results of the study while the mental model theory currently cannot fully account for the findings. Note that, the above application of PIT to both the first experiment, i.e., a task commonly considered a mental model reasoning task, and the second experiment, i.e., a task commonly considered a visual mental imagery task, indicates that mental model theory and visual mental imagery could be unified using the framework of PIT.

References

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*, 249–277.

Anderson, J. R. (2005). *Cognitive psychology and its implications*. New York: Worth Publishers. (6th Edition)

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, *48*, 57–86.

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

Barsalou, L. W. (2008, August). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645.

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science*, *2*(4), 716–724.

Bartolomeo, P. (2002). The relationship between visual perception and visual mental imagery: A reappraisal of the neuropsychological evidence. *Cortex*, *38*(3), 357–378.

Bartolomeo, P. (2007). Visual neglect. *Current Opinion in Neurology*, *27*, 381–386.

Bartolomeo, P., Bachoud-Levi, A.-C., Azouvi, P., & Chokron, S. (2005). Time to imagine space: a chronometric exploration of representational neglect. *Neuropsychologia*, *43*(9), 1249–1257.

Bartolomeo, P., & Chokron, S. (2001). Levels of impairment in unilateral neglect. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 4, pp. 67–98). Amsterdam: Elsevier.

Bartolomeo, P., & Chokron, S. (2002). Orienting of attention in left unilateral neglect. *Neuroscience and Biobehavioral Reviews*, *26*(2), 217–234.

Bartolomeo, P., D'Erme, P., & Gainotti, G. (1994). The relationship between visuospatial and representational neglect. *Neurology*, *44*, 1710–1714.

Behrmann, M., Watt, S., Black, S., & Barton, J. (1997). Impaired visual search in patients with unilateral neglect: an oculographic analysis. *Neuropsychologia*, *35*, 1445–1458.

Bisiach, E., Capitani, E., Luzzatti, C., & Perani, D. (1981). Brain and

conscious representation of outside reality. *Neuropsychologia*, *19*(4), 543 - 551.

Bisiach, E., Luzzatti, C., & Peranid, D. (1979). Unilateral neglect, representational schema and consciousness. *Brain*, *102*(3), 609–618.

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, *23*, 638–663.

Borst, G., Niven, E., & Logie, R. H. (2012). Visual mental image generation does not overlap with visual short-term memory: A dual-task interference study. *Memory & cognition*, 1–13.

Bourlon, C., Oliviero, B., Wattiez, N., Pouget, P., & Bartolomeo, P. (2010). Visual mental imagery: what the head's eye tells the mind's eye. *Brain Research*, *1367*, 287–297.

Brandimonte, M., & Gerbino, W. (1993). Mental image reversal and verbal recoding: When ducks become rabbits. *Memory & Cognition*, *21*, 23-33.

Brandimonte, M., Hitch, G., & Bishop, D. (1992a). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 157–165.

Brandimonte, M., Hitch, G., & Bishop, D. (1992b). Verbal recoding of visual stimuli impairs mental image transformations. *Memory and Cognition*, *20*, 449–455.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, *9*(1), 27–38.

Bridge, H., Harrold, S., Holmes, E. A., Stokes, M., & Kennard, C. (2012). Vivid visual mental imagery in the absence of the primary visual cortex. *Journal of Neurology*, *259*(6), 1062–1070.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 317–328.

Chambers, D., & Reisberg, D. (1992). What an image depicts depends on what an image means. *Cognitive Psychology*, *145–174*(24).

Cornoldi, C., Logie, R., Brandimonte, M., Kaufmann, G., & Reisberg, D. (1996). *Stretching the imagination: Representation and transformation in mental imagery*. Oxford: Oxford University Press.

Coslett, H. B. (1997). Neglect in vision and visual imagery: a double dissociation. *Brain*, *120*(7), 1163-1171.

Demarais, A. M., & Cohen, B. H. (1998). Evidence for image-scanning eye movements during transitive inference. *Biological Psychology*, *49*(3), 229–247.

Denis, M., & Cocude, M. (1992). Structural properties of visual images con-

structed from poorly or well-structured verbal descriptions. *Memory & Cognition*, *20*, 497-506.

Denis, M., & Cocude, M. (1997). On the metric properties of visual images generated from verbal descriptions: Evidence for the robustness of the mental scanning effect. *European Journal of Cognitive Psychology*, *9*(4), 353-380.

Denis, M., & Kosslyn, S. (1999). Scanning visual mental images: A window on the mind. *Cahiers Psychologiques Cognitives*, *18*, 409–465.

Finke, R. A., & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(2), 142–147.

Finke, R. A., & Pinker, S. (1983). Directional scanning of remembered visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(3), 398 - 410.

Finke, R. A., Pinker, S., & Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science*, *13*, 51–78.

Fodor, J. (1975). *The language of thought*. New York: Crowell.

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

Freksa, C. (1991). Qualitative spatial reasoning. In D. M. Mark & A. U. Frank (Eds.), *Cognitive and linguistic aspects of geographic space* (pp. 361–372). Kluwer, Dordrecht.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2009). *Cognitive neuroscience: The biology of the mind*. New York: W.W. Norton & Company. (3rd Edition)

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale: Lawrence Erlbaum Associates, Inc.

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

Goldenberg, G. (1993). The neural basis of mental imagery. *Baillieres Clinical Neurology*, *2*, 265–286.

Goldenberg, G. (1998). Is there a common substrate for visual recognition and visual imagery? *Neurocase*, *4*, 141–147.

Goldston, D., Hinrichs, J., & Richman, C. (1985). Subjects expectations, individual variability, and the scanning of mental images. *Memory & Cognition*, *13*(4), 365-370.

Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Research*, *1428*(0), 71 - 79.

Husain, M., Mannan, S., Hodgson, T., Wojciulik, E., Driver, J., & Kennard, C. (2001). Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect. *Brain*, *124*, 941–952.

Hyman, I., & Neisser, U. (1991). Reconstruing mental images: Problems of method. *Emory Cognition Project Technical Report #19*.

Intons-Peterson, M. J. (1983). Imagery paradigms: How vulnerable are they

to experimenters' expectations? *Journal of Experimental Psychology: Human Perception and Performance*, *9*(3), 394 - 412.

Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, *35*(8), 2075–2087.

Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K. (2011). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1289–1314.

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, *30*(6), 1053–1079.

Johansson, R., Holsanova, J., & Holmqvist, K. (2010). Eye movements during mental imagery are not reenactments of perception. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society.* Austin, TX: Cognitive Science Society.

Johansson, R., Holsanova, J., & Holmqvist, K. (2011). The dispersion of eye movements during visual imagery is related to individual differences in spatial imagery ability. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society.* Austin, TX: Cognitive Science Society.

Johnson-Laird, P. N. (1998). Imagery, visualization, and thinking. In J. Hochberg (Ed.), *Perception and cognition at century's end* (pp. 441–467). Academic Press.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*(10), 434-442.

Jolicoeur, P., & Kosslyn, S. (1985). Is time to scan visual images due to demand characteristics? *Memory & Cognition*, *13*, 320-332.

Kanizsa, G. (1955). Margini quasi-percettivi in campi con stimolazione omogenea. *Rivista di Psicologia*, *49*(1), 7–30.

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Luedtke, J. (2007). Experiential simulations of negated text information. *The Quarterly Journal of Experimental Psychology*, *60*(7), 976–990.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.*, *9*, 718–727.

Knauff, M., & Johnson-Laird, P. (2002). Visual imagery can impede reasoning. *Memory & Cognition*, *30*(3), 363–371.

Kosslyn, S. M. (1973). Scanning visual images - Some structural implications. *Perception & Psychophysics*, *14*(1), 90-94.

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology*, *7*(3), 341–370.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: The MIT Press.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 46–60.

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, *129*(5), 723–746.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2002). Mental imagery doesn't work like that. *Behavioral and Brain Sciences*, *25*(02), 198–200.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.

Kozhevnikov, M., Blazhenkova, O., & Becker, M. (2010). Trade-off in object versus spatial visualization abilities: Restrictions in the development of visual-processing resources. *Psychonomic Bulletin & Review*, *17*, 29–35.

Laeng, B., & Teodorescu, D. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scence. *Cognitive Science*, *26*, 207–231.

Logie, R. (1995). *Visuo-spatial working memory*. Hillsdale, NJ: Lawrence Erlbaum.

Logie, R. H. (2003). Spatial and visual working memory: A mental workspace. In D. E. Irwin & B. H. Ross (Eds.), *Cognitive vision* (Vol. 42, pp. 37–78). Academic Press.

Mast, F. W., & Kosslyn, S. M. (2002). Visual mental images can be ambiguous: Insights from individual differences in spatial transformation abilities. *Cognition*, *86*, 57–70.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Mellet, E., Petit, L., Mazoyer, B., Denis, M., & Tzourio, N. (1998). Reopening the mental imagery debate: Lessons from functional anatomy. *NeuroImage*, *8*(2), 129–139.

Meulen, M. van der, Logie, R. H., & Sala, S. D. (2009). Selective interference with image retention and generation: Evidence for the workspace model. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1568-1580.

Mitchell, D. B., & Richman, C. L. (1980). Confirmed reservations: Mental travel. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 58 - 66.

Moeller, R., & Schenk, W. (2008). Bootstrapping cognition from behavior

– a computerized thought experiment. *Cognitive Science*, *32*(3), 504–542.

Neisser, U. (1976). *Cognition and reality*. San Franciso: Freeman.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

O'Regan, K. J., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(05), 939–973.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. L. Lloyd (Eds.), *Cognition and categorization* (pp. 259–302). Hillsdale, N.J.: Erlbaum.

Peterson, M. A., Kihlstrom, J. F., Rose, P. M., & Glisky, M. L. (1992). Mental images can be ambiguous: Reconstruals and reference-frame reversals. *Memory and Cognition*, *20*(2), 107–123.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2011). The mechanics of embodiment: a dialog on embodiment and computational modeling. *Frontiers in Psychology*, *2*.

Pinker, S., Choate, P. A., & Finke, R. A. (1984). Mental extrapolation in patterns constructed from memory. *Memory and Cognition*, *12*(3), 207–218.

Posner, M., Walker, J., Friedrich, F., & Rafal, R. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience*, *4*, 1963–1874.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, *80*, 1–24.

Pylyshyn, Z. W. (1979). Validating computational models: A critique of Anderson's indeterminacy of representation claim. *Psychological Review*, *86*(4), 383–394.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, *88*, 16–45.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25*(2), 157–238.

Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world (jean nicod lectures)*. Cambridge, MA: The MIT Press.

Reed, S., Hock, H., & Lockhead, G. (1983). Tacit knowledge and the effect of pattern configuration on mental scanning. *Memory & Cognition*, *11*, 137-143.

Reisberg, D., & Chambers, D. (1991). Neither pictures nor propositions: What can we learn from a mental image? *Canadian Journal of Psychology*, *45*(3), 336–352.

Richman, C. L., Mitchell, D. B., & Reznick, J. S. (1979). Mental travel:

Some reservations. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(1), 13–18.

Rode, G., Cotton, F., Revol, R., Jacquin-Courtois, S., Rossetti, Y., & Bartolomeo, P. (2010). Representation and disconnection in imaginal neglect. *Neuropsychologia*, *48*, 2903–2911.

Rode, G., Rossetti, Y., Perenin, M.-T., & Boisson, D. (2004). Geographic information has to be spatialised to be neglected: A representational neglect case. *Cortex*, *40*(2), 391 - 397.

Schall, J., & Hanes, D. (1993). Neural basis of saccade target selection in frontal eye field during visual search. *Nature*, *366*, 467–469.

Schill, K., Umkehrer, E., Beinlich, S., Krieger, G., & Zetzsche, C. (2001). Scene analysis with saccadic eye movements: Top-down and bottom-up modeling. *Journal of Electronic Imaging*, *10*(1), 152–160.

Shapiro, S. C. (1992). *Encyclopedia of artificial intelligence (second edition)*. New York: John Wiley. (Section 4 is on AI-Complete Tasks)

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*(171), 701–703.

Sholl, M. (2001). The role of a self-reference system in spatial navigation. In D. R. Montello (Ed.), *Spatial information theory: Foundations of geographical information science* (pp. 217–232). Berlin: Springer.

Sima, J. F., Schultheis, H., & Barkowsky, T. (2013). Differences between spatial and visual mental representations. *Frontiers in Psychology*, *4*(240).

Slezak, P. (1995). The 'philosophical' case against visual imagery. In P. Slezak & T. Caelli (Eds.), *Perspective on cognitive science: Theories, experiments, and foundations* (pp. 237–271). Norwood, NJ: Ablex.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235–241.

Sun, R. (2009). Theoretical status of computational cognitive modeling. *Cognitive Systems Research*, *10*(2), 124–140.

Theeuwes, J., Belopolsky, A., & Olivers, C. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica*, *132*(2), 106–114.

Thomas, L. E., & Lleras, A. (2009). Covert shifts of attention function as an implicit aid to insight. *Cognition*, *111*, 168–174.

Thomas, N. J. T. (1999). Are theories of imagery theories of imagination? An active perception approach to conscious mental content. *Cognitive Science*, *23*, 207–245.

Thomas, N. J. T. (2002). A note on "schema" and "image schema". (Retrieved from http://www.imagery-imagination.com/schemata.htm on February, 12th 2012)

Thomas, N. J. T. (2013). Mental imagery. In E. N. Zalta

(Ed.), *The stanford encyclopedia of philosophy* (Spring 2013 ed.). http://plato.stanford.edu/archives/spr2013/entries/mental-imagery/.

Trojano, L., & Grossi, D. (1994). A critical review of mental imagery defects. *Brain and Cognition*, *24*, 213–243.

Tye, M. (1991). *The imagery debate*. Cambridge, MA: MIT Press.

Wallgrün, J. O., Frommberger, L., Wolter, D., Dylla, F., & Freksa, C. (2007). Qualitative spatial representation and reasoning in the SparQ-toolbox. In T. Barkowsky, M. Knauff, G. Ligozat, & D. Montello (Eds.), *Spatial cognition v: Reasoning, action, interaction: International conference spatial cognition 2006* (Vol. 4387, p. 39-58). Springer-Verlag Berlin Heidelberg.

Wang, R., & Spelke, E. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Science*, *6*(9), 376–382.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Science*, *2*(9), 338–347.

Ziemke, T., Jirenhed, D. A., & Hesslow, G. (2005). Internal simulation of perception: a minimal neuro-robotic model. *Neurocomputing*, *68*, 85–104.