The Nature of Mental Images – An Integrative Computational Theory

Jan Frederik Sima (sima@sfbtr8.uni-bremen.de)
SFB/TR 8 Spatial Cognition, University of Bremen, Germany

Abstract

We shed new light on the long-debated question about the nature of mental images, that is, the underlying structures and processes, with a new theory of mental imagery. This theory is formalized as a computational cognitive model and provides an integrated account of the three prevalent theories of mental imagery, i.e., the descriptive, the quasi-pictorial, and the enactive theory. It does so by offering a consistent explanation for a set of empirical results, which are not plausibly provided by any of the theories individually. We give a brief review of the three theories and summarize their core commitments from a computational modeling perspective. We present a set of empirical results, the different explanations offered by the three theories, and deficiencies of their explanations. The proposed theory and model are introduced and the model's explanatory power is evaluated using the previously identified set of phenomena.

Keywords: Mental Imagery; Cognitive Modeling; Imagery Debate; Mental Representations

Introduction

It is a widely accepted assumption, that we cannot prove the nature of the mental structures underlying mental imagery based on behavioral data (Anderson, 1978). Nevertheless, the large and growing stock of empirical data forces refinement or extension of existing theories to plausibly explain as many results as possible. Recent findings include results from eye tracking studies and brain imaging methods, for example. The three major accounts of explaining mental imagery, namely the descriptive theory (e.g., Pylyshyn, 2002), the quasi-pictorial theory (e.g., Kosslyn, 1994), and the enactive theory (e.g., Thomas, 1999), gain their theoretical relevance by explaining certain experimental results which the other theories cannot with the same degree of plausibility. In this paper, we will focus on a set of such distinguishing experiments. We will show how our new theory, that is implemented as a computational cognitive model, explains, in particular, these empirical results and offers an integrated account of the three different approaches.

We will first identify the core commitments of the existing theories and the relevant experimental studies. We will then describe our theory, its main assumptions, and present the resulting computational model. Afterwards, we will be able to evaluate the new theory against the previously selected experimental results.

Theories of Mental Imagery

We focus on the identification of the core commitments of each theory and do not aim at offering a comprehensive overview, as each of the following theories has several proponents, who themselves shape and interpret the respective theory in different ways. Particularly, we emphasize the implications of these commitments for a potential computational implementation.

Quasi-Pictorial Theory

An example from Kosslyn (1980) illustrates the basic idea of this theory as follows: the answer whether a fox has pointy or round-shaped ears is solved by retrieving the necessary (encoded) visual information from long-term memory and generating a "picture-like" representation of a fox in an internal representation structure, called the visual buffer. This mental image is then inspected to make the information conscious, i.e., the answer is read off the depictive representation. This means, a spatio-analogical mental representation is holding depictive visual information during imagery.

We identify three core commitments of the quasi-pictorial theory: 1) *the existence and usage of a visual buffer (i.e., a spatio-analogical representation structure)*, 2) *the generation of a percept-like activation in this buffer*, and 3) *the active inspection of this "percept" to extract information by processes partly shared with visual perception*. Many different issues have been raised regarding the idea of quasi-pictures being mentally inspected by processes shared with visual perception (e.g., Slezak, 1995; Pylyshyn, 2002; Thomas, 1999). A general problem of the theory, that becomes unavoidable in a computational implementation, is the lack of formalization of the apparent ambiguous nature of mental images: Empirical data indicates that mental images are much like actual images in some respects (e.g., linear scanning time, see Denis & Kosslyn, 1999), but different in other respects (e.g., difficulty of reinterpretation, see Slezak, 1995).

Descriptive Theory

The descriptive or propositional theory is most prominently defended by Pylyshyn (e.g., Pylyshyn, 2002) as the null hypothesis contrasting the quasi-pictorial theory. The main point of the theory is the rejection of a spatio-analogical, i.e., "depictive", representation and the claim that the format of the representations underlying mental imagery are purely propositional. Thus, proponents of this theory interpret empirical data that potentially contradicts a picture-like representation as arguments for the descriptive theory. The descriptive theory was extended with the concept of *tacit knowledge* (Pylyshyn, 1981) to explain (at that time) new chronometric data, e.g., in mental rotation or mental scanning tasks, which arguably pose strong support to the idea of an analogical "percept". It is hypothesized that humans use their tacit knowledge of what it would be like to see something in actual visual perception during certain mental imagery tasks to produce the linear reaction time patterns during mental scanning, for example.

We conclude the core commitment of the descriptive theory to be *the involvement of only non-analogical, propositional representation structures in mental imagery*. From a computational perspective, a propositional representation can in principle represent all necessary visual and spatial information and allow performance of common imagery tasks. Furthermore, specific linear reaction times can of course also be modeled this way. The major problem is the lack of formalization about how, why, and when tacit knowledge is used to emulate analogical reaction times. In general, the descriptive theory suffers not only from the vagueness of the concept of tacit knowledge but also from the fact, that tacit knowledge seems like an ad-hoc extension of a theory to explain new data, which would not be expected nor predicted by the original theory (see, e.g., Kosslyn, Thompson, & Ganis, 2002).

Enactive Theory

The enactive or perceptual activity theory of imagery is based on the idea that visual perception does not employ an internal representation but that the experience of vision is created by the active process of exploring the world, for example, proposed by O'Regan and Noe (2001) as a sensorimotor account of vision. Unfortunately, there is not much literature on this theory compared to the previous two theories and we mostly rely on the overview of Thomas (1999) and the computational model of Blain (2007). Enactive theory assumes that we have sets of inspection processes, commonly called schemata, that are associated with seeing or imagining concepts. For example, we go through the execution of a sequence of schemata that identify the concept "cat" whenever we look at and recognize a cat. The reenactment of these perceptual processes during the absence of a cat is what the theory claims to cause the experience of mentally imagining a cat. These inspection processes partly occur covertly, i.e., internally, and partly overtly, i.e., observable as eye movements.

Thomas (1999) emphasizes that there is no mental representation of the mental image in enactive theory, which contrasts the other two discussed theories. From a computational perspective, we have to clarify that the need for some kind of internal representation for mental imagery cannot be eliminated. However, the proposed kind of representation clearly differs compared to the other theories. Blain (2007) uses sets of inspection processes in his computational model, which give rise to the mental image of a given concept if the respective sequence of certain inspection steps is executed. A computational model does, however, require some kind of memory of the state of the system, i.e., the inspection steps executed so far. Irrespective of whether this is implemented as the state of a dynamical system or in a symbolic way, it does constitute an internal representation, which is linked to the experience of mentally "seeing" the respective object.

Summarizing, we interpret enactive theory to offer us another take on how visual shape information could be stored in long-term memory and processed during imagery: instead of storing and interpreting either a proposition describing a shape or a quasi-picture of a shape, one can *store, retrieve and execute the inspection processes corresponding to the respective shape*. Due to the fact that the few publications on enactive theory are rather vague regarding how exactly the theory accounts for common imagery phenomena and the model of Blain (2007) also does not focus on reproducing these phenomena, it is hard to judge the theory's explanatory power.

Selected Phenomena and Explanations

In the scope of this paper, we cannot pay attention to all relevant empirical results. We picked a limited, yet representative, set of experimental studies. We regard the following phenomena as particularly relevant for the evaluation of our integrative theory, as the contemporary theories differ significantly in their respective explanation (or lack thereof).

Mental scanning

We picked mental scanning as a representative operation for tasks that produce linearly increasing reaction times, e.g., mental zooming. Given a mental image, the time it takes to mentally scan from one entity in this image to another increases linearly with their distance. **Quasi-pictorial theory** straight-forwardly explains mental scanning given that an image-like "percept" is generated in the visual buffer and inspected locally. Kosslyn (1994) assumes an attention window which is successively shifted across the visual buffer. **Descriptive theory** explains these reaction times using tacit knowledge, i.e., subjects emulate what it would be like to scan from one point to another during visual perception and thus take longer the further the distance. **Enactive theory** claims that the executed inspection processes naturally take longer if their purpose (in vision) is to scan further distances. More specifically, this follows from the assumed connection between these inspection processes and the motor processes responsible for movement of the head/eyes.

Mental reinterpretation

Mental reinterpretation is an important phenomenon as it reveals particularly strongly how mental images differ from physical images. It proved to be very hard to reinterpret ambiguous pictures mentally, i.e., by using the mental image of them, while the same images are very easy to reinterpret when shown as an actual image (e.g., Chambers & Reisberg, 1985). Apparently, mental reinterpretation is often possible for very simple shapes, which mostly do not inherit any meaning, i.e., a first interpretation (Finke, Pinker, & Farah, 1989; Slezak, 1995). The **quasi-pictorial** explanation is that parts of a mental image in the visual buffer quickly start to fade so that a reinterpretation of more complex pictures fails due to the inability to hold the complete image. Given that stimuli used in mental scanning and mental rotation experiments are of similar complexity, this argument is not satisfactory, especially since the theory lacks concrete information about the threshold up to which reinterpretation is possible. These results can be interpreted as supportive of **descriptive theory** mainly by the fact that they seem to contradict the quasi-pictorial account. To our knowledge descriptive theory includes no further explanatory elaborations on this issue though. Thomas (1999) describes the **enactive** take on reinterpretation as follows: "If we are only exposed to an unfamiliar figure long enough for us to successfully undertake one of the possible ways of looking at and so interpreting it, only that one way will be stored, and our subsequent imagery will involve recapitulating only that one way, giving us access only to the one interpretation". The successful reinterpretation of simple images, like those of Finke et al. (1989), is only briefly explained by familiarity with the used stimuli, e.g., letters, in different contexts, which then allow different interpretations.

Eye movements

It has been shown that eye movements occur during mental imagery tasks (e.g., Johansson, Holsanova, & Holmqvist, 2006). Furthermore, these eye movements reflect the currently processed spatial relations of the mental image, i.e., if we make an internal attention shift between entity A and entity B, which is on the far left of A, we are likely to make a saccade towards the left of our visual field. **Quasi-pictorial theory** has no problems incorporating eye movements in principle due to the assumed common structures of mental imagery and vision. Mast and Kosslyn (2002) mention that it is possible that eye movements are stored along with the image information. The concrete role between the observed eye movements and the mental image or the shifts of the attention window are not elaborated though. Pylyshyn (2002) as the main advocate of the **descriptive theory** promoted the idea of spatial indexing, which hypothesizes that certain features in our visual field, e.g., a chair or a stain on a wall, are used as indices for parts of a mental image. This could potentially explain eye movements during imagery as well as phenomena like mental scanning. It has been shown, however, that eye movements do occur during imagery tasks in complete darkness (Johansson et al., 2006). **Enactive theory** seems to be able to explain eye movements quite naturally following from its assumption that the execution of schemata employs sensory instruments like the eyes and is thus linked to the respective motor areas (Thomas, 1999). Unfortunately, the literature does not go into detail about this relationship for imagery conditions.

Attention-Based Quantification Theory: An Integrative Model of Imagery

In the following, we introduce our integrative theory of mental imagery, called attention-based quantification theory, and its corresponding computational model. The theory explains the experience of mental imagery in terms of attentional processes that quantify spatial and visual information. We assume two distinct working memory structures: the Qualitative Spatial Representation (QSR), which represents concepts, parts, and their spatial relations on a qualitative level, and the Visuo-Spatial Attention Window (VSAW), which corresponds to an internal attention focus. Three main properties reflect our theory's integrative character and allow it to cover a wider range of phenomena within a consistent explanatory framework compared to the contemporary theories on their own. The first property is that every non-trivial mental image is based on qualitative information in the QSR. It follows that a mental image corresponds to one possible instantiation of that information and is further at least partially limited by the semantics and concepts it is linked to. The second property is the lack of a quasi-pictorial percept, which allows processing of inconsistent mental images and circumvents common flaws of assuming quasi-pictures. The third property are the spatio-analogical attention shifts executed by the VSAW explaining known analogical properties of mental imagery.

As we have seen in the previous two sections, a major problem all three contemporary theories have in common is the lack of formalization regarding their exact mechanisms and structures. As a result, it is possible to extend them with adhoc hypotheses in the light of new empirical data, rather than questioning the core commitments. Given that our theory is, in contrast, implemented in a running computational model, its structures and processes are formalized. Furthermore, a computational cognitive model can be an instrumental source to drive further empirical research as it allows to pin-point open questions and offer concrete assumptions and predictions. The model introduced in this section is not a complete model of mental imagery. It is questionable to which extent such a comprehensive model could be implemented today, as this would most likely require, for example, human vision to be "solved". The model is for these reasons limited in its focus and, for example, only deals with shape information and does not make definite statements about other "visual" information types[1]. Nevertheless, the core commitments of the proposed theory are all implemented in the computational model.

Note that we distinguish two types (and formats) of information in long-term memory (LTM): spatial information, i.e., configurations of complex objects/scenes based on purely qualitative relations structured as a graph, and visual information, i.e., shapes. Shape information is retrieved as a set of vectors of relative length connecting and defining visual features, e.g., a right angle. That is, shapes are represented by how we would look at them. These two information types are accessed independently by the QSR and the VSAW, respectively. In this respect, our theory can be seen as a specific implementation of the dual-coding theory (Paivio, 1971).

Representation Structures

Qualitative Spatial Representation The Qualitative Spatial Representation (QSR) holds active content retrieved from LTM and is implemented as a hierarchical graph structure. It contains the minimal necessary information to generate a mental image. This comprises a qualitative configuration of the imagined scene or complex object, e.g., spatial relations, part-of relations, and relative sizes. The QSR is extended on demand when more details are required, e.g., the concept *house* might initially consist of only *roof* and *wall*, but can be

---

[1] Color, depth and similar other "visual" features could, however, be included quite easily by, for example, propositionally linking them to the retrieved shape information.

elaborated by retrieving, for instance, subparts of *roof*, such as *chimney*, from LTM. The information retrieved with the concept *chimney* would, for example, contain a spatial relation and size relative to its super concept *roof*. Other properties, e.g., orientation, are by default inherited from the super concept. The QSR is used to guide the Visuo-Spatial Attention Window (VSAW) during processing of mental images. Furthermore, it temporarily stores information provided by the VSAW such as concrete coordinates, distances, or new spatial relations between imagined entities. This information can be used for solving a task, feeding it back to the VSAW later, or storing it into LTM.

Visuo-Spatial Attention Window The Visuo-Spatial Attention Window (VSAW) operates as an internal focus of attention and is controlled by the QSR. The VSAW can be best imagined as a circular window. Its implementation comprises a pair of cartesian coordinates and a resolution. The higher the current resolution is, the smaller the radius of the window becomes. High resolution is required to process shape information spanning over a small extent, e.g., detailed textures. High and low resolution can be compared to local and global attention in visual perception, respectively (Shulman & Wilson, 1987). When executing an attention shift the VSAW's coordinates are changed successively as if it was moving in an imagined visual field. We argue that these attention shifts at least partly use processes and structures of visual perception, in particular, motor processes responsible for saccades.

The VSAW serves two main functions during imagery: 1) making qualitative spatial relations or shape information concrete by determining the location, i.e., coordinates, of entities or features of a shape, 2) inferring spatial relations or shape information. We further distinguish the application of these functions on what we term *scene level* and *shape level*. Figure 1 shows an example of these two levels.

On the *scene level*, i.e., the imagination of a scene or complex object, the qualitative relations are made concrete by linking the concepts and their parts to coordinates. This is done by changing the VSAW's coordinates to these respective locations. Consider, for example, the QSR contains that the concept *house* is "to the close right of" *tree*. The concrete coordinates of the center of *house* relative to those of *tree* are calculated by taking into account default[2] translations of "close" and "right" into a vector. The VSAW's coordinates are changed accordingly and the location of *house* as well as the resulting distance are returned and thereby made conscious. The inference of spatial relations can be described as an inversion of the above process. Given the previously calculated locations of two entities, a shift of the VSAW returns the corresponding spatial relation between them.

On the *shape level*, i.e., the imagination of a specific shape, attention shifts are executed based on shape information from LTM in contrast to the qualitative spatial relations given by the QSR. Shapes are represented by a set of vectors that indicate the relative positions between the features of a shape. Concrete metrics, such as the height of a building compared to that of a tree, are made conscious by attention shifts along the given vectors. That is, shapes are "imagined" by making their properties (e.g., height, width, features) available through the execution of attention shifts that define the respective shape. The resulting information is then temporarily stored in the QSR.

We distinguish between covert and overt attention shifts. We predict only the latter to correspond to eye movements during imagery. Covert attention shifts are executed if the to-be-reached coordinates are within the current extent of the VSAW, i.e., the VSAW does not have to move. It follows that a high resolution of the VSAW, as needed for processing detailed shapes, leads to more eye movements during an imagery task as the extent of the VSAW is smaller.

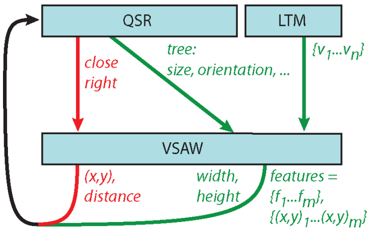

Figure 1: Simplified examples for the scene level (left) and shape level (right) function of the VSAW. On the left a qualitative spatial relation is translated into a concrete distance and coordinates by a shift of the VSAW. On the right the VSAW is provided with a set of vectors representing the shape of *tree*. Together with information from the QSR, e.g., qualitative size, the shape is made conscious by executing the appropriate attention shifts. Metric information (e.g., width) and visual features (e.g., edges) with their respective coordinates are returned.

Explanatory Power of the Model

Mental Scanning and Cognitive Penetration

Extending the classical mental scanning paradigm presented above, we additionally consider an example of cognitive penetration. A task is referred to as cognitively penetrable if a subject's knowledge can affect their behavior in this task. In general, cognitive penetrability of imagery tasks poses an argument against the idea of quasi-pictures. If, for example, reaction times in mental scanning are penetrable this way, then this suggests that the initial linear reaction times could not have resulted from the structural properties of the mental image and the visual buffer. Richman, Mitchell, and Reznick (1979) show how the scanning time of subjects is affected by additional information about distances between places within an image, despite the fact that the distance information is inconsistent with the image itself, as displayed in Figure 2. Neither quasi-pictorial nor enactive theory can straight-forwardly cope with this type of phenomena.

Our model is able to not only reproduce the traditional mental scanning reaction time pattern in Table 1 with a strong correlation of $r = 0.94$ but also the difference in reaction

---

[2] These default values might vary individually and by task.

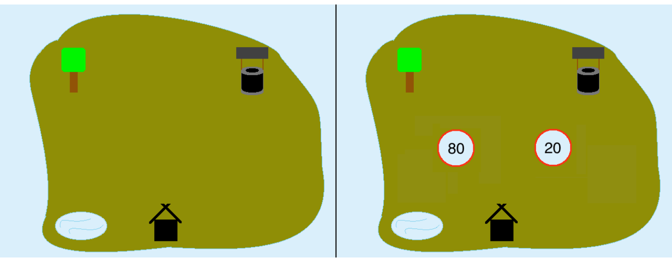

Figure 2: Left: classic island; right: example of an island similar to those used in (Richman et al., 1979)

Table 1: Mental scanning: classic island. The model's reaction times (RT) are affected by noise and averaged over 10 trials. 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

times described in (Richman et al., 1979) in Table 2. But most importantly, the model offers us some understanding as to how cognitive penetration in such imagery tasks might come about. In the presented case, the sign posts are represented as qualitative distance information between the corresponding entities along with the given relation between these same entities and the surrounding island in the QSR. That is, the QSR contains possibly conflicting information and can also generate inconsistent mental images. If the whole island is imagined, each entity is "placed" according to the part-of relations, e.g., *tree* is "in the top left" of *island*. But if an attention shift between *house* and *tree* is triggered, the stored direct relation between these entities is used and the qualitative distance relation, in the case of *tree* "far" (this distance is derived from the "80" sign), is translated into other coordinates.

When Mental Reinterpretation is Possible

The above mentioned distinction between *shape level* and *scene level* processing leads to a concrete explanation for the difference between reinterpretable mental images and non-reinterpretable ones. Figure 3 shows one example for each of the two classes. The crucial difference between these two stimuli, when represented in our model, is that the rabbitduck consists of a set of distinct shapes which are linked together by the spatial and semantic configuration given in the QSR. The heart-shaped image, in contrast, would be represented as a single shape. More specifically, it is not a part of a higher concept like a multi-part object and therefore not bound by semantics. But, in contrast, the shape that is linked with *ears* of *rabbit* in the QSR cannot be *beak* of *duck* (which is a nonexistent concept in the QSR at that moment) due to this exact semantic binding. The reinterpretation of the heart-like shape is realized by a sequence of attention shifts that identify some features and their relative position to each other with a new shape.

Table 2: Mental scanning: (Richman et al., 1979). Reaction times (RT) of the model are averages over ten trials.

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

Figure 3: Left: rabbitduck as used in (Chambers & Reisberg, 1985); right: reinterpretation stimulus from (Slezak, 1995): The right part of the heart-like shape is to be inspected to discover the "2". The left stimulus is very hard to reinterpret mentally, whereas the one on the right proved to be mentally reinterpretable by most subjects.

Eye Movements and Levels of Processing

There are experiments showing the occurrence of eye movements along the processed spatial relations in mental imagery. Our model reproduces these eye movements by overt attention shifts of the VSAW, therefore we want to extend this topic by also asking when meaningful eye movement *do not* occur. Looking through the literature, most reported studies using eye tracking during imagery tasks contain stimuli of relatively great detail, e.g., a fully fledged-out scene (Johansson et al., 2006) or rich descriptions like "Imagine that you are standing across the street from a 40 story apartment building. At the bottom there is a doorman in blue." in (Spivey & Geng, 2001).

We will, however, go beyond the established connection of such detailed imagery and eye movements and look for where to draw the line between imagery tasks which do elicit and those which do not elicit meaningful eye movements. Sima, Lindner, Schultheis, and Barkowsky (2010) report two experiments using three-term series problems of the following form: "A is west of B; B is north of C; infer the relation between A and C". When subjects were only told to solve the problems, eye movements along the given directions were not significant. The second experiment consisted of the very same task, but with the instruction to imagine the entities as red squares like cities on a map with the respective letter next to it. This resulted in a significant amount of eye movements along the given directions. The answer as to why such a minor change in instructions can trigger eye movements can be found in the properties of the VSAW. The more details, i.e., shape information, are processed, the more likely are overt attention shifts. The reason is the smaller extent of the VSAW in its high-resolution mode. The smaller the covered area of the imaginary visual field, the more likely the VSAW has to move to cover new coordinates. For the experiment reported first, no shape information might be processed at all, i.e., entities are merely linked to single coordinates, which allows for a low resolution of the VSAW with a therefore larger extent. The task can in principle be solved without any functional eye movements.

Additionally, it is worth noting that some tasks might also be processed entirely on the level of the QSR. That is, we can distinguish three levels of processing in the model: 1) processing only on the level of the QSR, 2) processing with quantified spatial relations (retrieved from the VSAW) between parts of a scene but without shape information (*scene level*), 3) processing with quantified shape information (*shape level*) either alone or within the context a scene.

Conclusion

In this paper, we have shown how the attention-based quantification theory formalized as a computational cognitive model is able to successfully offer a consistent and plausible explanation for a set of relevant phenomena in mental imagery, which cannot be satisfyingly explained by any of the contemporary theories on their own. The representation structures and resulting different processing levels of our model give new insight on how the mental structures underlying mental imagery can be understood and modeled. Furthermore, the different processing levels offer an interesting new take on the differences and similarities of what is often referred to as, for example, spatial mental models vs. mental imagery or visual/object imagery vs. spatial imagery. Lastly, our model exemplifies a concrete and satisfactory compromise between the problems of assuming quasi-pictures and the questionable assumption of purely propositional reasoning.

Acknowledgments

This paper presents work done in the project R1-[ImageSpace] of the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition. Funding by the German Research Foundation (DFG) is gratefully acknowledged. We thank the anonymous reviewers for their helpful comments.

References

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

Blain, P. J. (2007). *A computer model of creativity based on perceptual activity theory.* Ph.D. thesis, Griffith University.

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

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

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.

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

Mast, F. W., & Kosslyn, S. M. (2002). Eye movements during visual mental imagery. *Trends in Cognitive Sciences*, *6*(7), 271–272.

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

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart and Winston.

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.

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.

Shulman, G. L., & Wilson, J. (1987). Spatial frequency and selective attention to local and global information. *Perception*, *16*(1), 89–101.

Sima, J. F., Lindner, M., Schultheis, H., & Barkowsky, T. (2010). Eye movements reflect reasoning with mental images but not with mental models in orientation knowledge tasks. In C. Hölscher, T. Shipley, M. Olivetti Belardinelli, J. Bateman, & N. Newcombe (Eds.), *Spatial Cognition VII* (Vol. 6222, pp. 248–261). Berlin / Heidelberg: Springer.

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.

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.