On Mental Imagery and its Relevance for AI

Benjamin F. Maier*
Final paper in the course "Philosophy of AI", Department of Philosophy, Utrecht University

(Dated: June 29, 2012)

During the last fifty years there has been a lively debate on the topic of mental imagery, where this term relates colloquially to "thinking in mind's pictures". In this work, the positions of the major parties in this debate are briefly presented, an existing model of mental imagery for symbol manipulators is recapitulated and a slightly new model for mental imagery processing is proposed. Working out the meaning of mental imagery for developments in artificial intelligence (AI) based on the existing positions, we will come to the conclusion that a theory of visual perception may propose advantages for weak AI, while mental imagery itself might be a trick of nature helping humans to solve problems and hence would be unnecessary for computers.

I. INTRODUCTION

Before any statements on mental imagery can be made, we have to present a first indication of the definition of the term. Roughly spoken, the term "mental imagery" how it will mostly be used in this article relates to an eyes-closed picture one has in mind when thinking of a certain object or scenery, without the actual object or scenery being present for the senses to perceive [1, 2]. It is therefore often described as "quasi-perceptual" [3]. For example, given the question "How many windows does your house have?", a subject may first picture her house in the mind and then proceed to count the windows in order to answer the question, which would be some kind of a perceptual experience without the house being present. Although the literature mostly describes images rather vague as sensual experiences without being specific of the kind of interpretation, images will mainly be referred to pictures in this article.

The question whether or not humans think in images arose in ancient philosophy already, being discussed until and throughout early modern philosophy. Using the method of introspection, philosophers tended to declare images as literal pictures in the head, sometimes even more substantial than thinking in language. Later, in the early 20$^{\text{th}}$ century, philosophers could not agree on certain positions, until the debate had generally been declared meaningless since the method of introspection did not provide enough empirical insight [4].

However, during the last 50 years, cognitive scientists/psychologists started the debate again, forming mainly two views, 1) the pictorial and 2) the structural descriptive, where the first view is mainly supported by S. Kosslyn and the second by Z. Pylyshyn. They worked out their views in great detail and supported and disputed them by empirical experiments.

In section II, their views will be summarized briefly, while indications of experiments and explanations of some of the implications will be given. Afterwards a third model of mental imagery is introduced, developed bye M. Tye, which is also a model of pictorial kind, but concentrates more on a connection between pictures and symbol representations. Based on these models I will present my own view on the whole subject, mainly referring to mental imagery as a method to solve problems and give ideas on further experiments to analyze the potential for such.

Afterwards, in section III, the whole topic will be connected to actual and hypothetical developments in weak and strong AI and a conclusion for the value of the mental imagery debate for AI will be drawn.

*Electronic address: bfmaier@physik.hu-berlin.de

II. DIFFERENT VIEWS ON MENTAL IMAGERY

A. The Structural Descriptive View

Defenders of the structural descriptive view claim that images are of an abstract nature. The main defender of this view is Z. Pylyshyn. According to them, mental images have to be interpreted as complex representations of smaller parts, called "propositions", where propositions are informations with truth value, object information and spatial relation and carry semantic as well syntactic information (i.e. a set of propositions connected in two different ways delivers two different representations) [5]. It is unnecessary for the mind to process visual information, since all the information is present in form of propositions. Any occurrence of picturesque perception in the head is epiphenomenal. Pictorial data is not as fundamental as propositional.

To dispute that mental images are induced from something like a visual memory (even a propositional), Pylyshyn performed experiments, such as the "inclined beaker" experiment [6]. He showed an inclined fluid-filled beaker to children of the age of 4, where naturally the fluid's surface was not perpendicular to the beaker's sides. Afterwards he asked the children to draw a sketch of the beaker, which resulted in drawings where the fluid's surface was actually drawn perpendicular to the sides. He concludes that the children were just not familiar enough with the propositional concept of a constant angle of fluid's surfaces, yielding the next close concept of "perpendicular to the side". On the other hand, assuming there would be something like a visual memory, the children would have been able to memorize the actual picture of the inclined beaker and would have drawn it right. Mental image indeterminacy is also explained, using the model of "lack of concepts".

A similar view by G.E. Hinton and L.M. Parsons pro-

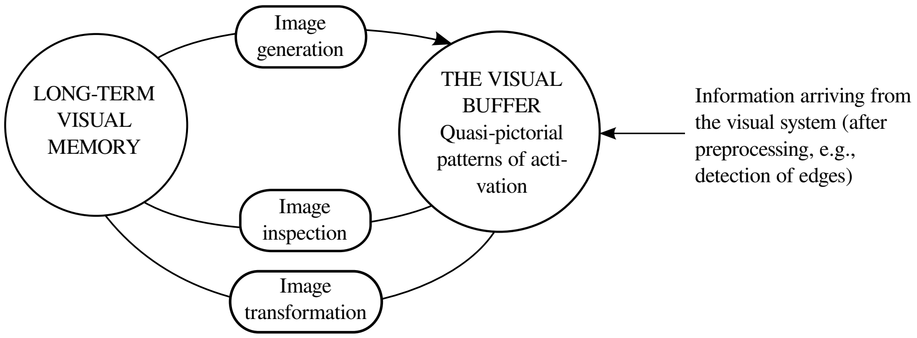

FIG. 1: Chart including the basic components of Kosslyn's theory, recreated from [9]

poses a hierarchical structure for propositional memory [7]. Given a question for a detail of a whole, the mind has to start with a mind-map-like representation of the whole, working its way through the branches to find detailed information, connected to the epiphenomenal activation of picturesque perception. E.g., asked "Do horses have toes?", the mind would start with the whole representational network of a horse, go to the branch "leg" and would eventually find the information "hoof" enabling it to answer the question with "no". Using this view he explains the picturesque perception one encounters through introspection when asked a question of this kind, stating that propositional content is directly used, rather than visual.

The propositional or descriptive view of mental imagery is described very well by J.-P. Sartre, who states "I can keep looking at an image for as long as I wish: I will never find anything but what I put there" [8].

B. The Pictorial View

The pictorialists claim that mental images do have a pictorial representation in the head which is as fundamental as propositions. While some even propose a strict distinction between verbal and picture-like processing, S. Kosslyn, as the main defender of this view, sees mental imagery as both, processing propositions and picture-like information. He introduces so called "surface representations" of information using visual propositional long-term memory with information about spatial relations, which is not available for the mind *per se*. These representations are constructed in a visual buffer and can be seen and interpreted by a functional mind's eye (see figure 1) [10]. Furthermore they may be transformed inside the visual buffer. Asked a question which is connected to a picture, the mind uses that procedure to actually work with the surface representation in order to obtain answers for the problem. The surface representations are described as quasi-pictorial, where M. Tye describes this term as follows.

"A representation R is a quasi-picture of an object O as seen from a point of view V if, and only if,

(i) every part of R that represents anything represents a part of O visible from V;

(ii) a sufficient number of apparent relative surface distance relationships among parts of O visible from V are represented in R;

(iii) for any three represented O parts, X, Y and Z, if X appears at a greater surface distance from Y than from Z, then this fact is represented in R if and only if there are more R parts representing apparently adjacent O parts that are connected by the shortest apparent path on the surface of O between X and Y and that are each of the same length L as measured along that path than there are R parts representing the corresponding O parts of apparent length L between X and Z." [11]

Using this view, a quasi-pictorial is attached with spatial properties, which is a main difference to the descriptive view. To support this view, various experiments have been performed of which I want to introduce two. The first is the rotation experiment [12]. Subjects were asked to look at two pictures of three-dimensional objects, which looked similar, most of them equal but rotated. Then, they were asked whether one of the objects can be transformed into the other. They discovered that the relation between the angle of a rotation and the time it took the subjects to give an answer was linear, indicating that a processing of the image had occurred in order to give an answer.

Some other experiments investigated the scanning of mental images [13]. E.g., subjects were asked to build a mental image of a map with certain objects on it they have been shown before. Then, they had to put a mental marker on one of the map's objects and to push a button. Afterwards, they scanned the shortest path between two objects with the inner marker and pushed the button again, the button responsible for a time measurement. The experimenters observed that the time needed to scan the mental image was linear with respect to the seperation of the objects. Again, they conclude that mental images apparently have spatial properties and are processed by a visual perception system.

With experiments of this kind, Kosslyn tries to dispute the structural descriptive view, because there, all necessary information would be available, without the need of processing information, thus there would be no correlation between answering time and spatial properties.

The pictorial interpretation can be summed up to the view that visual long-term information is as fundamental as purely propositional. Mental images are not only epiphenomenal but actually used by the visual system to work on problems of picturesque kind.

C. Interpreted Symbol-Filled Arrays

In order to assign mental imagery an importance in artificial intelligence, it is necessary to give indications of realizability of mental images within symbol system manipulators (i.e., computers). Obviously, without given realizability, the whole topic becomes redundant. However, Tye gives an explanation of how to create a mental imagery system on a computer in [14]. This subsection is used for a brief recapitulation of his explanation.

Tye proposes such a system based on the visual perception theory by Marr and Nishihara [15]. This perception works basically as described in the following but is described more detailed in the cited sources.

A lattice of sensors (i.e. an eye) measures the intensity of light falling through a pinhole on different positions (thus measures the projection of a three-dimensional (3D) scenery). It then uses edge detection to determine edges, ridges and areas of certain positions of edges and ridges. In the next step, the obtained image is compared with a second image from a nearby second sensor lattice (i.e. from the second eye) in order to obtain information about depth and orientation. Afterwards, the whole image is sectioned into surface patches, each of them carrying symbolic information about the relation to edges and ridges, the orientation of the surface and the depth of its position. Since this is not yet a 3D model of a scenery but rather obtained from two-dimensional images, Marr and Nishihara call it a $2^1/_2$D sketch. In the last step they assume that a proper 3D model can be inherited from the $2^1/_2$D sketch, which is then saved in a database with hierarchical structure, similar to that of Hinton.

So far for the visual perception. Tye proposes that mental imagery is an inverse process. Gathering information from memory, a visual storage is filled with an array similar to a $2^1/_2$D sketch. Each cell contains a surface patch in form of a symbol and additional information such as position, orientation, depth and color. He also proposes that some cells may be empty and others may only be filled with a color and a position in order to cope with the indeterminacy objection on mental imagery. Finally, which is an important point of Tye's theory, the whole $2^1/_2$D array is connected to an interpretation, such as "This is my house", or "This is a frog jumping over a dog".

This model of pictorial representation is rather different from the one of quasi-pictorials, as it adds more information and interpretation to the surface patches. Furthermore Tye claims that it is not necessary that every part of a mental image represents a part of the real object, which is a necessary condition for quasi-pictorials, enabling mental images to be indeterminate.

The advantages of this proposal over the quasi-pictorial view are lying in the additional information of the patches, such as depth. An experiment by S. Pinker indicates that depth is an essential part of mental imagery [16]. He presented test subjects a box, where inside he put object stacks of different size. Asked to remember the box and to scan the mental image with an inner eye, the subjects needed more time if the differences between the stacks's height varied more.

D. Solving Problems with Mental Imagery

In this section I would like to present my own interpretation of mental imagery, which makes use of Tye's interpreted symbol-filled arrays as representations of mental images as well as Kosslyn's mechanism of mental image processing, but is based on a solely propositional view. However, I will rather concentrate on the meaning of processing requests with mental images, namely solving problems. Afterwards I present an objection on former experiments and present an experiment which could indicate answers to the question when and how people use mental imagery processing to solve problems.

In my point of view, the empirical evidence that mental image processing behaves very much as if operations would be performed on actual pictures is strong enough to assume that the pictorial view of mental images has some truth in it: problems in these experiments are solved using the visual perceptual system. However, I do not see any argument to claim that it would not be *possible* to solve the problem without generating mental images, but only by using a system of propositional logic (the possibility is indicated by experiments with congenital blinds [17] and people who reported to have lost their mental imagery [18]). Tye claims that mental images are symbol-filled arrays, associated with interpretations, thus are propositions (in form of symbols). In Kosslyn's view, mental images in the sense of quasi-pictorials are also associated with some propositional value. Both views connect the actual visual perception of a mental image to semantics and symbolics. Information which is obtained by image processing can also be expressed as propositions.

That leads me to the conclusion that long term memory may only contain symbolic data in propositional form, which is on a level of abstraction that the human mind can not work with it *per se* (similar to the view of Kosslyn but with the property of descriptions being the *only* kind of information). Thus, consider the mental imagery processing as described in the following and shown in Fig. 2a. Being asked to perform a certain task, the mind sends a request to its fastest computational system. This computational system is assumed to be the visual system, since it is the most powerful evolutionary tool for humans to coordinate in their environment. However, the best computational system may differ from person to person due to training in a subject's early development. But let us now stick with the visual system. The request is sent to a visual processor, which needs data to perform a computation. This data request is sent through an image interpreter, forwarding it to the long term memory, which delivers data in propositional form. The data received from the long term memory has to be interpreted, which is done by the image interpreter inside the visual system, loading a $2^1/_2$D array into the visual buffer. The visual processor may now perform operations on the image in the visual buffer, and, if necessary, send requests to the interpreter, which reinterprets the request into propositional form, gathers the necessary propositional information and creates a new or varied image in the visual buffer. The visual buffer is observed by the functional mind's eye and reflected in the conscious or subconscious mind, which can now send further or varied

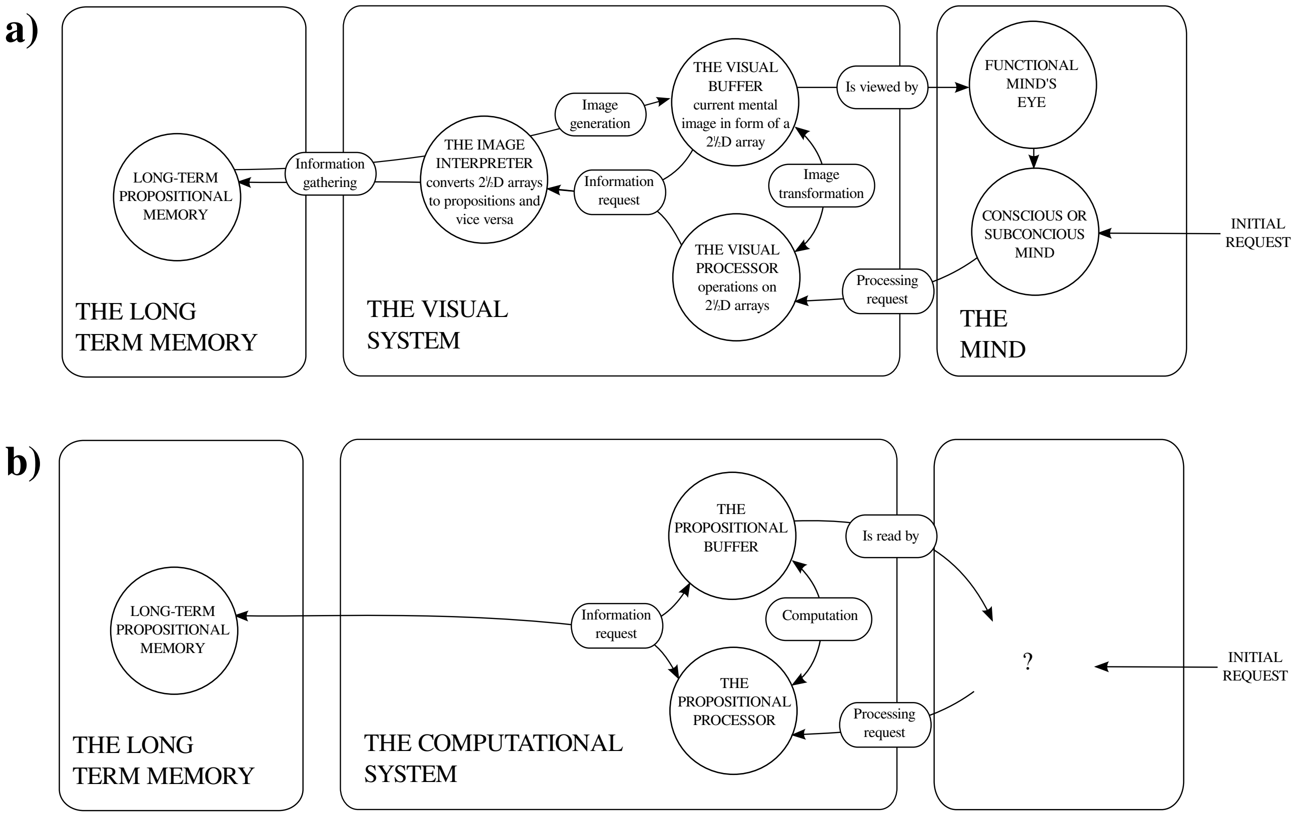

FIG. 2: Chart illustrating **a)** my proposal for a model of mental imagery processing and **b)** the amputed model for an implementation on a computer.

processing requests or do nothing.

A central question is how the interpreter actually yields the image from the propositional data. If this is done via an extra visual memory, the whole process becomes just a complication of the Kosslyn-Tye model. An argument which speaks for my proposal of mental imagery is that data storage would be efficient. Other perceptual systems could use the same data storage and the same mental processing (yielding mental images of another kind of perception) to solve problems. Furthermore it explains most of the experiments. Scanning and rotating mental images are picture operations, thus it is no surprise that they are solved by mental imagery processing. The $2^1/_2$D array as representation of the mental image yields indeterminacy and depth [19]. Pylyshyn's rotated beaker experiment is explained by the lack of concepts available for children – no visual long term memory is there to create the picture, only graphical interpretations of propositions. Objections by Kosslyn disputing the fundamental nature of propositional memory with rotation experiments are invalid, since no propositional memory is directly available for the mind. A representation has to be obtained, given by the visual system.

However, a problem that one encounters during work with the imagery debate is the following. The literature only presents experiments in which subjects are explicitly requested to solve problems with pictures (such as Kosslyn's rotation experiments or Pylyshyn's beaker experiment). Hence, all subjects were set to solve a pictorial problem. The method of mental imagery processing seems to be the most likely one to solve these. However, the question whether my proposed model of mental imagery processing is right, as well as the question whether we *actually* think (i.e. solve problems in everyday life) with mental images, being already an issue of discussions in the late 1800s [20], remains unsolved.

To give indications of the answer to these questions, I propose experiments which could be designed as follows. Given a number of subjects, ask them to perform some of the experiments mentioned above and record the brain activity while doing so. I would expect that the visual cortex is involved in solving the problems. Save the data for comparison.

In a second experiment, ask the subjects to answer more general questions, which cover experience from more perceptions than the visual, for example

(a) "Do men have a deep voice?", "What is the capital of [subject's home country]?"

(b) "What is the result of 3+6?", "You buy a shirt which usually costs 50$, but there is a discount of 50%. How much money do you save?"

(c) "How many windows does your house have?", "Do horses have kneecaps?"

Questions of kind (a) usually would not need any deep processing, since the answer should be given from everyday experience and therefore accessible quite quickly. Nevertheless it would be interesting to see whether or not the visual cortex becomes active (maybe at the words "men" or "home country"). It would indicate that memory activation is connected to visual experience. However, this activity could just be of epiphenomenal nature.

Questions of kind (b) are expected to only need processing in form of propositions. However, graphical representations might be handy to solve the problem (for example imagining to actually stand in a shop and consider the problem being real. This would actually connect the problem to other qualia, too, such as a good feeling for saving money). The region of brain activity during the problem solving should be measured, as well as the time to deliver an answer. This could provide an answer whether or not those problems are solved in connection to visual perception, or whether there are different kinds of people, whose approach of solving problems is different, too. Eventually one could obtain a correlation between the speed of processing and the different approaches, meaning to perform a correlation analysis similar as it was done in [21], where a correlation between the vividness of a visual perception during mental image processing and the brain activity was analyzed using *fMRI*.

Eventually, the (c) kind of questions are expected to be solved using mental imagery, such as imagining a house and counting the windows. However, other solutions for the problem are possible. One could perform these tests first with people who are able of sight and second with congenital blinds or people who lost their mental imagery system (as described in [18]). This method would give information of which imagery system (visual, non-visual) is actually faster in solving these problems, or whether there is any difference in brain activity at all (there have been indications that the imagery system of congenital blind people works differently [17]).

III. MENTAL IMAGERY IN RELATION TO AI

Obviously, the problem of mental imagery is one of cognition of the human mind. Thus analyzing it with respect to AI means that we have to answer two questions.

1. What is so essential about mental imagery for the human mind?
2. Will it help a system of AI to perform problem solving using the concept of mental imagery?

First, we can state that mental imagery is apparently a broadly used tool to solve problems of a certain kind (referring to the mentioned experiments). We should continue to analyze both question for both remaining mental imagery systems.

In the following I want to give arguments on the redundancy of mental imagery for AI when using my model of mental imagery processing, as it is represented in figure 2b.

Supposing that my proposed model of mental imagery is right, nature can be blamed to use tricks to save data storage and uses visual mental imagery as a fast computational process to obtain answers to problems.

Imagine a machine constructed on purpose (i.e. not a "natural machine"). It could be constructed to be able to interpret propositional data immediately. Thus, an "image interpreter" becomes redundant, and a processor as well as a buffer would be able to handle the necessary processes immediately. Furthermore a "mind" as in the former model becomes redundant as well (because the propositional data is available and understood all over the system, the machine does not need an "inner eye" to inspect the data).

However, the machine would have to handle a request. Because this request is probably asked by a human being or by the environment, a perception system has to cope with the natural, non-propositional communication. To this end, it actually would have to simulate perception systems such as visual or linguistic ones, to translate it into propositions first and to perform problem solving afterwards.

On the other hand, if Kosslyn is right and pictures are as fundamental as propositions, we as humans may possess something like visual memory. That means that we actually use pictorial representational content which might be represented in Tye's arrays. But what does that mean for a computation? The arrays themselves are symbols, thus they can be stored in a computer. But that only means that the computer acts on propositional values of another representation.

How does this differ from the implementation of the case before? It differs in a way that we would have to decide which of the two models solves problems faster. If we would implement propositional content in a graphical representation, a computation on it is still a computation on propositions. Thus, I would expect the pure propositional content to be processed faster, since it seems to be an easier construct of data and thus easier to interpret for computers. Nevertheless there could be a loss of speed in implementing the necessary weak AI systems for contact with the environment (which can be implemented with visual perception systems similar to those of Marr – a topic of current research!).

IV. CONCLUSION

To sum up, I would state that the insights of dealing with mental imagery may help weak AI to cope with its environment. For the development of strong AI it seems unnecessary since computations are either way performed on propositional (i.e. symbolic-semantic) data.

Conferring to the different models of mental imagery processing, performing the experiments described in the last section could give indications of which model is rather true and whether "thinking in images" is a usual approach to solve problems.

[1] Wikipedia Foundation, "*Mental image*", available here: `http://en.wikipedia.org/wiki/Mental_image`, last view: June 2012

[2] MacroVU, Inc.,"*Mapping Great Debates: Can Computers Think? – Map 5B*", available here: `http://www.macrovu.com/CCTWeb/CCT5B/CCTMap5B.html`, last view:

June 2012
[3] see [22] §1.1
[4] see [22] §1
[5] see[19], p. 61-64
[6] see[19], p. 68-71
[7] Hinton, G.E.; Parsons, L.M., (1981), "*Frames of reference and mental imagery*", J. Long, A.D. Baddeley Eds., Attention and Performance IX, Erlbaum, HillsDale, pp. 261-277
[8] Kossyln, S., Thompson, W.L., Ganis, G., (2006), "*The Case for Mental Imagery*", Oxford University Press, New York, p. 63
[9] see[19], p. 43
[10] see[19], pp. 33-60
[11] see[19], p. 36
[12] Shephard, R.N.; Metzler, N., (1971), "*Mental Rotation of Three-dimensional Objects*", Science 171, pp. 701-703
[13] 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, Vol. 4, No. 1, pp. 47-60
[14] see[19], , pp. 77-102
[15] Marr, D.; Nishihara, H.K., (1978), "*Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes*", Proceedings of the Royal Society of London. Series B, Biological Sciences, Vol. 200, No. 1140, pp. 269-294
[16] Pinker, S., (1980), "*Mental Imagery and the Third Dimension*", Journal of Experimental Psychology: General 109, pp. 354-371
[17] Kosslyn, S.; Arditi, A.; Holtzman, J.D., (1987) "*Mental Imagery and Sensory Experience in Congenital Blindness*", Neuropsychologia. 1988;26(1):1-12.
[18] Thomas, N.J.T, (2001-2012), "*Are There People Who Do Not Experience Imagery? (And why does it matter?)*, Addendum (x)", `http://www.imagery-imagination.com/non-im.htm#BrainDamage`, last view: June 2012
[19] Tye, M., (1991), "*The Imagery Debate*", MIT Press, Cambridge, Massachusettes
[20] see [22] §1.2
[21] Cui, X.; Jeter, C.B.; Yang, D.; Montague, P.R.; Eagleman, D.M., (2007). "*Vividness of mental imagery: Individual variability can be measured objectively*", Vision Research, 47, pp. 474-478
[22] Thomas, N.J.T., (2003/2005), "*Mental Imagery, Philosophical Issues About*", Encyclopedia of Cognitive Science, Volume 2, pp. 1147-1153, London/Hoboken, NJ