A Cognitive Framework for Modeling Mental Space Construction and Switching During Situation Assessment

James L. Eilbert, James Hicinbothom

CHI Systems
1035 Virginia Dr, Suite 300
Fort Washington, PA 19034
jeilbert@chisystems.com, jhicinbothom@chisystems.com

Abstract

A process is described for selecting the subset of internal models needed to perform situation assessment and organizing them in a representation we designate a mental space. A cognitive framework is proposed that supports the interacting processes of mental space construction and situation assessment. This framework is consistent with a set of design principles derived from a range of psychological and physiological evidence. The framework describes processes for assessing situations from noisy, incomplete, and incrementally arriving sensations.

Introduction

Most of the time, people have an awareness of current context, i.e. they are aware of where they are, what are the important or valuable objects near them, useful landmarks, and the boundary beyond which events can probably be ignored. They also understand when events have occurred that force them to switch to a new current context. In the psychological literature (Fauconnier and Turner 2002), the internal representation of current context is called a mental space. The concept of situation awareness is closely related to the mental space concept. According to Endsley (2000) situation awareness has three components – recognition of the elements in the current situation, comprehension of the current situation, and prediction of future situations. To comprehend or predict situations, situation awareness must include the knowledge of one or more partially executed stories (Lehnert, and Loiselle 1989), plans or episodes in the current context. We argue that stories can cut across contexts and that contexts can contain multiple stories, so that selecting a current context and assessing a situation are really distinct but interacting processes. Our objective in this paper is to suggest a cognitive framework that supports these interacting processes, and to describe how the framework deals with the various problems people face when estimating the situation from sensations. The framework we seek needs to work with noisy, incomplete, and incrementally arriving information; to specify requirements for internally stored information and mechanisms for learning them; and to characterize the processes that lead to the selection of current context and situation assessment.

Hard, Interacting Problems in Situation Assessment

All information coming to a person from the external world is sensed. Analysis of sensations, which could come from imagery or intelligence reports or other sources, requires the simultaneous solution of a number of hard interacting problems. Some of these problems are associated with characteristics of the data, while others are associated with using internally stored models to either perform recognition or make predictions based on the data.

The problems associated with the data itself, that analysts or data fusion systems must deal with include:

- Extracting useful signal from the noisy, cluttered, incomplete information from sensations that arrives incrementally over time.
- Segmenting pixels into objects (or at least surfaces), and people in intelligence reports into groups. People must do this even when there are unfamiliar or novel objects, groups, or cases
- Correlating objects in pairs of images (or reports) arriving at different times.

Recognition of objects in incomplete, noisy data is known to be an inverse problem (Granlund & Moe, 2004). This means that there is no general way to determine whether a pixel in an image (or a person in a report) is part of a threat object (or a threat organization) without recognizing the object (organization). The models needed to solve inverse problems can be considered descriptive models that include information about the attributes of the modeled object that distinguish it from its background. Some of the problems associated with using models of known objects or events to recognize occurrences of those objects or events in the stream of evidence include:

- Making partial matches between sensation and models
- The very large number of models people can learn.

The basic issue in making predictions is, given a set of objects that were recognized and relationships among them, how would the situation change if a particular behavior were executed. However, prediction requires generative models of both the object and the behavior. Generative models are significantly different from the descriptive models needed for recognition. They can ignore attributes that are primarily useful for separating an object from its background, and focus attributes that are important for the behaviors that must be simulated. Thus, a descriptive cat model may focus on visual texture and shape, while the generative model associated with jumping behavior might focus on weight and leg strength. Generative models for planned behavior are far more complicated, since they involve many branches where an acceptable (if not optimal) behavior must be selected from a number of options. As planned behaviors are executed, events occur in the world that force replanning or the selection of contingent branches of the plan. After a short while, it becomes impossible to determine whether any particular event is related to the execution of the original plan or not. Thus, problems with prediction include:

- Modeling the contingent and flexible nature of planned behavior.
- The large number of unknown objects and behaviors, although there may also be a large number of known ones.

People provide the only working implementation that can deal with all of the hard problems listed in this section. We therefore spend much of our effort on examining psychological and physiological evidence about how selection of current context and situation assessment work in people.

Mental Spaces as a Way to Organize Context Information

The work on cognitive maps and mental spaces provides an extensive literature on how people organize and use the models commonly considered to be context information. Tolman (1948) developed a theory of Cognitive Maps to model human spatial reasoning. Fauconnier and Turner (2002) have written extensively about a related concept that they call a mental space. The information stored in a mental space includes the boundary of the relevant region, relevant objects and landmarks within the region, and the spatial relations among them. The mental space will also include routine behaviors that store control knowledge about how to move around the region and strategies for collecting evidence. A set of plans or stories that can be carried out in the region is learned as part of the mental space. Studies of the plot units that make up stories (Lehnert & Loiselle, 1989) give a good indication of the level of detail and the type of situations that should serve as nodes in a story.

People can switch among many overlapping mental spaces of various sizes and scales. Analysts are aware that people often get stuck in the wrong mental space. For example, in explaining an analytic error Shlaim (1976) that, “Since the facts do not speak for themselves but need to be interpreted, it is inevitable that the individual human propensities of an intelligence officer will enter into the process of evaluation”. On occasions when people are dropped into situations where they have no appropriate mental space, they encounter problems estimating the situation and making appropriate decisions. When this occurs people will quickly search for a more appropriate mental space. Plots and stories that come from large, more abstract mental spaces can cross many more spatially and temporally limited mental spaces. Thus, stories can be used to direct the transition between more localized mental spaces where perceptions are really being analyzed.

Some mental spaces are much simpler than others. When surprised or terrified, people tend to drop into simpler reactive states, where everything except what they are scared of is background or obstacle, and the only behaviors accessible are simple variants of flight or fight reactions. After a few moments, and some distance is placed between the frightening object, people will switch back to a richer mental space. However, their analysis of the situation, and their choice of behaviors, is likely to be biased by the frightening experience well after they are able to do more normal situation estimation.

One last property of mental spaces that Fauconnier and Turner (2002) discuss is the notion of blends, where two or more mental spaces are combined into a single space. He claims that the origin of language results from the ability to form complex blends and run them.

Principles and Requirements for Mental Space Construction and Situation Assessment

In this section, we will present some requirements for building and switching among mental spaces. We will provide supporting psychological and neuroscience evidence for various aspects of our model of mental spaces.

We begin with the non-controversial assumptions that people store spatial memories (SM) of the world, and that they can learn at least simple behaviors. SMs are populated with models of landmarks, and objects of interest or value. Storing a SM of the environment that is independent of an animal’s focus of attention is a biologically old capability. There is evidence that an actual map is stored in the brain, at least in mammals, (Gallistel, 1990; Kaplan and Kaplan, 1982). People seem able to form a number of distinct SMs, and to augment them over time (rather than learn them all at once). At the cellular level, we see evidence of this capability in hippocampal neurons that respond to your place in the world or a relationship among several objects independent of which way you are facing (Hargreaves, et al., 2005).

Our first principle for mental space representation and construction is that SMs provide an information source and grounding for mental spaces, and object models in an SM really consist of paired descriptive and generative models.

This notion of paired descriptive and generative object models is consistent with the dissociation of place and recognition systems within the limbic system (Morris and Parslow, 2004). The primary functions of descriptive models of objects are to recognize occurrences of the object from sensations (Granlund and Moe, 2004), to register the SM with the world, and to determine if the world agrees with predictions made by running generative models.

In order for an SM to be useful to a person during situation assessment, there must be a process for matching a subset of descriptive models from an SM to the world and a registration process that uses these matches to fully register the SM to the world. In a sense, the registration process allows the world to act as an external memory store (Smith, 2003) that can fill-in important information that is missing in the SM. On the other hand, registration supports learning important descriptive models of objects and storing them in the appropriate location within the SM. There is also some evidence about how this registration process works in humans during everyday perceptual activity. Pylyshyn (2000) has argued that human vision makes a direct, pre-conceptual connection between objects in the visual world and their internal representations. In others words, people can track and locate a small set of landmarks or "deictic pointers" across eye movements.

The primary functions of the generative object models are to predict the future and to explain why past decisions were made. To produce predictions there must be a simulator that can run behavioral models by manipulating generative object models from the SM. We believe that it is the simulation of behavioral models working on generative object models from the SM that are stored in episodic memory. We conjecture that object models are learned and added to an SM while the SM is registered to the world, if the objects impact the behavioral simulation remembered in episodic memory. We would also expect that the links between the behavioral model and the objects that affect it are learned at the same time as those objects are added to the SM.

Brain lesion data (Aggleton & Brown, 1999) indicates that spatial memory is learned separately from episodic knowledge. However, acquiring episodic knowledge seems to require an SM (Tulving 1985). Episodic memory, in turn, seems to be required to form mental spaces (Fauconnier and Turner, 2002).

Given this view of SM formation and its interaction with internal behavior simulations (whose results are stored in episodic memory), situation assessment requires a second type of registration, i.e. registering behavioral models with the generative objects in the SM. Before discussing this second type of registration, we need to distinguish behavior from behavioral models. In this paper, behavior is defined as an action in the world that can be modeled as a sensory-guided control algorithm. We argue that mammals learn behaviors in a particular mental space, and the sensations that drive the behavior are limited by the sensations that are relevant in that mental space. Recent work on the learning curve for individual animals during conditioning experiments shows that, at least for mammals, once they "get it" they complete learning in just a few trials (Gallistel, Fairhurst, and Balsam, 2004). We interpret this to mean that conditioning occurs very quickly, once the right mental space and the right subset of sensations to focus on have been found.

We define an internal, generative model of a behavior called a mission planning template (MPT). A simple MPT is defined as a triple, where the triple has the form [start state, agent behavior template, and stop states]. The behavior template in the triple is an operator that simulates the effect of carrying out a sensation-driven behavior on an object. Each start or stop state must be a generative model of an object, place, or condition that can be registered to a generative model in the SM. In addition, for a triple to be a valid MPT, it must be possible to execute the behavior in the start state. The start state can be expressed as a conjunction of conditions on state variables, while the stop state can be expressed as a conjunction or a disjunction of conditions (a success state and various failure states). Thus, there may really be a number of distinct stop states associated with one MPT triple.

A mental space can be generated from an MPT, where the spatial boundaries of the mental space are determined by the positions reachable while executing the MPT, and the object models that must be included in the mental space are those objects whose presence can modify the execution of the behavior. A local context can be grown by adding MPTs to it, where the new start and stop states are objects or places in the previous mental space. We argue that this process will tend to divide an SM into nearly isolated sub-spaces. There will be many behaviors performed in a bedroom, but only a few composite MPTs that link bedroom objects to ones outside the room.

A composite MPT is a model of planned behavior that includes an appraisal process that allows one behavior to be selected from among the set of eligible behaviors at a decision point. To be an eligible behavior the stop state of the last triple must match the start state of the eligible triples. The boundary of the mental space associated with a composite MPT is the union of all the boundaries of the eligible MPTs it contains. Similarly, the objects contained in the composite MPT are the union of all the objects contained in the eligible MPTs.

In order to run or simulate an MPT, it must be possible to register the start of the MPT with an object in the SM and to recognize a stop state when it is reached. Running an MPT would allow a person to predict the effects of a behavior or generate expectations. In order to determine if a prediction is accurate or an expectation has been satisfied, the system must be able to retrieve the descriptive model of the predicted state and match it directly against the world.

Mental spaces can be built on-the-fly based on the behaviors that need to be run in them. Fauconnier and Turner (2002) give numerous examples of peoples' ability to do this. We believe that the mental space is built when a behavior to run is selected. The objects that affect the behavior to be run when it was learned are added to the mental space and the spatial boundary of the mental space is set large enough to contain the objects. The difference between the SM and the mental space of the same room is that the SM is really a spatial store while the mental space involves a dynamic retrieval of SM information using links to behavioral models that the person or agent wants to simulate. When using an internal model of a room as a mnemonic device, the behavior one simulates is a walk through of the room. Thus, a mental space temporarily calls up a subset of the objects in an SM. Mental spaces associated with very complete SMs, such as your own office or rooms at home, can be called up very quickly compared to, mental spaces that must be constructed from scratch. A mental space can continue to grow as alternative behaviors are run and appraised in order to decide on the best choice of behavior. In addition, if some of those objects are agents in their own right, then the mental space may need to be expanded to incorporate the objects affecting the behaviors of those agents.

In considering the terrorist bombing domain, we see that the mental space of a terrorist trying to blow up a well-guarded building is quite different from those of a terrorist trying to inflict casualties near the building. Different characteristics of the world matter to the different terrorists and even the same characteristics may trigger different decisions. The thickness and material in the building walls is important to the terrorist attacking the building, but irrelevant to the one going after people outside the building.

The underscored principles and requirements in this section, when taken together, describe a bootstrapping process for dynamically constructing complex mental spaces that contain all the internal models needed to do situation assessment. All that a system following this mental space construction approach needs to learn are descriptive and generative models of objects and control laws for behaviors and their corresponding simple MPTs.

There are some significant differences in constructing mental spaces from imagery as opposed to reports. When starting a conversation, people generally pass information about the mental space they are operating in as part of the dialogue. Thus, when a person makes a query about a situation, they will generally provide information about the context in which the assessment is supposed to be made. Clearly, this is not the case with imagery where objects and behaviors must be extracted before a mental space explaining them can be constructed.

We also believe there are some important connections between the concept of a mental space and ontologies. The ability to learn categories, such as tools or house pets, does not seems to be an innate human capability. Luria (1976) found that adults with no formal education could not do this type of classification and had little ability to learn how to do it. Learning ontologies seems to involve the ability to generalize across mental spaces. Godel's theorem essentially tells us we can't construct global ontologies that are both complete and consistent. On the other hand, we conjecture that you can find a consistent and complete ontology that will cover the contents of any mental space.

Human Strategies for Doing Situation Assessment

Since people can recognize objects (much of the time) and do situation assessment, they must be able to solve the ubiquitous inverse problems found in recognizing objects in sensation, and the issues in making plausible predictions. Perceptions carry only partial information about object appearance, movements, and activities; the rest of the information about the current situation is filled in from memory or knowledge. Intelligence analysts have long been aware that good analysis requires adding their own a priori knowledge to the evidence they receive and considering the meaning of the evidence in a very specific context. For example, Heuer (1999) gives the following description of the role of context in doing analysis -- "According to [one] view, objectivity requires the analyst to suppress any personal opinions or preconceptions, so as to be guided only by the 'facts' of the case. To think of analysis in this way overlooks the fact that information cannot speak for itself. The significance of information is always a joint function of the nature of the information and the context in which it is interpreted. The context is provided by the analyst in the form of a set of assumptions and expectations concerning human organization behavior. These preconceptions are critical determinants of which information is considered relevant and how it is interpreted." The question we would like to answer is what are the strategies and techniques the people use to solve these problems.

Underlying our framework for automated situation assessment is the assumption that rather than rebuilding an understanding of the current situation each time new evidence arrives, humans update and refine their prior understanding of the situation as new evidence arrives (Eilbert, Hicinbothom, & Karnavat 2005). Further, the knowledge about the situation, which is maintained from update to update, impacts even the early stages of sensory processing. This update process provides a type of regularization technique that various researchers have shown can be used to solve inverse problems (Poggio and Edelman, 1990). In our framework, a situation update is based on a combination of previous estimates of the situation, events occurring since the last estimate of the situation, and the current context.

People can segment the world into homogeneous regions based on various combinations of visual field attributes (brightness, color, disparity, edge or line orientation, velocity, etc.). In some cases, the segmented regions will correspond to objects and in other cases they won't. The chances of the segmented regions being complete objects are much better in familiar than in unfamiliar environments. We also know that inferotemporal brain damage can prevent people from assembling the component surfaces of an object into a complete object (Luria 1973). On the other hand, people can suppress a segmentation and search for individual objects as they do in puzzles where you have to find hidden objects. Since the puzzles violate standard segmentation rules, people need pictures of the hidden objects to be found. This is very different from pictures with anomalies where people just need to know there are anomalies, and the inappropriate regions seem to pop out.

Perceptions also arrive sequentially, so even what a person currently perceives is sewn together from discrete perceptions arriving over some period of time. People use process control knowledge, which is organized into routines, to determine the most useful sequence of perceptual views of the environment. The process control knowledge captured in routines, along with the knowledge about plots or plans that are used to make inference when updating or revising our estimate of the current situation are only valid in particular spatial locations. People who study situation awareness (SA) are aware that there is a cyclic process involving what you know and what you perceive. “People are very active participants in the situation assessment process, with SA guiding the process and the process resulting in the SA” (Endsley, 2000).

In addition to psychological and neuroanatomical evidence about the steps involved in each situation update cycle, there are important computational reasons for this particular sequence of steps. Performing processing in a bottom-up fashion, starting with rare indicators rather than working top-down from situations, greatly reduces the amount of object matching that needs to be done. Thus, the cycle of situation update requires comparing the relatively small number of segmented regions (Step 1 in Figure 1) to a relatively small number of low-level models relevant in the current situation (Step 2 in Figure 1). This greatly reduces the complexity of the partial matching problem. Simple components can then be boot-strapped to find more complex objects. Semantic labels for objects and actions can then be compared to relevant stories or plans in the mental space, in order to build global descriptions of the situation. Thus, we claim that it is the feedback between our projection of the current situation and its support from sensory evidence that allows us to make use of our a priori knowledge and to perceive the world as consistent.

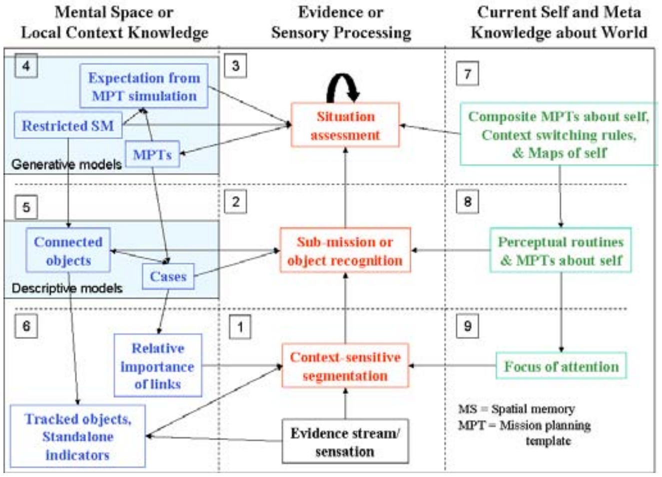

Figure 1: Framework for the Situation Update Cycle

The specific steps in the situation update cycle are:

1. Select a set of tracked indicators, and use a context-sensitive segmentation technique (e.g., Udupa, Saha & Lofuto, 2001) to decide which events and/or objects should be associated with which indicator or with the background. The regions that emerge from segmentation depend on the indicators and the relative strengths associated with different kinds of links.
2. Register the regions or the strong indicators themselves with indicators in images or maps, with a large area of overlap with the current image. Match models of behaviors or objects of interest against the regions resulting from the segmentation process, as opposed to matching them against the whole image. Check for predicted states and objects first. Completed behaviors are expressed as [Start, Behavior, Stop] triples. For on-going behaviors, Stop would be replaced with Continuing.
3. At this level, the system is looking for coordinated activity among objects that do not have to move in a coordinated fashion. Matching at this level is not done directly against sensations. Instead, the descriptive models are converted to generative models; the generative models are then used to model the decisions or the coordinating forces that are causing the coordinated activity. When the coordinated activity matches the predictions from the previous update cycle, extend predictions. For states that don’t match predictions, determine if the deviation is due to an incorrect prediction about behavior selection, an incorrect behavior ID during the previous cycle, an object starting a new behavior, or an unforeseen interaction. Since sensations arrive sequentially, a person’s estimate of the current situation is sewn together from a number of cycles of situation update.
4. Convert the predicted states coming from the generative model into descriptive models within the current mental space that are the expected states for the situation on the next cycle.
5. Based on the expected situation and spatial memory, predict expected objects and behaviors for the next cycle.
6. Based on the important objects in the current mental space, set relative link importance for segmentation on the next cycle.
7. (8 and 9) Select visual routines that can be used to distinguish between competing explanations of the current evidence. Note that one visual routine may direct attention over many situation update cycles.

In Figure 2, we show in more detail the process of mental space construction and switching. An MPT and associated generative models can be simulated, leading to an expectation. This in turn leads to feedback through the mental space to find the descriptive models associated with the expectation. The current mental space will switch if there are large discrepancies between the current and predicted sensations. or to predicted changes in the situation. Even if the sensation matches the prediction, the mental space may need to switch if the MPT simulator predicts that the current behavior will move the agent outside of the bounds of the current mental space.

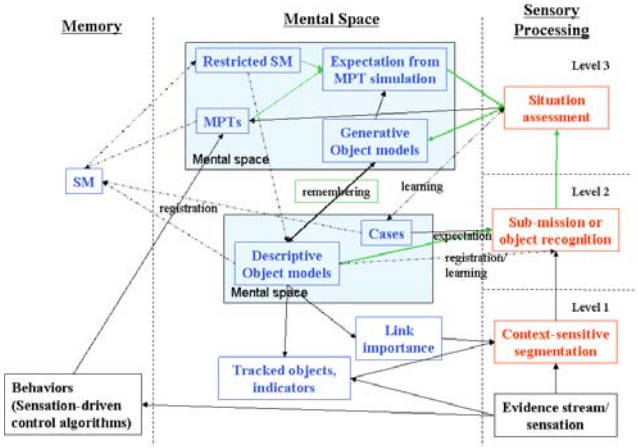

Figure 2: Combined Mental Space – Situation Assessment Framework

The current situation assessment will depend on the previous situation assessment, discrepancies between the current and predicted sensations, and predicted changes in the situation. For example, the mental space would need to be changed substantially when the MPT simulator predicts that the behavior being carried out will move the person (or agent) outside of the bounds of the current mental space.

While mental spaces are built on-the-fly, in familiar environments like your own bedroom, both the spatial memory of environment and the behaviors that the objects in it will support can be remembered together. The bedroom will become a rich and permanent mental space rather than one that is built on-the-fly in order to run a set of MPTs.

Conclusions

We do not claim that the framework we have described is a complete model of the human analytic process. However, we believe that the design principles underlined above are well supported by the biological evidence. Further, we have shown that major components of the framework can be implemented (Eilbert, Hicinbothom, and Karnavat 2005). The framework also applies to a wide range of information sources, including intelligence reports and imagery. We are currently working on developing an end-to-end implementation of the framework.

Acknowledgments

This research was sponsored and managed by Air Force Research Laboratory, Rome Research Site (AFRL/IF/RRS) under contract F051-104-1318. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of Air Force Research Laboratory, or the United States Government

References

Aggleton, J.P., Brown, M.W. 1999. Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. Behavioral and Brain Science 22(3).

Eilbert, J.L., Hicinbothom, J., Karnavat, A. 2005. A Cognitive Architecture Focusing on Situation Estimation: Merging Perceptual and Cognitive Processes, Behavior Representation in Modeling and Simulation. In the Proceedings of BRIMS 2005.

Endsley, M.R 2000. Theoretical underpinnings of situation awareness: A critical review, In: Situation Awareness Analysis and Measurement. Endsley, M.R and Garland, D.J. (eds.), Mahwah,NJ, Lawrence Erlbaum Assoc.

Fauconnier. G., Turner, M. 2002. *The Way We Think*. New York: Basic Books.

Gallistel, C.R. 1990. The Organization of Learning. The MIT Press, Cambridge, MA.

Gallistel, C. R., Fairhurst, S., Balsam, P. 2004. The learning curve: Implications of a quantitative analysis. PNAS. 101(36):13124–13131.

Granlund, G.H., Moe, A. 2004. Unrestricted recognition of 3D objects for robotics using multilevel triplet invariants. AI Magazine 25(2):51-

Heuer, R. 1999 *The Psychology of Intelligence Analysis. Center for the Study of Intelligence*, CIA. http://www.cia.gov/csi/books/19104/index.html.

Lehnert, W.G., Loiselle, C.L. 1989. An introduction to plot units. In Waltz, D.L. (ed.) Se-mantic Structures: Advances in Natural Language Processing. Lawrence Erlbaum Associates, Inc. Hillsdale, NJ.

Luria, A.R. 1976. *Cognitive Development: Its Cultural and Social Foundations*. Cambridge, MA. Harvard University Press.

Luria, A. R. 1973. *The Working Brain: An Introduction to Neuropsychology*. New York: Basic Books

Pylyshyn, Z.W. 2000. Situating vision in the world. Trends in Cognitive Sciences. Volume 4(5): 197-207.

Smith, B. 2003. The Ecological Approach to Information Processing, in Kristóf Nyíri (ed.), Mobile Learning: Essays on Philosophy, Psychology and Education. Vienna: Passagen Verlag, 17–24.

Tolman, E.C. 1948. Cognitive maps in rats and man. Psychological Review. 55:189.

Tulving, E. 1985. *Memory and Consciousness*. Oxford. Clarendon Press.

Udupa, J.K., Punam K. Saha, P.K., & Lofuto, R. A. 2002. Relative Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation. IEEE Trans. PAMI 24(11):1485-1500.