# Imagery and Mental Models in Problem Solving*

Yulin Qin and Herbert A. Simon
Carnegie Mellon University
Pittsburgh, PA 15213

## Abstract

Six subjects were asked to read and try to understand the first part of Einstein's 1905 paper on special relativity. On the basis of the diagrams drawn by the subjects while they read the text, their thinking aloud protocols, their responses to questions, and other observations, we have inferred the processes they used to understand the paper, and the role that mental images (mental representations) and their paper-and-pencil drawings played in achieving the understanding.

In particular, we discuss the characteristics of the representations the subjects formed and the processes they used to form them, as well as the ways in which they derived or reached an understanding of the equations of relativity, using the mental images to assist them. From our analysis we infer that subjects translating the natural-language text into images are able to manipulate, control, and observe these representations to run simple "mental experiments," and by these means draw out the qualitative and quantitative implications of the text.

The mental representations and drawings appear to mediate between the initial natural language text and the final equations. In no case did any subject achieve an understanding of the equations without using this kind of intermediate representation.

For some decades, imagery has been an active research area in cognitive science (c.f., Paivio 1971; Shepard and Metzler 1971; Simon 1972; Pylyshyn 1973; Kosslyn 1980, 1983; Shepard and Copper 1982; Pinker 1985; Finke 1989). Research on imagery's neuro psychological base is also progressing (Farah 1985; Kosslyn 1987; Kosslyn, Sokolv and Chen 1989; etc.). And the role of imagery in education has been examine (e.g., Martel 1991). However, there are only a few publications of experimental work concerned mainly with imagery in problem solving (e.g., Simon and Barenfeld 1969; Hayes 1973; Novak 1976; and some aspects in Larkin and Simon 1987).

Our study, in which subjects were asked to read and try to understand the first part of Einstein's 1905 paper on special relativity, focuses on imagery in understanding and problem solving, in mental modeling (c.f. Gentner and Stevens 1983; Johnson-Laird 1983; Norman 1988) and in qualitative reasoning (c.f. Iwasaki 1989; Weld and de Kleer 1990). In Qin and Simon (1990, 1991), we discussed some results of our project relating to problem solving success. In this paper, we summarize very briefly our major findings about imagery and mental models.

## Experiment

### Reading material

Except for a few modifications, the reading material used in this experiment is simply a copy of the first three sections of the first part, the kinematical part, of Einstein's 1905 paper: "On the

---

* This research was supported by the Personnel and Training Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-86-k-0678; and by the Defense Advanced Research Projects Agency, Department of Defense, ARPA Order 3597, monitored by the Air Force Avionics Laboratory under contract F33615-81-k-1539. Reproduction in whole or in part is permitted for any purpose of the United States Government. Approved for public release; distribution unlimited.

Electrodynamics of Moving Bodies" (Einstein, 1905). Einstein employed a "divide then integrate" strategy and motivated each of his initial equations by inviting the reader to form a mental picture that exemplifies it. The equations can, so to speak, be read directly from these mental pictures, so that the reader who has found the picture understands the corresponding equation. Interestingly enough, the published paper contains no diagrams to guide or assist this process.

In section 1 the major point is the definition of simultaneity:

"Let a ray of light start at the 'A time' $t_A$ from A towards B, let it at the 'B time' $t_B$ be reflected at B in the direction of A, and arrive again at A at the 'A time' $t'_A$. In accordance with definition the two clocks synchronize if

$$t_B - t_A = t'_A - t_B \qquad (1.0)\text{"}$$

In section 2, one of the key points, after introducing the principle of relativity and the principle of the constancy of the velocity of light, is the relativity of times:

"We imagine further that at the two ends A and B of the rod, clocks are placed which synchronize with the clocks of the stationary system, that is to say that their indications correspond at any instant to the 'time of the stationary system' at the place where they happen to be. These clocks are therefore 'synchronous in the stationary system.' "We imagine further .... Let a ray of light depart from A at the time $t_A$, let it be reflected at B at the time $t_B$, and reach A again at the time $t'_A$. "Taking into consideration the principle of the constancy of the velocity of light we find that

$$t_B - t_A = r_{AB} / (c-v) \qquad (2.1)$$

and

$$t'_A - t_B = r_{AB} / (c+v) \qquad (2.2)$$

where $r_{AB}$ denotes the length of the moving rod -- measured in the stationary system."

Then he concluded:

"Observers moving with the moving rod would thus find that the two clocks were not synchronous, while observers in the stationary system would declare the clocks to be synchronous. So we see that we cannot attach any absolute signification to the concept of simultaneity,..."

In section 3, he introduced two systems: the stationary system (system K) and the moving system (system k) at first. Then he pointed out:

"To any system of values x, y, z, t, which completely defines the place and time of an event in the stationary system, there belongs a system of values $\xi, \eta, \zeta, \tau$ determining that event relatively to the system k, and our task is now to find the system of equations connecting these quantities." "If we place $x' = x - vt$, it is clear that a point at rest in the system k must have a system of values x', y, z, independent of time. We first define $\tau$ as a function of x', y, z, and t...." "From the origin of system k let a ray be emitted at the time $\tau_0$ along the X-axis to x', and at the time $\tau_1$ be reflected thence to the origin of the co-ordinates, arriving there at the time $\tau_2$; we then must have

$$\frac{1}{2}(\tau_0 + \tau_2) = \tau_1 \qquad (3.0)$$

"or, by inserting the arguments of the function $\tau$ and applying the principle of the constancy of the velocity of light in the stationary system:-

$$\frac{1}{2}[\tau(0,0,0,t) + \tau(0,0,0,t + \frac{x'}{c-v} + \frac{x'}{c+v})] = \tau(x',0,0,t + \frac{x'}{c-v}) \qquad (3.1)$$

"To simplify the mathematical operations, losing no generality, let t = 0 and $\tau_0=0$. We get:

$$\frac{1}{2}[\tau(0,0,0,0) + \tau(0,0,0,\frac{x'}{c-v} + \frac{x'}{c+v})] = \tau(x',0,0,\frac{x'}{c-v}) \qquad (3.2)\text{"}$$

Based on the above key equations, by algebraic reasoning, he inferred the Lorentz transformation equations.

## Method and subjects

To understand Einstein's paper, subjects need to read it with comprehension, to modify or build concepts (e.g., time, simultaneous events, synchronous clocks), and to understand and/or derive the equations. The latter involves building images, and doing qualitative and quantitative reasoning.

We obtain information about the images from the diagrams drawn by the subjects, the subjects' gestures, and their protocols. That is, we postulate a consistency between these observable data and their mental images. A good deal of work must be done on the relation between internal and external representations to check this postulate and the conditions under which it holds. For the present, we can use the data from these three sources to see what processes are used to form an image from the reading material and how the image is used to understand the material.

Most of our subjects were undergraduate or graduate students in electrical engineering or computer science at Carnegie Mellon University. None of them were familiar with derivations of the Lorentz equations or could derive them. None were aware that the reading material was from Einstein (1905), until, at the end of the experiment, they were told by the experimenter.

| Group | Subject | Sessions | Time spent | Position |
|---|---|---|---|---|
| UseD | $S_m$ | 8 | 840min | Research assistant, Psy<br>Master, Computer Science |
| | $S_r$ | 3 | 259min | 2nd year Grad, ECE |
| NoD | $S_j$ | 4 | 512min | Senior Undergrad, ECE |
| | $S_b$ | 3 | 315min | 1st year Grad, SCS (Under, Math) |
| UnusedD | $S_s$ | 12 | 1180min | Senior Undergrad, ECE |
| | $S_g$ | 7 | 680min | 1st year Grad, SCS |

ECE: Electrical and computer engineering department at CMU;
Psy: Psychology department at CMU; SCS: School of computer science at CMU;
Undergrad: Undergraduate student; Grad: Graduate student;

Table 1

We divided our subjects into three groups, on the basis of what we asked them to do: (a) describe images, draw diagrams, use diagrams (UseD group); (b) describe images, do not draw diagrams while deriving the equations (NoD group); (c) describe images, draw diagrams, but do not retain the diagrams (The experimenter took each diagram away from the subjects as soon as it was drawn, and they could not draw diagrams while deriving the equations.) (UnusedD group).

Sessions extended for around 1.5 to 2 hours. Subjects who had not completed the material at the end of a session returned for one or more additional sessions.

Six subjects' protocols have been analyzed. The times the subjects spent understanding the material, the number of sessions required, and summary information on their backgrounds are shown in Table 1.

## Results

The major findings in our experiment are as follows:

1. All of our subjects could form mental images to derive or justify the essential equations of special relativity, including those subjects who, when meeting difficulties in reasoning, usually claimed not to be able to do so;

2. Subjects could watch the processes in an evolving image as they would watch processes in the real world. By watching the processes, subjects could draw qualitative conclusions, and then inferred quantitative relations, i.e., the equation to be derived. Without images or with incorrect images, subjects found the task very difficult. They were not able to derive the equation or derived a wrong result;

Kosslyn et al (1990) discussed the uses of imagery in everyday life, and found that "(m)ost of the images the subjects reported had no recognizable purpose; most images occurred in isolation, not as part of a sequence..." It seems that when imagery is used in problem solving, the behavior is different from when they are used in everyday life. While deriving the equations, imagery was used with obvious purpose: as the representation of the process of light traveling.

Subjects could use images in two ways:

(1) To simulate the process and derive the equations that described it;

For example, to derive the equations (2.1) and (2.2), the process looks like this:

a) Guided by the goal (i.e., why was the equation right?), they formed their images;

b) By watching the images and comparing the related quantities, they derived the qualitative relations. For example, the distance the light traveled was greater than the length of the moving rod while light was going in the same direction as the moving rod doing, and it was smaller while it was moving in the opposite direction;

c) By assigning the known values to the given quantities and deriving the values of the unknown quantities, they inferred the quantitative relations among the quantities, i.e., the equations. In this process, some subjects, with the help of their images, could create new concepts, quantities, and form a semi-quantitative equation.

(2) To simulate the process and identify its type.

In the process of inserting the arguments of function $\tau$ in equation (3.0) to get equation (3.1), some subjects reported their images. But they did not really use their images to derive the values of these arguments as mentioned in (1). Instead, by watching the image, they identified the process as identical with that discussed in the previous section and then simply used the results derived there to get this new equation.

Both of these processes combine simulating and recognition processes (cf. Kosslyn 1990, 1987).

3. One of the differences between watching an image and watching a real-world object is that the components of the image are better understood, because they are formed by the observers themselves, who know what they should observe.

The basic pieces of information for forming a mental image are the components of the situation, their organization or structure, the kinematic relations among them, and the process (if the image is used to represent a process). There are two basic types of processes for forming images: one used while reading for comprehension; the other used for problem solving, such as deriving an equation. The difference between them is that the first is based on attention to the reading matter, and the second is guided by the goal of problem solving.

When an image was used in problem solving, there appeared to be a heavy STM load while keeping the image and reasoning simultaneously. Subjects shifted frequently between forming and watching images and reasoning. There was active interaction between the processes of forming images and of problem solving. On the one hand, the way subjects derived the equations was closely related to their images. On the other hand, subjects might totally change their images when they met trouble in using their original images to solve problems, or they might modify their images and shift their attention in the process of problem solving.

5. Understanding relativity required some basic changes in the subjects' mental models before they could construct and build mental images to help them draw the necessary inferences about the situation and derive the equations describing it. By a mental model we mean the subjects' knowledge structure about the world, held in long-term memory. This appears to be close to what Norman (1988) had in mind in speaking of "the models people have of themselves, others, the environment, and the things with which they interact." . It is deeper and more stable, systematic and general than an image formed while performing a specific task, and provides a source of information for imaging. As shown by Novak (1976), the mental model influences the process of getting information from the reading material, and determines what kind of image will be formed. Bugs in the mental model may cause subjects to form an image that does not correspond to the real situation, or to form the right picture but with wrong values of the quantities of the components (e.g., the time for transmission of the light, measured in the stationary system);

6. Understanding reading material is a process of changing the reader's mental model. It is not easy to change one's mental model totally. Furthermore, a mental model is not very complete and may be inconsistent. So, when reading a paper that fundamentally changes their world model, as our reading material does, subjects may for some time hold a confused mental model. They were learning non-Galilean transformations, but still held the concept of Galilean transformations in their minds. They could give the superficial meaning of an equation, but found it difficult to accept a deep one (e.g., the relativity of two events being synchronous). We have seen three instances of the gap between Einstein's relativity model and the readers' mental models that commonly occurred with our subjects: the relativity of simultaneity in section 2, the relativity of the length when Einstein transforms x to x', and the constancy of the velocity of light. The first and the third problems became foci of explicit conflict, the second one had not yet been noticed by the subjects when deriving equations (3.0) and (3.1). Encountering and resolving these kinds of conflicts may be like experiencing and reacting to surprises in scientific discovery (Kulkarni and Simon 1988), and it may be a basis for changing mental models.

## Discussion and conclusion.

In this paper, we have discussed the processes used by subjects in forming and watching dynamic mental images, drawing qualitative conclusions and then deriving quantitative equations. We saw that the reading and understanding process required subjects to change their mental models. Mental models in long-term memory provide the substrate for the more transient mental images that subjects form. They consist of the relevant knowledge already held in memory that shapes attempts to form the new representations needed in problem solving.

Mental models, even when appropriately modified, are not enough for understanding the concepts of relativity. Our subjects also had to construct mental images in short-term memory that could be manipulated to reveal the relations among the successive events that Einstein introduces to describe the transmission of light in the world of special relativity. When the subjects were successful in constructing appropriate images, they could derive from them both the qualitative properties of the situation and the equations that described it quantitatively, and in this way, were able to change their mental models. However, deriving an equation does not mean understanding the equation fully. In some cases subjects encountering conflicts with their models, did not fully resolve them. In other case, they did not even notice the conflict. In these situations their mental models were not fully revised to fit relativity.

## References


Einstein, A. (1905). On the electrodynamics of moving bodies. In A. Einstein, et al. *The Principle of relativity.* Dover Publications, 1923.

Farah, M. J. (1985). The neurological basis of mental imagery: a componential analysis. In S. Pinker (ed.), *Visual cognition*. Cambridge MA: MIT press.

Finke, R. A. (1989). *Principles of mental imagery*. Cambridge MA: MIT press.

Gentner, D. and Stevens, A.L. (Eds.). (1983) Mental model. Hillsdale, NJ: Erlbaum.

Hayes, J.R. (1973). On the function of visual imagery in elementary mathematics. In W. G. Chase (ed.), *Visual information processing*. New York: Academic Press.

Iwasaki, Y. (1989). Qualitative physics. In A. Barr, P.R. Cohen, and E. A. Feigenbaum (ed.), The handbook of artificial intelligence. Vol., 4. Menlo Park, CA: Addison-Wesley Publishing Comp.

Johnson-Laird, P. N. (1983). *Mental model.* Cambridge, MA: Harvard University Press.

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

Kosslyn, S. M. (1983). Ghosts in the mind's machine - Creating and using images in the brain. New York: W. W. Norton & Company.

Kosslyn, S. M. (1987). seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, 94, 148-175.

Kosslyn, S., Seger, C. Pani, J. R., and Hillger, L. A. (1990). When is imagery used in everyday life? A diary study. *Journal of Mental Imagery*, 14(3 & 4) 131-152.

Kosslyn, S. M., Sokolov, M. A. and Chen, J. C. (1989). The lateralization of BRAIN: A computer theory and model of visual hemispheric specialization. In D. Klahr and K. Kotovsky (Eds.) *Complex information processing - the impact of Herbert A. Simon.* NJ: LEA publishers.

Kulkarni, D. and Simon, H. A. (1988). The process of scientific discovery: The strategy of experimentation. *Cognitive Science*, 12, 139 - 175.

Larkin, J. H. and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65-100.

Martel, L. D. (1991). The role of guided imagery in educational reform: the integrative learning system. In R. G. Kunzendorf (ED.), *Mental imagery.* NY: Plenum press.

Norman, D. A. (1988). The psychology of everyday things. NY: Basic Books, Inc., Publishers.

Novak, G. S. (1976). Computer understanding of physics problems stated in natural language. Ph.D. Dissertation, The university of Texas at Austin.

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

Pinker, S. (Ed.) (1985). *Visual cognition.* Cambridge MA: MIT press.

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

Qin, Y. and Simon, H. A. (1990). Imagery and problem solving. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society.* 646 - 653. NJ: LEA publishers.

Qin, Y. and Simon, H. A. (1991). Imagery as process representation in problem solving. Unpublished manuscript.

Shepard, R. N. and Copper, L. A. (1982). *Mental Images and Their Transformations*. Cambridge MA: MIT press.

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

Simon, H.A. (1972). What is visual imagery? An information processing interpretation. In L. W. Gregg (ed.), *Cognition in learning and memory*. New York: John Wiley & Sons.

Simon, H.A. and Barenfeld, M. (1969). Information processing analysis of perceptual processes in problem solving. *Psychological review* 76, 473-483.

Weld, D. S. and de Kleer, Johan. (Eds.). (1990). Readings in qualitative reasoning about physical systems. San Mateo, CA: Morgen Kaufmann publishers, Inc.