Foundations of Science (2005) 10: 89–106

J. GREGORY TRAFTON, SUSAN B. TRICKETT and FARILEE E. MINTZ

CONNECTING INTERNAL AND EXTERNAL REPRESENTATIONS: SPATIAL TRANSFORMATIONS OF SCIENTIFIC VISUALIZATIONS

ABSTRACT. Many scientific discoveries have depended on external diagrams or visualizations. Many scientists also report to use an internal mental representation or mental imagery to help them solve problems and reason. How do scientists connect these internal and external representations? We examined working scientists as they worked on external scientific visualizations. We coded the number and type of spatial transformations (mental operations that scientists used on internal or external representations or images) and found that there were a very large number of comparisons, either between different visualizations or between a visualization and the scientists' internal mental representation. We found that when scientists compared visualization to visualization, the comparisons were based primarily on features. However, when scientists compared a visualization to their mental representation, they were attempting to align the two representations. We suggest that this alignment process is how scientists connect internal and external representations.

KEY WORDS: diagramatic reasoning, graph comprehension, scientific reasoning, scientific visualization, spatial transformations

There are numerous examples, anecdotes, and stories about scientists using internal mental imagery and episodic mental representations. For example, Einstein said that imagining himself traveling through space next to a beam of light helped him discover the special theory of relativity (Einstein, in Hadamard, 1945, p. 142). Maxwell professed to have envisioned the lines of magnetic force (Newman, 1955, p. 67). Watson reported he had visualized adenine residues from DNA (Watson, 1968). Kekulé imagined atoms combined to form molecules (Findlay, 1948). These and many other scientists all claimed they used mental imagery to help them make scientific discoveries (Kaplan and Simon, 1990; Shepard, 1988; Thomas, 1999).

In addition to this kind of internal visualization process, many of these same scientists used self-generated diagrams (external visualizations) to aid their thinking. Many of these diagrams survive in the scientists' notebooks and diaries, such as those left by Faraday, Maxwell, and Newton, to name a few. A number of historical psychological analyses have been conducted on these notebooks, and as a result, several researchers have proposed that the diagrams themselves had a profound effect on the scientist's reasoning (Gentner et al., 1997; Nersessian, 1999; Tweney, 1992). In many of these cases, there is a clear link between the scientist's internal representation and the external diagram: the scientist was drawing, or making explicit, his own mental representation, in order to explain it and its implications to himself, as well as to engage in further exploration. Insofar as they explicitly represent abstract information, can be redrawn and re-envisioned, and perhaps most importantly, enable the scientist to explain and investigate particular phenomena, these diagrams can be seen as a precursor to today's world of complex scientific visualization.

Many contemporary scientists use complex visualization tools to facilitate their scientific investigation. In fact, scientific visualization is a major part of current scientific work and discovery, as evidenced by the large number of universities and research laboratories that have extensive scientific visualization laboratories – Georgia Institute of Technology, Virginia Tech, Brown University, New York University, the Naval Research Laboratory, and the University of Minnesota to name a few. These visualization tools provide a means whereby extremely complex, multi-dimensional data can be represented in tangible ways. For example, virtual reality labs are currently able to display three and four dimensional data without collapsing any of the dimensions, allowing scientists to manipulate the viewable data and see its effects and interactions on other visible data.

An important difference between these computer-powered visualization tools and the type of external visualization (diagram) discussed earlier, is the power and ease with which the former can be redrawn and manipulated by the user. The scientist can rather readily examine different views of the data and can consequently explore the implications of different assumptions. What, then, is the role of internal mental representation, when external representation offers such a rich array of options? When scientists use these complex visualization tools, do they continue to rely on internal representations to investigate their data, or do they focus their attention on the external display of the data? If internal representations remain important, as is suggested by the historical studies discussed above, what kind of interaction is there between these internal representations and the scientists' use of these sophisticated external visualization tools? We present a study to examine the importance of scientists' own mental imagery and mental representations while using complex external visualization tools, and to investigate the relationship between these internal and external visualizations.

METHOD

In order to investigate the issues discussed above, we have adapted Dunbar's in vivo methodology (Dunbar, 1997; Trickett, Fu, Schunn and Trafton, 2000). This approach offers several advantages. It allows observation of experts, who can use their domain knowledge to guide their strategy selection. It also allows the collection of "on-line" measures of thinking, so that the scientists' thought processes can be examined as they occur. Finally, the tasks the scientists do are fully authentic.

Two sets of scientists were videotaped while conducting their own research. All the scientists were experts, having earned their PhDs more than 6 years previously. In the first set, two astronomers, one a tenured professor at a university, the other a fellow at a research institute, worked collaboratively to investigate computer-generated visual representations of a new set of observational data. At the time of this study, one astronomer had approximately 20 publications in this general area, and the other approximately 10. The astronomers have been collaborating for some years, although they do not frequently work at the same computer screen at the same time to examine data.

In the second dataset, a physicist with expertise in computational fluid dynamics worked alone to inspect the results of a computational model he had built and run. He works as a research scientist at a major U.S. scientific research facility and had earned his PhD

23 years ago. He had inspected the data earlier but made some adjustments to the physics parameters underlying the model and was therefore revisiting the data. Both sets of scientists were instructed to carry out their work as though no camera was present and without explanation to the experimenter (Ericsson and Simon, 1993). The relevant part of the astronomy session lasted about 53 minutes, and the physics session, 15 minutes. All utterances were later transcribed and segmented according to complete thought. All segments were coded by 2 coders as on-task (pertaining to data analysis) or off-task (e.g., jokes, phone interruptions, etc.). Inter-rater reliability for this coding was more than 95%. Off-task segments were excluded from further analysis. On-task segments (N = 649 for astronomy and N = 176 for physics) were then coded for external imagery and internal mental representations (spatial transformations).

The Tasks and the Data

Astronomy. The data under analysis were optical and radio data of a ring galaxy. The astronomers' high-level goal was to understand its evolution and structure by understanding the flow of gas in the galaxy. In order to understand the gas flow, the astronomers must make inferences about the velocity field, represented by contour lines on the 2-dimensional display.

The astronomers' task was made difficult by two characteristics of their data. First, the data were one- or at best 2-dimensional, whereas the structure they were attempting to understand was 3-dimensional. Second, the data were noisy, with no easy way to separate noise from real phenomena. Figure 1 shows a screen snapshot of the type of data the astronomers were examining. In order to make their inferences, the astronomers used different types of image, representing different phenomena (e.g., different forms of gas), which contain different information about the structure and dynamics of the galaxy. In addition, they could choose from images created by different processing algorithms, each with advantages and disadvantages (e.g., more or less resolution). Finally, they could adjust some features of the display, such as contrast or false color.

Physics. The physicist was working to evaluate how deep into a pellet a laser light will go before being reflected. His high-level goal

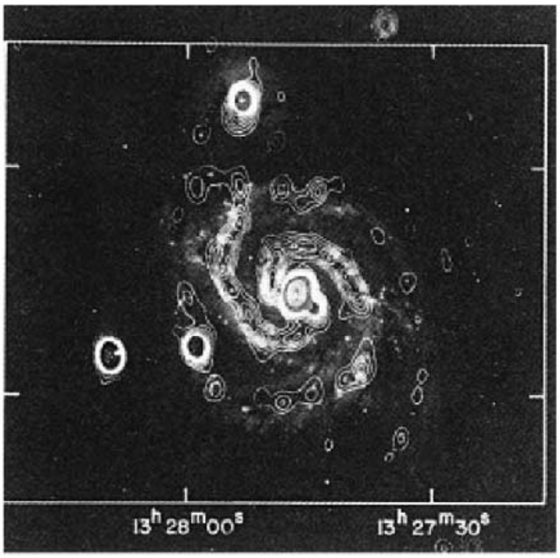

Figure 1. Example of data examined by astronomers. Radio data (contour lines) are laid over optical data.

was to understand the fundamental physics underlying the reaction, an understanding that hinged on comprehending the relative importance and growth rates of different modes. The physicist had built a model of the reaction; other scientists had independently conducted experiments in which lasers were fired at pellets and the reactions recorded. A close match between model and empirical data would indicate a good understanding of the underlying theory. Although the physicist had been in conversation with the experimentalist, he had not viewed the empirical data, and in this session he was investigating only the results of his computational model. However, he believed the model to be correct (i.e., he had strong expectations about what he would see), and in this sense, this session may be considered confirmatory.

The data consisted of two different kinds of representation of the different modes, shown over time (nanoseconds). The physicist was able to view either a Fourier decomposition of the modes or a representation of the "raw" data. Figure 2 shows an example of the physicist's data. He could choose from black-and-white or a variety of color representations, and could adjust the scales of the displayed image, as well as some other features. He was able to open numerous views simultaneously.

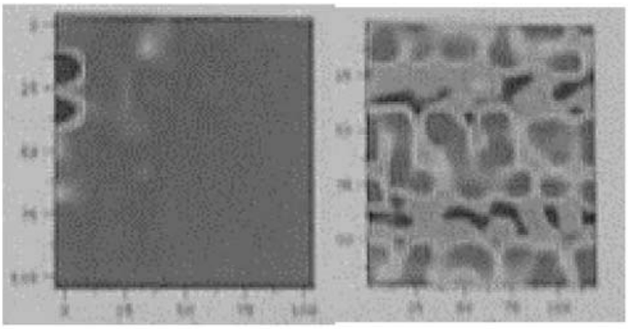

Figure 2. Example of data examined by physicist. Fourier modes (left) and raw data (right).

Coding Scheme

Our goals in this research were first, to investigate scientists' relative use of both external and internal imagery, and second, to explore the interaction between the two. In order to investigate the first question, we developed a coding scheme, described below, that allowed us to evaluate the scientists' use of both types of representation. We discuss later our approach to investigating the issue of their interaction.

Both protocols were coded independently by two different coders. Initial inter-rater reliability for each code was greater than 85%. Disagreements were resolved by discussion. Any coding disagreements that could not be resolved were excluded from further analysis.

External imagery coding

We examined external imagery by coding the number of times the scientists created a new visualization or adjusted an existing visualization. For example, a new visualization was coded when the visualization was completely different from the one before (Fourier modes vs. raw data in Figure 2). Adjusting a visualization typically entailed "tweaking" a visualization, for example, by changing the resolution or contrast.

Internal representation coding

In order to investigate the scientists' mental representations, we identified a specific type of mental representation which we call spatial transformations. Since these scientists are working with external scientific visualizations, we needed a coding scheme to examine how these scientists reason with spatial and imagistic representations. We are assuming that many of the internal representations that the scientists are using are spatial or imagerial because the scientists are working explicitly with diagrams and visualizations.

Spatial transformations are cognitive operations that a scientist performs on an internal representation (like an image) or an external visualization. Even when a spatial transformation is being performed on an external visualization, it is generating or using an internal representation. Sample spatial transformations are mental rotation (Shepard and Metzler, 1971), creating a mental image, modifying that mental image by adding or deleting features, mentally moving an object, animating a static image (Hegarty, 1992), making comparisons between different views (Kosslyn, Sukel and Bly, 1999; Trafton and Trickett, 2001), and anything else a scientist *mentally* does to a visualization in order to understand it or facilitate problem-solving. Note that a spatial transformation can be performed on either an internal (i.e., mental) image or representation or an external image (e.g., a computer-generated visualization). Statements by which the scientists directly extracted information from the visualization were *not* considered spatial transformations. A more complete description can be found at http://www.aic.nrl.navy.mil/~trafton/st.html.

For every utterance in each protocol we evaluated whether there was a spatial transformation. Spatial transformations were further divided into two major categories, "pure" spatial transformations, which involved some type of mental operation on a single representation, and comparisons, in which the mental operation consisted of comparing two or more representations. Pure spatial transformations were further coded as Create or Manipulate (Add or Delete). Subsequent coding of the comparisons is discussed below. Table I shows examples of each type of spatial transformation (note that the first two utterances are independent of one another and do not represent a sequence; however, the last three utterances were spoken in sequence by the same speaker working on a specific visualization).

TABLE I

Examples of spatial transformations

Utterance	Spatial transformation	Explanation
That's about 220 km per second	N/A	Direct extraction of information from display
Although if it's, if it's, if the arm's detaching here and sort of flowing away, . . .	Create & Mentally Manipulate	Mental image of arm detaching is first created; then scientist manipulates that image by imagining the arm flowing away
I mean, in a perfect, in a perfect world, in a perfect sort of spider diagram	Create	Spider diagram does not exist on display; scientist creates mental image of it
If you looked at the velocity contours without any sort of streaming motions, without streaming motions . . .	Mentally Manipulate: Delete	Scientist is mentally removing streaming motions from internal representation of spider diagram created above
You'd probably expect these lines here [gesturing to screen] to go all the way across, you know	Comparison	Scientist compares adjusted mental representation of spider diagram without streaming motions with external visualization

RESULTS

Our first question was what is the relative use made by scientists of both internal and external representations? If the scientists focus their attention on the external representation, we would expect to see more new visualizations and adjustments to current visualizations than spatial transformations. On the other hand, the extent to which internal representations are important to the scientists' thinking will be reflected in the relative frequency with which they use spatial transformations to create new mental images or to perform mental operations on existing images (internal or external).

Figure 3 shows the number of physical transformations to the external visualizations (new and adjusted visualizations) and the number of spatial transformations (broken down by type). Inter-rater reliability was 95% for this coding between two independent coders.

As Figure 3 shows, there are many more spatial transformations than physical transformations for both the astronomy dataset, $\chi^2(1) = 40.3$, $p < 0.001$ and the physics dataset, $\chi^2(1) = 24.6$, $p < 0.001$. In both datasets, the scientists took advantage of the visualization tools to create and modify visualizations of the data. (This is particularly true in the longer astronomy dataset.) However, the significantly larger number of spatial transformations suggests that, although generating and adjusting visualizations on the external display is an important aspects of scientific visualization work, generating mental representations and performing internal operations on all these representations is also extremely important.[1]

It appears, then, that mentally manipulating both external imagery and internal representations is extremely important to these scientists. Scientists do not seem to use the computer's visualization capability *instead of* their own mental imagery and representations, but rather seem to use both their own mental representations and the computer's visualizations. Not only do they use features of the software to tweak the visual image, but they also use spatial transformations to make mental adjustments to that image.

Our second question concerned the relationship between the scientists' use of internal and external representation. In other words, how do scientists connect their mental representations and the external scientific visualizations? There are two possibilities. First, the "pure" spatial transformations could have been used on

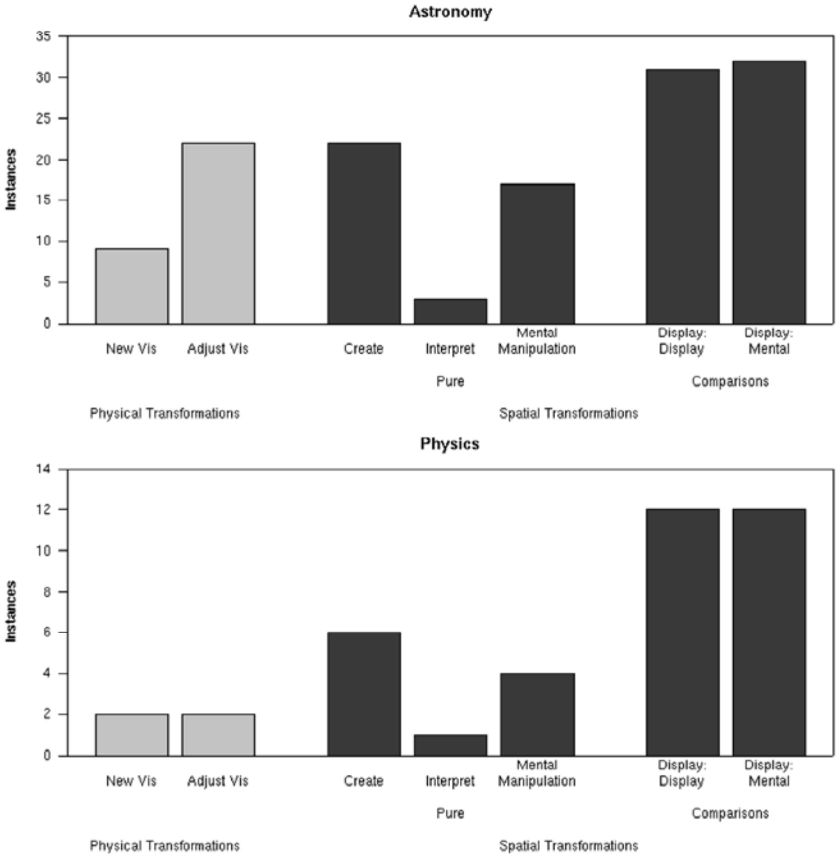

Figure 3. Spatial transformation breakdown for both datasets. The leftmost (lighter) bars show the number of physical transformations (new visualizations and the number of times an on-screen visualization was adjusted or modified) and the darker bars show the number of spatial transformations.

the external imagery (e.g., the utterance "Although if it's, if it's, if the arm's detaching here and sort of flowing away . . ." creates a mental representation (spatial transformation) that is based on the current external visualization). There is, in fact, some evidence for this hypothesis: when a scientist mentally manipulated a representation, 71% of the time the source was a visualization, and only 29% of the time was it a "pure" mental representation (e.g., the "spider diagram" example shown in Table I). Thus, some of the time, scientists seem to create and interpret mental representations that are different from the images in the visual display.

The second possibility is that the scientists could use a comparison process to connect their internal representation with the external visualizations. As Figure 3 shows, there are many more comparisons than the other types of spatial transformations, and in fact, this difference is significant for both the astronomy dataset, $\chi^2(1) = 4.2$, $p < 0.05$ and the physics dataset, $\chi^2(1) = 4.8$, $p < 0.05$. This result suggests that after the scientists created and/or mentally manipulated a representation, they were not finished with that representation. Instead, they frequently compared it to something else. What role did this comparison process play in connecting the internal and external representations? In order to investigate how comparisons were used, we further coded all comparisons in two ways. First, we identified the source of the two representations being compared. When a comparison was made, one of the phenomena being compared was invariably on the visualization (display). Whether this phenomenon was being compared to something in another visualization (display) or some internal representation (mental) was coded. Second, we coded what kind of operation was performed by the comparison – Identify (ID), Feature Comparison, or Alignment. An ID was coded when a scientist made a comparison to determine the identity of one of the objects. A Feature Comparison was made when a scientist compared two things in terms of their relative features (size, shape, color, etc.). An Alignment was made when a scientist was trying to determine an estimation of fit of one representation to another (e.g., visually inspecting the fit of a regression line to a scatterplot). Table II shows examples of this coding scheme.

We performed inter-rater reliability on 34% of the comparison data; our agreement was 88% and our kappa was 0.77 ($p < 0.001$). Figure 4 shows the results of this coding from both datasets. Again, note the remarkable similarity between the datasets. Clearly the same relationship between the source of the comparison and the operation it performs pertains in both datasets. Thus, we combined both datasets in order to perform statistical analyses. Because the ID category was relatively small, the combined dataset contains only Alignments and Feature Comparisons. Figure 5 shows the results for both datasets combined.

TABLE II

Examples of comparison coding scheme: source and operation

Utterance	Source of comparison	Operation
. . . except for oh-two who marginally wasn't seeded . . . which is two-oh's symetric counterpart	Display:Display	ID
Well, it's actually, it's, well, it's not as high as you see down here.	Display:Display	Feature Comparison
Uh, the one-three is actually smaller than the oh-three	Display:Display	Feature Comparison
You'd probably expect these lines here [gesturing to screen] to go all the way across, you know	Display:Mental	Alignment
If it was stuff out of the plain in all directions [creates mental image] – I mean, the fact you see such a strong concentration of gas in the ring, um . . .	Display:Mental	Alignment

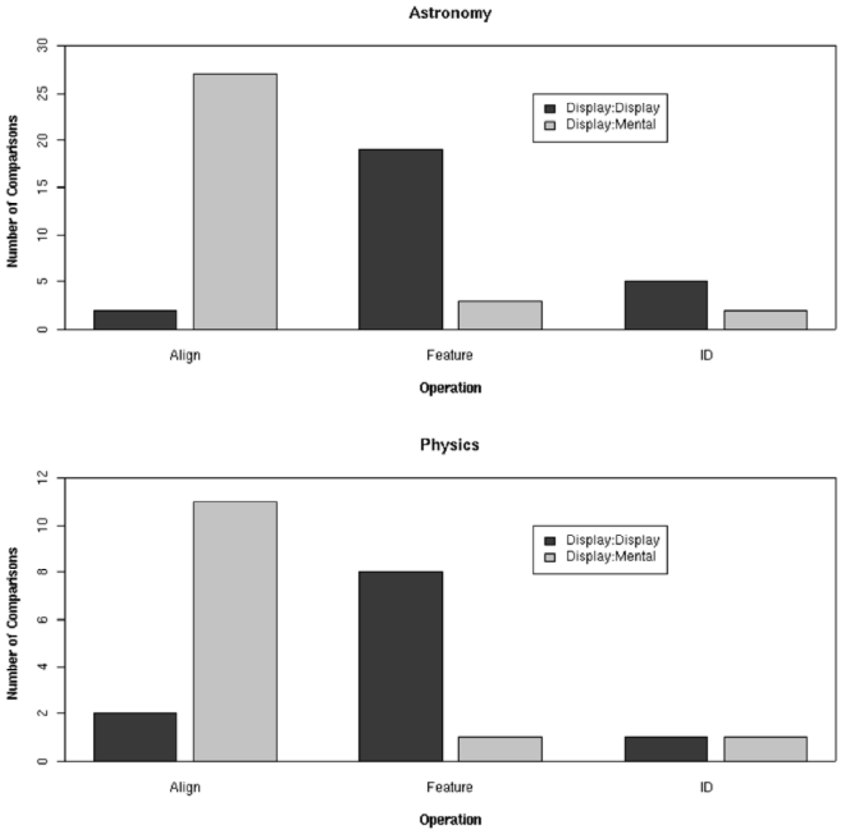

Figure 4. Breakdown of comparisons by source and operation for astronomy and physics datasets separately.

As Figure 5 shows, there is a very strong relationship between the source of the comparison and the operation. When the scientist was comparing an image on the screen with an internal representation, he was most likely performing an Alignment; however, when the scientist was comparing two images on the external display, he was most frequently performing a Feature Comparison. This relationship is highly significant, $r_\Phi = 0.726$, $p < 0.001$.

What are the scientists trying to do when they make a comparison between two different displays? One possibility is that they are looking for similar patterns in the data. They could be then investigating these similarities at the theoretical level. Another possibility is that they could be looking for discrepancies or anomalies (Trickett, Trafton, Schunn and Harrison, 2001).

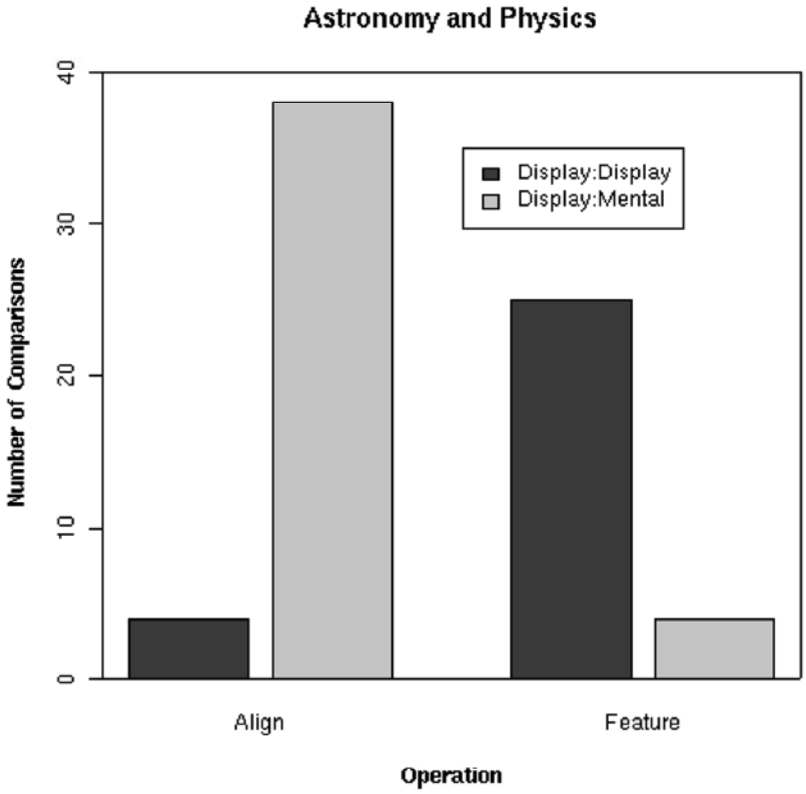

Figure 5. Breakdown of comparisons by source and operation in astronomy and physics datasets combined.

GENERAL DISCUSSION

Scientists seem to perform many mental operations while examining external scientific visualizations. In fact, they seem to perform many more spatial transformations on their own mental representations than creating new visualizations or changing the visualizations. This is somewhat surprising since the scientific visualizations that these scientists (like most scientists) used were tailor made for them in their own area of expertise. The main question we started off with, "How do scientists connect their internal and external representations" seems to have two answers. First, scientists perform spatial transformations on external visualizations. These spatial transformations on the external visualizations connect the internal and external representations by forcing the scientist to merge their two representations.

The second way that scientists connect their internal and external representations is by comparing their own internal representations with an external visualization. This comparison process is quite interesting because scientists seem to be making an estimation of fit between their own internal representation and the external scientific visualization. In contrast, when scientists compare two visualizations, they seem to compare different features.

MENTAL REPRESENTATION VS. IMAGERY

Throughout this paper, we have discussed internal representations and spatial transformations. What exactly do we mean by an internal representation, and how are spatial transformations represented mentally? Unfortunately, there is no direct physical or experimental evidence (e.g., fMRI) other than self-report, so the actual representation is debatable. We believe, however, that spatial transformations are represented by a visual component and a spatial component. In addition, many of the spatial transformations seem to be imagistic, and the scientists themselves frequently describe these operations as imagistic. In addition, Kosslyn and his colleagues make a very strong argument that the types of comparisons described here are based on imagery (Kosslyn et al., 1999).

WHY ALIGN?

Why do scientists engage in this alignment process between a mental representation and an external scientific visualization? There has been an enormous amount of study of the alignment process (e.g., Gentner, 1983; Gentner, 1989) and we believe that the alignment process described in this paper is very similar to the process described by Gentner.

We believe that the estimation of fit process we describe is part of the process of theory construction. Scientists seem to be constructing "models on the fly" as they look at these visualizations. We believe that these models (which we call Qualitative Mental Models, Trafton et al., 2000) are then used to build formal theories or models (e.g., computer models). For example, the following conversation comes near the end of the astronomy dataset:

Observationalist: Well, let's see. So do we understand this perfectly now?
Theoriest: Oh, sure. Except for the um, all the stuff we don't understand.
Observationalist: Yeah.
Theorist: Well, but actually, it's pretty useful 'cause I can see where I need to make some [computer] models . . .

The theorist believes that this particular session has helped him understand several aspects that he can then model in a formal computational modeling system. Thus, the qualitative mental models he has built will now be put to use as he is building his computer models.

ACKNOWLEDGEMENTS

This research was supported in part by grant 55-7850-00 to the first author from the Office of Naval Research. 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 the U.S. Navy.

We would like to thank Chris Schunn, Paula Raymond, and Lelyn Saner for comments and suggestions on this research.

NOTE

1. The similarity between the two datasets is striking, especially since there were so many differences between the sessions. There were different number of scientists (1 vs. 2), different stages of research (confirmatory vs. exploratory), different domains, and completely different visualization tools used.

REFERENCES

Dunbar, K.: 1997, How Scientists Think: On-Line Creativity and Conceptual Change in Science. In T.B. Ward and S.M. Smith (eds.), *Creative Thought: An Investigation of Conceptual Structures and Processes*. Washington, DC, USA: American Psychological Association, 461–493.

Ericsson, K.A. and H.A. Simon: 1993, *Protocol Analysis: Verbal Reports as Data*, 2nd edn. Cambridge, MA: MIT Press.

Findlay, A.: 1948, *A hundred years of Chemistry*, 2nd edn. London: Duckworth.

Gentner, D.: 1983, Structure Mapping: A Theoretical Framework for Analogy, *Cognitive Science* 7: 155–170.

Gentner, D.: 1989, The Mechanisms of Analogical Learning. In S. Vosniadou and A. Ortony (eds.), *Similarity and Analogical Reasoning*. New York, NY: Cambridge University Press.

Gentner, D., S. Brem, R.W. Ferguson, A.B. Markman, B.B. Levidow, P. Wolff and K.D. Forbus: 1997, Analogical Reasoning and Conceptual Change: A Case Study of Johannes Kepler, *Journal of the Learning Sciences* 6(1): 3–40.

Hadamard, J.: 1945, *The Psychology of Invention in the Mathematical Field*. Princeton, NJ: Princeton University Press.

Hegarty, M.: 1992, Mental Animation: Inferring Motion from Static Displays of Mechanical Systems, *Journal of Experimental Psychology: Learning, Memory and Cognition* 18(5): 1084–1102.

Kaplan, C.A. and H.A. Simon: 1990, In Search of Insight, *Cognitive Psychology* 22: 374–419.

Kosslyn, S.M., K.E. Sukel and B.M. Bly: 1999, Squinting with the Mind's Eye: Effects of Stimulus Resolution on Imaginal and Perceptual Comparisons, *Memory and Cognition* 27(2): 276–287.

Nersessian, N.J.: 1999, Model-Based Reasoning in Conceptual Change. In L. Magnani, N.J. Nersessian and P. Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*. New York: Kluwer Academic/Plenum Publishers, 5–22.

Newman, J.R.: 1955, James Clark Maxwell, *Scientific American* 192(6): 58–71.

Shepard, R.: 1988, The Imagination of the Scientist. In K. Egan and D. Nadaner (eds.), *Imagination and Education*. New York/London: Teachers College Press, 153–185.

Shepard, R. and J. Metzler: 1971, Mental Rotation of Three-Dimensional Objects, *Science* 171: 701–703.

Thomas, N.J.T.: 1999, Are theories of Imagery Theories of Imagination? *Cognitive Science* 23(2): 207–245.

Trafton, J.G., S.S. Kirschenbaum, T.L. Tsui, R.T. Miyamoto, J.A. Ballas and P.D. Raymond: 2000, Turning Pictures into Numbers: Extracting and Generating Information from Complex Visualizations, *International Journal of Human Computer Studies* 53(5): 827–850.

Trafton, J.G. and S.B. Trickett: 2001, A New Model of Graph and Visualization Usage. In *The Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Trickett, S.B., W.-T. Fu, C.D. Schunn and J.G. Trafton: 2000, From Dipsy-Doodles to Streaming Motions: Changes in Representation in the Analysis of Visual Scientific Data. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Trickett, S.B., J.G. Trafton, C.D. Schunn, and A. Harrison: 2001, That's Odd! How Scientists Respond to Anomalous Data. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Tweney, R.D.: 1992. Inventing the Field: Michael Faraday and the Creative "Engineering" of Electromagnetic Field Theory. In R.J. Weber and D.N. Perkins (eds.), *Inventive minds: Creativity in Technology*. New York, NY, USA: Oxford University Press, 31–47.

Watson, J.D.: 1968, *The Double Helix*. New York: New American Library.

J. GREGORY TRAFTON
Naval Research Laboratory
NRL, Code 5513
Washington, DC 20375-5337
USA
E-mail: trafton@itd.nrl.navy.mil

SUSAN B. TRICKETT
George Mason University
E-mail: stricket@gmu.edu

FARILEE E. MINTZ
ITT Industries
E-mail: mintz@aic.nrl.navy.mil