From Dipsy-Doodles to Streaming Motions: Changes in Representation in the Analysis of Visual Scientific Data

Susan B. Trickett
(stricket@gmu.edu)
Department of Psychology
George Mason University
Fairfax, VA 22030

Wai-Tat Fu
(wfu@gmu.edu)
Department of Psychology
George Mason University
Fairfax, VA 22030

Christian D. Schunn
(schunn@gmu.edu)
Department of Psychology
George Mason University
Fairfax, VA 22030

J. Gregory Trafton
(trafton@itd.nrl.navy.mil)
Naval Research Laboratory
NRL Code 5513
Washington, DC 20375

Abstract

This paper investigates the change in scientists' representation of phenomena of interest during the exploratory analysis of visual data. The scientists initially represented expected findings in formal, scientific terms, whereas they represented anomalies in informal terms. Over time, these representations shifted from informal to formal. We propose that this shift in representation is the result of an increased understanding of the individual phenomena, rather than of greater understanding of the data at a global level.

Introduction

A strong and perhaps foundational theme in cognitive science is the issue of representation. From both empirical and computational perspectives, performance has been found to depend heavily on how information is internally represented (Kotovsky, Hayes, & Simon, 1985; Larkin & Simon, 1987; Newell & Simon, 1972; Zhang & Norman, 1994).

One area in which representation is likely to be especially important is scientific discovery (Schunn & Klahr, 1995). There are many formal and informal methods for representing data, even within the same discipline and narrow subarea. The choice of representation of the data is likely to have a large impact on what can and will be discovered.

An additional twist on the issue of data representations in science is the difference between goals of scientific discovery and goals of communication of the discoveries. The external representations that are best for discovery are not necessarily the representations that are best for communication of the discovery to others. For example, issues of historical convention are likely to be more important in communication, whereas issues of ease of generation and manipulation are going to be more important for the original discovery.

The goal of this paper is to examine how scientists represent data internally to themselves while they are analyzing their data. In particular, do they tend to think of their data in formal, discipline-specific terms, or do they rely on more informal and simple perceptual terms? One might expect them to use formal terms because of their expertise and extensive domain knowledge. On the other hand, they may use perceptual terms because in many areas of science, the data are presented in fairly complex visual displays that make heavy use of spatial metaphors—or indeed represent spatial dimensions directly (Trafton et al, under review).

One dimension that we hypothesize would influence the choice of internal representation is the degree to which the data are as the scientist expects. That is, perhaps scientists are more likely to represent apparently anomalous data in informal, perceptual terms and expected data in formal, conceptual terms.

Another related dimension that we investigated was time: How do scientists' representations of their data change over time as they explore their data? One might imagine that the representations become more formal as scientists develop an understanding of the dataset as a whole. Alternatively, the changes in representation may occur at a more item-specific level—the representation of each item changes separately as understanding of the item changes.

A wide variety of methodologies has been used to study scientific reasoning and scientific discovery, each with their advantages and disadvantages (see Klahr & Simon, 1999, for a review). For this research project, we adopted a modified form of Kevin Dunbar's "in vivo" methodology (Dunbar, 1995, 1997, in press). The "in vivo" methodology involves observing scientists as they are doing their research. Dunbar focused on the activities that occur in lab group meetings. Because we were interested in the processes of data analysis, we focused, instead, on pairs of scientists working at their computers, analyzing their data. Like Dunbar, we perform a form of protocol analysis (Ericsson & Simon, 1993), analyzing the speech produced by the scientists to make inferences about the underlying cognitive processes.

The reason for focusing on pairs of scientists rather than on an individual scientist is that dyads produce speech naturally as part of their data analysis activities. By contrast, forcing an individual scientist to give a think-aloud protocol may change the very representations that we seek to study. For example, the individual scientist may change her focus to aspects of the data that are more easily verbalized, or she may change her representations from visio-spatial representations to more verbal representations.

Our methodology also contrasts with the retrospective analyses of historical cases from science (e.g., Gentner et al., 1997; Nersessian, 1985; Thagard, 1999). By focusing on the activities of non-famous (albeit expert) scientists working on a problem that may or may not lead to an important discovery, we may obtain a more representative view of how scientists reason.[1]

[1] Of course, if one's goal is to understand how large conceptual leaps are made in science, the historical case-study approach may be more fruitful.

Because our methodology is extremely labor-intensive, it lends itself most readily to case studies. However, the use of case studies always raises the question of generalizability: does the pattern found with these scientists at this particular time in this particular domain generalize to other scientists in other domains? To address this issue, we gathered data from two sets of scientists working in different disciplines on different kinds of problems. The first set of scientists was a pair of astronomers examining radio and optical data of distant galaxies. The second set of scientists was a pair of neuropsychologists examining fMRI imaging data of brain functioning under different experimental conditions. Thus we included both observational and experimental research from disciplines differing widely in types of training and age of the discipline. One should note, however, that both situations involved preliminary examinations of complex data visualizations presented on computer screens.

Method

Participants

The participants in the first domain were two expert astronomers, one a tenured professor at a university, the other a fellow at a research institute. The astronomers had earned their Ph.D.s six years and ten years respectively before this study; one has approximately 20 journal publications and the other approximately 10 in this area. One of the astronomers, hereafter referred to as A1, focuses on conducting and analyzing astronomical observations, and has an expertise in ring galaxies; the other, hereafter referred to as A2, combines teaching with primarily theoretical astronomical research and model construction. The astronomers have been collaborating for some years, although they do not frequently work physically alongside one another (i.e., work simultaneously at the same computer screen to examine data).

The participants in the second domain were two scientists in neuropsychology, one a postdoctoral researcher (B1) who has been in the field over 3 years, the other a graduate researcher (B2) who has been in the field for 1 year. The scientists work in a renowned national US research institute and are involved in developing a new methodology for analyzing fMRI brain data. They frequently work simultaneously at the same computer screen to examine data.

Procedure

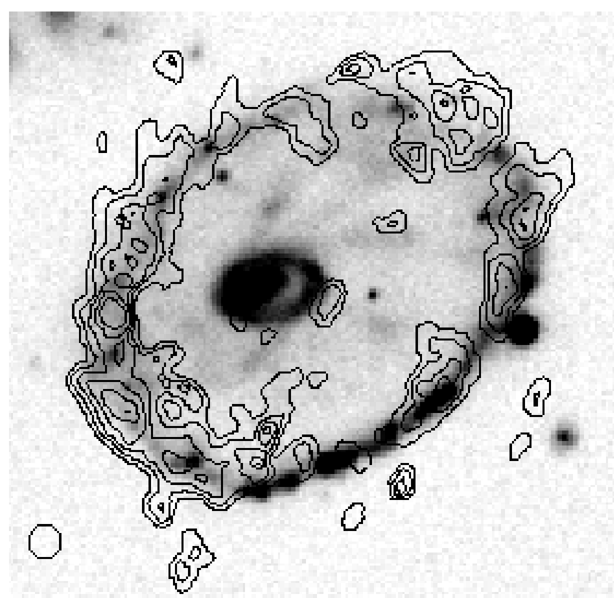
In both studies, the scientists were video- and audio-taped as they explored computer-generated visual representations of a new set of data. For the first study, A1 was in charge of the keyboard and mouse and sat directly in front of the screen; A2 sat slightly to his left. For the second study, B2 was in charge of the keyboard and mouse and sat directly in front of the screen while B1 sat slightly to her right. In both studies, all scientists had the shared monitor in their clear line of sight. They were instructed not to explain or interpret their comments to the researchers, but to carry out their work as though no camera were present. For each study, the relevant part of the session lasted about 1 hour. The scientists' interactions were transcribed and coded as described below. At a later date, we interviewed the scientists in both domains in order to obtain clarification of some domain-related issues.

The Tasks and the Data

The astronomical data under analysis were optical and radio data of a ring galaxy. A ring galaxy forms as the result of a collision between two galaxies, and such collisions are relatively frequent cosmic events; consequently, ring galaxies *per se* are not uncommon. Both astronomers had conducted research and published scholarly articles on other ring galaxies, but this particular galaxy was relatively new to them. Nor had they examined this data set before; consequently, they considered this session exploratory.

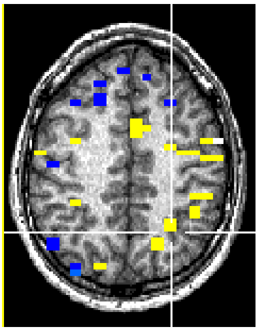
The astronomers' high-level goal was to understand the evolution and structure of the ring galaxy, by a complex sequence of inferences that began with interpreting contour lines on the display in terms of the 3-dimensional flow of gas in the galaxy. The astronomers' task was made difficult by two characteristics of the data: First, the data were one- or at best two-dimensional, whereas the structure they were attempting to understand was three-dimensional. Second, the data were noisy, and there was no easy way to distinguish between noise and real phenomena. Figure 1 shows a screen snapshot of the type of data they were examining.

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



The fMRI data were obtained to understand how activation patterns inside the brain would change when people are anticipating some events to happen. There were two experimental conditions and one control condition. The scientists had to examine the data and compare them across the three conditions. This was the first time they had conducted the experiment and examined the data. The session was considered exploratory. Figure 2 shows an example of the fMRI data that they were analyzing. Similar to the astronomical data, fMRI data are inherently noisy and can only be displayed in two dimensions although the activation patterns under analysis were mostly three dimensional.

Figure 2. Example of fMRI data (color removed).



Coding Scheme

The protocols were divided into 829 (astronomy) and 370 (fMRI) segments. As each scientist spoke in turn, a new segment was established. Then the scientists' individual utterances were further segmented by complete thought.

A coding scheme was developed to examine how the scientists explored the data. The entire astronomy protocol was coded independently by 2 different coders in order to establish the reliability of this scheme. Inter-rater reliabilities for each code are reported below. Because we found high agreement in coding the astronomy protocol, we expect the agreement in the neuropsychology protocol to be high also.

On/Off Task In order to allow us to focus our analysis only on those utterances that were relevant to the scientists' task of data analysis, we coded each segment as on-task or off-task. All segments that addressed matters external to the data analysis were coded as off-task; these segments included external interruptions (e.g., the telephone ringing), extraneous comments by the scientists (e.g., jokes or banter between them), comments relating to the software, specific details about plans for future observations, and so on. All segments that addressed issues of data analysis were coded as on-task. These included comments relating to the selection of a display type (as opposed to comments about how to implement that display) as well as decisions about obtaining additional data in the future (as opposed to details about how to obtain those data). Initial agreement between the coders was 90%. All disagreements were resolved by discussion.

Noticings In order to establish which phenomena the scientists attended to, we first coded for the scientists' *noticing* phenomena in the data or features of the display. A noticing could involve merely some surface feature of the display, such as a line, shape, or color, or it could involve some interpretation by the scientists, for example, identifying an area of star formation or concentration of gas for the astronomers or activation in a particular area of the brain (e.g. thalamus) for the neuropsychologists. Only the first reference to a phenomenon was coded as a noticing; coding of subsequent references to the same phenomenon is discussed below. Agreement between the coders was 95%. Disagreements were resolved by discussion.

Because our investigation focused on the change in representation of anomalies in the data, we further coded these noticings as either "anomalous" or "expected," according to one or more of the following criteria: a) in some cases the scientists made explicit verbal reference to the fact that something was anomalous or expected; b) if there was no explicit reference, domain knowledge was used to determine whether a noticing was anomalous or not; c) a phenomenon might be associated with (i.e., identified as either like) another phenomenon that had already been established as anomalous or not; d) a phenomenon might be contrasted with (i.e., identified as unlike) a phenomenon that had already been established as anomalous or not; e) the scientists might question a feature, thus implying that it is unexpected. Table 1 illustrates these codes. Agreement between the coders was 87%. Those noticings for which disagreement could not be resolved were excluded from further analysis.

Subsequent References Our investigation focused on the astronomers' representation of phenomena over time. Whereas the coding of the noticings captured the first reference the astronomers made to a phenomenon of interest, we also needed to establish how they made subsequent reference to each noticing. Consequently, all subsequent references to each phenomenon were also identified.

Because the scientists were sharing a computer monitor, frequently the first interaction between them after a noticing was to establish that they were both looking at the same thing. Subsequent references that served purely to establish identity were *not* included in the analyses.

Not all subsequent references immediately followed a noticing; frequently, the scientists returned to a phenomenon of interest after investigating other features of the data. The

Table 1. Noticings (italicized) coded as unusual or expected.

Criterion	Code	Example - Astronomy	Example - fMRI
Explicit	Anomalous	What's that *funky thing*…That's odd	Bunch of *stuff here* … Yeah, that's weird
Domain Knowledge	Expected	You can see that all the *H1* is concentrated in the ring	So there is a *subcortical activation* that is probably caudate.
Association	Anomalous	You see similar kinds of *intrusions* along here	So there's the *thing* we've been seeing consistently.
Contrast	Expected	That's odd…As opposed to *these things*, which are just the lower contours down here	So *it's lateral*, which means its not in the midline… on our incentive task we see midline, but not lateral, so that's why that's not a spot.
Question	Anomalous	I still wonder why we don't see any *H1* up here in this sort of northern ring segment?	[None found]

scientists made frequent gestures to the feature of the image under discussion; by constructing a map of the noticings, and cross-referencing it with these gestures, the coders were able to determine the specific noticing to which a subsequent reference referred. Tables 2a and 2b illustrate the coding scheme for subsequent references in each domain.

Entity Coding To investigate the initial and changing representations of the phenomena the scientists noticed, we first identified what characteristics of each noticing (anomalous and expected) first caught the scientists' attention. We then noted what characteristics the scientists attended to in their subsequent references to each noticing. We coded each noticing and subsequent reference as either "formal" or "informal" as follows. Formal references are those for which the scientists referred to the underlying phenomenon, using the terminology of the domain—for example, to a specific gas, star formation, the stellar continuum, or the like in the astronomy domain, and to the thalamus or caudate nucleus, for example, in the neuropsychology domain. Informal references include references to some generic feature of the display, such as a blob, a bulge, or a “dipsy-doodle" in the astronomy domain, and a neuron "lighting up" in the neuropsychology domain. They also include references to a phenomenon by its location (e.g., "lower right," "northwest") and anaphoric references, (e.g., pronouns). A few references combined characteristics of more than one code (e.g., "big blob of H1" combined the informal reference to a "blob" with the formal reference to H1 gas). Such references were coded as "mixed" references, and were excluded from subsequent analysis. Coder agreement on this coding was 100%.

Table 2a. Subsequent references in astronomy domain.
Noticing: First reference to phenomenon
Establish identity: Reference excluded from analysis
SR: Subsequent reference included in analysis

Code	Utterance
Noticing (N9)	A1: What's that funky thing…
Establish identity	A2: Left center, you mean…
Establish identity	A2: This stuff? [points to screen]
Establish identity	A1: Yeah
Establish identity	A2: Yeah
SR to N9	A1: What is that?
Noticing (N10)	A2: You can see there is some gas here [points to different area] inside the ring, but not much…
SR to N9	A1: Except for that little knot there.

Table 2b. Subsequent references in neuropsychology domain.

Code	Utterance
Noticing (N23)	B1: There, did you see that?
Establish identity	B2: Yah did you see that? [points to screen]
SR to N23	B1: That was near the thalamus.
SR to N23	B1: That might be spurious.
Noticing (N24)	B2: So the z-score of that one is 4.22.
SR to N24	B1: It's right up there [points to the threshold on screen]

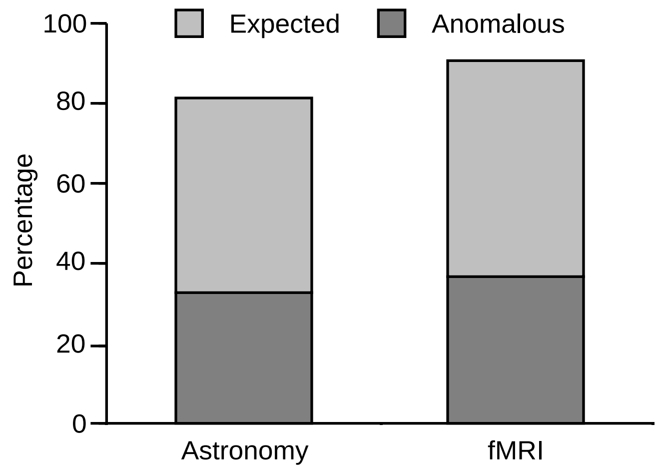
Results and Discussion

There were 619 (75%) (astronomy) and 317 (85%) (neuropsychology) on-task segments. Subsequent analyses do not include off-task segments.

Noticing Anomalies and Expected Phenomena

There were 27 (astronomy) and 35 (fMRI) noticings. In the astronomy data, 9 (33%) were anomalous, 13 (48%) were expected, and 5 (19%) were uncoded because either the astronomers or the coders disagreed. In the fMRI data, 13 (37%) were anomalous, 19 (54%) were expected, and 3 (9%) were uncoded. Figure 3 shows that the proportion of anomalous and expected noticings was similar in each dataset. Uncoded noticings were excluded from subsequent analysis.

Figure 3. Percentage of anomalous and expected noticings.

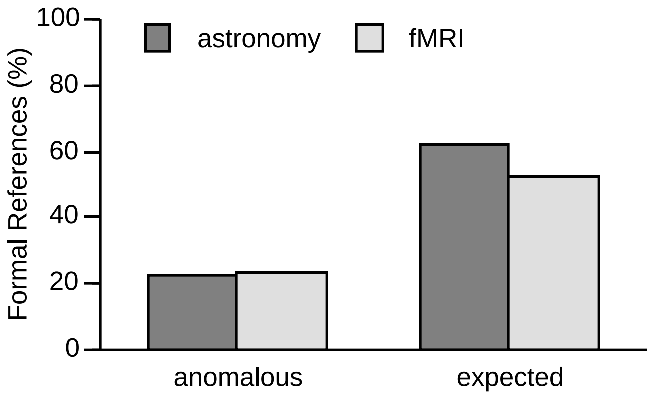


Representation of Noticings

Our first question concerned how the scientists initially represented the phenomena they investigated. In the astronomy domain, 8 of the 13 (62%) expected phenomena were first identified by formal references and the remaining 5 (38%) by informal references. In contrast, most of the anomalies (78%) were initially identified by informal references, with only 2 of the 9 anomalies (22%) identified formally. Interestingly, both formal references were negative—they referred to the *absence* of the astronomical phenomenon (e.g., "I still wonder why we don't see any H1 up here.") A similar pattern was observed in the neuropsychology domain. Ten of the 19 (53%) expected phenomena were first identified by formal references, and 9 (47%) by informal references. Ten of the 13 (77%) anomalies were identified informally, with 3 (23%) identified by formal references. Again, 2 of the 3 formal references were negative (e.g. "There's nothing on the thalamus either, that's surprising"). Figure 4 shows the percentage of formal references to these initial noticings.

Thus it appears that in general, the scientists initially represented the expected phenomena in the formal, scientific terminology of that domain. However, their initial representations of unexpected or anomalous features of the data were highly informal. Recall that these informal references were based primarily on irregular features of the display rather than the underlying phenomena that these features represented. Occasionally, it was the *absence* of a phenomenon that first drew the scientists' attention to these anomalies.

Figure 4. Percentage of formal references to initial noticings (anomalous and expected) in two domains.



Local Changes in Representation

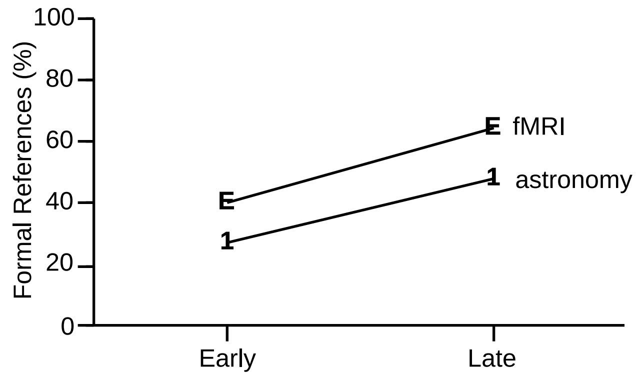
Next, we examined whether the scientists' representation of these phenomena changed as their investigation of the data progressed. The analyses that follow depend on the subsequent references to the noticings. In order to ensure a sufficient basis on which to judge change, we include only those noticings that received more than the mean number of subsequent references (i.e., more than 8 subsequent references for the astronomy and more than 3 for the fMRI data).

Five of the noticings in the astronomy data and 15 in the fMRI data received more than the mean number of subsequent references. In the astronomy data, the subsequent references to these 5 noticings account for 66% of all segments that made any reference to a phenomenon noticed by the astronomers. In the fMRI data, the subsequent references to these 15 noticings account for 69% of all such segments. Thus by confining our analyses to these 20 noticings, we focus on the majority of the data. It should also be noted that, because in general expected phenomena received little further attention, especially in the astronomy domain (Trickett, Trafton, & Schunn, 2000), most noticings included in these analyses are anomalies.

In order to examine change over time, we divided the period of attention to each individual noticed object into two phases, early and late. We tallied the total number of subsequent references for each and divided it by 2. For noticed objects with an odd number of subsequent references, we discarded the midpoint reference, to insure an even split. We then compared the numbers of formal and informal references in the early and late phases of the scientists' investigation.

In the astronomy data, in the early phase of investigation, 17 of the 63 (27%) subsequent references were formal compared with 30 (48%) in the later phase. By contrast, 39 (61%) of the subsequent references were informal in the early phase compared with 25 (39%) in the later phase, $\chi^2(1) = 6.65$, $p < .01$. (These percentages do not sum to 100% because of the mixed references excluded from the analysis.) In the fMRI data, in the early phase of investigation, 18 of the 45 (40%) subsequent references were formal compared with 29 (64%) in the later phase. By contrast, 27 (60%) subsequent references were informal in the early phase compared with 16 (36%) in the later phase, $\chi^2(1) = 5.39$, $p < .05$. Thus, in both domains, the number of formal representations increased, while the number of informal representations decreased. Figure 5 shows the increase in formal references in the later phase.

Figure 5. Changes in representation of noticed objects.



These results show that the scientists' representations changed significantly over time, as they investigated these anomalies. In the early phase of analysis, their representations were informal and display-based, most likely because they did not have a precise understanding of the phenomenon under investigation. The scientists needed a label by which they could identify, discuss, and refer to the phenomenon, and this label tended to be based on the visual appearance of the feature. As their investigation proceeded, however, these visually-based labels decreased. The reduction in display-based and anaphoric references suggests that the scientists became more specific, and points to an increased understanding of these anomalous phenomena.

Global Changes in Representation

It is possible that the shift toward a formal representation occurred not because the scientists' understanding of individual anomalies increased, but because their global understanding of the data increased over time. In order to investigate this possibility, we divided each entire analysis session into early and late phases, based on overall time spent. Thus, in this analysis, there were unequal numbers of reference in each phase, but the time spent on each phase was the same.

We counted the number of formal and informal references to these well-referenced phenomena in each phase. In the astronomy protocol, 63% of the references in the early phase were informal, compared with 45% in the late phase; 35% of the early references were formal, compared with 39% in the late phase. This difference was not significant, $\chi^2(1) = 1.4$, $p > .2$. Although the proportion of informal references did drop off, the number of formal references remained constant. In the fMRI protocol, 53% of the references in the early phase were informal, compared with 52% in the late phase; 47% of the early references were formal, compared with 48%

in the late phase. This difference was also non-significant, $\chi^2 (1) < 1$. Thus, it does not appear that the shift toward a more formal representation occurred as a result of a more general, global understanding of the data.

General Discussion

Our results show that both groups of scientists initially represented expected and anomalous phenomena quite differently. Whereas they represented the expected phenomena in the formal terms appropriate to their domain of expertise, they represented the anomalous phenomena in highly informal terms that referred to salient features of the visual data. These results also show that these internal representations changed over time, shifting from informal to formal representations. However, this shift in representation did not appear to be caused by a global increase in understanding of the data under analysis, but was instead local, and associated with the individual phenomena under investigation. This shift in representation appears to have affected primarily the scientists' representation of anomalous or unexpected findings in their data. We have investigated elsewhere the key role of anomalies in the exploratory stages of data analysis (Trickett, Trafton & Schunn, 2000).

Our focus in this study has been on the exploratory stages of data analysis. We believe that including two independent data analysis sessions in quite different scientific domains strengthens our claims about these changes in representation. However, clearly we need to ascertain whether our results generalize to other situations and scientific domains.

In this paper we examined fairly small changes in representation at the item-specific level. Much research in cognitive science on the topic of conceptual change has focused on relatively larger scale changes in representation (e.g., Chi, 1997; Thagard, 1999). One may wonder what the relationship is between the micro-level changes that we have reported in this paper and the more macro-level changes reported in the conceptual change literature. Some researchers (e.g., Chi, 1997) have speculated that some macro-level changes are not the result of many small changes; instead Chi has argued that some macro-level changes are the result of a complete conceptual reorganization. Similarly, some historians of science have noted that some scientific changes appear to be more radical or revolutionary than others appear to be (e.g., Kuhn, 1967). We believe that the relationship between the micro-level changes in representation and the macro-level changes that are thought to constitute conceptual change remains an open question, and that this question could be fruitfully studied by observing the same scientists over a much longer time scale than we have done so far. We are currently planning such longitudinal studies.

Acknowledgments

This research was supported in part by student fellowships from George Mason University to the first and second authors and by grant number 55-7850-00 from the Office of Naval Research to the Naval Research Laboratory. We thank Georgia Seeley and Audrey Lipps for assistance with coding.

References

Chi, M. T. H. (1997). Creativity: Shifting across ontological categories flexibly. In T. B. Ward & S. M. Smith (Eds.), *Creative thought: An investigation of conceptual structures and processes.* Washington, DC, USA: American Psychological Association.

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science, 17*(3), 397-434.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. E. Sternberg & J. E. Davidson, (Eds.), *The nature of insight.* Cambridge, MA, USA: MIT Press.

Dunbar, K. (in press). What scientific thinking reveals about the nature of cognition. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), *Designing for Science: Implications from everyday, classroom, and professional settings*. Mahwah, NJ: Erlbaum.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Gentner, D., Brem, S., Ferguson, R. W., Markman, A. B., Levidow, B. B., Wolff, P., & Forbus, K. D. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *Journal of the Learning Sciences, 6*(1), 3-40.

Klahr, D. & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin, 125*(5), 524-543.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*(2), 248-294.

Kuhn, T. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science, 4*, 317-345.

Nersessian, N. J. (1985). Faraday's field concept. In D. Gooding & F. James (Eds.), *Faraday rediscovered.* London: Macmillan.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Schunn, C. D., & Klahr, D. (1995). *A 4-space model of scientific discovery.* In the proceedings of the 17th Annual Conference of the Cognitive Science Society.

Thagard, P. (1999). *How scientists explain disease.* Princeton, NJ: Princeton University Press.

Trafton, J. G.., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (under review). Turning Pictures into Numbers: Use of Complex Visualizations.

Trickett, S. B., Trafton, J. G., & Schunn, C. D. (2000). *Blobs, dipsy-doodles, and other funky things: Framework anomalies in exploratory data analysis.* In the proceedings of the 22nd Annual Conference of the Cognitive Science Society.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*(1), 87-122.