Blobs, Dipsy-Doodles and Other Funky Things: Framework Anomalies in Exploratory Data Analysis

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

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

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

Abstract

This study investigates the role of anomalies in the exploratory analysis of visual scientific data. We found that anomalies played a crucial role as two experts analyzed astronomical data. Not only did they pay significantly more attention to anomalies than expected phenomena, both immediately and over time, but also anomalies provided a framework within which they investigated the data.

Introduction

Attention to the unexpected may be an important component of scientific discovery. Exploring anomalies can lead to theory development and even conceptual change. Philosophers of science (e.g., Kuhn, 1962) have argued that unusual findings play a key role in scientific revolutions, and scientists themselves have claimed that investigating anomalies lies at the heart of scientific innovation (e.g., Knorr, 1980).

Within cognitive psychology, response to anomalous data during scientific inquiry has been noted in a variety of studies, including historical reconstructions of actual scientific discoveries (e.g., Kulkarni & Simon, 1988), on-line studies of scientists (e.g., Dunbar, 1997), laboratory studies in which participants "rediscover" a scientific phenomenon (Dunbar 1993), and studies of those with little scientific training as they perform abstract scientific reasoning tasks (e.g., Tweney, Dowerty, & Mynatt, 1982; Klahr & Dunbar, 1988). These studies have not yielded a consistent pattern of response to unexpected data, possibly because of the range of scientific training and knowledge among the participants.

Recognizing this variety of responses to anomalous data, Chinn and Brewer (1992, 1993), propose a taxonomy of seven reactions to unusual findings, from ignoring the data and upholding the theory to accepting the data and changing the theory. This taxonomy is derived from anecdotal examples from the history of science and from empirical studies of scientific reasoning in the psychological literature. Although Chinn and Brewer propose that this taxonomy applies to scientists and non-scientists alike, they have tested it only among undergraduates with little scientific training.

Thus, despite the general belief that anomalous data is important in scientific discovery, no clear picture has emerged of how scientists (as opposed to laypersons performing scaled-down scientific discovery tasks) respond to unexpected findings. On one hand, there is a well-established tradition in studies of scientific thinking that shows people overlook data inconsistent with their hypothesis, looking *only* for support for their theories (e.g., Wason, 1960). Within this tradition, scientists have been found to be as susceptible to this confirmation bias as laypeople (e.g., Mahoney & DeMonbreun, 1977; Mitroff, 1974.) Similarly, studies of complex visualization usage have shown that expert meteorologists do not pay much attention to unusual or anomalous features. Instead, they seem to extract information in a very goal directed manner, rarely following up on features that are not directly relevant to their immediate task (Trafton et al., under review). This evidence—of confirmation bias, even among scientists, and of the goal-directed nature of complex visualization usage—suggests that scientists may overlook unexpected results or anomalies.

On the other hand, however, Dunbar has recently questioned the validity of the studies of confirmation bias on the grounds that they employ arbitrary experimental tasks that involve no scientific knowledge and therefore bear little relationship to tasks that real scientists perform (Dunbar, 1997). Dunbar has argued that in order to investigate how scientists reason, one must observe scientists as they perform their scientific tasks.

Using an "in vivo" methodology that involves observing actual scientists at work, Dunbar has suggested that scientists do attend to unusual results (Dunbar, 1997). He found not only that scientists attended to unexpected results more than they did to expected findings, but also that individual scientists were quick to discard a hypothesis when faced with results that were inconsistent with it. Furthermore, he noted that in lab meetings, the group of scientists tended to focus on a surprising result until they had constructed a plausible hypothesis to account for it. Dunbar concluded that attending to anomalous findings is an important strategy that contributes to successful scientific inquiry (Dunbar 1997). Similarly, Kulkarni and Simon (1988) identified an "attend to surprising result" heuristic as crucial to Hans Krebs' discovery of the urea cycle.

Both Chinn and Brewer's and Dunbar's studies have involved participants, whether trained scientists or not, who were evaluating data to test a specific theory. However, there are many phases of scientific inquiry, and response to anomalous data might be quite different during an exploratory phase from when a theory is firmly established. During exploratory data analysis, theories may be only partially defined. Nonetheless, given their extensive domain knowledge, scientists doubtless have general frameworks which lead to expectations that may or may not be met by the data. They may therefore pay more attention to unusual results, because such framework anomalies may provide insights for interpreting data and developing theories.

Similarly, there are many forms of data, but previous studies have focused on data that were either presented textually or required direct, relatively simple perceptual judgments. However, scientists in many domains employ complex visualization techniques in order to inspect their data. Little is known about the role of unusual or unexpected findings in either exploratory or scientific visualization.

Our goal is to investigate the role of anomalies during early, primarily exploratory phases of visual data analysis. Specifically, we investigate whether scientists notice anomalies in this type of data and, if so, the extent to which they attend to them, both immediately and over time. We also investigate how the visual nature of the data affects the detection of and attention to anomalies.

There are many methodologies available by which to examine the processes of scientific inquiry, and there are strengths and weaknesses associated with each (see Klahr & Simon, 1999 for a review). Our approach has been to combine features of several methodologies in order to take advantage of their respective strengths.

First, we have chosen to conduct a case study of actual scientists at work because this methodology offers an extraordinarily rich set of observations of high face validity. Most case studies of scientific inquiry have focused on famous scientists who have made discoveries of great historical importance (e.g., Gentner et al, 1997; Kulkarni & Simon, 1988). We have chosen instead to focus our investigation on more "ordinary"—albeit expert—scientists, the ultimate significance of whose work is currently unknown. We believe this focus on the more mundane aspects of scientific inquiry may yield results that are more representative of scientists' everyday activities.

Second, we collected verbal and visualization data of the scientists working together and conducted a verbal protocol analysis of these data in order to gain insight into the scientists' concurrent thought processes (Ericsson & Simon, 1993). Verbal protocols have frequently been collected in laboratory studies of non-scientists performing scientific discovery tasks; however, this methodology has rarely been used with practicing scientists. Furthermore, because we collected a protocol of a work session involving two scientists, there was no need for an experimenter to prompt the participants to keep talking. By focusing on a dyad, we were able to obtain a more natural account of the scientists' thinking than is possible with an individual.

Finally, we have adopted Dunbar's (1995, 1997) "in vivo" methodology because, as Dunbar points out, it affords a unique opportunity to observe "how scientists really reason." Instead of observing a lab group as Dunbar did, however, we chose to study a dyad, for two reasons. First, the two scientists we observed were of equal professional status, thus we avoid social issues that might make junior scientists reluctant to question the interpretations of a senior colleague. Second, we think that the verbal protocols of a dyad might represent each scientist's thinking more completely than those of a group. In a group setting, with more people "jumping into" a discussion, individuals may be less likely to pursue lines of thought in significant depth.

Method

Participants

The participants in this study 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 publications in this general area and the other approximately 10. 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 astronomer, 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 at the same computer screen at the same time to examine data).

Procedure

The astronomers were video- and audio-taped as they explored computer-generated visual representations of a new set of observational data. They were working in one astronomer's office at a shared computer monitor. One astronomer was in charge of the keyboard and mouse and sat directly in front of the screen; the other astronomer sat to his left, with the monitor clearly in view. They were instructed not to explain their comments to the researchers, but to carry out their work as though no camera were present. The relevant part of the session lasted about 53 minutes and generated 7676 words. The astronomers' interactions were later transcribed and coded as described below. At a later date, we interviewed A2 to obtain clarification of domain-related issues.

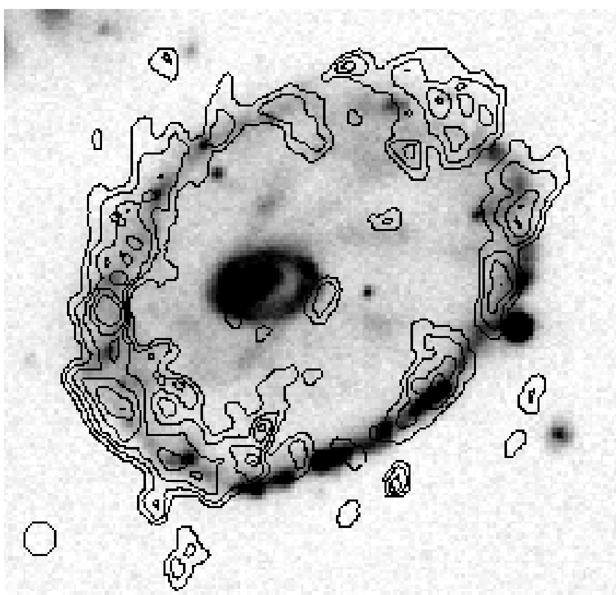
The Task and the Data

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: one galaxy is thought to have passed through another, leaving both a doughnut-shaped ring of stars and gas (the ring galaxy) and a smaller galaxy nearby. Such galactic 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. This understanding emerges from an understanding of where, how, and why star formation occurs within the galaxy, which rests on an understanding of the flow of gas in the galaxy. In order to understand the flow of gas, the astronomers must understand the kinematics (the velocity and position) of the system, by inferring the 3-dimensional streaming motions of the gas. They make inferences about streaming motions by interpreting the velocity field, represented by contour lines on the 2-dimensional display. Examining the velocity contours is thus the lowest level task in this chain of inferences.

The astronomers' task was made difficult by two characteristics of their data. First, the data were one- or at best twodimensional, whereas the structure they were attempting to understand is 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 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 represent different information about the structure and dynamics of the galaxy. Some of these images could be overlaid on each other. In addition, the astronomers could choose from images created by different processing algorithms that result in different weightings of the data, each with advantages and disadvantages (e.g., more or less resolution). Finally, they could adjust different features of the display, such as contrast or false color.

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



Coding Scheme

The protocol was divided into 829 segments: as each astronomer spoke, a new segment was coded; then their utterances were further segmented by complete thought.

A coding scheme was developed to examine the astronomers' attention to anomalous phenomena in the ring galaxy. The protocol was coded independently by two different coders. Inter-rater reliabilities for each code are reported below.

On/Off Task In order to allow us to focus our analysis only on those utterances relevant to the scientists' task of data analysis, we coded each segment as on- 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 astronomers (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 segments 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.

Episodes Next, we divided the protocol into discrete, non-overlapping episodes that would allow us to study the astronomers' shifting focus of attention. The protocol was segmented into 19 exhaustive episodes. An episode began with the astronomers' focus on a feature or point of discussion and lasted until their attention switched to another phenomenon or theoretical point; at this switch of attention, a new episode was coded. Although the focus of most episodes was a feature of the galaxy, this was not necessarily the case; for example, one episode consisted of a discussion about a future observation session and the data to be obtained from it. Agreement between coders was 98%.

A new episode frequently coincided with a display change, but did not necessarily do so. Sometimes the astronomers switched their focus of attention to another galactic feature visible on the same display, thus beginning a new episode without changing the display. At other times, they changed the display in order to explore another representation of a feature, thus changing the display within the same episode.

Noticings In order to establish which phenomena—unusual or not—the astronomers attended to, we first coded for the astronomers' *noticing* phenomena in the data. 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 astronomer, for example, identifying an area of star formation or concentration of gas. 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.

Table 1: Noticings (in italics) coded as unusual or expected

Criterion	Code	Example
a) Explicit	Anomalous	What's *that funky thing*…That's odd
b) Domain Knowledge	Expected	You can see that *all the H1* is concentrated in the ring
c) Association	Anomalous	You see *similar kinds of intrusions* along here
d) Contrast	Expected	That's odd…As opposed to *these things*, which are just the lower contours down here
e) Question	Anomalous	I still wonder why *we don't see any H1 up here* in this sort of northern ring segment?

Subsequent References One of our questions was the extent to which the astronomers continued to investigate anomalies. Whereas the coding of the noticings captured the first reference the astronomers made to a phenomenon of interest, we needed to establish how frequently they made subsequent reference to each noticing. Consequently, all subsequent references were also identified and coded.

Because the astronomers 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.

Table 2: Coding of subsequent references
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.

Not all subsequent references immediately followed a noticing; frequently, the astronomers returned to a phenomenon of interest after investigating other features of the galaxy. The astronomers made frequent gestures to the feature of the image under discussion; by constructing a map of the noticings on the galaxy, and cross-referencing it with these gestures, the coders were able to determine the specific noticing to which a subsequent reference referred. Table 2 illustrates the coding scheme for (sequential) subsequent references.

Results and Discussion

There were 619 on-task segments (75%). Subsequent analyses do not include off-task segments.

Noticing Framework Anomalies

Our first question was did the astronomers notice anomalies in the data? Recall that a “noticing” is a first-time reference to a phenomenon of interest. There were 27 noticings during this session. Of these, 9 (33%) were anomalous, 13 (48%) were expected, and 5 (19%) were uncoded, because the astronomers themselves or the coders disagreed. This analysis shows that at least one-third of the phenomena the astronomers identified were unusual in some way. It appears then that the astronomers *did* notice anomalies in this dataset.

Interestingly, most of the anomalies (78%) were identified in highly informal terms or by features of the display, rather than by underlying astronomical phenomena. Thus, the astronomers usually identified anomalous phenomena as "blobs," "bulges," or "dipsy-doodles" rather than in formal astronomical terms (such as a specific type of gas). Not only were anomalies important to the astronomers, but their attention to these anomalies appears to be drawn primarily by visual features of the data. We investigate the relationship between representation and anomalous/expected results elsewhere (Trickett, Fu, Schunn, & Trafton, 2000).

Relationship between Episodes and Noticings

Next, we investigated whether the anomalies played any part in guiding or structuring this exploratory session, that is, whether there was any relationship between the noticings and the episodes, and if so, whether this relationship was different for the anomalies than for the expected phenomena.

In order to investigate this relationship, we noted how each episode began. Nine of the 19 episodes began with a noticing, 7 began with a subsequent reference, and 3 episodes began with something other than a noticing or subsequent reference to a noticing. Thus, out of 19 episodes, only 3 were initiated by theoretical or other similar considerations. Noticing and subsequent references, while common, only account for 61% of the segments. Thus episodes are more likely to start with a data-driven event (noticing or SR) than one would expect from the base rates, X^2 (1) = 3.88, $p < .05$. This result suggests that the most likely focus of attention was some feature of the data rather than some theoretical or other matter. This exploratory session analyzing visual data appears to have been driven primarily by the data themselves rather than theoretical considerations.

What features of the data were likely to attract the astronomers' attention? Of the noticings that sparked an episode, an equal number (3) were anomalous and expected. But whereas *no* episodes began with a subsequent reference to an expected phenomenon, 6 subsequent references to anomalies launched a new episode. This analysis suggests that at a first pass, the astronomers were equally likely to attend to expected as to anomalous phenomena in the data. However, as they explored the data further, it was the anomalies, not the expected phenomena, that directed their investigations. Table 3 summarizes these results.

Table 3: Noticings and subsequent references beginning an episode

	Notice	SR
Anomalous	3	6
Expected	3	0

Initial Attention to Anomalies

Once the astronomers had identified something unusual in the data, what did they do with this observation? There are several possible reactions: they could pursue the anomaly in order to try to account for it, they might temporarily disregard it but return to it later, or they might move on to explore some other, better understood, aspect of the data. A related question is whether their response to anomalies was different from their response to expected phenomena.

First, we investigated this issue by considering the extent to which the astronomers made subsequent reference to a noticing immediately upon identifying it. If anomalies and expected phenomena are of equal interest, we would expect them to make a similar number of subsequent references to both the anomalous and expected patterns. However, if anomalies play a more important role in their efforts to understand the structure of the galaxy, we would expect them to pay more attention (measured by the number of subsequent references) to anomalies than to expected observations.

Although there were fewer anomalies identified in this session, collectively these anomalies received over 3 times as many subsequent references within the same episode as the expected phenomena. The total number of subsequent references to anomalies was 68 (mean = 7.6), compared with 19 (mean = 1.5) for expected phenomena. A t-test on these data was significant, t (20) = 2.27, SE = 2.69, $p < .05$. These results show that the astronomers did pay more attention to the anomalies immediately upon noticing them, or soon thereafter, than they did to the expected phenomena. This in turn suggests that the anomalies were more important to the astronomers than those phenomena they expected to find. Table 4 summarizes the results of this analysis.

Table 4: Subsequent references (SRs) within episodes

	Total SRs	Mean SRs	Range
Anomalous (N=9)	68	7.6	1 - 30
Expected (N=13)	19	1.5	0 - 4

Furthermore, as Table 4 shows, the range of subsequent references was also much greater for the anomalies than for the expected phenomena. *All* the anomalies received at least one subsequent reference soon after the astronomers noticed it. In contrast, 5 of the 13 expected phenomena (38%) received no subsequent references, i.e., no immediate further attention. In addition, 5 of the 9 anomalies (56%) received more than 5 subsequent references; none of the expected phenomena was referred to so frequently. This analysis provides further support for our claim that overall, the anomalies were more important to the astronomers' goals than the expected phenomena. In addition, it suggests that the anomallies themselves were not of equal importance, with some anomalies receiving much more attention than others.

Long-Term Attention to Anomalies

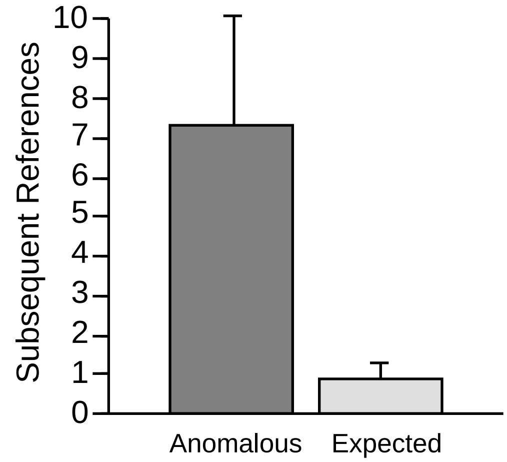
It appears, then, that as the astronomers explored the data about the ring galaxy, they paid more attention immediately to the anomalies in this data than they did to expected phenomena. These results say nothing, however, about the continuing role of the anomalies in the astronomers' analysis. Possibly, having explored an anomaly, the astronomers might "consider the matter closed" and switch their attention to another phenomenon. As with the astronomers' immediate attention to phenomena, we compare their treatment of anomalies with their response to the expected phenomena.

In order to investigate the extent to which the astronomers revisited the phenomena they noticed, we examined the number of subsequent references to both anomalies and expected findings *across* episodes. Recall that an episode ended when the astronomers switched attention to another feature or point of discussion. Thus, a reference to a feature across an episode indicates a switch of attention *back* to that feature, after having focused attention on something else.

One of the expected phenomena was first noticed in the last episode; because it was not possible for it to be referenced in a later episode. Thus the number of expected phenomena for these analyses is reduced from 13 to 12.

Seven of the 9 anomalies (78%) were referenced across episodes, compared with 6 (50%) of the expected phenomena. Overall, the total number of subsequent references across episodes to anomalies was 66 (mean = 7.3), compared with 11 (mean = 0.9) to expected phenomena. A t-test on these data was significant, t (19) = 2.66, SE = 2.41, $p <$ .05. This result shows that the astronomers continued to pay more attention to anomalies than to expected phenomena, revisiting them even after switching their attention to other features of the data. Figure 2 summarizes these results.

Figure 2: Mean subsequent references (SRs) across episodes



Five of the 9 anomalies (56%) received more than 5 subsequent references across episodes. None of the expected phenomena was referenced so frequently. Furthermore, the astronomers persisted in returning to some anomalies, in 3, 4, 5, or even 6 episodes. The spread of episodes during which subsequent references were made was quite extensive and in several cases spanned almost the entire session. For example, Noticing 2 was first identified in episode 1 and was further referred to in episodes 2, 5, 6, 13, 15, and 17. Noticing 11 was first identified later in the session, in episode 9, and was further referenced in episodes 11, 13, 15, and 17. These results suggest that some anomalies were very puzzling to the astronomers and that they were sufficiently important to the exploration of the data that they returned to them repeatedly, even long after they had first noticed them.

General Discussion and Conclusion

This study was conducted to investigate the role of anomalies in the exploratory stages of visual data analysis. We found that the astronomers did notice and pay attention to anomalies. They paid significantly greater attention to the anomalies in the data than to the expected phenomena. Furthermore, they found some anomalies sufficiently intriguing that they returned to them later in their exploration, in some cases repeatedly and over relatively long stretches of time. None of the expected phenomena received this type of prolonged attention. We conclude, therefore, that anomalies played an important role in the exploration of these data.

In addition, we found that the astronomers' attention was initially drawn by features of the data rather than theoretical considerations. Although at first an expected phenomenon was as likely as an anomaly to become the focus of attention, as the analysis proceeded, the anomalies were more likely to hold the astronomers' attention. Furthermore, attention to the anomalies was initially drawn by irregular features of the visual representation rather than by the underlying phenomenon itself. This suggests that their approach was highly perceptual, because they identified anomalies primarily on the basis of unusual curves, lines, etc.

It is possible that anomalies played a significant role in this data analysis session *because* of the visual nature of the data. The anomaly was visible on the display at all times; consequently, it is possible that the astronomers were cued primarily by the display rather than memory to revisit the anomaly. However, this does not seem to be the case. If the display were the only means by which the astronomers were cued to make subsequent reference to the anomaly, we would expect them to make subsequent references to *all* anomalies. As our results indicate, though, they were selective in the anomalies they continued to investigate. Although visibility on the display may have helped to keep a particular anomaly activated in the astronomers' memory, this alone does not seem to have been sufficient to prompt them to revisit it. Rather, it appears that some anomalies were "tagged" as worthy of further investigation, and that the astronomers continued to search for a satisfactory way to explain them.

In contrast to the widely-held belief that scientists are susceptible to confirmation bias and seek chiefly to confirm what they already expect, our results present a picture in which investigating framework anomalies is a central activity in exploratory data analysis. We propose that the anomalies were instrumental in guiding the structure and content of the data analysis session.

We acknowledge that this is a case study of particular scientists in one domain, working at a specific phase of their research. However, our results are part of a growing body of evidence that attention to anomalies may be an important component of scientific inquiry (cf. Dunbar, 1997). Moreover, the scientists in this study were engaged in a task—exploratory data analysis—that is undertaken in all scientific domains. They neither employed unusual techniques nor used specialized equipment unique to their domain. We therefore expect our results to generalize to scientists in other domains. Whether or not they apply to later stages of data analysis (such as hypothesis-testing) remains an open question. We are currently extending this research by applying our methodology to a variety of scientists working with scientific visualizations in several domains. We are also planning to conduct longitudinal observations of scientists, in order to investigate the role of anomalies in their work at different stages of data analysis.

Acknowledgments

This research was supported in part by a student fellowship from George Mason University to the first author 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 coding assistance and Erik Altmann, Melanie Diez, Anthony Harrison, William Liles, and Lelyn Saner for comments.

References

Chinn, C. A., & Brewer, W. F. (1992). Psychological reponses to anomalous data. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching, 35*(6), 623-654.

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

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & J. Vaid, (Eds.), *Creative thought: An investigation of conceptual structures and processes*. Washington, DC, USA: American Psychological Association.

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., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1-48.

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

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211-228.

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

Kulkarni, D., & Simon, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science, 12*(2), 139-175.

Mahoney, M. J., & DeMonbreun, B. G. (1977). Psychology of the scientist: An analysis of problem-solving bias. *Cognitive Therapy and Research, 3*, 229-238.

Mitroff, I. (1974). *The subjective side of science: A philosophical inquiry into the psychology of the Apollo moon scientists.* Amsterdam: Elsevier.

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., Fu, W-T., Schunn, C. D., & Trafton, J. G. (2000). From dipsy-doodles to streaming motions: Changes in representation in the analysis of visual scientific data.

Tweney, R. D., Doherty, M. E., & Mynatt, C. R. (Eds.). (1982). *On scientific thinking.* New York: Columbia University Press.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129-140.