The Instantiation and Use of Conceptual Simulations in Evaluating Hypotheses: Movies-in-the-Mind in Scientific Reasoning

Susan B. Trickett (stricket@gmu.edu)
Department of Psychology, George Mason University
Fairfax, VA 22030-4444 USA

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

Abstract

This study investigates the strategies used by expert scientists to evaluate hypotheses when they analyze data. We used an *in vivo* methodology to observe experts' on-line thinking. In contrast to the results of laboratory studies of scientific reasoning, we found that the scientists rarely used experimentation but relied on a variety of other strategies, including conceptual simulation. This strategy was most prevalent in evaluating a hypothesis about a phenomenon that violated the scientists' expectations.

Introduction

How do scientists test and evaluate hypotheses? One obvious answer is that they design and conduct experiments. The canonical method of scientific inquiry is represented by a cycle of hypothesis generation, experimentation, data analysis and hypothesis refinement that has its roots in the philosophy of science (Popper, 1956) and is frequently taught explicitly to students (Okada & Shimokido, 2001).

Psychologists investigating the processes of scientific reasoning have also been influenced by the "scientific method" and so have focused on experimentation in investigating hypothesis-evaluation strategies,. There have been numerous laboratory studies of scientific reasoning in which participants are asked to find the cause of a given effect (e.g. Dunbar, 1993; Schunn & Anderson, 1999), or to identify the role of a causal mechanism (e.g., Klahr & Dunbar, 1988; Trafton & Trickett, 2001a; Trickett, Trafton, & Raymond, 1998; Vollmeyer, Burns, & Holyoak, 1996). In these studies, participants propose hypotheses, then design and run experiments to test them

There are several reasons why participants in laboratory studies of science use experimentation to evaluate hypotheses. The instructions in these studies explicitly tell participants to run experiments. Participants have little choice—they are provided with limited time, equipment, and materials. Moreover, they are frequently asked to reason in a domain about which they have no relevant knowledge. Running an experiment is also "cheap"—the variables are already identified, it involves a few mouse-clicks, and the results are almost instantaneous and easy to interpret.

However, practicing scientists have a wider array of options. They can select their own methods and equipment, and, as experts, they have domain knowledge to guide their problem-solving. Experimentation may *not* be the best strategy, as it is expensive in terms of planning, paperwork, personnel, the need for special equipment, the complexity of data interpretation, and the high cost of errors.

What strategies besides experimentation might scientists use to evaluate hypotheses? Prior research on scientific thinking suggests several possibilities. One likely strategy is extracting information from data, whether by reading off information, transforming data, replotting data), or looking at data that is not currently on view but that is available. Trafton found that expert meteorologists spent considerable time on information extraction (Trafton et al., 2000).

Given the cost of experimentation, it is also likely that scientists use different strategies to reason about hypotheses before committing to an experiment. Analogical reasoning has been shown to be a powerful strategy in science (Clement, 1988; Dunbar, 1997; Gentner et al., 1997). It allows people to make inferences about an unknown entity based upon their knowledge of a different, known entity (Gentner, 1983) and has been proposed as a mechanism of conceptual change in numerous historic scientific advances (Gentner et al., 1997; Nersessian, 1992; Thagard, 1992). It is also a strategy used by successful contemporary scientists in scientific problem-solving, such as hypothesis generation (Clement, 1988), experimental design (Dunbar, 1997), and discovery itself (Ueda, 1997). Given its widespread use in other aspects of scientific reasoning, it seems plausible that analogy may be used as a hypothesis-testing strategy; however, whether this is the case remains an open question.

Conceptual simulation has also been shown to be a means of successful scientific reasoning (Nersessian, 1999; Qin & Simon, 1990; Schraagen, 1993). A conceptual simulation is a mentally constructed model of a phenomenon or data representation that is manipulated in such a way that there is a resulting change of state (a formal definition is provided below). As with analogy, conceptual simulations have been proposed as a strategy used by both historical and practicing scientists. In historical reconstructions, Ippolito and Tweney have developed a model of insight that involves the construction of a dynamic, "runnable" mental model (Ippolito & Tweney, 1995), and Nersessian proposes that scientists construct and conduct mental experiments that yield usable data, in a process that mirrors an empirical experiment (Nersessian, 1999). In contemporary scientific problem-solving, Hegarty has found that people develop sequences of mental animations (Hegarty, 1992). Qin and Simon (1990) found that people used a series of mental processes of manipulation, control, and inspection in order to extract information that was only implicit in their initial mental image. Similarly, participants in Schraagen's study of experimental design used a strategy of mental simulation to project what experimental procedures would look like under particular circumstances (Schraagen, 1993). As with

analogy, how much scientists use conceptual simulations in evaluating hypotheses remains an open question.

One can imagine several other means whereby scientists might evaluate a hypothesis. For example, a scientist might consult a colleague or other expert or attempt to tie the hypothesis into current theoretical understanding of the domain. A scientist might also defer evaluation until some later time or even abandon a hypothesis altogether.

The purpose of this research is to investigate the means by which scientists evaluate hypotheses. In order to investigate this issue, we adapted Dunbar's *in vivo* methodology (Dunbar, 1997), an observational technique developed to study creative and complex thinking in a real-world context. The main advantage of Dunbar's method is that it allows the collection of on-line measures of thinking by experts engaged in authentic scientific tasks.

Method

We chose to investigate scientists at work during the data analysis phase of their research because it is a stage at which a great deal of scientific reasoning takes place. Scientists must integrate their expectations about the data with the actual data; it is thus likely to be rich in hypotheses.

We analyzed 8 different datasets from 9 scientists working in one of 4 domains—neuroscience, astronomy, computational fluid dynamics (CFD), and psychology. Each dataset consists of a recorded session in which one or more scientists analyzed their data.

Participants were all working scientists recruited through personal connection of the experimenters. Either they were expert scientists who had earned their PhDs more than 6 years previously, or they were graduate students working alongside one of these experts. Only experts with a Ph.D. worked alone; in the group sessions involving graduate students, the scientist in charge always had a Ph.D.

Participants agreed to contact a member of the research team when they were ready to conduct some analysis of recently acquired data, and an experimenter visited the scientists at their regular work location. Participants working alone were trained to give talk-aloud verbal protocols. For scientists working in groups, we recorded their conversation as they engaged in scientific discussion about their data. All participants were instructed to carry out their work without explanation to the experimenter (Ericsson & Simon, 1993). It is important to emphasize that all participants were performing their usual tasks in the manner in which they typically did so. At the beginning of the session, some participants gave the experimenter an explanatory overview of the data and the questions to be resolved, and after the session, the experimenter interviewed the participants to gain clarification about any uncertainties. During the analysis session itself, however, the experimenter did not interrupt the participants.

All utterances were later transcribed and segmented according to complete thought. All segments were coded by 2 coders as on-task (data analysis) or off-task (e.g., software management, phone interruptions, jokes, etc.). Inter-rater reliability for this coding was more than 95%. Introductory comments from the scientists to the experimenter and post-session interviews of the scientists were excluded from analysis. The number and percentage of on-task utterances, the number of participating scientists, and the duration of the relevant portion of each individual session are reported in Table 1. Finally, a coding scheme (described below) was developed to examine how the scientists evaluated hypotheses they developed in the course of analyzing data.

Table 1: Characteristics of datasets

Domain	Utterances: On-Task	Total	Time (mins)	# scientists
Astronomy	649	859	49	2
CFD sub	430	954	39	1
CFD laser 1	172	400	15	1
CFD laser 2	184	249	13	1
fMRI	317	373	55	2
Neuroscience	219	343	54	2
Psychology 1	482	541	31	3
Psychology 2	914	1426	75	2

Although each scientist or group used different tools, their tasks shared several characteristics. All the scientists were analyzing data that they themselves had collected, from observations, from a controlled experiment, or from running a computational model. They displayed this data using their regular tools, whether custom-built visualization programs, while others used widely available commercial products, such as Microsoft's Excel. Figure 1 shows an example of the type of data examined by the astronomers. Visualizations used in other domains were similarly complex.

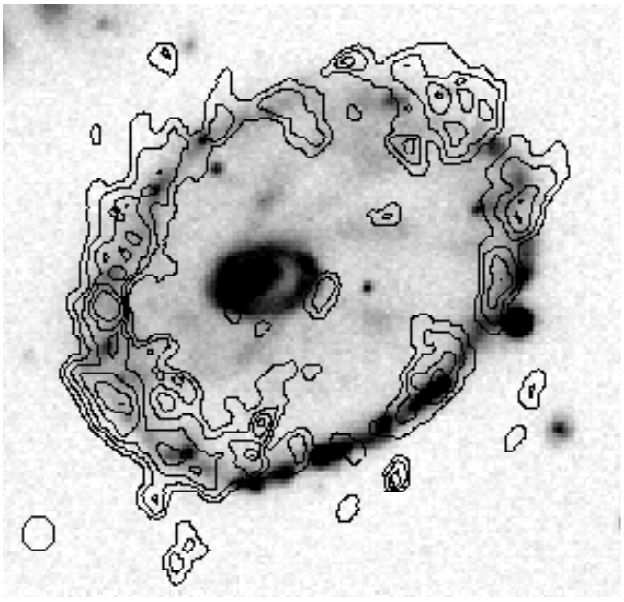

Figure 1: Example of data examined by astronomers. Radio data (contour lines) are laid over optical data.

Almost all sessions represented the initial investigation of this data (the exception was the second CFD session, which was a follow-up to the first session). Although in some sessions the scientists did not have strong *a priori* beliefs about the data (these sessions were thus exploratory), in others, the scientists did approach the task with particular hypotheses that they expected to be supported by the data. It is interesting to note, however, that none of the scientists performed any statistical analyses

Coding Scheme

In addition to coding all segments as on- or off-task, we coded the following (see Table 2 for examples):

Table 2: Examples of coding scheme
(Coded utterance in italics)

Code	Utterance
Hypothesis	You'd think [the number of reclassifications] would go up for condition C, but it didn't…*So maybe the subjects are having a better memory of the ones they've already done* (Psychology 2)
Data collection	*Do you think it's worth getting some more time, just to do an offset plane, or offset velocity?* (Astronomy)
Information extraction	*Well, that's a really clean neuron,, uh it goes down and up and away from the edges* (Neuroscience)
Consult colleague	*I'm gonna have to discuss it with ah, Robbie when he gets back.* (CFD submarine)
Tie-in with theory	*OK, so how do these Fourier modes work?* (CFD laser 1)
Analogy (general)	*Think of this as a spiral arm* (Astronomy)
Analogy (alignment)	And, if I've got a scaling problem, then it should show up here too, *but it doesn't show up here* (CFD submarine)
Conceptual simulation	*In a perfect sort of spider diagram, if you looked at the velocity contours without any sort of streaming motions, no, what I'm trying to say is, um, in the absence of streaming motions, you'd probably expect these lines here to go all the way across, you know, the ring* (Astronomy)

Hypotheses All statements that attempted to explain or account for a phenomenon identified in the data were coded as hypotheses. After a hypothesis, utterances that pertain to (elaborate) that hypothesis were identified. Such utterances constitute further investigation of the hypothesis and may be support or oppose the hypothesis. All subsequent utterances pertaining to a hypothesis were coded as follows:

Data collection Utterances in which the scientist proposed to collect more data were coded as data collection strategies. These include statements that propose an experiment, plans to run such an experiment, or plans to collect additional data for an experiment that has already been run (e.g., increasing the sample size or making some other adjustment) or to collect more observational data. Data collection strategies also include plans to build and run computational models.

Information extraction Statements that "read off" data from the visible display (i.e., extract information) were coded as information extraction (Trafton et al, in press). In addition, we coded as information extraction strategies statements that refer to looking at data in a different way (e.g., replotting the data or displaying it in a different visualization), to "tweaking" data (by transformation, removing outliers, etc.), or to looking at data that is not currently on view but that is available.

Consult a colleague Utterances that refer to showing the data to or asking the opinion of a co-worker or other expert were coded as consulting a colleague.

Tie-in with theory We expected that expert scientists with a vast array of domain knowledge stored in memory were likely to apply that theoretical domain knowledge to their hypotheses. We coded as "tie-in with theory" utterances that refer to theoretical underpinnings of the data.

Analogy/Alignment Although different theories of analogy specify different processes by which the mapping between source and target occurs (Gentner, 1983; Holyoak, 1985), all theories share these elements: source, target, and a process of mapping or alignment. During alignment, the relevant parts of the source are "applied" to the target. It is thus during this phase that inferencing occurs, and hence we expected that scientific reasoning would occur during this part of the analogical process.

We coded analogies using the definition and coding scheme developed by Dunbar (1997). According to this scheme, analogy is coded when a scientist either refers to another base of knowledge to explain a concept or uses another base of knowledge to modify a concept. Analogies were coded at both a "general" level (e.g., "The atom is like the solar system") and at the level of the actual mapping or alignment. Statements of similarity (i.e., "X is like Y") were *not* considered analogies; they do not provide explanations nor result in mapping features from the source to the target.

Conceptual Simulations Recall that a conceptual simulation is a mentally constructed model of a phenomenon or data representation. The initial representation may be grounded in memory (e.g., theoretical knowledge of the phenomenon) or in a mental modification of the displayed image. The key feature of a conceptual simulation is that it involves a simulation "run" that alters the representation, such that there is a change of state.

To code conceptual simulations, we adapted Trafton's spatial transformation framework (Trafton & Trickett, 2001b; Trafton, Trickett, & Mintz, in press;). We conducted

a spatial transformation analysis to determine for each on-task utterance whether the speaker was extracting information from the display ("read-off") and which mental operations, if any, were applied to a representation. Some possibilities include rotation, modification, moving an image, creating a mental representation, animating features, and comparison. Conceptual simulations may be defined formally as a specific sequence of spatial transformations:

1. Create representation: The scientist creates a mental representation that is not the same as the currently displayed representation. This representation creation may occur via the display (it modifies the display), via theory, (a theoretical construct); or via memory (the scientist recalls a previously viewed representation).
2. Simulation Run: The scientist builds on the created representation by spatial transformation (e.g., extend, add, delete) such that its state is changed.

Note that these codes are not mutually exclusive, and that the created representation and explicit run can occur in the same utterance. Approximately 20% of the data has been coded for conceptual simulations by 2 independent coders, and initial inter-rater reliability was greater than 90%.

Results

Eight *in vivo* datasets, comprising 330 minutes of relevant protocol and 3508 on-task utterances were analyzed. We coded 68 hypotheses, an average of approximately 1 hypothesis every 5 minutes. 57 hypotheses (84%) were elaborated; that is, the scientist made some follow-up utterance(s) that further explored the hypothesis.

How did the scientists evaluate the hypotheses?

We identified and counted the type of utterance following each hypothesis. Table 3 summarizes this count. Counts were performed in the following manner: Each individual instance of information extraction was included in the count. For example, the sequence "If I look at the average of that, it's a nice clean spike" (utterance 1) "and I can look at the standard deviation around that and it's pretty tight right in the middle where it needs to be" (utterance 2) was coded as two instances of information extraction. Each utterance identifies a different piece of information extracted (average, standard deviation). In all other cases, the count was based on the number of instances of the coded phenomenon. For example, the sequence "In a perfect sort of spider diagram" (utterance 1) "if you looked at the velocity contours without any sort of streaming motions, (utterance 2) "no, what I'm trying to say is, um, in the absence of streaming motions," (utterance 3) "you probably would expect these lines here [gestures] to go straight across, you know, the ring" (utterance 4) was coded as one conceptual simulation because each utterance contributed to, but did not constitute, one conceptual simulation.

As Table 3 shows, the most frequent strategy used for evaluating hypotheses was information extraction. This result is unsurprising, in that the scientists' task was to examine and analyze the data; one would therefore expect them to devote a significant amount of time to extracting information directly from the data itself. Similarly, the second most frequent strategy, tie-in with theory, might also be predicted from an understanding of the general procedures of science. These scientists have significant expertise and knowledge of the theories relevant to their domains, and one would expect them to consider new data in the context of current theoretical understanding of the domain. One might also expect data collection strategies (which include plans to design or conduct experiments) to occur frequently; however, these were one of the *least* frequent strategies used by these scientists.

Table 3: Frequency of hypothesis-evaluation strategies

Strategy	Frequency
Information extraction	268
Tie-in with theory	36
Conceptual simulation	34
Analogy/Alignment	30
Data collection	3
Consult a colleague	1

The use of analogy is also of interest. Of the 30 uses of the analogy/alignment strategy, only one consisted of a "general" analogy. The remaining 29 were alignments in which the mapping between source and target actually took place. This result is consistent with findings of other studies in which analogy use has been found to be more "local" than "global" (Dunbar, 1997; Saner & Schunn, 1999). The use of alignment is discussed in more detail below.

Of particular interest is the relative frequency of the conceptual simulation strategy. Specifically, this strategy was linked with the alignment strategy in a sequence that took the form of conceptual simulation followed by alignment. There were 34 conceptual simulations and 29 alignments; out of these, there were 27 Conceptual Simulation —> Alignment sequences. Thus most (79%) of the conceptual simulations were immediately followed by an alignment, and most (93%) of the alignments immediately followed a conceptual simulation.

The frequency of the Conceptual Simulation—> Alignment sequence suggests a tight coupling between the two strategies. It appears that the scientists used conceptual simulation to build a "mental model" of the data, based on assumption that the hypothesis under evaluation was true. The scientists used the data on display and their domain knowledge to investigate the implications of the hypothesis, by dynamically constructing a mental simulation of a series of processes. The result of this conceptual simulation was an inspectable mental model that was used as the source of comparison with the actual data in the alignment process. To the extent that the two models aligned, the hypothesis was supported; if there were relevant differences between the models, the hypothesis would be rejected. Figure 2 illustrates this process of model-building and alignment.

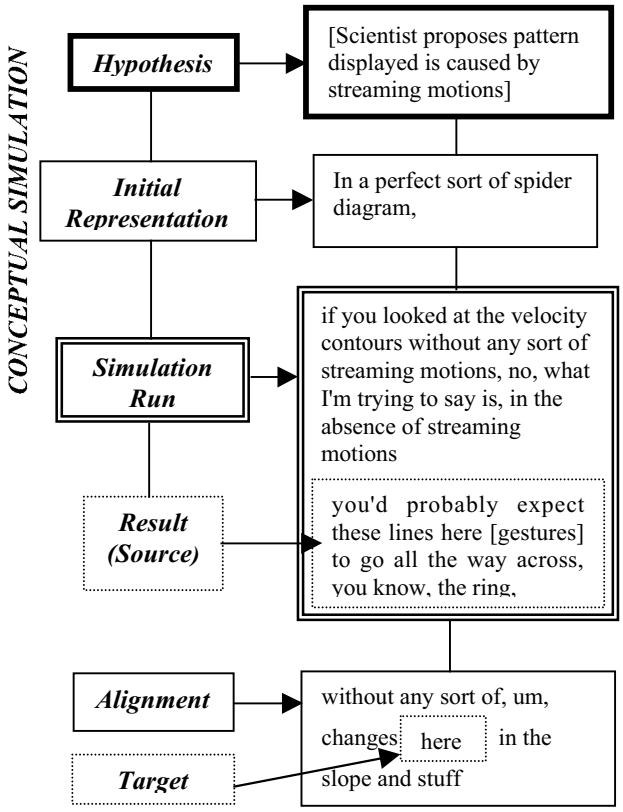

Figure 2: Conceptual simulation as source of comparison in alignment process

Why were conceptual simulations used?

There were 57 elaborated hypotheses in these datasets, and 34 conceptual simulations. The high frequency with which conceptual simulation was used as an evaluation strategy indicates that its use is important and significant. Under what circumstances did the scientists use this strategy? Conceptual simulations were used across a variety of criteria: in both group and individual settings, when the data consisted of either images or numerical tables, in exploratory and confirmatory analysis sessions, and across a variety of domains. It seems, therefore, less likely that conceptual simulations were motivated by characteristics of the data than by some characteristic of the task.

An examination of the structure of a conceptual simulation reveals that its dynamic nature allows an understanding of the *processes* involved in constructing the revised mental representation of the relevant phenomenon. Understanding process may be particularly important when there is significant uncertainty. For example, a poorly understood phenomenon is likely to evoke more investigation than one that is well understood (Trickett, Trafton, & Schunn, 2000). Thus we conjectured that the use of conceptual simulation, with its associated construction of underlying process, was associated with attempts to account for a phenomenon that violated the scientists' expectations.

In order to investigate this possibility, the hypotheses in this dataset were categorized into those that attempted to account for some expectation that wasn't met, and those that pertained to some expected phenomenon. The coding criteria for this categorization were adapted from Trickett et al., 2000. In some cases, the scientists made explicit verbal reference to the fact that something was expected or unexpected. If there was no explicit reference, domain knowledge was used to determine whether a phenomenon was expected or not. A phenomenon might be associated with (i.e., identified as similar or dissimilar to) another phenomenon that had already been established as expected or not, or the scientist might question a phenomenon, thus implying that it was not what was expected. This coding scheme was applied by two independent coders to a subset of the data (the entire astronomy protocol), and agreement between those coders was 87%. Table 4 provides examples.

Table 4: Examples of expectation-violation hypotheses (hypotheses in italics)

Domain	Utterance
CFD (submarine)	Computational model does not agree with the experiments in the least…*It could be that the turbulence is all screwed up too.*
Astronomy	That, that's odd…Why isn't there star formation going on there?…*It may be because of the large velocity dispersion.*

After we coded the hypotheses as associated with expectation violation or confirmation, we counted the use of conceptual simulation and information extraction strategies to evaluate each type of hypothesis. Note that the purpose of the analysis was to determine the circumstances under which each strategy was used, not the frequency with which the strategy followed a hypothesis; thus, only the first instance of each strategy use was counted. We performed a *phi* coefficient association measure. The correlation between hypothesis type and conceptual simulation was significant, $r_\phi = .487$, $p < .01$. There was no correlation between hypothesis type and information extraction, $r = .006$. Table 5 summarizes the results of this analysis.

Table 5: Strategy use and hypothesis type

	Violate Expectation	Confirm Expectation
Conceptual Simulation	22	3
Information Extraction	27	20

General Discussion and Conclusion

The protocol data discussed above have provided a rich dataset by which to investigate the on-line thinking of working scientists analyzing data. The scientists develop hypotheses to account for the data and then evaluate those hypotheses in light of theoretical knowledge and the data itself. In contrast to results of laboratory studies of scientific reasoning, the analyses presented above reveal that the scientists *rarely* chose to evaluate hypotheses by

experimentation (including planning experiments). They frequently used a strategy of conceptual simulation followed by alignment. In particular, they used the conceptual simulation-alignment strategy most often to evaluate a hypothesis about something that violated their expectations.

Conceptual simulation is a process of mental model-building and manipulation that results in a revised mental model, or "Qualitative Mental Model" (QMM) (Trafton et al., 2000). This QMM serves as the source of an analogy that allowed the scientists to compare the QMM with the observed data and from there to evaluate the scientist's current hypothesis. Insofar as the QMM matched the data, the scientist found evidence for the hypothesis; in the absence of a match, the scientist needed to revise the hypothesis. The alignment between source (QMM) and target (data) occurred as a series of mental processes, which amount to a recreation of the *processes* that underlie the external manifestation of the phenomenon of interest.

Acknowledgments

This research was supported in part by grants N00014-00-WX-20844 and N00014-00-WX-4002 to the 2nd author. We thank Christian D. Schunn for comments on this research.

References

Clement, J. (1988). Observed methods for generating analogies in scientific problem solving. *Cognitive Science, 12*(4), 563-586.

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 (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 461-493). Washington, DC, USA: APA

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* (2nd ed.). Cambridge, MA: MIT Press.

Gentner, D. (1983). Structure Mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

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

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: LMP, 18*(5), 1084-1102.

Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation (Vol. 19)* (pp. 59-87). New York: Academic Press.

Ippolito, M. F., & Tweney, R. D. (1995). The inception of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 433-462). Cambridge, MA, USA: MIT Press.

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

Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. Giere, (Ed.), *Cognitive models of science* (pp. 3-44). Minneapolis, MN: University of Minneapolis Press.

Nersessian, N. J. (1999). Model-based reasoning in conceptual change. In L. Magnani, N. J. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 5 - 22). New York: Kluwer Academic/Plenum Publishers.

Okada, T., & Shimokido, T. (2001). The role of hypothesis formation in a community of psychology. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), *Designing for Science: Implications from everyday, classroom, and professional settings* . Mahwah, NJ: Erlbaum.

Popper, K. R. (1956). *The logic of scientific discovery (rev. ed)*. New York: Basic Books.

Qin, Y, & Simon, H. A. (1990). Imagery and problem-solving. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society.* Pp. 646-653.

Saner, L., & Schunn, C. D. (1999). Analogies out of the blue: When history seems to retell itself. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society.*

Schraagen, J. (1993). How experts solve a novel problem in experimental design. *Cognitive Science, 17*(2), 285-309.

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science, 23*(3), 337-370.

Thagard, P. (1992). *Conceptual revolutions.* Princeton, NJ: Princeton University Press.

Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (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. & Trickett, S. B. (2001a). Note-taking for self-explanation and problem-solving. *Human-Computer Interaction, 16*(1), 1-38.

Trafton, J. G. & Trickett, S. B. (2001b). A new model of graph and visualization use. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Erlbaum

Trafton, J. G., Trickett, S. B., & Mintz, F. E. (in press). Connecting internal and external images: Spatial transformations of scientific visualizations. *Foundations of Science.*

Trickett, S. B., Trafton, J. G., & Raymond, P. D. (1998). *Exploration in the experiment space: The relationship between systematicity and performance.* In *Proceedings of the 20th Annual Meeting of the Cognitive Science Society*.

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

Ueda, K. (1997). Actual use of analogy in remarkable scientific discovery. In *Proceedings of the 19th Annual Meeting of the Cognitive Science Society.*

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*(1), 75-100.