

How Do Scientists Respond to Anomalies? Different Strategies Used in Basic and Applied Science

Susan Bell Trickett,^a J. Gregory Trafton,^a Christian D. Schunn^b

^a*Naval Research Laboratory*

^b*University of Pittsburgh*

Received 10 July 2008; received in revised form 23 March 2009; accepted 5 May 2009

Abstract

We conducted two in vivo studies to explore how scientists respond to anomalies. Based on prior research, we identify three candidate strategies: mental simulation, mental manipulation of an image, and comparison between images. In Study 1, we compared experts in basic and applied domains (physics and meteorology). We found that the basic scientists used mental simulation to resolve an anomaly, whereas applied science practitioners mentally manipulated the image. In Study 2, we compared novice and expert meteorologists. We found that unlike experts, novices used comparison to address anomalies. We discuss the nature of expertise in the two kinds of science, the relationship between the type of science and the task performed, and the relationship of the strategies investigated to scientific creativity.

Keywords: Anomalies; Creativity; Diagrammatic reasoning; Expertise; Mental models; Mental simulation; Scientific reasoning; Causal reasoning

1. Introduction

An anomaly, loosely defined, is any phenomenon that deviates from a common form, that displays inconsistency with what is expected, or that is generally considered “odd” or “peculiar” in some way. Psychologists studying scientific reasoning have placed a great deal of emphasis on anomalies. Kuhn proposed that anomalies can lead to a rethinking of current theoretical understanding within an entire field, for example, the shift from the Ptolomaic to the Copernican view of the universe (Kuhn, 1970). Kulkarni and Simon (1988) suggested that investigating an anomalous result led Hans Krebs to discover the urea cycle. Dunbar (1995) has shown that scientists attending to

Correspondence should be sent to Susan Bell Trickett, 1480 S. Columbine St., Denver, CO 80210. E-mail: sbtrickett@aol.com

anomalies in the daily grind of laboratory life are more likely to make progress than those who ignore them. Overall, anomalies can help advance science at a number of different levels.

The focus in these prior studies has been the role of anomalies in scientific discovery. However, there is a different kind of science that uses general scientific understanding of phenomena to build predictive models of particular situations in order to solve practical, rather than theoretical, problems. The goals of such applied science are different from those of basic science. In basic science, the goal is to develop, refine, and advance general theoretical scientific understanding. In contrast, in many applied sciences, such as in practicing meteorology, medicine, and nutrition, for example, the goal is to build models of specific situations to support decision making.¹

Because of the different goals of basic and applied science, anomalies may serve different functions, and consequently, may be treated differently. Whereas anomalies represent an opportunity for discovery in basic science, in applied science they may function as a nuisance to be resolved; in devising a model of a specific problem to support decision making, there is little room for ambiguity. Moreover, the general scientific theories are not called into question in applied science; only the applications of general theories to the specific situation can be fruitfully questioned. Nonetheless, anomalies in both types of science must be attended to and resolved. The fundamental question we investigate in this paper is: “Are anomalies treated differently in basic and applied science?”

Previous work has shown that basic scientists are likely to attend to anomalies rather than ignore them, suggesting that scientists are likely to enter a period of uncertainty when anomalies are encountered. It has also been shown that when basic scientists are uncertain, they use conceptual simulations more frequently than or as frequently as any other strategy (other than focusing on the data) (Trickett & Trafton, 2007). Christensen and Schunn (2009) have shown a similar use of mental simulation among design scientists (engineers) during periods of uncertainty.

It is possible that strategy differences in the handling of anomalies are more locally grounded, that is, that they are driven by the task being performed rather than the type of science being performed. The distinction we have drawn between basic and applied science is, in fact, more likely to be a continuum than a strict dichotomy. Some applied cases might involve significant new conceptual challenges that require the generation of new knowledge. For example, Pasteur’s work assisting industrialists who were trying to make alcohol from beets—an applied-science task—led to the identification of microorganisms responsible for fermentation and to an understanding of how they function—a theoretical advance (Stokes, 1997). In cases where applied science practitioners have to seek causal explanations, they might behave more like theoreticians. In general, however, we anticipate that basic scientists are more *likely* to engage in understanding deep process, whereas applied science practitioners are more *likely* to be seeking a practical solution to a problem. Consequently, we begin our investigation by observing the response to anomalies of both basic scientists and applied science practitioners.

1.1. Conceptual simulations

By conceptual simulation, we mean a common reasoning strategy by which a person imagines a situation and mentally plays out the implications in order to “see what happens.” As a famous example, Galileo imagined what would happen if two rocks of different weights fell as they were lashed together by a rope, and *mentally* determined that both would fall at the same rate.

A conceptual simulation is a form of mental simulation, or “what if” reasoning that consists of three phases. It begins with an initial representation of a system or part of a system. This representation is (mentally) modified by a series of mental operations in order to produce a simulation “run.” Two key features of this “run” are that it is (a) hypothetical, that is, does not require actual physical behaviors be enacted during the simulation or a currently present start state, and (b) that it leads to an altered representation of the phenomenon itself. This final representation, or “result” can be mentally inspected to draw inferences from it about the validity of the hypothetical conditions involved in the “run.” Table 1 shows an

Table 1

Examples of conceptual simulation (CS) from astronomy and spatial transformations (ST) (in italics) from meteorology

Utterances	Code	Explanation
Look at the little sort of, er, sort of intrusion of the velocity field here...What can it mean?		Scientist looks at image of velocity contours
<i>In a perfect sort of spider diagram</i>	CS	Scientist is <i>not</i> looking at a spider diagram. This is a reference to new representation (spider diagram)
<i>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</i>	CS continued	Reference to transforming representation (mentally removing existing streaming motions)
<i>you'd probably expect these lines here</i> [gestures] <i>to go all the way across, you know, the ring</i>	CS continued	Reference to result (sees what happens)
so that would lead me to believe, based on this pattern		Looks at upper air map
based on the location of these guys here <i>we're going to have good southwesterly flow over these parts of South Carolina</i> [points to location on map]	ST	Looks at upper air map Looks at different map; mentally adds southwesterly flow inferred from upper air map, not marked on current map
<i>more of a maritime influence here</i> [points]	ST	Mentally adds area of maritime influence (not marked on map)
<i>this is going to be high here</i> [points]	ST	Mentally adds high (not marked on map)

example of conceptual simulation (see Trickett & Trafton, 2007 for more on conceptual simulation and its somewhat complex relationship to mental models).

There are both costs and benefits to using conceptual simulations. On the cost side, it places heavy demands on working memory and requires significant domain knowledge both to play out the steps in the “run” and to draw valid inferences from the result. Because it is a qualitative reasoning strategy, its results are likely to be incomplete or imprecise (Forbus, 1997). On the other hand, as a qualitative reasoning strategy, it allows the reasoner to reason with partial knowledge, and hence to accommodate ambiguity (Forbus, 1997), which may be especially useful in situations of uncertainty (Christensen & Schunn, 2009). Furthermore, conceptual simulation provides a “quick and dirty” method of evaluating different scenarios that is cheap compared with the cost of actually constructing such alternatives (e.g., running an experiment or building a computational model).

These characteristics of conceptual simulation are likely to be particularly relevant for basic science, because they map well to its goals. Manipulating representations of phenomena meshes with the need to understand such phenomena within a theoretical framework. Inspecting the result of such manipulations can allow inferences to be made about underlying causes, by evaluating whether the conditions specified in the “run” are valid. The very act of generating the simulation requires the scientist to specify the relationship between theory and data, at least within the constraints imposed by the conditions of the simulation. Finally, the capacity of conceptual simulation to accommodate ambiguity makes it especially appropriate for scientists operating under significant uncertainty. We hypothesize, then, that basic scientists will use conceptual simulation after an anomaly.

In contrast, because conceptual simulation maps well to the investigation of underlying *causes* and the relationship between theory and data, and because these goals are for the most part not applicable to applied science, we hypothesize applied science practitioners will use fewer conceptual simulations following the detection of an anomaly. The goals of applied science are less likely to be supported by a strategy whose strength is to allow for situational ambiguity. In meteorology, for example, there may be a great deal of ambiguity in the data from which the scientist must construct a forecast, but the goal is to *resolve* that ambiguity in order to make a specific prediction about the weather. Strategies other than conceptual simulation are likely to support that goal.

We have found that, particularly for visual domains, in situations of informational uncertainty scientists often mentally transform the visualization that contains the source of their uncertainty (Trickett, Trafton, Saner, & Schunn, 2007). By mentally transforming the visualization, the scientist is able to add his or her own representation of uncertainty by mentally manipulating individual aspects of data that may be misrepresented (e.g., two discrepant displays).

1.2. Spatial transformations

A common test of spatial ability requires people to match a target figure consisting of stacked cubes with a rotated version of the target, such as the stimuli used by Shepard and Metzler (1971). In order to perform this task, one has to mentally transform the spatial

display of either the target or the candidate matching configurations. This mental rotation activity is a form of spatial transformation.

In domains that use complex visualizations, the most frequent strategy by which mental transformations happen is spatial transformations. Spatial transformations occur when a spatial object is transformed from one mental state or location to another mental state or location. They take place in a mental representation that is an analog of physical space. They can be performed purely mentally with an imagined object or “on top of” an existing visualization. Common examples of spatial transformations are creating or modifying a mental image, mental rotation (Shepard & Metzler, 1971), animating a static image (Bogacz & Trafton, 2005; Hegarty, 1992), transforming a two-dimensional into a three-dimensional image (St. John, Cowen, Smallman, & Oonk, 2001), and making comparisons between different views (Kosslyn, Sukel, & Bly, 1999; Trafton, Trickett, & Mintz, 2005). Table 1 provides examples of spatial transformations.

Because the goals of applied science are practical—to understand a specific set of circumstances and to use that understanding in problem solving—we hypothesize that applied science practitioners will resolve anomalies using spatial transformations. By mentally manipulating features of the display, these scientists can build a more complete picture of the data to use in their problem solving—for example, in the meteorological domain, a meteorologist might mentally need to redraw a weather map by placing a front in a different location.

Although they are related, in that they both involve mental manipulation of a representation, conceptual simulation and spatial transformation are distinct. Comparing the examples in Table 1 highlights this difference. The spatial transformations are discrete units. Even when they occur in sequences, they remain individual manipulations. Thus, in the example in Table 1, the meteorologist looks at an upper air map and uses that representation to make inferences about weather features (air flow, maritime influence, areas of high air pressure) at surface level. He then mentally adds those individual features to the surface level map. Conceptual simulations, on the other hand, are more complex, involving sequences of manipulations that are not only interrelated but also build on each other to create a completely new mental representation that is several steps removed from the original. Conceptual simulations not only involve this series of interrelated spatial transformations that comprise a simulation “run,” but in addition they entail both a starting representation and an inspectable ending state that reflects the changes to the original representation engendered by the simulation itself. Thus, the astronomer illustrated in Table 1 begins by looking at a representation of velocity contours overlaid on a galaxy, which display anomalous “intrusions.” He thinks that these might be caused by streaming motions. He constructs a mental representation of the theoretical appearance of the velocity contours (“a perfect spider diagram”). He mentally deletes any streaming motions from this representation (“if you looked at the velocity contours without any sort of streaming motions”) and identifies how the lines would, then, hypothetically appear (“you’d probably expect [them] to go all the way across the ring”); that is, he is able to “see what happened.” Thus, although conceptual simulations are likely to include spatial transformations, a series of spatial transformations alone does not constitute a conceptual simulation. The starting

representation, the changed final state, and, critically, the simulation run must all be present for a conceptual simulation to occur.

Whereas some spatial transformations involve mentally manipulating an aspect of the display so that a new mental image is generated, others involve only comparing two images. For example, a meteorologist might look at two maps displaying data generated by different weather models, each predicting a front in a different location, and compare the differences. We call the former type *pure spatial transformations*, in contrast to the latter, which we term *comparison spatial transformations*.

In comparison spatial transformations, the images might be two external images, two internal images, or an internal and an external image. These comparisons are considered spatial transformations because they involve mentally overlaying one image on top of another (Trafton et al., 2005). (Other types of comparisons, such as comparing numbers, do not involve spatial working memory and are not considered spatial transformations.) When encountering an anomaly, it is likely that scientists will change the display in order to view the anomaly from a different perspective and make comparisons between the different views. In addition, both conceptual simulations and pure spatial transformations result in a new mental image that invites comparison with the displayed image, in order either to make inferences or to evaluate the display. Consequently, we anticipate that both basic scientists and applied science practitioners will use comparison spatial transformations in responding to anomalies.

We conducted an in vivo study to investigate how scientists respond to anomalies. We compared expert basic scientists and applied science practitioners in order to investigate the differences between these two types of science. We hypothesized that, whereas both groups would use comparison spatial transformations, the basic scientists would use more conceptual simulations than applied science practitioners, and that the applied science practitioners would use more pure spatial transformations than basic scientists.

2. Study 1

2.1. Method

We used Dunbar's in vivo methodology for on-line observation of scientific thinking, in which participants perform their regular tasks and the experimenter observes and records their verbalizations (Dunbar, 1995, 1997). We collected concurrent verbal protocols (Ericsson & Simon, 1993). According to Ericsson and Simon, the verbal stream can offer a window onto the cognitive processes in use. In this way, we were able to obtain authentic data about how, as Dunbar puts it, "scientists really reason." We used the verbal protocols to identify anomalies and the strategies scientists used to deal with them.

We collected data in three domains, two basic sciences (astronomy and computational fluid dynamics) and one applied science (meteorology). We observed four sessions of basic science, involving three expert scientists, and five sessions of applied science, involving five expert meteorologists. All the basic scientists had earned their Ph.D. more than 6 years previously; consequently, each had at least 10 years of experience. The meteorologists were

Navy forecasters, each with over 10 years experience. Ten years of experience is a common threshold for using the term “expert” (Hayes, 1985).

The task for both groups was to carry out their normal work. For the basic scientists, the task involved analyzing radio telescope data about a galaxy, analyzing computer simulation data from a model of submarine motion, or analyzing data from an experiment involving laser pellets. For the meteorologists, it meant creating either a local or a long-term regional weather forecast.

Participants were trained to give talk-aloud verbal protocols. All sessions were videotaped. The sessions were later transcribed and segmented according to complete thought (see Trickett & Trafton, 2007; for more on the in vivo methodology). Data from the basic scientists was a subset of the data presented in Trickett and Trafton (2007).

2.1.1. Coding scheme

2.1.1.1. Inter-rater reliability: One coder coded all the protocols. A second coder, blind to the hypotheses under investigation, coded a subset of the data in order to establish inter-rater reliability, which is reported for each code below.

2.1.1.2. Anomalies: Because of differences in the two kinds of science, we coded anomalies differently for the basic and applied science data. For the basic science, we first identified instances of the scientists noticing a phenomenon of interest. We then identified which phenomena were considered anomalous by the scientists, according to the following criteria: (a) the scientist made an 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 phenomenon was anomalous;² (c) a phenomenon might be associated with (i.e., identified as similar to) another phenomenon that had already been established as anomalous; (d) a phenomenon might be contrasted with (i.e., identified as unlike) another phenomenon that had already been established as expected; (e) a scientist might question a feature of a phenomenon (see Table 2).

For the anomaly coding for the basic science, the second coder coded 10% of the data, and agreement was good (Kappa = 0.77).

Table 2
Coding of phenomena as anomalous or expected in basic science

Criterion	Code	Example
Explicit	Anomalous	What’s <i>that funky thing</i> ... <i>That’s odd</i>
Domain knowledge	Expected	You can see that <i>all the HI</i> is concentrated in the ring
Association	Anomalous	You see <i>similar kinds of intrusions</i> along here
Contrast	Expected	That’s odd...As opposed to <i>these things</i> , which are just the lower contours down here
Question	Anomalous	I still wonder why <i>we don’t see any HI up here</i> in this sort of northern ring segment?

Table 3
Coding of phenomena as anomalous or expected in meteorology (indication of anomaly in italics)

Utterance	Anomaly
The old watch had put 35 to 40 saying that it would sustain off of the coast of Greenland <i>I don't see that</i>	Discrepancy between previous data (old watch) and current data
But I guess the ETA kinda has some moisture there too, so <i>but not quite as much</i> <i>Hmm, and then the GFS has, has much less</i>	Discrepancy between models Discrepancy between models
Umm, looks like there's gonna be some precip coming through a little later in the week like couple days through like 42 hr so maybe there will be some precip in the forecast <i>unlike what I thought before</i>	Discrepancy between model and forecaster's expectation

This coding scheme was not appropriate for the meteorology domain, where the data were forecast model data, and was not questioned per se by the meteorologists. Instead, we identified discrepancies in the data, where two or more models disagreed or where a model disagreed with the meteorologist's expectations (see Table 3). These discrepancies were explicitly mentioned in the protocols and were straightforward to identify; consequently, we did not double-code for anomaly identification here.

2.1.1.3. Conceptual simulations: We coded all utterances pertaining to an anomaly for conceptual simulations and spatial transformations. We coded conceptual simulations according to the coding scheme established in Trickett and Trafton (2007). A conceptual simulation spans several utterances and consists of a specific, three-step sequence (see Table 1):

1. reference to a new representation of a system or mechanism;
2. reference to transforming that representation spatially, in a hypothetical manner;
3. reference to a result of the transformation (seeing what happens).

For the conceptual simulation coding, the second coder coded 33% of each the basic science data protocols and agreement was good, kappa = 0.75.

2.1.1.4. Spatial transformations: We coded spatial transformations according to the coding scheme established in Trafton et al. (2006), that is, any time a participant mentally transformed one spatial object from one state or location into another. Kappa for this coding was 0.79.

We further categorized the spatial transformations as either *pure* or *comparison* spatial transformations (see Table 4). *Pure* spatial transformations involve a mental manipulation of a single image, without reference to a second image. *Comparison* spatial transformations involve an explicit or implicit comparison between two images.

Table 4
Coding of spatial transformations as ‘‘comparison’’ or ‘‘pure’’

Utterance	Code
Yeah, OK, so they have precip coming in 48 hr from now	
Let me try to go back to GFS and see what they have	
Well, OK, <i>they don't differ</i>	Comparison (two model maps of precipitation compared)
They have a little bit at 54 even a little bit	
<i>and they have that storm passing further to the south</i>	Comparison (two model maps)
You also have a 12 max 14, winds are not supporting that	
<i>The next chart has it moving down further to the south</i>	Pure (adds representation of high sea area to current chart, but places it further south as second chart suggests)
Here's the low and here's the warm front see it right here	
<i>it comes around, comes around, comes around</i>	Pure (mentally adds movement to static representation)
<i>it comes around here</i>	Pure (mentally adds movement to static representation)
<i>see it dips like that</i>	Pure (mentally adds movement to static representation)
that's exactly what that thing's doing	
You can see the high	
<i>See how it's going here</i>	Pure (mentally adds movement to static representation)
<i>And the front's back in here</i>	Pure (mentally adds front to map it is not represented)

2.2. Results

We coded 1,449 on-task utterances for the basic scientists and 2,202 on-task utterances for the applied science practitioners (utterances irrelevant to data analysis were excluded). All participants found anomalies, 20 in the basic science and 25 in the applied (five per session in both domains). In the basic science, some anomalies were so closely related that the scientists referred to them together; consequently, we combined them, resulting in 17 basic science anomalies.

The 10 utterances before each anomaly were coded to explore baseline differences in conceptual simulations and spatial transformation. A one-way analysis of variance showed that there were no differences in this base-rate use of any of these strategies between the basic and applied sciences, all $F_s < 1$.

Second, for each strategy we conducted a mixed-factor ANOVA with timing (before or after the anomaly) as the within-subjects factor and group (basic or applied science) as the between-subjects factor.³ For conceptual simulation, more conceptual simulations were used after an anomaly than before it, $F(1, 7) = 25.88, p < .01$. Also, basic scientists used more conceptual simulations than applied science practitioners, $F(1, 7) = 8.43, p < .05$. Fig. 1A shows the significant interaction, $F(1, 7) = 18.53, p < .01$.

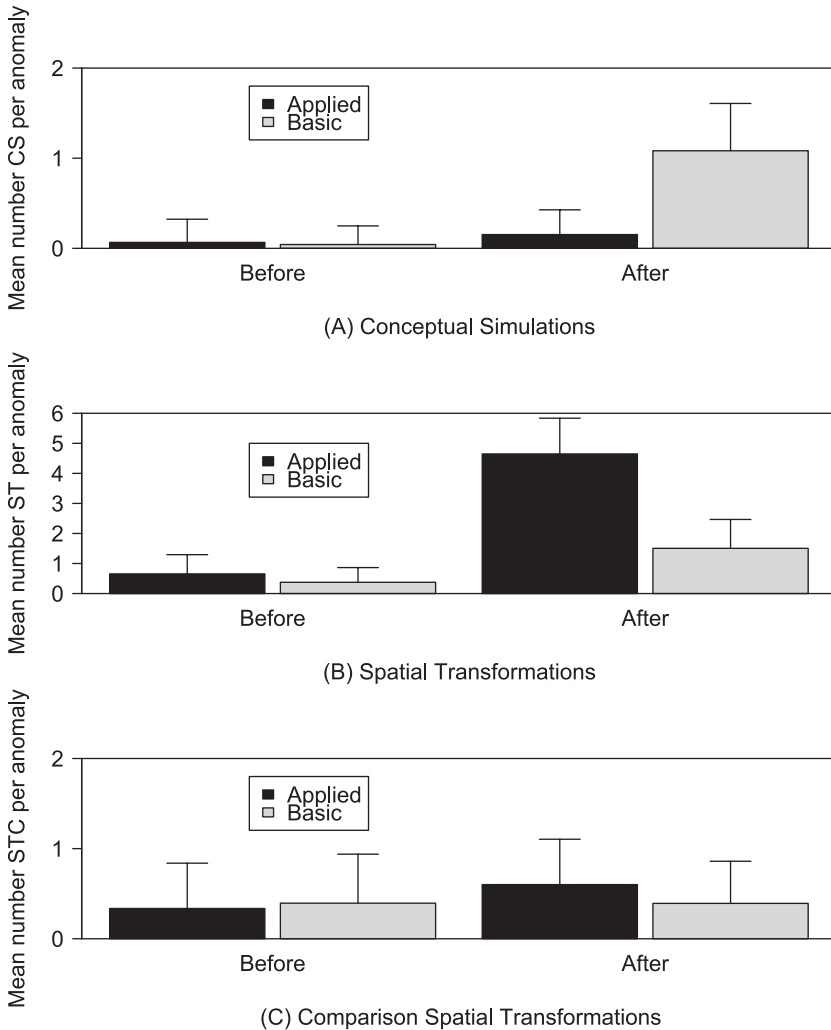


Fig. 1. Mean number (with standard error bars) of conceptual simulations, spatial transformations, and comparison spatial transformations before and after each anomaly for applied and basic science.

More pure spatial transformations were used after an anomaly than before it, $F(1, 7) = 9.82, p < .05$ (see Fig. 1B). The use of pure spatial transformations by the two types of scientist did not differ, $F(1, 8) = 2.7, p = .14$; nor was there a significant interaction, $F(1, 7) = 3.07, p = .12$.

Comparison spatial transformations did not differ in terms of timing, $F < 1$, or domain, $F < 1$ (see Fig. 1C).

Because there were two different domains in the basic science, we examined the data for each session to make sure that the pattern of results was the same for each domain. For both the astronomy and computational fluid dynamics domains, each scientist used more

conceptual simulations after an anomaly than before and the same (one session) or more (three sessions) pure spatial transformations after than before. Use of comparison spatial transformations was mixed, with more used after the anomaly in two sessions, and more used before in the other two.

Taken together, these results suggest that both conceptual simulation and pure spatial transformation are strategies scientists use to respond to anomalies, since for both these strategies there was greater use *after* the anomaly than before it. In contrast, the use of comparison spatial transformations was approximately the same before as after an anomaly, and therefore it does not appear to be especially associated with the scientists encountering an anomaly.

The results also suggest that there are procedural differences in how experts in basic and applied science deal with anomalies. Before an anomaly, both groups use conceptual simulation, pure and comparison spatial transformations equally and infrequently. However, *after* anomaly, in basic science, experts use conceptual simulation, whereas in applied science, they tend to use pure spatial transformations. Although the difference in use of pure spatial transformations by the applied science practitioners did not reach statistical significance, the applied science practitioners used three times as many pure spatial transformations following an anomaly as the basic scientists (4.65 vs. 1.5), suggesting there was a strong trend in this direction. (The lack of statistical significance is likely caused by the lack of power frequently experienced in *in vivo* research.) Not surprisingly, given the visual nature of the data in this study, both basic scientists and applied science practitioners make equal, albeit sparse, use of comparison spatial transformations (e.g., by comparing different visualizations of the anomaly).

How did the scientists use these different strategies to help them resolve the uncertainty fostered by the anomaly? Table 5 shows an example of a conceptual simulation following an anomaly in one of the basic science protocols. The scientist had built and run a computational model of the flow of fluid around a submarine and was comparing the model's output with experimental data. The scientist was quite confident that there would be a good match between model and data, but to his dismay the match was "not even close." He was baffled

Table 5
Example of conceptual simulation (CS) used to resolve anomaly in basic science (CS in italics)

Utterances	Coding	Explanation
<i>It is conceivably possible that this curve is floating around all over the place, and what they're showing is an average</i> (scientist is looking at a graphical representation [a curve] that represents the turbulence)	CS	Reference to new representation (this curve)
<i>so if this thing is really floating around that much, just up and down, and I'm at the extreme end, and if I average all of this stuff,</i>	CS continued	Reference to transforming representation
<i>then I may actually still get the curve right</i>	CS continued	Reference to result (sees what happens)

as to the cause of the discrepancy, declaring on several occasions, “I have no idea.” He proposed the hypothesis that the problem lay with the experimental data: “It is conceivably possible that this curve (a flow curve represented graphically on the visualization) is floating around all over the place, OK, and what they’re showing is an average.” He then used conceptual simulation to generate the implications of this hypothesis for the discrepancy: “So if this thing is really floating around that much, just up and down, and I’m at the extreme end, and if I average all of this stuff (computational data), then I may actually still get the curve right.” Apparently, he constructed a mental representation of the hypothetically averaged computational data and compared it with the experimental data, because he concluded: “But I don’t think that’s right. I just don’t see it, right off hand”—even if he performed the necessary averaging operations, the model would still not match the experimental data.

In another, similar example from a different basic science protocol, the scientists were jointly trying to understand an anomalous “blob” that had been puzzling them for some time in the displayed image. (See Table 6 for the details of this example.) They had considered (and rejected) several hypotheses, when one scientist recalled an earlier model he had run. In a complicated sequence of steps, he reconstructed relevant features of that model and mapped them to the current data concerning the blob, reinterpreting these data in the light of the model data. He imagined that the puzzling data might be “a completely different sort of kinematic population” and mentally redrew the image with two groups of stars “bending” in different directions. He then inspected the result of this transformation and found a separation similar to what he had observed in his model data. This match between previously viewed model data and the mentally transformed current data led him to conclude that he may have resolved the mystery surrounding the anomalous blob.

These examples are typical of the way in which conceptual simulation functioned for the basic scientists. It allowed them to mentally “play out” in detail the implications of some

Table 6

Example of conceptual simulation (CS) used to resolve anomaly in basic science (CS in italics)

Utterances Scientist 1	Coding	Explanation
OK, OK, one of the things that show up in at least the preliminary models that I did run are this thing sort of breaks apart and this thing sort of goes...		
so you have a separation of the ring into a, an outer arm and another arm		
<i>so this could be actually be a completely different sort of kinematic population</i>	CS	Reference to new representation (a completely different sort of kinematic population)
<i>This could actually, this, these stars could be bending inward</i>	CS contd	Reference to transforming representation
<i>While these stars are bending outward</i>	CS contd	Reference to transforming representation
<i>So you actually have a separation of the two like that</i>	CS contd	Reference to result
That’s where the blob could really be coming from		Conclusion regarding anomaly

possibility, assuming that it was true, and thus to determine the outcome of that mental simulation. This process allowed the scientists to draw inferences about the data in relation to the possible explanation they were considering.

A comparable example of the applied science practitioners using spatial transformations to resolve anomalies is illustrated in Table 7. The meteorologist notices that the model data shows a temperature increase that he does not believe is accurate, creating a discrepancy that must be resolved. He performs a series of mental adjustments to the forecast map, thus adding his own representation to the map that he thinks is incomplete. These adjustments are not hypothetical, except insofar as they are not literally drawn on the image, nor is there any simulation involved. Instead, he adds missing information to the data representation. From this mentally redrawn map, the meteorologist is able to “read off” the information that, with these weather features in place, the temperature will be lower than the current map suggests, thus resolving the anomaly and justifying his decision to disagree with the model’s prediction.

Thus, in this study, we found differences between the two groups of scientists. However, our results do not fundamentally address the source of these differences—whether they are due to the type of science or the type of task undertaken. There are at least two ways to think about how the strategy might be affected by the task. First, it is possible that discrepancies between hypothesis and data are more likely to be resolved by conceptual simulation, whereas discrepancies between two sets of data are more likely to be resolved by spatial transformation. In this case, basic scientists with two conflicting datasets would be expected to use spatial transformation, and applied science practitioners attempting to resolve a theoretical discrepancy would be more likely to use conceptual simulation. Unfortunately, there are not enough instances in our dataset of this type of interaction between domain and task to test this hypothesis.

A second avenue of interest is to consider the role of expertise in the use of strategy. If the strategy is directed by the task—or even the domain—it is possible that any problem

Table 7

Example of applied science practitioner using spatial transformations (ST) (in italics) to resolve anomaly

Utterance	Code	Explanation
They really want to drive some warm air in there I just can’t buy that	Anomaly	Discrepancy between model data and forecaster’s expectation
What did I do for the 5th? 82 I’m gonna stay with 82 there even though the thickness now shows it’s in here <i>the front is back in here somewhere</i> <i>you’ve got warm moist air</i> <i>you’ve got the high over here that’s off Bermuda</i> <i>and you got this one in here...</i> ...so the temperature, the max temperature’s going to be pushed down	ST ST ST ST	Mentally adds front to map (not represented) Mentally adds weather feature (not represented) Mentally adds high pressure (not represented) Mentally adds weather feature (not represented) Resolves anomaly: justifies forecasting lower temperature than model predicts

solver would approach the task in this manner. However, it seems likely that there are particularly useful ways of solving the task that perhaps require higher levels of skill and knowledge in order to be implemented. In order to test this hypothesis, we conducted a second study to examine expert/novice differences in one of the domains. The purpose of this second study is to provide clarification about the use of the various strategies in the applied domain.

3. Study 2

In order to investigate strategy differences between experts and novices, it is important that both groups be asked to perform equivalent tasks; consequently, we restricted this study to the meteorology domain. In this domain novices can be asked to perform the exact same task as experts (create a weather forecast). At the same time, the task remains challenging for experts.

Although both experts and novices are capable of performing the task, we predicted that there would be differences in the strategies by which they did so. Specifically regarding the two types of spatial transformation, we predicted that novices would use fewer pure spatial transformations following an anomaly than the experts, because we hypothesized that this strategy requires skill and knowledge to implement (Sims & Mayer, 2002). Comparison spatial transformations also require some domain knowledge, because the scientist must be able to discern which are the relevant points of comparison to be made. It is likely that novices will have sufficient domain knowledge to focus their comparisons on relevant data and thus to use comparisons effectively. Consequently, we predicted that the novices would use more comparison spatial transformations than pure spatial transformations to handle an anomaly. Since the expert meteorologists in Study 1 used conceptual simulation only rarely, we did not expect to find significant use of this strategy by the novices.

Participants in this study were 10 undergraduate juniors or senior meteorology students with 1 to 2 years experience making forecasts. In addition to coursework, all the students regularly created forecasts for the campus weather site and participated in team-based national forecasting competitions. We compared the new novice data with the data from the expert meteorologists in Study 1.

The novices were asked to create a 3-day weather forecast. They performed the task in the university weather center, using the tools that they regularly used. They were trained to talk aloud while performing the task, and each session was videotaped, transcribed, and segmented, as in Study 1. The verbal protocols were coded for anomalies (discrepancies), conceptual simulations, and pure and comparison spatial transformations, as described earlier.

4. Results

The novices' transcripts comprised a total of 2,340 utterances. They noticed a total of 23 anomalies, with each novice noticing at least one anomaly.

As in Study 1, we established that there were no differences in the base-rate use of the strategies of interest by experts and novices. We compared the use of conceptual simulation and both types of spatial transformation during the 10 utterances before each anomaly. The novices used no conceptual simulations at all, compared with a very slight use by the experts (a mean of 0.067 per anomaly), $F(1, 13) = 2.16, p = .17$. Before an anomaly, novices and experts did not differ in their use of either pure spatial transformations ($F < 1$) or comparison spatial transformations, $F(1, 13) = 2.88, p = .11$.

For conceptual simulation, pure spatial transformation, and comparison spatial transformation we conducted a mixed-factor ANOVA with timing as the within-subjects factor and expertise as the between-subjects factor. For conceptual simulation, the effect of timing was marginal, $F(1, 13) = 3.89, p = .07$. Experts used more conceptual simulations than novices, $F(1, 13) = 6.48, p < .05$. The interaction was marginal, $F(1, 13) = 3.89, p = .07$. Fig. 2A shows these results. In fact the novices used no conceptual simulations at all, either before or after the anomaly, and the experts' use was quite low. The very small number of conceptual simulations overall in the applied science domain, even by experts, suggests that this is not a particularly relevant strategy in this domain. The complete lack of conceptual simulations among the novices suggests that when it is used, it is an expert strategy for which novices lack the skill or knowledge.

For pure spatial transformations more pure spatial transformations were used after the anomaly than before it, $F(1, 13) = 17.35, p < .01$. Experts used more pure spatial transformations than novices, $F(1, 13) = 10.54, p < .01$. Finally, the interaction between timing and expertise was significant, $F(1, 13) = 17.99, p < .01$. As Fig. 2B suggests, the experts used many times more pure spatial transformations than the novices *after* the anomaly, whereas there was no difference in their rate of use before the anomaly. Fig. 2B also illustrates how there was very little difference in the novices' use of pure spatial transformations after an anomaly (mean = 0.48) compared with before an anomaly (mean = 0.52). Overall, the results suggest that pure spatial transformation is an important *expert* strategy for handling anomalies in this domain.

For comparison spatial transformations, marginally more comparison spatial transformations were used after an anomaly than before it, $F(1, 13) = 4.06, p = .06$. The effect of expertise was not significant, $F < 1$, and the interaction was not significant, $F(1, 13) = 2.22, p = .16$. Although these results were not statistically significant, as Fig. 2C shows, the novices used many more comparison spatial transformations after an anomaly (mean = 1.8) than before (mean = 0.03), suggesting at least a trend on their part to use comparison spatial transformations when encountering an anomaly. The experts, in contrast, used a similar number of comparison spatial transformations before and after an anomaly.

Overall, these results suggest expert applied science practitioners use pure spatial transformations but not conceptual simulation when they encounter an anomaly. Novices, however, used many times more comparison spatial transformations after an anomaly. Coupled with the relative lack of pure spatial transformations by novices, this result suggests that the novices were sensitive to anomalies but focused on further identifying features of the anomaly by comparing different representations of the data, rather than on actually resolving the anomaly. Possibly the novices lacked the knowledge and skill to determine which, if either,

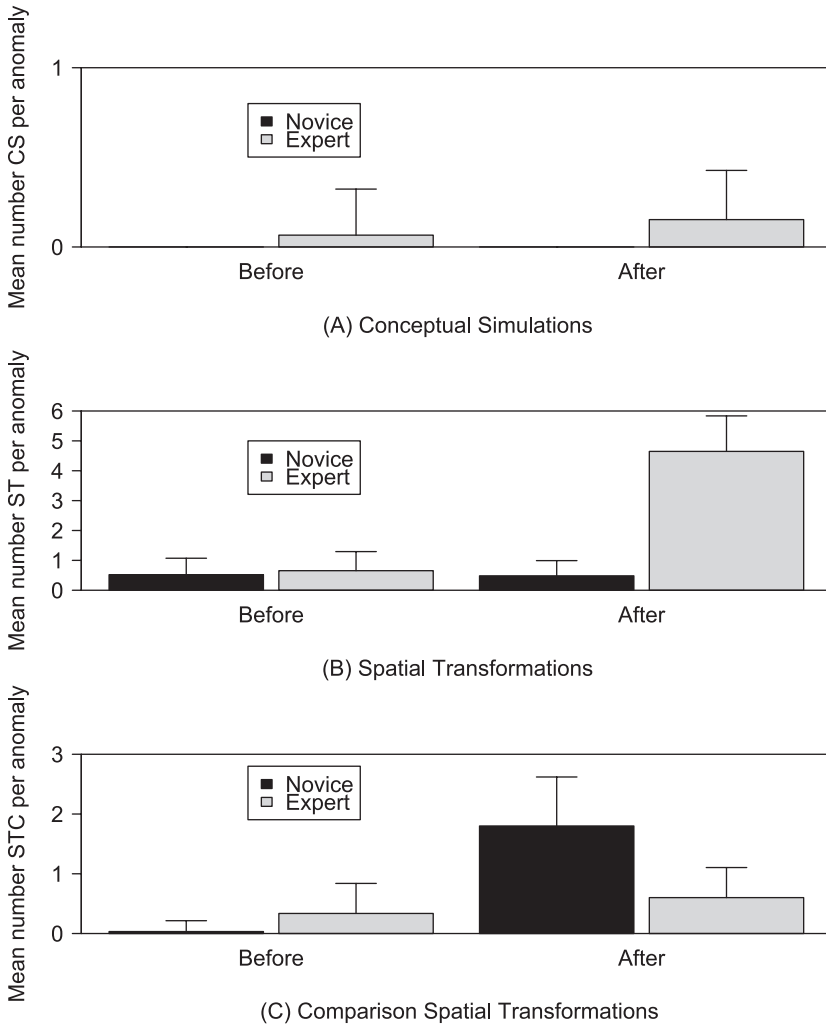


Fig. 2. Mean number (with standard error bars) of conceptual simulations, spatial transformations, and comparison spatial transformations before and after each anomaly for expert and novice applied science practitioners.

of the discrepant models was likely to be more accurate and thus got “stuck” on identifying aspects of the discrepancy itself.

5. General discussion

We have identified three distinct, albeit related, problem-solving strategies used in scientific reasoning, particularly among scientists using complex visual displays of data: conceptual simulation, pure spatial transformation, and comparison spatial transformation, and we

have explored their role in addressing anomalies among different types of scientist with different levels of expertise.

The results of these two studies show differences in the way anomalies were handled by both basic scientists and applied science practitioners. The expert basic scientists used conceptual simulations, whereas the expert applied science practitioners used pure spatial transformations. However, it is not clear whether the different strategies were due to the type of science or the task undertaken. The results of Study 2 clarify some aspects of this difference: Novices in applied science seem to lack the expertise to perform pure spatial transformations and instead use comparison spatial transformations. The tasks examined in Study 1 were not such that any problem solver would have responded in the ways that were observed in the experts, but rather some level of expertise was required to produce the observed pattern (at least for the applied science task).

Another way to interpret this difference between expert and novice behavior is that novices focus on the data representations themselves, whereas experts move beyond the data. Both the basic scientists and the applied science practitioners mentally manipulated the visual image. The applied science practitioners did this by means of spatial transformations, mentally redrawing a visualization to make it a more accurate representation. The basic scientists used conceptual simulation, constructing whole scenarios to allow them to quickly test hypotheses they developed about the anomaly.

There are several possible reasons for these strategy differences. One reason, as discussed above, may be the different goals of the different tasks. Experts by definition use the appropriate strategy for their task; consequently, we conclude that conceptual simulation, while useful in basic science, is less applicable to the demands of applied science.

Another possibility is that the type of anomaly likely to be encountered is different in the two domains. In the meteorology domain, at least, the fundamental phenomena of the domain are generally well understood, and anomalies are unlikely to be of the type that would threaten that understanding. In basic science, practitioners are trying to develop or further theories; consequently, anomalies may concern phenomena that do not easily fit into a current theoretical understanding. As a first pass, conceptual simulation might enable the scientist to resolve some of these anomalies and thus “weed out” the more trivial from the more significant.

A third possibility is that the immediate task determines the strategy. In some instances in our own data, as well as in the archives of science, applied science practitioners use conceptual simulations—for example, Tesla (Hegarty, 2004), Orville and Wilbur Wright (Johnson-Laird, 2006), as well as contemporary engineers (Christensen & Schunn, 2009). Similarly, the basic scientists in our study also used spatial transformations after an anomaly, without necessarily constructing a conceptual simulation. It is possible that when the task is to resolve anomalies concerning surface-level phenomena, such as discrepancies in data, scientists of either persuasion use spatial transformations, but that when the task is to resolve deep-seated incompatibilities between theory and data, scientists use conceptual simulations. It is further likely that domain and task are interrelated, in that theoreticians are more *likely* focused on deep process and consequently more apt to attend to those deep-seated discrepancies and applied science

practitioners, and especially the meteorologists in our study, are more likely to encounter discrepancies between pieces of data.

A final possibility concerns the creative nature of basic science. Christensen and Schunn (2009) argue that conceptual simulation is commonly used by creative engineers to turn uncertainty into approximation. In meteorology and other applied science, most problem solving is relatively routine. Often the expert forecaster is able to rely on prior experience, as he or she recognizes weather patterns previously seen and stored in memory. In resolving discrepancies between models, the forecaster can use the different models' known strengths and weaknesses to evaluate their likely accuracy. Spatial transformations are used to adjust these individual models to fit, based on a global model.

Anomalies represent an important aspect of the scientific process, because of the opportunity for problem solving that they present. Whereas most studies of scientific reasoning consider science as a single-stranded enterprise, in this article, we have teased apart two approaches to doing science, the basic and the applied. We have validated this distinction by identifying separate strategies by which practitioners in each domain dealt with anomalies in the domain. We have proposed that, when faced with the uncertainty posed by anomalies, applied science practitioners responded to a specific situation in order to resolve a specific problem, and did so by using pure spatial transformations. This strategy, though by no means undemanding or effortless (as shown by its infrequent use by novices in the domain) nonetheless makes fewer demands on the scientist's imagination, because it requires only a discrete mental adjustment of an external visual representation. Basic scientists, on the other hand, in response to the more creative demands of the domain, engaged in the more creative strategy of conceptual simulation, which requires not only multiple, interdependent mental adjustments of an external visual representation but also the highly creative step of drawing inferences from the effects of those manipulations. We suggest that conceptual simulation may thus be a hallmark of creative problem solving.

Notes

1. Of course, in research on such areas (i.e., meteorological research, medical research, nutrition science research), the goals are generally basic science goals.
2. The coders' domain knowledge came from textbooks and interviews with the scientists.
3. Although there were issues in the data with lack of homogeneity of variance, a log transform shows the results are in the same pattern, with the same significance. Consequently, we report the raw data, with comparable significance values.

Acknowledgments

Work on this project was supported by Office of Naval Research grant N0001406WX20091 to Greg Trafton and Office of Naval Research grant N000140610053

to Chris Schunn. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Navy. In addition, we would like to thank Mike Gorman, Phil Johnson-Laird, and Jeff Shrager for their helpful comments.

References

- Bogacz, S., & Trafton, J. G. (2005). Using dynamic and static displays: Using images to reason dynamically. *Cognitive Systems Research*, 6(4), 312–319.
- Christensen, B. T., & Schunn, C. D. (2009). The role and impact of mental simulation in design. *Applied Cognitive Psychology*, 23(3), 327–344.
- Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 365–395). Cambridge, MA: MIT Press.
- 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: American Psychological Association.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Cambridge, MA: MIT Press.
- Forbus, K. (1997). Qualitative reasoning. In A. B. Tucker (Ed.), *The computer science and engineering handbook*. (pp. 715–733) Boca Raton, FL: CRC Press Online.
- Hayes, J. R. 1985. Three problems in teaching general skills. In S. Chipman, J. W. Segal & R. Glaser (Eds.), *Thinking and learning skills, Vol. 2* (pp. 391–406). Hillsdale, NJ : Erlbaum.
- Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18(5), 1084–1102.
- Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Science*, 8(6), 280–285.
- Johnson-Laird, P. N. (2006). *How we reason*. Oxford, England: Oxford University Press.
- Kosslyn, S. M., Sukel, K. E., & Bly, B. M. (1999). Squinting with the mind’s eye: Effects of stimulus resolution on imaginal and perceptual comparisons. *Memory and Cognition*, 27(2), 276–287.
- Kuhn, T. S (1970). *The structure of scientific resolutions* (2nd ed). Chicago: University of Chicago Press.
- Kulkarni, D., & Simon, H. A. (1988). The process of scientific discovery: The Strategy of experimentation. *Cognitive Science*, 12, 139–176.
- Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703.
- Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology*, 16(1), 97–115.
- St. John, M., Cowen, M. B., Smallman, H. S., & Oonk, H. M. (2001). The use of 2D and 3D displays for shape understanding vs. relative position tasks. *Human Factors*, 43(1), 79–98.
- Stokes, D. E. (1997). *Pasteur’s quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.
- Trafton, J. G., Trickett, S. B., & Mintz, F. E. (2005). Connecting internal and external representations: Spatial transformations of scientific visualizations. *Foundations of Science*, 10, 89–106.
- Trafton, J. G., Trickett, S. B., Stitzlein, C. A., Saner, L., Schunn, C. D., & Kirschenbaum, S. S. (2006). The relationship between spatial transformations and iconic gestures. *Spatial Cognition and Computation*, 6(1), 1–29.
- Trickett, S. B., & Trafton, J. G. (2007). “What if...” The use of conceptual simulations in scientific reasoning. *Cognitive Science*, 31(5), 843–875.
- Trickett, S. B., Trafton, J. G., Saner, L., & Schunn, C. D. (2007). “I don’t know what’s going on there”: The use of spatial transformations to deal with and resolve uncertainty in complex visualizations. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 65–85). Mahwah, NJ: Erlbaum.