The Effects of Supportive and Anomalous Information on the Evaluation of Explanations

A Thesis

Presented to the Faculty of the Graduate School of Cornell University

in Partial Fulfillment of the Requirements

for the Degree of Master of Arts

by

Andrew Jefferson

January 2012

ABSTRACT

Extensive work has been done looking at how people reason causally and respond to information anomalous to their beliefs. However, little work has looked at how supportive and anomalous information interact when people evaluate the plausibility of a possible explanation. In this study, participants were given an event and either one or two possible explanations. Participants evaluated the plausibility of the explanation(s) repeatedly after exposure to combinations of supportive/non-supportive information and a strong/weak anomaly, producing a 2 (number of explanations) x 2 (presence of support) x 2 (strength of anomaly) design. The results showed that support initially increased plausibility and generally decreased the effect of anomalies. This was qualified by a 3-way interaction, which may be driven by how alternative explanations are evaluated relative to each other. The implications for future work, to understand how available information affects causal reasoning in everyday and professional contexts, are discussed.

BIOGRAPHICAL SKETCH

Andrew Jefferson is currently a fourth year PhD student in Cornell University's Department of human Development. He earned his B.S. in Neuroscience in 2007 from the College of William and Mary. He grew up in Lynchburg, VA and now resides in Ithaca, NY. His current research looks at the role of anomalies in causal reasoning as well as the educational potential of video games as a medium.

To my family:

my parents

and both Jessicas

ACKNOWLEDGEMENTS

This work would not have been possible without the work of many people. First, I'd like to thank Barbara Koslowski for her patience and guidance. I'd like to thank Barbara and Steve Ceci for contributing their knowledge, experience, time, and energy.

A number of other people contributed directly or indirectly. I'd like to thank Francoise Vermeylen for her statistical expertise; Alexandra Hildreth for her time on the project; the members of the Koslowski Lab for their insight and advice, including Wanda Casillas, Briana Robustelli, Jessica Sue-Wern P'ng, Lindsay Jessica Freid, and Camille Waldron; and Nadia Chenyak, Daphna Ram, and Charlotte Sweeney for their personal support and experience.

I'd also like to thank my family, my parents, sister, and loving wife. Without you supporting me, believing in me, and preparing me for challenges (personal and professional), none of this would have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

"Scientific reasoning" calls to mind professionals in lab coats thinking carefully in their laboratories, but refers to everyday situations as well: a mechanic determining the problem with a car or an employer judging if an employee is late due to laziness or traffic would be examples. In fact, anytime conclusions are being produced based on data, or one explanation is being compared to another, the process of scientific reasoning is taking place. The study of scientific reasoning then gives us insight into both the processes we use to investigate the world and the conclusions people reach everyday, based on the information available to them.

To date, most studies of scientific reasoning have been concerned with people's causal reasoning, or how people determine what causes something.

Covariational and Mechanistic Causal Information

One particularly influential approach to scientific causal reasoning is based on the work of David Hume, who characterized the evaluation of causality as based on "Humean indices" of priority, contiguity, and covariation(Koslowski, 1996; Proctor & Capaldi, 2006; Wilson & Keil, 1998). That is, the cause of an event must precede the event (priority), be present whenever the event occurs (contiguity), and changes in the cause must be associated with changes in the event (covariation). The use of such indices is quite appealing from a research perspective, both by being based on a number of concrete dimensions data and for appearing to be "theory-independent", or theoretically applicable to any given domain (Koslowski, 1996).

In studies that look at the use of Humean indices, covariational information is often presented in the form of statistical rates (Koerber, Sodian, Thoemer, Nett, 2005; Kuhn, D., Amsel, & O'Loughlin, 1988), usually either ratios or percentages of the time the event and possible cause would covary. For example, in one study children generated a theory about which factors were causal and which were not, and then were given covariational information that either supported (confirmed) or undermined (disconfirmed) their theory. For example, one event was that children were getting sick and one factor the participants were given was whether the children ate chocolate cake or carrot cake (Kuhn, D., 1989). If they said this factor (the type of cake) *was not* causal and were in the "disconfirm" condition, they would be shown pictures of children who had eaten chocolate cake getting sick and children who at carrot cake being healthy, visually depicting 100% correlation. Similarly, if the participants said the factor *was* causal and were in the disconfirm condition, they would be shown pictures of children who ate carrot cake getting sick and who ate chocolate cake getting sick, as well as both kinds of children being healthy, visually depicting 0% correlation between the factor and the outcome.

However, there are several drawbacks to the Humean indices approach. For example, much of the work that looks at reasoning based on Humean indices focuses on covariational information and then counts only responses based on this information as 'correct' (Inhelder & Piaget, 1958; Kuhn, D. , 1989). This means that responses based on a child's own theory, are not scored as good reasoning. Even if the child proposed mechanism for how the event is or isn't linked to the factor, or cited previous experience that contradicts the presented information, if the child doesn't base their answer on the presented information, it is not scored as proper reasoning. Some (Koslowski, 1996) have suggested this is an overly narrow view of scientific reasoning and the kinds of reasoning that are disregarded have an important place in scientific and everyday thinking. For example, some participants might offer reasons like "carrots are good for you, so carrot cake should be healthier." When presented information that there is no covariation, if the participant disregards it, even if offering other factors that could have explained the results or saying it needed to be repeated, this is interpreted as persisting in a flawed theory in the face of contradictory evidence, which the researcher subsequently treats as being unscientific. However, this kind of appeal to prior experience on the part of participants suggests that the evidence presented isn't strong enough to overcome this foundation, likely formed by personal experiences and information from parents, teachers and other 'experts'. “Neo-humeans” and Bayesians even specifically address this concern by including prior experience in their models (Young, 1995). Studies of successful laboratories have also found that when experimental results are unexpected, the response is generally to repeat the experiment, assuming the result was due to a procedural error or other mistake, (Dunbar, 2000) rather than initially accepting the result and seeking a new explanation. It is only after repetition that the result is seen as valid and addressed, perhaps similarly to the children assuming other factors are behind the covariational results that are not consistent with their beliefs and theories.

Additionally, others have highlighted the important role of non-covariational information, such as information about possible mechanisms, or *how* a factor causes a given event (Ahn & Kalish, 1995; Koslowski 1996; Young, 1995; Tenenbaum & Griffiths, 2009). This mechanistic information may provide information about the world (which will be revisited later) or may specifically address possible causal chains or processes that lead to the event of interest. In fact, the ubiquity of covariations in the world makes it necessary to filter which covariations are even considered plausible candidates for causing the event in question (Quine, 1969). Mechanistic information, which conveys not how often something happens but how it can happen, has been proposed to play a role in filtering candidates and interpreting which covariations are actually meaningful.

One of Koslowski’s experiments (1996) illustrates the interaction of mechanistic and covariational information. In the study, some participants were told that the only difference between two cars with different gas mileage was the color that the cars were painted. These participants were less likely to believe that color was causal than they were to believe a more plausible factor, such as how new the factory was that had made the cars. Even though the covariation evidence for those factors was the same, participants were less likely to see the paint color as causal because they had no mechanism for how the paint color might affect the gas mileage. However, when participants were supplied with such a mechanism (color might affect driver behavior, such as red making them more alert and better driver, and this behavioral change would affect mileage) they were then more likely to rate color as causal. This example highlights the filtering role of mechanistic information; when participants lacked a plausible mechanism, they tended to disregard the covariational information while those with a mechanism were more likely to be influenced by the same covariational information.

Another reason offered for the importance of mechanistic information is that covariational evidence is not always available (Koslowski, 1996). If an event is historical, or a unique occurrence, there won't be enough information to form a covariational account, but mechanistic information about the factors that caused the event may be available. In everyday instances as well, some findings suggest that people may be more comfortable dealing with mechanistic information than rates of covariation. This can be seen in rationales participants provide, which typically reference mechanistic arguments even when asked about covariational information (Koslowski, 1996; Kuhn, D., 1989) and particularly in studies in which participants are asked to convince others that a certain explanation is correct. In such studies participants almost exclusively suggested mechanistic arguments rather than statistical appeals (Ahn & Kalish, 1995; Brem & Rips, 2000; Kuhn, D., 1989). While mechanistic reasoning may or may not be a sound or appropriate form or reasoning, these findings demonstrate that it is a kind of reasoning that people regularly use; in order to understand how people reason about problems, one must have some understanding of mechanistic reasoning.

Anomalous Information

No explanation is perfect. There will be information that is problematic for the explanation, and this type of information is called an anomaly. It can be an unexpected finding or a piece of information that is difficult to explain within one's current framework. An anomaly does not necessarily contradict the explanation, though it can; rather, anomalies are just information that does not easily fit into the current explanation. T.S. Kuhn (1962), noted the central role of anomalies in bringing about a paradigm shift; because no theory is perfect, anomalies to that theory will emerge over time. It is through the process of exploring these inconsistencies and refining one's theory until the findings that were anomalies are expected results that a paradigm shift occurs.

It is important to note that anomalies are very clearly defined if one only considers things purely from a covariational perspective: An anomaly is either a failure of the possible cause and event to covary when the covariation is expected, or a clear covariation when the covariation is not expected. The examples of undermining covariational information discussed previously (i.e. children getting sick or not after eating cake) would be examples of such anomalies.

However, Koslowski has suggested that defining what information is or isn't anomalous is much more difficult if one takes a broader perspective that considers mechanistic as well as covariational information. Since mechanisms are based on beliefs about how the world works, types of information other than a failure to covary can be problematic. For instance, the previously discussed covariational anomalies are still problematic from this broader perspective (ie children getting sick after eating cake), but now information that relates to beliefs about the mechanism (what makes one sick, what makes one healthy) could be anomalous as well. It is often more difficult to evaluate how these other types of information affect the plausibility of a theory than it is to analyze the impact of an observed covariation. In this perspective information may even be ambiguous, and whether the information is anomalous depends on how it is integrated with one's already present beliefs and information. This makes identifying anomalies much less straightforward in the broader perspective than in one that only considers covariational information.

One strategy in dealing with these complexities is referred to as Inference to the Best

Explanation (IBE). A more in depth treatment of this approach has been done by Lipton (2004) but to summarize: The general approach is to evaluate which of the plausible available explanations best fits the available information, is consistent with well established background beliefs, and best fits a number of explanatory dimensions. These dimensions include scope (how broad a category of phenomena the explanation applies to), precision (how accurately it reflects the information), and simplicity (how many assumptions must be made). Lipton argues it is not about which explanation is the most likely, but which is "loveliest" in terms of balancing the concerns of these dimensions and fitting the evidence. Some have argued that this model better fits actual scientific process and everyday causal reasoning (Lipton, 2004; Magnani, 2001; Proctor & Capaldi, 2006; Thagard, 1989) than other models, such as hypothetico-deductive reasoning, or Popperian falsification.

One important implication of this model is that an explanation is evaluated on how well it fits the information available, which is not limited to information in a particular domain. In the classic example from the history of science, one reason natural selection was so compelling was not just that it explained differentiation arguably as well as other theories (such as divine intervention), but that it explained differentiation in a way that was consistent with cell theory, and tectonic shifts, etc. Another reason was its predictive power, as it could make predictions about how a population that is isolated in new conditions for long periods of time will diverge from its parent population.

This example also demonstrates the importance of context in evaluating explanations: in IBE, the explanation is evaluated *with respect to the available alternative explanations*, not in a vacuum or in an absolute sense. This key aspect of reasoning will be revisited later.

As an elaborated framework of how to evaluate a theory, IBE also provides a way to evaluate which information is problematic and how that information is problematic. That is, since IBE details a way to evaluate a theory it also allows us describe how that evaluation changes in response to new information. If a theory would be "less lovely" when a piece of information is available than when that piece is not, such a change in the evaluation can be used as a measure of how anomalous that piece of information is to that theory. This makes IBE useful for the previously mentioned difficulty of defining what information is problematic for theories when one considers mechanistic information in addition to covariational information. IBE provides a system and framework to evaluate how problematic a piece of information is in the broader perspective and exactly why it is problematic.

Past Work on Anomalous Information

Turning from abstract reasoning to more concrete questions, previous work has also asked “how do people respond when presented with anomalies?” Chinn and Brewer (1993; 1998) have generated and attempted to empirically test a framework of the various ways in which people respond to anomalies, including shifting from a held explanation to an alternative theory. To summarize, the authors detailed eight possible responses to anomalies, six of which may be generally grouped into dismissing the information or making it less problematic, and while the remaining two are making peripheral changes to the theory and actually changing which theory they believe. These results may indicate something about how resistant most participants are to changing their evaluations of a theory in response to an anomaly.

Chinn and Brewer's work provides a picture of possible response to anomalies, but was not designed to look at what factors lead to individual differences in these responses. Their studies presented participants with theories and anomalies to those theories, then looked for common structures in the responses. This was done to develop (and later test the completeness of) their taxonomy of responses, but neglected to more thoroughly examine their anomalies. If there is meaningful variation in the anomalies, this could be an important factor in how people will respond to it, and research exploring possible distinctions between anomalies has found some interesting results. Koslowski, Libby, O’Connor, Rush, and Golub, (in prep) have looked at what makes some anomalies “stronger” or more compelling than other anomalies; for example, if you have the theory “penicillin kills germs” then experimental results showing “penicillin makes germs grow” will be a stronger anomaly than “penicillin only kills some germs”(Koslowski, 1996). What little work has been done on the topic (Koslowski et al., in prep) has shown that people can generate approximately the same number of reconciliations for both strong and weak anomalies, finding some way to account for the problematic information while maintaining a held theory. When participants are asked to rate the plausibility of resolutions they have generated, however, they tend to rate the reconciliations for strong anomalies as much less plausible than the reconciliations they have generated for weak anomalies. One interpretation of these findings is that there is always a way to account for an anomaly, after the fact, but strong anomalies are hard to account for in a plausible way; in other words, you can always say aliens did it, but this is rarely satisfying. This inability to resolve the anomaly in a satisfying, plausible way may be part of why strong anomalies are more problematic.

Although important, this work did not explicitly examine the role of background information in dealing with anomalies, though; The amount and content of previous experience should play an important role in whether an anomaly is seen as a trivial oddity or as a significant problem for a theory (Koslowski, 1996; Kuhn, D., 1989). If all one’s life balls fall down when dropped and suddenly one falls up, the expected response is not a diminished belief in gravity, but an inspection of the ball or an assertion that some other factors -- magnets or fishing line perhaps -- were at play. This is an extreme example of how past experience can influence response to an anomaly, but it serves to demonstrate how large a role such knowledge can play in dealing with new, anomalous information.

One way experience may have such an effect is because it acts not just on the explanation of interest but also on a number related beliefs. That is, the theory being undermined is part of what Quine and Ullian (1978) termed a network of related beliefs that support each other and make it more difficult to convince people of another explanation. Swiderek (1999) explored this by taking a large, central belief of the participants and presenting evidence that disconfirmed subsidiary beliefs that related to the central belief. In this case, the central belief was one's opinion on capital punishment and the subsidiary beliefs were factors such as how effective a deterrent it is, the economic impact, and how likely it is that there are wrongful convictions. She found that the sum of the change in the subsidiary beliefs predicted 85% of the variance of the change in the central belief about capital punishment in subsequent interviews, demonstrating the impact these related beliefs can have on each other.

Some feel that the large role of background knowledge in responding to new information can be a detriment to reasoning. For example, 'confirmation bias' or a tendency to discard or misinterpret information that does not support one's held theory, can frequently lead to errors in scientific reasoning (Nickerson, 1998). While such a bias can certainly be seen as flawed reasoning in some instances, such as, for example, the persistence in the belief the earth is flat, in other cases it is a reasonable response. To revisit the previous example of a ball falling up, in this case, looking for an alternate explanation that preserves the theory of gravity, while it may be confirmation bias, can be seen as good reasoning. The reluctance to accept information that contradicts beliefs that are based on numerous past experiences, are consistent with other knowledge of the world, and are well integrated with what else one knows about the world, is not poor reasoning. Studies of productive labs (Dunbar, 2000) have also found this kind of skepticism is quite prevalent; as already noted, researchers tend to replicate unexpected findings repeatedly to eliminate other possible causes before considering how the new finding could impact their theories.

The Current Investigation

While some work has investigated the effects of background information and anomalous information on theory change, little work has directly attempted to manipulate both factors to examine the interaction. Is it simply that background support makes people less likely to change their view in response to both strong and weak anomalies? Does such background support change the impact anomalies make differently based on the strength of the anomalies? Does support change anomalies effects in different ways if other explanations are present than if they are not?

To address these questions, the current experiment was designed to manipulate the supportive information available, the strength of the anomaly, and the presence of an alternative explanation. Background knowledge is difficult to manipulate, and work using such manipulations has typically employed quasi-experimental designs using experts and novices in a given field (Hmelo-Silver & Pfeffer, 2004 ; Schauble, Glaser, Raghavan, & Reiner, 1991), with the assumption that experts bring more background knowledge and relevant previous experiences to the experimental situation than novices. However, the current investigation attempts to manipulate the information available to participants directly: some participants were given background information that supported an explanation and then given an anomaly while others are simply given the same anomaly without the preceding background information. While the effect of supporting information may be weaker than the effect of the more elaborate background knowledge of an expert, this approach was selected because it allowed us to directly manipulate the presence of supportive information. Such an approach also controls for the amount of support, as each participant can receive the exact same information (as opposed to the varying experiences of experts). A two-step process, presenting the participants with the supportive information then presenting them with the anomaly, was used; this way the participants had the support and then encountered the anomaly, as one with background experience would.

METHODS

Participants

Participants were 65 Cornell University undergraduates (32 male, 33 female) between the ages of 18 and 23 (M = 19.95, SD= 1.11) recruited through a university website. They were interviewed during the Fall 2009 and Spring 2010 terms and compensated for their time.

Materials

Stories: Six stories were used which each consisted of an event to be explained, two possible explanations for the event, supportive information, and two anomalies. The content of these stories varied from explaining biological phenomena (fish dying in an area, the changing height of populations) to explaining cultural practices (differences in ancient burial customs, rising age of marriage, religious conversion) or events that could be either (differences in the success rates of surgery). The full stories are included in Appendix A.

The stories were selected to be in generally familiar domains, but employed largely unfamiliar specifics; this was to allow the participants enough familiarity to intelligently reason about the explanation, but little relevant experience, such that the presented information would likely be novel for all the participants.

Design

A 2 (alternative present vs. absent) x 2 (supportive information presence vs. absence) x 2 (anomaly strength: strong vs. weak) design was used. It is important to note that anomaly strength was a within-subjects variable, while alternative and supportive information presence were between-subjects variables. That is all six of the stories that a given participant received either had alternatives (alternative present condition) or else did not (alternative absent condition) because presenting an alternative in one story might have primed the participant to expect or to generate their own alternatives in subsequent stories that did not explicitly present them with an alternative. This would have made the results more difficult to interpret. There was a similar concern regarding the presence of supportive information so it was made a between subjects variable. However, since all participants received anomalies, this was not a concern for that variable; it did not matter if previous stories prompted them to think there would be anomalies, because there always were anomalies. Participants were thus randomly assigned to one of 4 conditions (alternative present/absent x support present/absent), and received all six stories in random order. Half the stories received by each subject had strong anomalies and half had weak anomalies.

Anomaly strength was made a within subjects variable to address concerns about expectation and bias. Pilot studies raised concerns that if participants received all strong anomalies, they would anticipate problematic information coming and either reduce initial ratings or not fully read anomalies, and assume they were very problematic. This concern was addressed by giving participants both strong and weak anomalies in random order; participants would have to think about the implications of each piece of information, rather than sensing a pattern.

The additional information (both supportive and anomalous) was selected for its relationship to the target explanation. When the additional information is referred to as supportive or anomalous, this refers to it in relation to the target explanation not the alternative explanation or to both explanations. As much as possible, the additional information was intended to be neutral toward the alternative explanation. This allowed us to better compare results across the alternative present and alternative absent groups (since the "target explanation" is the explanation that was present in both conditions). Information could be considered anomalous to the target if it supported an alternative explanation (as discussed in Chinn and Brewer, 2001), but this creates issues of comparison when there is no alternative both because it is not there to be supported by this and because the information might suggest the alternative to the participant even if they would not otherwise have generated it. Therefore, for this task the anomalous information, as much as possible, only applied to the target and was neutral to the alternative.

The supportive information provided support by either presenting a possible mechanism for the theory to lead to the event (as in stories 1 & 5) or by providing information that would be expected if the theory was true, but did not provide conclusive evidence (stories 2, 3, 4, 6) and the support used is available in Appendix A. These kinds of information were used in an attempt to introduce related beliefs that could support the target explanation without making the explanation too strong to be affected by the anomalies.

Procedure

After signing the consent form, the participants were presented (shown and read by the experimenter) an event (e.g. coastal fish dying), as well as an explanation for the event (e.g. invasive species are killing them; referred to as the target explanation). They were then asked for both an initial (Time 1) rating of the plausibility of the target explanation and a rationale of why they chose that rating. Their ratings of plausibility were on a 1 (“not at all convincing”) to 11 (“completely convincing”) point Likert scale (Appendix B). Half the participants were also presented with a second possible explanation (referred to as the alternative) for the event as well as the target. The alternative was also rated each time the target was rated.

After the initial rating(s), half the participants were given additional information that supported the target explanation (such as: Invasive species are dangerous because native species have never encountered them before. They may produce substances that are very harmful to creatures). The others participants were told that researchers had not yet had a chance to search for additional information. All participants were then asked to rerate the explanation(s) (Time 2 ratings).

Finally, the participants were all given information that was problematic for (anomalous to) the target explanation and asked to rerate the explanation(s) (Time 3). However the strength of the anomalies was varied, so that some information was more problematic for the target than others (which anomalies were strong and weak was determined by pilot testing or previous studies). Half of the participants in each previous group (those who received support and those who did not) received "strong anomalies" which were very problematic for the target, while the other half received "weak anomalies" which were still problematic, but easier to reconcile with the explanation or resolve than the strong anomalies.

This procedure was then repeated for the other five stories.

An example of a full protocol has been included in Appendix B.

RESULTS AND DISCUSSION

For reasons previously discussed, the following analyses were done on the participants' ratings of the target explanation or on the changes in ratings of the target explanation between one time point and the next. Each of the analyses used a mixed linear model run in a PSAW statistical package. The effects of the presence of alternative, the presence of support, the strength of the anomaly, the story type (types 1-6), and the gender of the participant were included in the model as fixed effects. The subject was included as a random effect. Each of the final models only included interactions that were significant. In the first analysis of differences scores (Time 1 to Time 2), the rating at Time 1 was included as an additional fixed effect; in the second analysis (Time 2 to Time 3) rating at Time 2 was included instead.

Overview

A general overview of the participants ratings over time is given in Figure 1. This figure displays the average ratings when given the initial explanation(s) (Time 1), after being given the supportive information or not (Time 2), and after receiving the anomaly (Time 3). In this overview, there are a few clear results which will be discussed in more detail subsequently: For participants who received support (represented by the 4 solid lines) ratings of plausibility of the target explanation went up (from Time 1 to Time 2), while for those that did not receive support (represented by the 4 dotted lines) remained approximately the same during the same period. Additionally, we can see that the ratings decreased in all groups after the participants were presented with an anomaly (Time 3). These results show the basic manipulations were effective, but the more interesting part of the results are the variation in the slopes between Time 2 and Time 3, as this suggests there are interactions, which will be explored in more depth later.

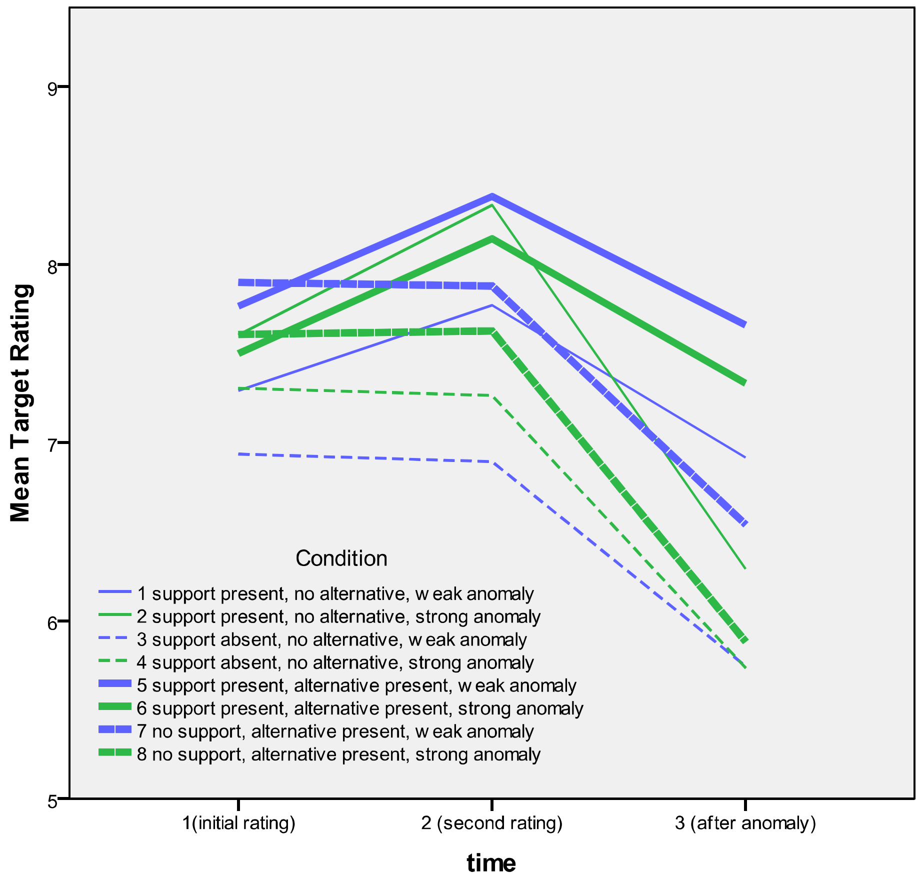

Figure 1; Overview of participants' ratings of the target explanation across conditions and time

Initial Ratings

At Time 1, initial ratings of the target explanation showed no significant differences among any conditions. The only significant effect was of story content (F(4,254=11.699, $P < .001$), with stories 1 and 4 (involving medical care and marriage age respectively) having slightly higher initial ratings of the target explanation than other stories. Story type, however, did not interact with any of the other variables at any of the timepoints and the effect of story type was included in subsequent models, so it could be partialled out.

Effect of Supportive Information

Difference scores between the ratings at Time 1 and Time 2 were used as the response measure in order to look at how ratings of the target changed when support was presented between the ratings at Time 1 and Time 2. The main effects of support presence ($F(1,352)=42.238$, $P<.001$) and the initial target rating at Time 1 ($F(1,370)=12.550$, $P<.001$) were qualified by an interaction of the two ($F(1, 372)=14.042$, $P < .001$). The estimated fixed effect of this interaction showed a negative relationship between presence of support and the initial target rating at Time 1. That is,

those who initially rated the target as more plausible at Time 1 responded less to the supportive information (ie did not increase their ratings as much at Time 2). This result may have been due to a ceiling effect.

One possible explanation for this interaction is that participants who initially rated the target as more plausible did so based on supportive information that was available to them based on prior experience, so the presented supportive information had either already been considered in the initial rating or was redundant with the information that had been considered. In contrast those who initially rated the target as less plausible may have had less background knowledge that was supportive of the target explanation or relevant to the story. Therefore the supportive information may have been particularly novel for these participants. This account is consistent with the some of the rationales that were collected; participants who initially rated the target more plausible responded to the support with “that's what I just said” and similarly, participants who had initially rated the target less plausible responded to the support with “I hadn't thought of that”.

Effect of Anomaly Strength

Difference scores between the ratings at Time 2 and Time 3 were then used to examine how the ratings changed when the anomalies were presented. There was a 3-way interaction between the presence of an alternative explanation, the presence of supportive information, and the strength of the anomaly ($F(1,312)=4.604$, $P < .05$). The difference scores for the different conditions are shown in Figure 2 (in contrast to the mean scores of Figure 1). To better understand the 3-way interaction, the conditions were broken down for additional analyses. Looking only at the subjects who did not receive support (the first four bars in Figure 2, from the left to right), the only significant main effect was the anomaly strength ($F(1, 122)=6.640$, $P <.05$); when there was no supportive information, the presence of an alternative explanation did not significantly affect ratings. When supportive information was present, however, (the other four bars in Figure 2), there was a significant interaction of alternative presence and anomaly strength ($F(1, 125)=8.302$, $P<.005$) such that the strong anomaly had a larger effect when no alternative explanation present.

The 3-way interaction also qualified a main effect of anomaly strength ($F(1, 208)=4.497$, $P<.05$). The presence of an alternative explanation and the presence of support did not have significant main effects in the full, final model. The rating at Time 2 was included in the model, and had a main effect ($F(10, 214)=3.185$, $P<.001$) but no significant interactions.

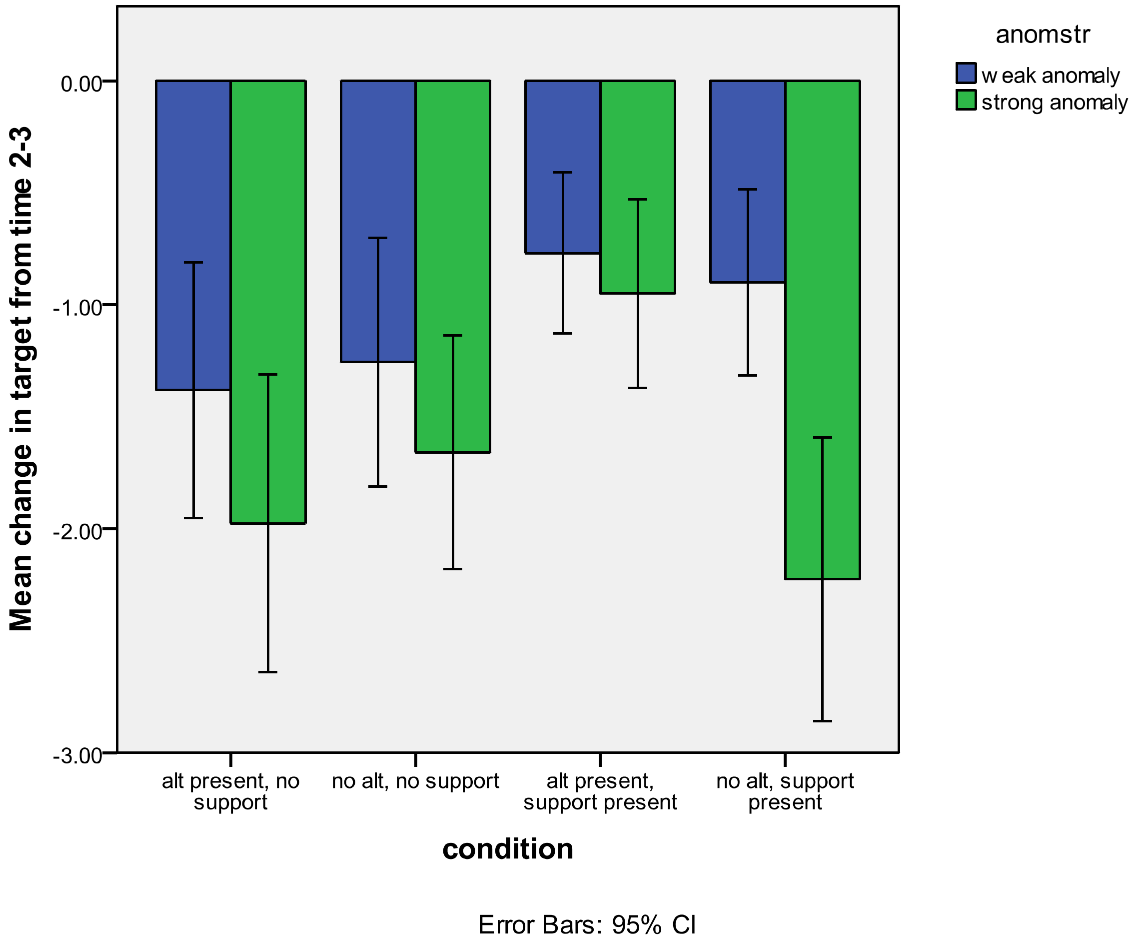

Figure 2; Mean differences in ratings from Time 2-3 for all conditions

To summarize, when there was no support, strong anomalies decreased target plausibility more than weak anomalies. The presence of an alternative explanation had no effect on this decrease. Thus it appears that when there was no supportive information reinforcing the target's plausibility, anomalies made the target less plausible regardless of the presence of an alternative. It is not surprising that when there was no alternative present the anomaly was seen as a weakness or flaw in the target. When there was an alternative, but no additional support to distinguish the target from the alternative, the anomaly was still a flaw in the target and had the same effect.

When support for the target was present, the effect of the anomaly depended on both its strength and whether an alternative was present. Weak anomalies reduced the plausibility of the target less when supportive information was present than when it was absent ($t(170)=-2.201$, $P<.05$). It appears that supportive information made the target more resistant to weak anomalies. The effect of support on strong anomalies interacts with the presence of an alternative however, and this may be due to a framing effect.

When support for the target was present and there was an alternative available, the effect of the strong anomaly was reduced to the point it did not significantly differ from the effect of the weak anomaly when support was present ($t(75.4)=.659$, $P=n.s.$), yet when there was support present but no alternative the strong anomaly had a substantial effect on the target's plausibility. It may be that the reason for this discrepancy is that having an alternative influenced *how* the participants rated the target's plausibility. For example, when there was an alternative present it was clear that there were other possible explanations in addition to the target. Although the anomaly was a problem for the target and not the alternative, participants may have thought that there were probably anomalies to the alternative that were simply not available to them yet. They did not know if there were also anomalies to the presented alternative, if there were other alternatives, or if those explanations were better or worse than the currently presented explanations. If participants had this mindset, the strong anomaly may not have reduced the plausibility of the target as much because while it was a problem for the target, it did not reduce the plausibility of the target much *compared to the plausibility of alternative explanations.* In contrast, when there was support available but no alternative, the target appears to be the only viable explanation; the supportive information reinforces this by making the target seem even more plausible. The strong anomaly is still problematic for the target however, and *on an approximation of an absolute scale* makes the target much less plausible.

It may be that the presence of an alternative altered how the participants evaluated the plausibility of the target rather than affecting the response to the anomaly. The presence of the alternative cued participants to rate the target's plausibility relative to alternatives, while the absence of an alternative left them to rate the target on more of an “absolute scale of plausibility”. Thus participants given an alternative changed their rating little in response to the strong anomaly, because all explanations have problems. However, participants without an alternative changed their ratings more in response because the anomaly was in fact very problematic for the target.

Change from initial to final rating

Difference scores from Time 1 to Time 3 were also analyzed, to look at the total change in ratings. These results are shown in Figure 3, which is in the same format as Figure 2. In fact, since there was no significant change from Time 1 to Time 2 when no support was presented, as discussed previously, the only difference between Figures 2 and 3 should be in the conditions where support was presented.

Statistically, there were strong main effects of anomaly strength ($F(1, 304)=8.689$, $P<.005$) and the presence of support ($F(1, 184)=12.811$, $P<.001$), but not of alternative presence, and there were no significant interactions between them. There was a significant interaction between the presence of support and the initial rating at Time 1 ($F(9, 340)=2.308$, $P<.05$), but this was already discussed under “*Effect of supportive information”*. It seems odd that the 3-way interaction from Time 2-3 does not appear here, but that may be due to the combination of higher variability in responses for this analysis withthe large number of conditions relative to *n*.

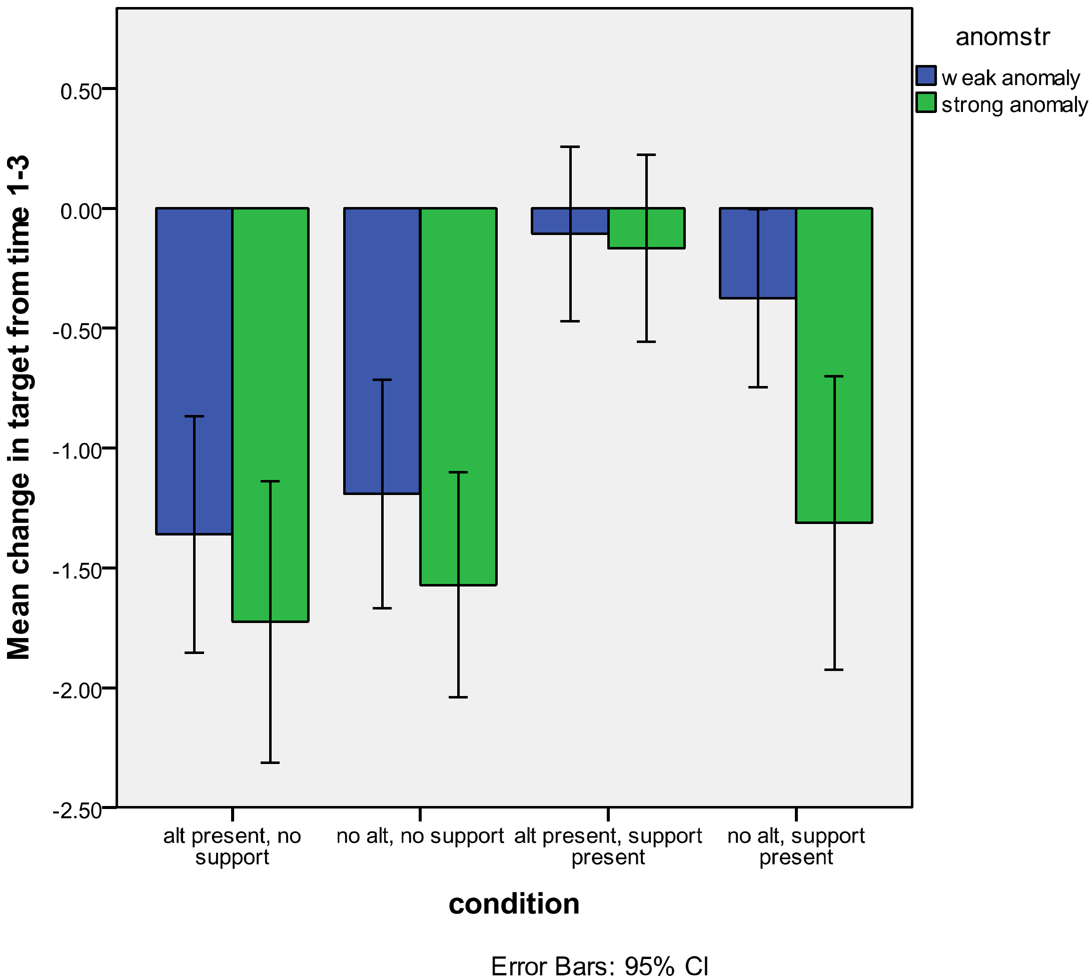

Figure 3; Mean differences in ratings from Time 1-3 for all conditions

However, considering these results as the sum of the component changes from Time 1 to Time 2 and from Time 2 to Time 3 provides a much clearer picture of the interaction of supportive and anomalous information. Supportive information first bolsters the plausibility of the target (the increase from Time 1-2) and then reduces the effect of the anomaly (from Time 2-3), and the sum of these effects almost completely cancels the effect of the anomaly when an alternative was present. It appears that when alternative explanations are present, as is very often the case, supportive information could be said to inoculate one against the effects of anomalous information. However, there are some cases, such as when there are no available alternatives and the anomaly is particularly strong, that an anomaly can still overcome these effects and reduce an explanation's plausibility.

GENERAL DISCUSSION

There are other possible explanations for the 3-way interaction between Times 2 and 3:

One possible explanation is its simply a response to the complexity of the stimuli. The participants who received support and an alternative received the most information to keep track of out of all the participants. Though they did not have to memorize the information (they could freely refer back to all the information presented), simply keeping track of how it related to the presented explanations and to their background knowledge created greater task demands. When they were then presented with a weak anomaly, they considered it problematic, but its effect was attenuated by the supportive information. The strong anomaly would have been more difficult to reconcile with what they had been previously given, however. It may be that rather than dealing with this difficulty they held the anomaly in abeyance. With the demands on them, it could be that they just registered that it was a problem, which could explain why the strong anomaly's effect was equivalent to a weak anomaly.

Reviewing the rationales provides some support for this type of explanation. The rationales do show deep processing but it is combined with a confusion about how to integrate the anomaly into the participant's current understanding. The participants have been relating the support and their own background knowledge to the two presented explanations, and the new information from the strong anomaly is not easily integrated; they know it is relevant, but are unsure how to make sense of it, and some participants explicitly question its validity. Knowing the anomaly is a problem, but unsure how to respond, they simply note that it is a problem, treat it as a weak anomaly, and maintain their current understanding as best they can. It is not that they don't know its a problem; it is that they don't know what to do about it.

Another possible explanation is that when an alternative was present, participants saw the target explanation as part of a larger compound explanation, which is less affected by the strong anomaly. Koslowski (2010) has suggested that having a non-mutually exclusive alternative present may cue participants that multiple factors could be causing the event, and the target explanation is only one of them. While the strong anomaly may be problematic for an explanation with the target as the sole causal factor, some participants responded with rationalizations in the form of "its not as important, but still a factor" indicating that they believed the target explanation was still part of a larger explanation but played a less central role in causing the event. If they see the strong anomaly as only undermining part of this larger explanation, it may explain why the strong anomaly has less impact and is treated as weak anomaly when an alternative is present.

Future Directions

One question related to these findings is whether only new information affects ratings of explanation plausibility or if activating background knowledge explicitly could cause participants to change from an initial rating. That is, does only new supportive information inoculate one against anomalies, does previously learned information have as strong an effect as novel information, and does reminding subjects of previously learned information cause it to have more of an effect. This would likely vary based on how deeply the participant originally considered the explanation; some participants would try to apply all their relevant background knowledge initially and thus have little reason to change their ratings after such a manipulation. Perhaps more interesting then would be explicitly asking them to generate support for an explanation before being presented with an anomaly and measuring if this attenuates the effect of anomalies in the same way as presenting new supportive information.

Another questions is what other conditions are relevant to determining when anomalies overcome supportive information. What factors other than the availability of plausible alternative explanations and the strength of the anomaly might play a role? Do these factors differ between novices that have a less developed understanding of the support than experts with more entrenched and interconnected concepts? As previously discussed, work with experts finds them more resistant to anomalies, and the effect of support discussed could play a role in explaining that. However, “resistant” means that change still occurs sometimes; closely examining when anomalies have an effect and if the factors relevant to an anomaly's impact on expert reasoning have a similar impact on everyday reasoning could be quite informative.

In addition, the rationales given by participants suggest more work needs to be done on when and how people reason using compound explanations composed of multiple causal agents to explain an event. Some work on compound explanations has been done in looking at the differences between experts and novices (Schauble, Glaser, Raghavan, & Reiner, 1991; Hmelo-Silver & Pfeffer, 2004) or in more computational environments (Krems & Johnson, 1995) but little has been done on their role in everyday reasoning. Looking at when non-experts use compound explanations, how they use them, and how these more complex explanations respond to supportive and anomalous information could give us an even more complete picture of how people reason about events in the real world.

Conclusion

Comparing explanations is nearly ubiquitous and this work shows that not only is it affected by many factors, but that sometimes the interactions of basic factors can be counter-intuitive. How people evaluate explanations is not only affected by what information is available and what alternatives are available, but the alternatives can significantly change the impact of that information. Continued work to expand our understanding of how the information available to people affects their beliefs and their responses to subsequent information is needed to better understand this complex process. Such work has great potential impact, as a better understanding of how people choose between possible explanations allows us to correct for possible biases: If doctors generally attend one type of information over another, compensating for this kind of bias could be critical. In schools, knowing how new information interacts with previous experience could allow teachers to more effectively convey how or why things work to students. This kind of insight is particularly useful in science education, where current methods take great effort to convey fundamental concepts. Finally, evaluating explanations is at the core of scientific inquiry; a better understanding of how people reason should help researchers reason better about their own findings.

APPENDICES

APPENDIX A: Story Content

Story 1:

Event:

When doing complicated and innovative surgeries (such as transplants, etc.), some hospitals clearly have a better success rate than others, even if the patients and procedures are comparable.

Possible Explanation:

(One) possible explanation is that successful hospitals have attracted surgeons and surgical nurses who are especially <u>skilled and well educated</u>.

Alternative:

(Another) possible explanation is that in successful hospitals each surgeon works with a particular group of surgical nurses, so that surgeon and nurses learn to work <u>together as a standard team</u>.

Support:

Surgeons and nurses with better skills and education have a better understanding of procedures and are more capable of anticipating possible complications to the success of an operation.

Anomalies:

Weak:

The success rates of many hospitals remain poor even after most of their surgeons have received training and practice in complicated surgical procedures.

Strong:

Some of the hospitals with the worst success rates have poor success rates even when they recruit especially skilled surgeons from other high-performing hospitals.

Story 2

Event:

Over the past fifty years or so, there has been a documented decrease in the average height of Americans, while Europeans have, on average, gotten taller. This difference is found even if one excludes recent immigrants.

Possible explanation:

One hypothesis is that the increasing popularity of "junk food" in the United States has led to a general decline in <u>nutrition</u> (reduced vitamins and micronutrients). This, in turn, has caused people not to grow to their full height potential.

Alternative:

One hypothesis is that there has come to be a growing percentage of people in the U.S. that receives little or no <u>health care</u>. People with no health care are more susceptible to various infections, which they transmit even to people who do have health care. This reduces the growth potential of everyone.

Support:

The past fifty years have also seen an increase in the number of snack and candy vending machines in the United States.

Anomalies:

Weak:

The height differences occur even though a larger percentage of people in the U.S. than in Europe take vitamins to compensate for nutritional deficiencies.

Strong:

The decrease in average height in the U.S. is found even when the population is divided into individual groups based on income, education, race, or ethnic origin. That is, the decrease in average height is found in each individual group, even in the groups that one would expect eat very little junk food.

Story 3

Event:

Prehistoric sites in Eurasia have found the graves of several women from a nomadic, herding tribe. One group of women was buried with swords, daggers, and bows and arrows, while another was buried with jewelry and luxury items. Why was one group buried with weapons?

Possible explanation:

This group of women actually fought alongside the men (for example, to help protect their herds from raiding parties) and they were buried with the weapons they had <u>used in life</u>.

Alternative:

This group of women was buried with these weapons as part of a <u>symbolic ritual</u>, to insure safety and protection in the afterlife.

Support:

Women buried with weapons had bowed legs, which often results from riding horses.

Anomalies:

Weak:

Most of the men had signs of some injuries, such as places where minor fractures had occurred and healed, but few of the women did, regardless of what they were buried with.

Strong:

Anthropologists have found evidence that the tribe had strong religious taboos against women fighting or hunting.

Story 4

Event:

The average age at which American women choose to get married for the first time has been increasing over the past few decades.

Possible explanation:

One possible explanation is that more women are pursuing advanced degrees after college and this means additional time spent in school. The suggestion is that women are delaying marriage because they are pursuing <u>advanced degrees</u>.

Alternative:

One possible explanation is that, since the 1960's, there has been a decreased negative stigma attached to couples who <u>live together</u> before marriage. Women are marrying later because they are choosing instead to live with their partners before they get married.

Support:

More women are achieving high positions in their chosen fields.

Anomalies:

Weak:

Some women are getting married while in their professional programs, not delaying it until they finish.

Strong:

The increase in marriage age is observed across economic status and education levels.

Story 5:

Event:

Over the past few years, there has been an increase in the rate at which fish are dying in the waters off the east coast of Central America.

Possible explanation:

The amount of invasive algae (plant like organisms) has been increasing in the coastal waters. On this view, <u>toxins</u> produced by the algae have been killing the fish.

Alternative:

The number of hurricanes and other violent storms in the area has been increasing and the strong winds have been affecting <u>ocean currents</u>. On this view, the rapidly changing ocean currents have churned up the water and disrupted the food supply of the area.

Support:

Invasive species are dangerous because native species have never encountered them before. While other creatures in a species natural environment have adapted to them, they may produce substances that are very harmful to creatures in other environments.

Anomalies:

Weak:

Swimmers visiting the coastline have shown no change in health.

Strong:

Large sea mammals that also live in these coastal waters have shown no change in health.

Story 6

Event:

When Rome conquered various parts of the world (such as Britain or Gaul), the indigenous people swiftly adopted Roman gods and began to celebrate Roman religious holidays.

Possible explanation:

It was simply a <u>matter of force</u>; the Roman Empire forced the conquered people to adopt their gods and religions or else suffer retaliation.

Alternative:

The Romans had an extensive network of trade. The indigenous people adopted the Roman religion for <u>economic reasons</u>, that is, to have better access to the wealth and goods that could result from joining the Romans in trade.

Support:

There is evidence that several sites of indigenous worship (such as temples) were destroyed by invading Roman armies.

Anomalies:

Weak:

Some Roman administrators themselves adopted the gods of the conquered people and suffered no punishment from their superiors.

Strong:

Records show that even when their children were seriously ill, the indigenous people would pray to the Roman gods, rather than the indigenous gods, for their child's health.

APPENDIX B: Full protocol

(Note: This protocol is for Story 1 and presents the target, then the alternative, no support, and a strong anomaly.)

Today's date: DOB: M F
C 4
St 1 T + Alt no sup T an str

Event:

When doing complicated and innovative surgeries (such as transplants, etc.), some hospitals clearly have a better success rate than others, even if the patients and procedures are comparable.

Possible Explanations:
Researchers have identified two possible explanations for this event:

One possible explanation is that successful hospitals have attracted surgeons and surgical nurses who are especially skilled and well educated.

Another possible explanation is that in successful hospitals each surgeon works with a particular group of surgical nurses, so that surgeon and nurses learn to work together as a standard team.

Now that you've thought about the possible explanations, please rate how convincing each one is.

How convincing is the skilled and well educated explanation?

1 2 3 4 5 6 7 8 9 10 11

not at all convincing	as convincing as not	completely convincing

Please tell me why you chose this rating.

How convincing is the together as a standard team explanation?

1 2 3 4 5 6 7 8 9 10 11

not at all convincing	as convincing as not	completely convincing

Please tell me why you chose this rating.

Researchers have not yet had a chance to search for any additional information.

Sometimes information makes you want to change your rating; sometimes it makes you want to keep it the same. Now that you've heard about the additional information, do you want to change your ratings of either explanation or keep them the same?

Do you want to change your rating of the <u>skilled and well educated</u> explanation or keep it the same?

1	2	3	4	5	6	7	8	9	10	11

not at all convincing — as convincing as not — completely convincing

Please tell me why you chose this rating.

Do you want to change your rating of the <u>together as a standard team</u> explanation or keep it the same?

1	2	3	4	5	6	7	8	9	10	11

not at all convincing — as convincing as not — completely convincing

Please tell me why you chose this rating.

More recently, researchers have found some additional information:

> Some of the hospitals with the worst success rates have poor success rates even when they recruit especially skilled surgeons from other high-performing hospitals.

In light of this information, do you want to change your rating of the <u>skilled and well educated</u> explanation or keep it the same?

1	2	3	4	5	6	7	8	9	10	11

not at all convincing — as convincing as not — completely convincing

Please tell me why you chose this rating.

Do you want to change your rating of the <u>together as a standard team</u> explanation or keep it the same?

1	2	3	4	5	6	7	8	9	10	11

not at all convincing — as convincing as not — completely convincing

Please tell me why you chose this rating.

If you had to select one explanation as the more convincing, which would it be:

<u>skilled and well educated</u>　　<u>together as a standard team</u>

Please tell me why you chose this explanation.

REFERENCES

Ahn, W., & Kalish, C. (1995). The role of covariation vs. mechanism information in causal attribution. In R. Wilson, & F. Keil (Eds.) *Cognition and explanation*, Cambridge, MA: MIT Press.

Chinn, C., & Brewer, W. (1993). The role of anomalous data in knowledge acquisition; A theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1-50

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*, 623-654.

Chinn, C. A. & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, 19, 323-393.

Dunbar, K. (2000) How scientists think in the real world; implications for science education. *Journal of Applied Developmental Psychology* 21(1); 49-58

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, 1, 127–138.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Koslowski, B., (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.

Koslowski, B., Libby, L.A., O'Connor, K., Rush, K., & Golub, N. (in preparation) Why some anomalies are more problematic than others.

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005) Scientific reasoning in your children: preschoolers' abilty to evaluate covariation evidence. *Swiss Journal of Psychology.* 64(3) 141-152

Krems, J.F. & Johnson, T.R. (1995). Integration of Anomalous Data in Multicausal Explanations. In J.D. Moore & J.F. Lehmann (Hrsg.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (S. 277-282). Mahwah, NJ: Lawrence Erlbaum.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674–689.

Kuhn, D., Amsel , E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* Orlando, FL: Academic.

Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press. ISBN

0-226-45808-3

Lipton, P. (2004) *Inference to the best explanation, 2nd edition*. New York, NY; Routledge

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175-220.

Magnani, L. (2001). *Abduction, reason, and science: Processes of discovery and explanation.* Dordrecht, Netherlands: Kluwer Academic Publishers.

Proctor, R. W. & Capaldi, E.J. (2006) *Why science matters: understanding the methods of psychological research.* Malden, MA; Blackwell publishing

Quine, W.V. & Ullian, J.S. (1978). *The Web of Belief.* New York: Random house

Quine, W.V.(1969) Natural kinds. In W. V. Quine, *Ontological relativity and other essays.* New York: Columbia University Press

Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences*, 1(2), 201–238.

Griffiths, T. & Tenenbaum, J. (2009)Theory-Based Causal Induction. *Psychological Review*, 116(4):661–716

Swiderek, M. (1999). Beliefs Can Change In Response To Disconfirming Evidence and Can Do So In Compicated Ways, But only If Collateral Beliefs Are Disconfirmed. *Cornell University.*

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*, 435-502.

Wilson, R. A. & Keil, F. C. (1998) The Shadows and shallows of explanation. *Minds and Machines,* 8; 137-159

Young, M. E. (1995) On the origin of personal causal theories. *Psychonomic Bulletin & Review,* 2(1), 83-104