

Science About What Scientists Do:

Distinguishing Between Explanations for Causal Events

Laura A. Libby

Cornell University

DISTINGUISHING BETWEEN EXPLANATIONS 2

Abstract

The causal reasoning literature suggests that hypothesis testing will only include tests that support a particular hypothesis (confirmation bias), rather than distinguish between possible hypotheses (Wason, 1960; Mynatt, Daugherty, & Tweney, 1977; Klayman & Ha, 1987). Self-generated hypotheses should elicit a stronger confirmation bias than other-generated hypotheses, possibly because the generation of a hypothesis requires an initial assessment of plausibility (Schunn & Klahr, 1993). Plausibility is determined by considering a possible cause within a network of prior knowledge about the world (Koslowski, 1996). Our study examines the testing of genuine explanations in non-emotion-laden, complex causal reasoning situations. We predict that the source of the explanation (self or other) and presence of alternative explanations will influence ability to distinguish between two hypotheses. Furthermore, we predict that the incorporation of prior knowledge into the hypothesis test will allow individuals to distinguish more successfully. Sixty subjects ($F = 32$, aged 18-22) completed a structured interview evaluating explanations that varied on number of explanations present and source of explanation. Ability to distinguish between a target explanation and its complement (but not a genuine alternative) was shown to differ based on the source (self or other) of the target and whether an alternative was provided. Prior knowledge was only used when distinguishing between the target and a genuine alternative. In general, the use of a contrast or covariation test is the best predictor of ability to distinguish between two explanations.

Science About What Scientists Do:

Distinguishing Between Explanations for Causal Events

Scientists across all fields employ the same process when they try to explain phenomena – the process known as the “scientific method.” A scientist generates a hypothesis about the causality of a particular factor; the hypothesis is subjected to tests; and, when the results of these tests are anomalous for the initial hypothesis, that factor is rejected as a cause, and a new hypothesis is generated and tested. By integrating hypotheses with what is already known about the world, scientists develop theories on not just *what* causes a phenomena, but also *how* it is caused. Non-scientists use the same method of reasoning about possible causes. After identifying a covariate that is plausibly causal, most individuals want to go beyond mere covariation and find a theory or explanation that accounts for causation. In our study, we investigate the ability of non-scientists to do what scientists do – to engage in scientific reasoning about genuine explanations for causal events.

Operationalizing Causal Reasoning

Early causal reasoning paradigms arose in tandem with endeavors to understand the development of scientific reasoning in children. Consistent with his formal-operational model, Piaget’s approach to causal reasoning was hypothetico-deductive, incorporating a basic understanding of cause and effect (as simple as, “If I cry, Mommy will give me what I want”). His initial causal tasks with children involved testing combinations of chemicals to find the components of a specific brew. Either a particular combination yielded the desired result or it did not, and children were successful in determining the makeup of the target chemical (Inhelder & Piaget, 1958). As in this task, the assignment of causation usually stems from systematic

DISTINGUISHING BETWEEN EXPLANATIONS 4

analysis of *covariation*. In another key covariation task, children had to differentiate between two possible hypotheses: that a light was turned on by one lever or by another. If the child knew that the activation of both levers turned on the light, and the activation of only the first lever did not turn on the light, the child would infer that the second lever is responsible for turning on the light. The covariation for the second lever is perfect, while the covariation for the first lever is flawed (Shultz & Mendelson, 1975).

Of course, covariation as evidence for causation is inherently problematic. In many situations, reverse causality poses a strong potential confound. Without a controlled experiment or time-sequence data, if A covaries with B, it is just as likely that A causes B as it is that B causes A. Alternatively, there could be a third confounding factor, C, that covaries with both A and B and may be the actual cause for both. The aforementioned studies included no consideration of such alternative hypotheses. In Inhelder & Piaget, the resulting chemical could not have caused its components because the participants saw that the components existed before they created the result – thus, reverse causality was not called into question. In Shultz & Mendelson, a third, potentially confounding hypothesis, that activation of both levers was necessary for turning on the light, was neither incorporated in the experiment nor considered by the children. Overall, covariation is necessary for causality, but to engage in sound reasoning means to discriminate between covariations that are causal and those that are merely artifactual (Koslowski, 1996).

Determining whether a covariation is causal requires an understanding of the overarching theory linking the proposed cause to its effect. An *explanation* for an event not only identifies the covariation but can explain *why* it is causal. With this causal mechanism, covariations can be assessed for relative plausibility, and, where mechanism does not exist, completely implausible

DISTINGUISHING BETWEEN EXPLANATIONS 5

covariations can be attributed to artifact and discarded (Brem & Rips, 2000). When individuals were asked to determine which of two covariations was the cause for better gas mileage in a car, either the car being small or the car being red, causation was likely to be attributed to the “red” hypothesis when the experimenter provided mechanism (that the color red facilitates alertness in drivers, causing them to drive more efficiently) (Koslowski, 1996). Experimenters presented this piece of information because it was not available to most participants based only on their prior knowledge or experience. When an explanation is consistent with background information that individuals have about the world, mechanism is easier to determine, and therefore causality is easier to detect. For example, natural selection as an explanation for speciation is consistent with information about plate tectonics and animal breeding. Spontaneous generation, on the other hand, is not consistent with these concepts. Furthermore, outside information that is consistent with the spontaneous generation hypothesis is vastly more difficult to obtain than it is for the natural selection hypothesis. The natural selection hypothesis is therefore the more plausible explanation for speciation (Weber & Depew, 1995). In all, when assessing the plausibility of any explanation, two types of information aside from covariation can be incorporated: results of a test that are consistent with one explanation but inconsistent with another; and information from the network of past experiences and knowledge that is consistent with the target hypothesis.

Hypothesis Testing and Rejection

When reasoning about multiple explanations, an initial assessment must be made to establish relative plausibility. Which of the possible explanations is the most likely to be correct? In hypothesis rejection, completely anomalous (undermining) information diminishes the plausibility of a particular explanation to practically (but not exactly) zero, and the

DISTINGUISHING BETWEEN EXPLANATIONS 6

explanation is discarded. The literature suggests that self-generated hypotheses are rated as more plausible than other-generated hypotheses. This effect may arise because, when generating their own hypothesis, individuals perform systematic mental tests of spontaneously generated alternatives, rejecting numerous alternatives before settling on the one that they “generate.” It also appears that other-generated hypotheses are tested more rigorously than self-generated, possibly due to the lack of spontaneous mental testing of those explanations (Schunn & Klahr, 1993). An explanation increases in plausibility as alternative explanations are discounted, so the explanation that is the most difficult to undermine is chosen as the most plausible explanation (Koehler & Harvey, 2004; Koslowski, 1996).

Many tasks operationalizing causation as covariation, such as those listed in the first section, have a limited hypothesis set that is small and uncomplicated. The hypotheses are derived from multiple plausible, mutually exclusive factors, and, of these hypothetical causes, one hypothesis has perfect covariation with the result. Therefore, if undermining evidence for any one factor arises, the appropriate response is rejection of that hypothesis. This variety of study incorporates little accounting for mechanism or incorporation of anomalous (incongruent) evidence into a hypothesis. Such studies provide information about reasoning in situations with a limited number of distinct, plausible hypotheses, but, in everyday life, causal inference often takes place in more complex arenas (Koslowski, 1996).

In causal reasoning situations with greater ecological validity, an initial hypothesis serves as a *working hypothesis* or *global theory* that is broad and imprecise. When incongruent data arises, this evidence can function to narrow a global theory down to a *local hypothesis* that accounts for anomalies. For example, a person, having only ever seen a set of blue Encyclopedia Britannicas, may hold the hypothesis, “All books are blue.” Upon entering a library and viewing

DISTINGUISHING BETWEEN EXPLANATIONS 7

other books, the person would modify the hypothesis to, “All encyclopedias are blue.” If he then sees a set of red encyclopedias, the hypothesis would become, “All Encyclopedia Britannicas are blue,” and so on (Koslowski, 1996).

The decision-making model grows in complexity when *tests* must be performed to differentiate between hypotheses. In seeking evidence that leads to the rejection of hypotheses, the first step is to identify the properties of the stimuli that could covary with a phenomenon. The next step is to conduct tests that could *undermine* the hypotheses of the causality of each property. Thus, information that is sought should ideally distinguish between hypotheses, rather than be consistent with more than one. Most studies have shown that people move through hypotheses based on a global theory on the mechanism of *how* causation is taking place, consecutively undermining them until one non-rejected covariation stands alone (Koehler & Harvey, 2004). However, when the overarching theory is incorrect, tests of hypotheses often fail to expose the accurate causal factor; repeated testing of a flawed theory without its eventual rejection will prohibit individuals from determining the true causal

When previous studies have investigated hypothesis testing, good reasoning has been framed as seeking undermining evidence. In one quintessential hypothesis testing study, participants were asked to identify the rule governing the sequence of numbers, “2, 4, 6.” Participants most frequently hypothesized that the rule was “increasing even numbers,” but failed to test sequences that, if they were consistent with the rule, would undermine the increasing evens theory. In actuality, the rule was “any three increasing digits.” Testing the sequence “8, 10, 12” both supports increasing evens and fails to distinguish between multiple possible rules. Conversely, testing the sequence “7, 6, 5,” fails to undermine the increasing evens hypothesis (Wason, 1960). Mynatt, Daugherty, & Tweney (1977) developed a computer-

DISTINGUISHING BETWEEN EXPLANATIONS 8

animated task to investigate the differential testing of hypotheses. Participants overwhelmingly chose tests that supported their (incorrect) target hypothesis, and therefore failed to discard it. Nevertheless, working from their flawed global theory, individuals were unable to identify the appropriate local hypotheses to test.

An important distinction exists between a *test* of an explanation and its *results*. Klayman and Ha (Klayman & Ha, 1987) note that the results of a test are what ultimately distinguish or fail to distinguish between hypotheses. For example, the results of the test “8, 10, 12” are congruent with both the target hypothesis and the experimenter’s rule, so that test is essentially meaningless for determining the correct rule. However, had the experimenter’s rule been “any three numbers below 10” (and, for all the participants knew, it could have been), the test result would have been that it is inconsistent with the rule. This result would undermine the ascending evens hypothesis. Good scientific thinking occurs only if that particular hypothesis is discarded and alternative hypotheses are tested sequentially. However, the rule used in the study (“three increasing digits”) was contrived to elicit tests that were consistent with both the rule and the target hypothesis, so alternative hypotheses were rarely tested. Nevertheless, because successful hypothesis testing is defined as seeking distinguishing evidence, participants in the Wason and Mynatt tasks were said to have engaged in what is known as *confirmation bias*.

Confirmation Bias

A large amount of disagreement exists over what, exactly, is confirmation bias. *Why* is it said that no undermining evidence sought for one particular explanation in the Wason task? Many researchers claim that alternative explanations are not adequately perceived or considered. In the Wason task, only certain alternative hypotheses were undermined; most of the tested

sequences only supported the generated hypothesis, and many alternative hypotheses were left unexamined. Others claim that alternative hypotheses are considered, but are quickly and often irrationally discounted as relatively implausible. More debate exists over what happens (and ought to happen) when an individual is presented with undermining (or *anomalous*) evidence – information that is incongruent with the target hypothesis. The evidence may be simply ignored, or there could be an attempt (often dubious) to reconcile the anomaly with the first hypothesis (Koslowski & Maqueda, 1993).

Overall, confirmation bias implies a “marriage” to one particular hypothesis. This dedication is especially apparent with self-generated hypotheses; other-generated hypotheses evoke more skepticism and thus are tested more rigorously (Schunn & Klahr, 1993). In much of the available literature, when given the opportunity to test their hypothesis, people invariably only seek out information that will either support the target hypothesis or fail to distinguish between any possible hypotheses (Koslowski & Maqueda, 1993). There are two types of valid alternatives to any explanation. The first is a genuine alternative, which introduces another specific possible cause with causal elements that are independent from the original explanation. The second is a complementary alternative, which does not identify any specific causal elements, and can only be described imprecisely as “not the target explanation.” Karmiloff-Smith’s studies (Karmiloff-Smith, 1984) have shown that people prefer a flawed explanation over merely its complement, so confirmation bias may result from avoiding the discarding of a specific explanation for lack of a more plausible alternative. Thus, our operational definition of confirmation bias is: *a failure to take account of information that successfully distinguishes between two hypotheses.*

Our Study

Where the causal reasoning literature focuses on the examination of covariation, we investigated the process behind evaluation of actual explanations for complex events. In addition, by allowing for relatively unconstrained reasoning in this task, we were able to assess what people do spontaneously when engaging in scientific thinking. In many tasks assessing reasoning about covariation, performance is treated as correct only when possibly relevant information is either limited by the participant to very few variables or completely excludes mechanism and corollary events. In one classic study, children were asked to state which of two types of cakes (carrot or chocolate) causes the common cold (Kuhn, Amsel, & O'Loughlin, 1988). Without obvious mechanism or corollary event provided by the experimenter, the researchers treated as correct reasoning responses that were limited to the information that the experimenters provided, namely information about covariation. When responses included a priori knowledge (such as previously acquired information about causal mechanism or corollary events), the participant was considered to have engaged in flawed reasoning. In scientific reasoning, the incorporation of outside information in the form of mechanism or corollary events for a covariation is a fundamental step that leads to causal inference. Our experiment allowed participants to choose the information that they believed to be pertinent from an unrestricted domain that included factors both presented by the experimenter and brought in from the participant's prior knowledge.

We examined confirmation bias in reasoning about explanations, rather than merely covariation, for events drawn from the natural and social sciences. The event stimuli that we constructed did not elicit a strong emotional response – with events that are emotion-laden, it would be impossible to differentiate between judgment based on reasoning and judgment based

on emotion. We gave participants free rein to seek information about explanations for events, and evaluated their ability to generate *information that could, in principle, distinguish between two explanations*, namely between the target explanation and either a genuine alternative or a complementary alternative. Then we investigated which aspects of a piece of information made tendency to distinguish more likely. These elements included the origin of explanations (self- or other-generated) and the number of explanations (one or two), as well as certain types of factors within a piece of evidence that are traditionally thought to be relevant to causal reasoning.

In order for a piece of information to be relevant for assessing causality, it must incorporate a variable, or *factor*, that may contribute to the true explanation. When trying to ascertain if Joe is an extravert or an introvert, one could ask if Joe enjoys social gatherings. In this case, “enjoyment of social gatherings” is a factor. (Since socializing is positively associated with extraversion, requesting this information would be an example of reasoning framed in a “positive test strategy” (Klayman & Ha, 1987) – if Joe is an extravert, then he likes parties) Sometimes, a factor can function as an alternative explanation. If the target explanation for the decline of a rare lizard population is global warming, and an individual asks if there has been recent human interference in the lizard’s habitat, then “human interference” is both a factor and a generated, genuine alternate explanation. When factors are relevant to assessing any explanation, we categorize them into one or more of the following three non-mutually exclusive kinds of information.

Causal Mechanism. Information on causal mechanism provides the “how” in a logical line of argument. In the hypothesis that cigarette smoking causes lung cancer, the factor that links smoking and cancer together (namely, the mutagenic effects of toxins found in cigarettes on lung cells) would function as a causal mechanism. However, in the hypothesis that unhealthy

behaviors lead to lung cancer, cigarette smoking, which links maladaptive behavior with cancer, could be considered a causal mechanism. In other words, a causal mechanism is *any cause on a causal pathway that is more proximal to the effect than the explanation*. Therefore, because an individual must self-generate the link between a proposed cause and the effect, causal mechanism must arise from a priori knowledge of information related to an explanation. An individual's prior experience limits the ways in which he or she evaluates an explanation; identifying mechanism allows for the treatment of covariation evidence as an explanation rather than noise, making the hypothesis more plausible (Ahn, Kalish, Medin, & Gelman, 1995). For example, most people know that the flu, which is most prevalent in the winter months, is not transmitted by cold weather (one covariation) but by contact with people who are already infected (another covariation). Individuals' prior experience allows them to understand that there is no good mechanism for the cold weather hypothesis, but knowledge about viruses leads to the establishment of a feasible mechanism for the human contact explanation. The literature also shows that the attribution of higher plausibility to a hypothesis will influence the rigor with which the hypothesis is tested (Schunn & Klahr, 1993).

Corollary Event. Explanations do not stand alone, but rather exist and are evaluated within an array of related information (Koslowski, 1996). It is possible that, in certain instances when individuals request evidence for an explanation, they are attempting to find consistencies between the evidence and *something that they believe to be true about the world*. In our experiment, these beliefs came from the participant's prior understanding of natural and social phenomena. For example, when testing global warming as the explanation for lizard population decline, a participant might ask for evidence about methane in the atmosphere. The participant is introducing a new factor that is based on his or her belief about the role of atmospheric methane

in global warming. Roughly, the participant is thinking: if global warming is at fault, then their hypothesis would be supported if methane was present in the atmosphere. Therefore, this participant has brought in a corollary event.

Contrast or Covariation Test. A contrastive test compares information from distinct data points, such as between times or populations; it asks, “Is there a difference?” A covariation test explicitly examines if the difference is associated with the event to be explained. These tests are akin to those in conventional, constrained causal reasoning tasks, such as when flipping one of two switches turns on a light bulb; both are tests of direct covariation. However, because our causal reasoning task is complex and more ecologically valid, involving infinite potentially contributing variables, covariation tests in this study can strengthen or weaken the plausibility of an explanation but never determine causality.

Nevertheless, due to the important distinction between a test and its possible results, contrast and covariation tests are valuable when attempting to distinguish between explanations. When the variable being tested must necessarily correlate with the event in order for an explanation to be true, if there is *not* a correlation, then the explanation cannot be true. For example, in testing the hypothesis that increased predator population has led to a decline in lizards, if 1) the contrastive test between predator populations before and after lizard decline shows that there was no increase in predator, or 2) the covariation test of a relationship between change in predator population and lizard decline is not significant, then the predator increase explanation is undermined. In our evaluation of tests generated by participants, we considered both possible results – positive and negative – in order to determine if the test could potentially distinguish between two explanations.

Hypotheses

Our evaluation approached spontaneous generation of potentially distinguishing hypothesis tests from two angles. We asked which types of explanation are most linked to the ability to request information that could result in distinguishing evidence. When individuals are able to generate distinguishing tests, they should be more likely to introduce factors from a network of related information. Therefore, we also asked if participants who are able to distinguish use causal mechanism, corollary events, and contrast/covariation tests differentially than those who fail to distinguish. We employed the following hypotheses:

H₁: The ability to generate distinguishing information will differ based on the origin (self- or other-generated) and the number (zero, one, or two) of evaluated explanations.

H₂: When people generate distinguishing information, they are more likely to use causal mechanism, corollary event, and/or covariation tests than when they fail to generate distinguishing information.

Methods

Participants

Sixty participants (female = 33), ages ranging from 18 years to 22 years, were recruited on a Northeast American university campus. Informed consent was obtained from all participants. Participants were randomly assigned to one of four conditions: No Explanation Present (NE), Self-Generated Explanation Present (SGE), One Explanation Provided (1Exp), and Two Explanations Provided (2Exp). (The data presented in this study are part of a larger sample,

which will be analyzed in its entirety in future research.) Instructions for protocol assembly and random assignment to condition by gender are included in Appendix A.

Materials

Story Problems. Six story problems from the natural and the social sciences were constructed. From the natural sciences: 1) Why the platypus population is declining; 2) Why only one particular species of tree would die while other species in the same area were not affected; 3) How the species of grape that is used to make European wines is flourishing in Europe when it experienced a devastating parasite infestation in 1900. From the social sciences: 4) Why homelessness skyrocketed in the 1990's; 5) How the entire population of an ancient city could vanish; 6) Why hard-to-place adopted children are more successful in homes with same-sex rather than opposite-sex parents.

Using a table of random numbers, the protocols were grouped into three pairs. Each pair included one story from the social sciences and one from the natural sciences: 1) Platypus and Cities; 2) Grapes and Adoption; 3) Trees and Homelessness. Story problem order was randomized within protocol pair. Each subject was presented with one protocol pair, thereby completing the same condition twice, once for each story problem in the pair.

Explanations. For each story problem, two possible explanations (Explanation A and Explanation B) were created by the experimenters. Within story problems, we controlled for length and syntactical complexity of the explanations. In both the 1Exp and 2Exp conditions, one of the experimenter's explanations was presented as the target explanation (T). In 2Exp only, the other explanation was presented as the alternate explanation (A). In both conditions, we counterbalanced which explanation was presented as T as a between subjects variable. In

order to establish a comparable target explanation in the NE and SGE conditions, participants were asked to generate what they believed to be the most plausible explanation. In SGE, they generated the target explanation directly after presentation of the story problem. In NE, which is conceptualized as the “unprompted” condition, participants generated the target explanation at the end of the interview so that they did not focus on any particular explanation at the outset of the study. All stimuli for each story problem are included in Appendix B.

Design

The between-subjects variables were age, sex, protocol pair, condition (NE, SGE, 1Exp, or 2Exp), and presented target hypothesis (in 1Exp and 2Exp). The within-subjects variable was story problem.

This study is based on a sub-set of a larger data collection in which 24 participants (12 males and 12 females) were randomly assigned to each condition for a total of 96 participants. In the 1Exp and 2Exp conditions, half of the subjects (6 males and 6 females) received Explanation A as the target explanation in both story problems, and the other half received Explanation B as the target. The first 60 participants to complete the task for the larger data collection are included in this study. Out of these participants, there were 15 in NE, 14 in SGE, 16 in 1Exp, and 15 in 2Exp.

Procedure

The semi-structured interview was conducted by trained experimenters who obtained informed consent at the outset of the study. Then the interview protocol was placed in front of both the interviewer and the participant so the participant could follow along visually while the

interviewer read through the stimuli. (Before encountering any story problems, all participants were provided instructions on recording emotionality on the Affect Grid (Russell, Weiss, & Mendelsohn, 1989), which would be presented to them at points throughout the interview. Participants were asked to use the Grid to report the arousal and valence of their emotionality based on provided information. Although the Affect Grid is often treated as a focal experimental measure, we include it in order to establish that our stimuli do not engender an emotional response. We found this grid to be appropriate in its measure of both the nature and intensity of emotion, as well as practical in its brevity and clarity. However, for the purposes of simplicity and straightforwardness in this study, we have not included analyses of the emotion measure; results on emotionality will be analyzed in a future study.)

After completing the instructions, the experimenter read through the entire protocol while the subject followed along silently. The structure of the interview protocol varied among the four conditions. Examples of information presented and questions asked in each of the conditions are available in Appendix C. Table 1 outlines the protocol construction of each of the four conditions. Every protocol consisted of three major sections (listed in bold type in Table 1): story problem, target explanation, and information seeking. In every condition, participants were presented with the story problem in the first section. In the SGE, 1Exp, and 2Exp conditions, the target explanation was identified (either provided or generated) directly after the story problem. In the NE condition, as previously discussed, the target explanation was identified at the end of the protocol in order to preclude having participants focus on particular explanation at the outset of the task. Additionally, in every condition, participants were asked to describe what information they would seek in order to better evaluate the story problem (in NE) or the target explanation (in SGE, 1Exp, and 2Exp). Generation of responses stopped when participants

DISTINGUISHING BETWEEN EXPLANATIONS 18

indicated that they would not seek any other evidence. After generating all of the requested information, participants were then asked to rate each piece of information that they generated on a seven-point scale of usefulness in understanding the story problem. Finally, they were asked to explain their ratings. At the end of the experiment, participants were debriefed.

Table 1. Interview arrangement for NE, SGE, 1Ex, and 2Ex conditions.

<i>No Explanation (NE)</i>	<i>Self-Generated Explanation (SGE)</i>	<i>One Explanation (1Ex)</i>	<i>Two Explanation (2Ex)</i>
<ul style="list-style-type: none"> • Story problem • Information seeking • Usefulness ratings • Identify most plausible explanation (use as T) • Affect Grid for T • Plausibility rating for T 	<ul style="list-style-type: none"> • Story problem • Identify most plausible explanation (use as T) • Affect Grid for T • Plausibility rating for T • Information seeking • Usefulness ratings 	<ul style="list-style-type: none"> • Story problem • Target explanation • Affect Grid for T • Plausibility rating for T • Information seeking • Usefulness ratings 	<ul style="list-style-type: none"> • Story problem • Target explanation • Affect Grid for T • Plausibility rating for T • Alternate explanation • Plausibility rating for A • Information seeking • Usefulness ratings

Experimenters recorded responses on the protocol pages, taking detailed notes of the participant’s answers to open-ended questions. Non-directive prompts (such as, “Can you tell me more?” and, “Anything else?”) were used to elicit clarity of participant response. Qualitative variables were categorically coded according to the schematics included in Appendix D. Each piece of evidence generated by each participant for both story problems was evaluated for all of the following variables:

- *Distinguishes Between Target and a Genuine Alternative (GenuAlt)* – The information requested could, depending on the results, distinguish a target from an alternative explanation that includes specifically defined factors. The genuine alternative could be either generated by the participant or provided by the experimenter (in the 2Exp condition).

- *Distinguishes Between Target and a Complementary Alternative (CompAlt)* – The information requested could, depending on the results, distinguish a target from its complementary alternative, but participant does not provide a genuine alternative explanation.
- *Does Not Distinguish (NoDisting)* – Regardless of the results, the requested information could not distinguish between the two explanations in question. Often, the information could increase or decrease the plausibility of one or more explanations, but would not completely disconfirm one explanation.
- *Causal Mechanism (CM)* – The requested information addresses a causal factor that is more proximal to the effect than the explanation being evaluated; in other words the responses include some information about the mechanism that mediates between a cause and its effect.
- *Corollary Event* – The requested information is consistent with something that the participant holds to be true about the world, such as evolution or a particular economic theory.
- *Contrast or Covariation Test* – In the requested information, the participant proposes either a contrastive or a covariation test of specific factors, including time trends and group comparisons. An important note for coding is that a contrast or covariation test must be explicitly stated; if a participant simply sought the presence or absence of a factor, even though the results of that search could potentially distinguish between alternatives, there is no contrast or covariation test generated. Cue words for a test are “before” and “after,” “associates with.” The test could be either empirical (by proposing an intervention) or observational.

The three types of distinguishing responses (GenuAlt, CompAlt, and NoDisting) are mutually exclusive; each piece of evidence was coded as only one of these three variables. The three types of factors (CM, corollary event, and contrast/covariation test) are not mutually exclusive; each piece of requested information could be coded for any combination (or none) of these three variables. The relationships between these variables are similar to how, for any person, gender is a mutually exclusive variable, but both genders can be coded for hair color, eye

color, and height. Throughout coding, we erred on the side of caution; we judged explanations, factors, and tests at face value, coding only based on what the participants explicitly said.

Results

Controlling for Confounding Variables

There was no significant effect of gender or target explanation on tendency to distinguish between any kinds of explanations at the .05 significance level. Differences in failure to distinguish (NoDisting) and distinguishing between the target explanation and a complementary alternative (CompAlt) were not significant between protocol pairs ($\alpha = .05$). However, generating evidence that could distinguish between the target and a genuine alternative (GenuAlt) does differ between protocol pairs, $F(2, 57) = 4.27, p = .019, \alpha = .05$. Because an approximately equal number of each of the protocol pairs per condition is included in analysis, we did not control statistically for the effect of protocol pair. Future analyses will implement such a control.

Verbal Fluency

For each subject, we computed the average number of pieces of information requested (verbal fluency) across the two story problems. The continuous variable was reduced to a categorical variable because the distribution was not normal. The frequencies of verbosity levels by condition are listed in Table 2. Verbal fluency differed significantly by condition, $\chi^2(1, 30) = 47.30, p = .023$. As Table 2 indicates, there were no subjects who generated above five pieces of information in SGE, 1Exp, and 2Exp. The maximum amount of information generated declined as we moved from NE to SGE, 1Exp, and 2Exp. Large numbers of responses are more likely to

DISTINGUISHING BETWEEN EXPLANATIONS 21

occur in NE than in other conditions; NE promotes higher verbal fluency. Verbal fluency did not differ by gender or by protocol pair at the .05 significance level.

Table 2. Frequency of verbal fluency level by condition.

		Average Count of Information Requests											
		1.50	2.00	2.50	3.00	3.50	4.00	4.50	5.00	5.50	6.00	6.50	Total
Condition	NE	1	2	1	1	0	1	2	1	4	1	1	15
	SGE	2	2	1	2	1	3	2	1	0	0	0	14
	1Exp	2	1	1	9	0	3	0	0	0	0	0	16
	2Exp	2	4	2	1	3	3	0	0	0	0	0	15
Total		7	9	5	13	4	10	4	2	4	1	1	60

Analysis for Hypotheses

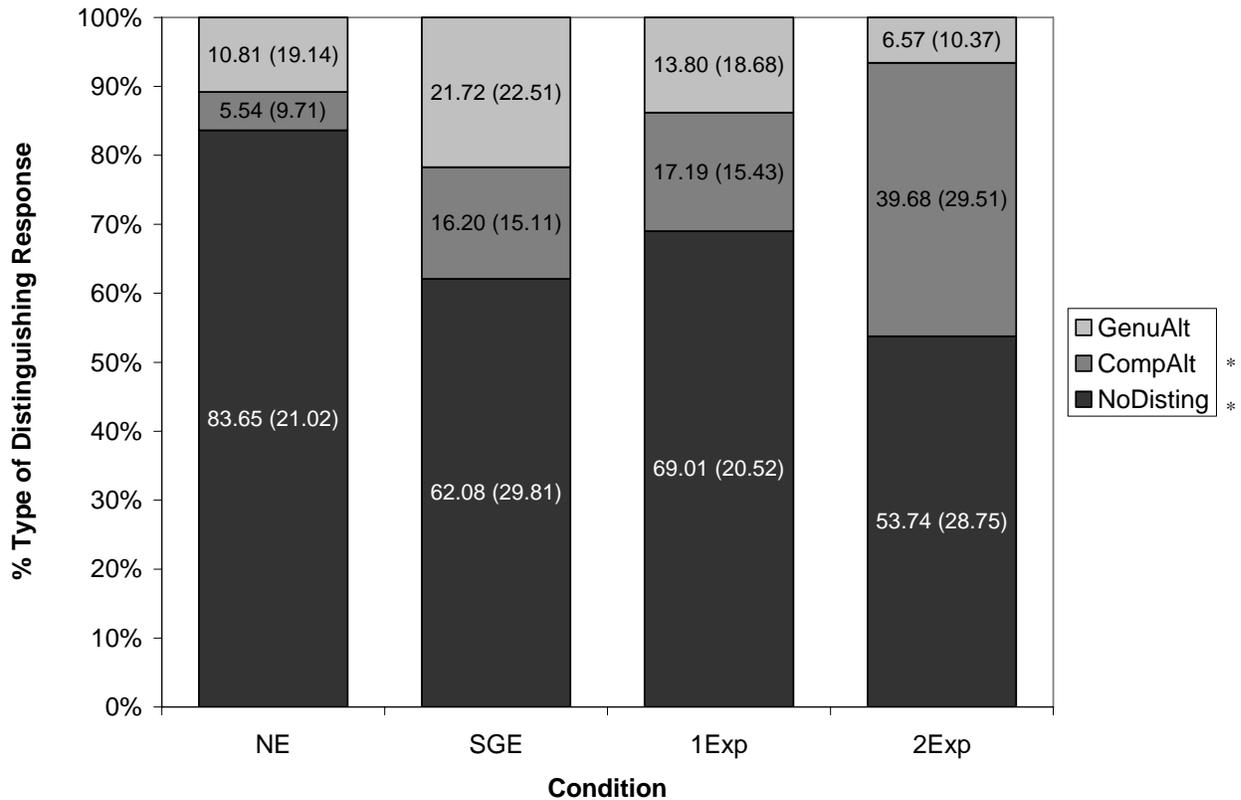
H_1 . Our first question was, are individuals better able to generate tests that could distinguish between hypotheses in certain conditions? Across the subjects in this sample, only four subjects failed to consider at least one genuine or complementary alternative. In future analyses, we will eliminate these subjects in order to reduce the noise from responses that are irrelevant to assessing explanations. In the present analysis, these subjects were included.

For the first analysis, for each participant, we computed the percentage of the total number of responses in each condition that distinguished between two explanations. This percentage was computed separately for distinguishing genuine alternatives, distinguishing complementary alternatives, and failing to distinguish altogether. The means (with standard deviations in parentheses) for the percentage of pieces of evidence generated by each participant in those three groups across condition, respectively, were 13.09 (18.51), 19.67 (22.29), and 67.24 (26.94). The distribution of each type of distinguishing response by condition is shown in Figure 1. For NoDisting, an ANOVA found an effect of condition, $F(3, 56) = 3.77, p = .02$. A follow-up with Fisher's LSD tests showed that NE was significantly different than both SGE and

DISTINGUISHING BETWEEN EXPLANATIONS 22

2Exp, $p = .02$. For CompAlt, an ANOVA also found an effect of condition, $F(3, 56) = 8.63, p < .001$. Follow-up LSD tests showed a significant difference between NE and 2Exp, between SGE and 2Exp, and between 1Exp and 2Exp, $p < .001$. For GenuAlt, however, the difference between conditions was not statistically significant at an α value of .05. In summary, condition affected tendency to distinguish between the target and a complementary explanation, as well as between the target and a genuine alternative.

Figure 1. Percentage of Types of Distinguishing Responses by Condition – Mean (SD).



* Statistically significant, $\alpha = .05$

H_2 . Our second question was, when individuals do generate distinguishing tests, are they more likely to use causal mechanism, corollary events, and contrast/covariation tests than when they fail to distinguish?

Incorporating causal mechanism (CM) ($M = 9.57$, $SD = 14.03$) and distinguishing between the target explanation and a genuine alternative (GenuAlt) ($M = 13.09$, $SD = 18.51$) was moderately positively correlated, $r(58) = .257$, $p = .048$. Including a corollary event ($M = 57.59$, $SD = 25.40$) and GenuAlt was also moderately positively correlated, $r(58) = .268$, $p = .038$. Using a contrast/covariation test was not significantly correlated with GenuAlt at an α value of .05.

Incorporating a contrast/covariation test ($M = 35.83$, $SD = 28.57$) and distinguishing between the target explanation and a complementary alternative (CompAlt) ($M = 19.97$, $SD = 22.29$) were strongly positively correlated, $r(58) = .682$, $p < .001$. Use of neither CM nor a corollary event were correlated with CompAlt ($\alpha = .05$).

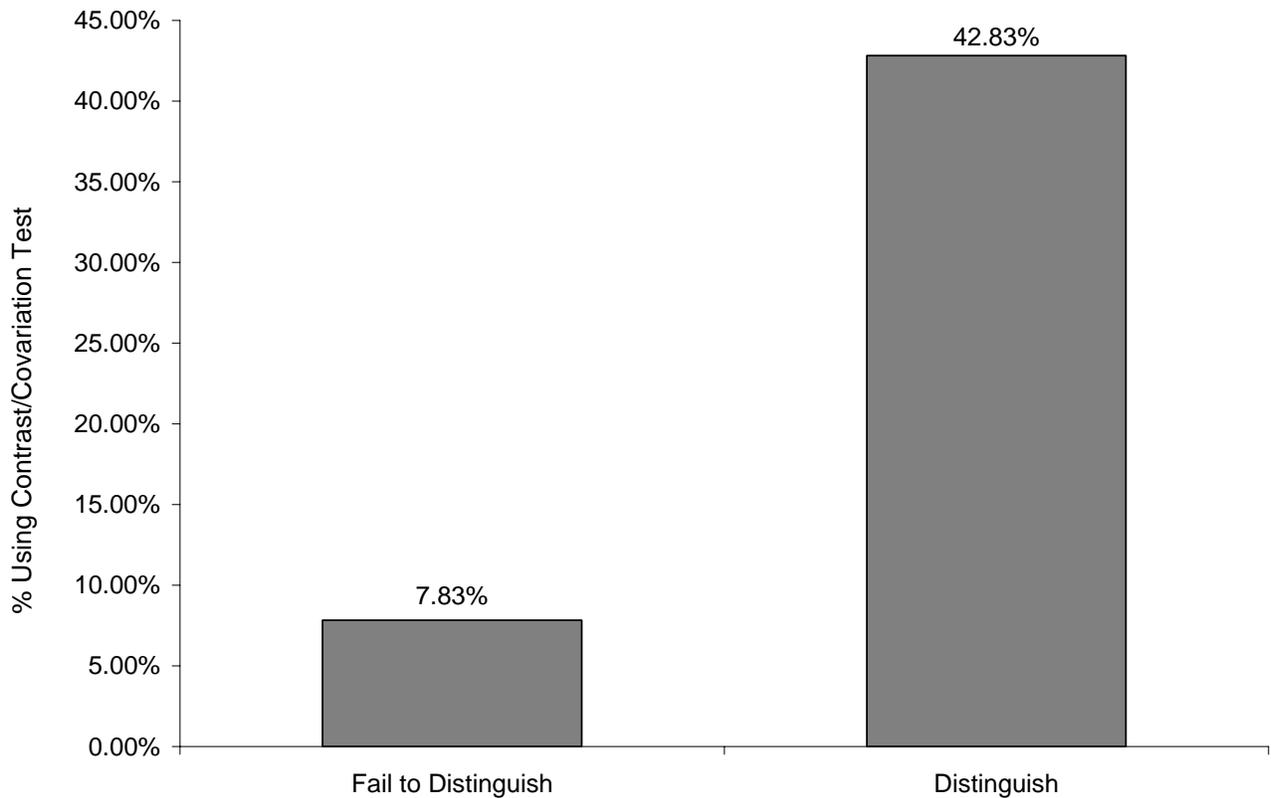
Failure to distinguish between two hypotheses (NoDisting) ($M = 67.24$, $SD = 26.94$) was strongly negatively correlated with use of a contrast/covariation test, $r(58) = -.718$, $p < .001$. NoDisting was not correlated with the use of either CM or a corollary event ($\alpha = .05$).

To further examine the relationship between distinguishing and using causal mechanism, a corollary event, or a contrast/covariation test, we created the categorical variables “Distinguish” (generated any distinguishing response, between the target and either a complementary or a genuine alternative) and “Fail to Distinguish” (never generated a distinguishing response). The average percentage of participants in Distinguish who used a contrast/covariation test was 42.83 ($SD = 26.74$). The average percentage of participants in Fail to Distinguish who used a contrast/covariation test was 7.83 ($SD = 15.92$). The difference in

DISTINGUISHING BETWEEN EXPLANATIONS 24

contrast/covariation test use between those who distinguished and those who did not was statistically significant, $F(1, 58) = 18.73, p < .001$. Figure 2 displays the differences in test use between failure and success at distinguishing. Use of neither causal mechanism nor corollary event differed based on ability to distinguish ($\alpha = .05$).

Figure 2. Incorporation of Contrast/Covariation Test by Ability to Distinguish



Discussion

Verbal Fluency

We asked participants to identify what information they would like to have in order to better evaluate either the story problem (in the NE condition) or the target explanation (in SGE, 1Exp, and 2Exp). Because we put no restriction on the number of responses that a participant could generate, participants were able to think out loud and cognitively pump out the maximum number of pieces of evidence that they would seek. The average number of responses (or verbal fluency), therefore, represents the “range of thought” permitted by each condition. Verbal fluency decreased as we moved from the no explanation condition to the self-generated explanation, one explanation provided, and two explanation provided conditions; the variables in each condition differentially constrained the possible range of thought. In each successive condition, the experimenter provided more information to participants, from no explanation, to a self-generated explanation, to one explanation, to two explanations. It appears that, as more information was presented to participants, they were able to focus more within the boundaries of that information, and they therefore generated less tangential information on their own.

The NE condition did not prime the participant to focus on one particular explanation while generating responses, which led to unconstrained thinking and the broadest possible range of responses. Conversely, the other three conditions prompted the participant to focus on one or more specific explanations, which seems to have limited the range of thought used in response generation. In conjunction with verbal fluency, assessing the content of the responses is vital for determining if the thought constraint is beneficial or detrimental to scientific reasoning. If participants generated fewer responses in the three other conditions but discriminated between

explanations better than they do in NE, then the high verbal fluency in NE can be attributed to generating responses that are an accumulation of information about the story problem, rather than evidence for determining the correctness of a causal factor. For example, if increased predator population is the target explanation for lizard decline, asking about the lizard's habitat would only yield additional information about the story problem, but would not help distinguish between possible explanations.

First Hypothesis

We investigated how the different conditions would influence ability to request information that could, in principle, distinguish between two possible explanations. In the no explanation condition, participants generated non-distinguishing responses significantly more often than they did in the self-generated explanation condition or the condition with two explanations provided. Suppose that a participant was generating information to evaluate the explanation the pollutants in the soil were responsible for lizard population decline. A non-distinguishing response can be composed of one of three kinds of information: 1) Information about the story problem that is not related to any explanation ("What kind of soil do they live on?"), 2) The generation of an alternative explanation without an attempt to test it ("Were the lizards hunted by humans?"), or 3) a proposed test of a covariate that could not yield distinguishing results ("Are there more factories producing industrial waste now than before the population decline?" – Even if this contrastive test showed no difference, pollutants could have originated from a different source, so pollution as a general explanation cannot be ruled out). We can conclude that, with a higher percentage of non-distinguishing evidence and a higher

verbal fluency, participants in the NE condition were spending much more time generating one of these three types of information, rather than successfully seeking distinguishing evidence.

A converging study showed that participants in NE spontaneously generated more genuine alternatives overall (Rush, 2007). It appears that an unrestricted range of thought leads to the spontaneous generation of more possible alternatives, but that those alternatives are rarely effectively tested and distinguished. On the other hand, in the 2Exp condition, where verbal fluency is the lowest, both the fewest genuine alternatives were spontaneously generated and a significantly lower percentage of non-distinguishing evidence was generated than in NE. The 2Exp condition also showed a significantly higher percentage of responses that distinguished between the target explanation and its complement than in any other condition. The literature varies in its definition of confirmation bias; some hold that people do not consider alternative explanations; others claim that people are just bad at looking for evidence that distinguishes between explanations. Our results suggest that, depending on the amount of information provided by the experimenter, those two measures of reasoning have an inverse relationship.

It is seemingly counterintuitive that, in a condition where a participant is directly presented with an alternative explanation and where he or she spends more time generating potentially distinguishing tests than other kinds of information, participants seldom distinguish between the target and any genuine alternative. To understand this result, we must consider what our operational definitions of “GenuAlt” and “CompAlt” imply about target and alternative explanations. Because our study incorporates a target explanation that is designed to be dominant in the participant’s thoughts, we discuss distinguishing evidence as discriminating between the target and another explanation, either a genuine or a complementary alternative. It is highly unlikely that the results of a test could both conclusively disconfirm one hypothesis and

explicitly confirm another genuine alternative – rather, undermining or ruling out one explanation would make the remaining alternative explanations slightly more plausible. When an individual is said to have distinguished between the target explanation and a genuine alternative (GenuAlt), what has actually happened is the individual has developed a test that could result in evidence that *undermines the genuine alternative*. If the information were to undermine only the target explanation, it would not explicitly support any particular genuine alternative (hence the use of the non-specific complement). Therefore, when an individual distinguishes between the target explanation and its complement, he or she has actually generated information that could simply *undermine the target*. This distinction allows for a new perspective on hypothesis testing: concentration on one hypothesis at a time.

In the 2Exp condition, where individuals are the most likely to distinguish between the target explanation and its complement, they are in fact focusing on undermining only the target explanation. Rush found that participants generate the fewest spontaneous explanations in this condition. It seems that the presence of a provided genuine alternative diminished the need to spend time on generating additional alternatives, and the individual could focus on developing good tests of the target explanation. The study instructions directed the participant to “determine if the [target] explanation is the correct one.” The results imply that individuals were the most able to follow those instructions in the 2Exp condition. In a future, similarly constructed condition, subjects will be asked to assess which of the *two* explanations (target or alternative) is correct to see if participants will perform differently under different instructions.

The SGE and 1Exp conditions did not show differences in any type of ability to distinguish. Koehler (Koehler & Harvey, 2004) maintains that individuals should test other-generated hypotheses more rigorously because there is no mental assessment of plausibility prior

to the selection of that hypothesis. We find that participants generate information to undermine the self-generated target in SGE and the other-generated target in 1Exp equally often. Because participants are asked to think out loud and articulate all pieces of information that they would seek to evaluate the target explanation, it is likely that participants are verbally reiterating the mental assessment of the plausibility of alternatives that they made prior to selecting their target explanation. While participants may generate more alternative explanations in the SGE condition than in 1Exp (Rush, 2007), which is consistent with Koehler's findings, the origin of the target explanation does not appear to affect its testing.

Second Hypothesis

The results from analysis of the first hypothesis demonstrate that it is very difficult to generate evidence that could successfully distinguish between two explanations; the majority of responses in every condition did not ask for distinguishing evidence. One reason that distinguishing is so challenging may be that individuals do not have enough background information on a topic (Ahn et al., 1995). If this idea is correct, then the use of more background information should correspond to more distinguishing evidence. Thus, for our second hypothesis, we asked about what factors in a response make distinguishing between explanations easier. Causal mechanism, corollary events, and contrast/covariation tests are indeed employed differentially depending on the ability of a response to distinguish between explanations. Our breakdown of the implications of distinguishing between a target hypothesis and either a genuine or a complementary alternative helps to clarify the results on the use of these factors.

When an individual distinguishes between the target and a genuine alternative, conceptually, he or she has identified evidence that undermines the alternative explanation.

Undermining an alternative would require either finding the results of a covariation test that are inconsistent with the explanation, or finding that the explanation is not consistent with any background information. Consistent with our hypothesis, the incorporation of both causal mechanism and corollary events are associated with the ability to distinguish between target and alternative. However, the use of a contrast or covariation test does not correlate with the ability to generate distinguishing evidence. Possibly, because individuals consider alternative explanations in light of the target explanations (by nature of the term “alternative”), and may have already considered possible mechanism for the target, they need the same kind of background information about alternatives in order to be able to compare the two explanations for plausibility. Future studies should investigate the seeking of causal mechanism and corollary events for both target and alternative hypotheses, and the order in which participants ask for these pieces of information. If individuals require comparable background information in order to be able to distinguish between two genuine explanations, then information on background information for the alternative should be requested after background information on the target.

Distinguishing between a target and a complementary alternative is equivalent to generating evidence that undermines the target explanation. The results for the use of background information and covariation tests in undermining the target are the inverse of the results for undermining a genuine alternative presented above. The use of contrast or covariation tests is highly correlated with ability to distinguish, while the incorporation of both causal mechanism and corollary events is unrelated to distinguishing. One possible reason for the strong correlation with covariation test is that, in the target explanations, certain factors were already present and available for the participant to test for covariation. Testing those factors required less cognitive exertion than generating an alternative hypothesis with its own factors to

test. In future analyses, we will investigate the origin of the factors used in each piece of requested information in relation to its ability to distinguish between explanations. Furthermore, because these factors are on hand for the participant to test, running covariation tests may have been chosen over finding mechanism or corollary events, which would again require more mental effort. To control for information provided by the experimenter, a future study could provide participants with both causal mechanism and covariates for an event and evaluate which set of factors they use more frequently when distinguishing.

There was also a strong negative correlation between failure to distinguish between hypotheses and use of contrast and covariation test; as the use of a test increases, failure decreases. Additionally, the difference in use of a contrast or covariation test between those who successfully distinguished between any two explanations and those who failed to distinguish was statistically significant. Causal mechanism and corollary event use was unrelated to failure to distinguish. It appears that the best predictor of avoiding a confirmation bias is the incorporation of a contrastive or covariation test. While it may be that, without access to any background information that would make causality likely, individuals have difficulty evaluating explanations (Kuhn et al., 1988), in causal reasoning situations that are more ecologically valid and inherently involve access to relevant background knowledge, tests of covariation seem to be the preferred method of distinguishing between explanations.

Conclusion

Theory delineates the range of evidence that we consider plausibly causal for an event, and bridges the gaps between background knowledge, a possible distal cause, and the event. From theory, we can generate distinguishing hypothesis tests, a crucial component to the

scientific method. Although complete adherence to a theory can cause trouble, such as when we infer causation from mere covariation or engage in confirmation bias, theory helps us to efficiently categorize covariations as either potentially causal or only artifact. In evaluating explanations for causal events, theory plays a fundamental role in good reasoning for both scientists and non-scientists.

References

- Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. [Electronic version]. *Cognition*, 54, 299-352.
- Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science: A Multidisciplinary Journal*, 24(4)
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Karmiloff-Smith, A. (1984). Children's problem solving. In M. E. Lamb, A. L. Brown & B. Rogoff (Eds.), *Advances in developmental psychology* (pp. 33-90). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis-testing. *Psychological review*, 94(2), 211-228.
- Koehler, D. J., & Harvey, N. (Eds.). (2004). *Blackwell handbook of judgment and decision making* (xvi ed.). Malden, MA, US: Blackwell Publishing.
- Koslowski, B. (1996). *Theory and evidence*. MIT.
- Koslowski, B., & Maqueda, M. (1993). What is confirmation bias and when do people actually have it. *Merrill-Palmer Quarterly-Journal of Developmental Psychology*, 39(1), 104-130.
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando: Academic Press.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment - experimental-study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29(FEB), 85-95.

Rush, K. A. (2007). *Spontaneous generation of alternative explanations*. Cornell University.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of personality and social psychology*, 57(3), 493-502.

Schunn, C., & Klahr, D. (1993). Self vs. other-generated hypotheses in scientific discovery. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*,

Shultz, T. R., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development*, 46, 394-399.

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

Weber, D., & Depew, B. (1995). *Darwinism evolving: Systems dynamics and the genealogy of natural selection*. Cambridge: MIT Press.

Appendix A

Protocol Assembly Schematic

Four Conditions

- 1) No Explanation (NE)
- 2) Self-Generated Explanation (SGE)
- 3) One Explanation (1Exp) Provided
- 4) Two Explanations (2Exp) Provided

Six Story Problems: Platypus, Trees, Grapes, Homelessness, Cities, Adoption.

- In condition 1Exp-A, the target explanation is Explanation A.
- In condition 1Exp-B, the target explanation is Explanation B.
- In condition 2Exp-A, Explanation A is target and Explanation B is alternative.
- In condition 2Exp-B, Explanation B is target and Explanation A is alternative.

Three Protocol Pairs:

- Platypus and Cities
- Grapes and Adoption
- Trees and Homelessness

No Explanation (SGE) Condition

Target Explanation	Males	Females	Total
<i>No explanation provided</i>	4 copies of each pair = 12 protocols	4 copies of each pair = 12 protocols	8 copies of each pair, half for M, half for F

= 24 protocols

Self-Generated Explanation (SGE) Condition

Target Explanation	Males	Females	Total
<i>No explanation provided</i>	4 copies of each pair = 12 protocols	4 copies of each pair = 12 protocols	8 copies of each pair, half for M, half for F

= 24 protocols

One Explanation Provided (1Exp) Condition

Target Explanation	Males	Females	Total
Explanation A	2 copies of each pair = 6 protocols	2 copies of each pair = 6 protocols	4 copies of each pair, half for M, half for F, = 12 protocols
Explanation B	2 copies of each pair = 6 protocols	2 copies of each pair = 6 protocols	4 copies of each pair, half for M, half for F, = 12 protocols

= 24 protocols

DISTINGUISHING BETWEEN EXPLANATIONS 36

Two Explanations Provided (2Exp) Condition

Target Explanation	Males	Females	Total
Explanation A (<i>Explanation B as alternative</i>)	2 copies of each pair = 6 protocols	2 copies of each pair = 6 protocols	4 copies of each pair, half for M, half for F, = 12 protocols
Explanation B (<i>Explanation A as alternative</i>)	2 copies of each pair = 6 protocols	2 copies of each pair = 6 protocols	4 copies of each pair, half for M, half for F, = 12 protocols = 24 protocols

→ 24 protocols x 4 conditions = 96 total protocols

Appendix B

Story Problems with Explanations

Events and explanations from each of these stories were inserted into the protocol design in Appendix C.

Platypus

Event to be explained

The platypus, an animal that lives in Australian waters, is a mammal that nevertheless lays eggs. In recent years, the platypus population has suffered a sharp decline. What might account for the decrease in the platypus population?

Explanation A

Environmental pollutants have introduced toxins into the water, killing many platypuses.

Explanation B

There has been an increase in the population of lizards that prey on platypus eggs.

Trees

Event to be explained

In the coastal regions of Eastern Asia, some species of deciduous trees are steadily dying out while others are still living. What might explain this difference?

Explanation A

Due to global warming, certain deciduous trees that grow in altitudes that are normally cooler are unable to adapt to the slight increase in temperatures.

Explanation B

Introduction of an invasive type of fungus that affects specific species of deciduous trees resulted in a differential decrease in deciduous tree population.

Grapes

Event to be explained

Varieties of *vinifera*, a European grape species, are the fruit involved in making European wines. At the end of the 19th century, most of the European *vinifera* plants died due to a parasite infestation. How are *vinifera* vines still thriving today?

Explanation A

A small number of vinifera vines had natural resistance to parasites, and those vines were selectively bred by winemakers and used to repopulate the vineyards over time.

Explanation B

Winemakers graft the fruit-bearing parts of vinifera vines onto rootstocks of a grape species that is resistant to the parasite, and the vines continue to produce vinifera grapes.

Homelessness

Event to be explained

In the mid-90's, the rate of homelessness in the US suddenly skyrocketed. Before the sharp increase, the level of homelessness remained relatively constant. What might explain this sudden increase?

Explanation A

Because of an economic recession in the 90's, many businesses engaged in massive layoffs. Many people in the lower socioeconomic brackets who lost their jobs could no longer afford to pay rent or make mortgage payments.

Explanation B

In the 90's, the federal government reduced funding for outpatient services for people with mental illnesses. Many patients were left with little clinical supervision and stopped taking their medications, making it difficult to maintain a stable life.

Cities

Event to be explained

In several locations scattered throughout the world, archeologists have discovered prehistoric cities that had been abandoned very suddenly by entire populations. What might explain the sudden abandonment of these ancient cities?

Explanation A

The cities experienced sudden, major changes in climate, which forced inhabitants to move to locations that had better climates.

Explanation B

The cities were invaded by enemy groups that killed the populations and plundered the cities before they moved on.

Adoption

Event to be explained

In the process of adoption, hard-to-place children are usually characterized by behavior problems (such as aggression) and/or infant drug addiction from addicted birth mothers. These children are more likely to have mental and physical problems later in life. However, hard-to-place children who are adopted by same-sex couples tend to do better later in life than hard-to-place children adopted by opposite-sex couples. What is the explanation for this improvement?

Explanation A

Same-sex couples are more committed to being parents, so they are more likely to provide environments that allow their children to overcome the obstacles of their adoptions.

Explanation B

Because of the obstacles to adoption faced by same-sex couples, they are often older at the time of adoption. Thus, they are able to provide more stable lives for their adopted children.

DISTINGUISHING BETWEEN EXPLANATIONS 41

Using the platypus story as an example, protocols in each of the four conditions had the following formats:

No Explanation (NE) Condition

Event to be explained:

The platypus, an animal that lives in Australian waters, is a mammal that nevertheless lays eggs. In recent years, the platypus population has suffered a sharp decline. What might account for the decrease in the platypus population?

DISTINGUISHING BETWEEN EXPLANATIONS 43

If you were trying to explain this event, what sort of information would you like to have to help you explain it? Imagine you had a very efficient research assistant. What sorts of information would you like your assistant to collect?

DISTINGUISHING BETWEEN EXPLANATIONS 44

You've suggested several pieces of information that might help you explain this event.

When people make suggestions, they often suggest some things that will be very useful and some that will be less useful.

Please rate how useful you think each piece of information will be.

Please also tell me why you chose each rating.

1	2	3	4	5	6	7
Not at all			As Useful			Completely
Useful			As Not			Useful

DISTINGUISHING BETWEEN EXPLANATIONS 45

Self-Generated Explanation (SGE) Condition

Event to be explained:

The platypus, an animal that lives in Australian waters, is a mammal that nevertheless lays eggs. In recent years, the platypus population has suffered a sharp decline. What might account for the decrease in the platypus population?

What do you think the most plausible explanation is for this event?

DISTINGUISHING BETWEEN EXPLANATIONS 47

How would you rate how plausible this explanation is?

Why did you choose this rating?

1	2	3	4	5	6	7
Not at all plausible			As plausible as not			Completely plausible

DISTINGUISHING BETWEEN EXPLANATIONS 48

Once any explanation is proposed, it needs to be evaluated to see how likely it is to be correct. How would you evaluate this explanation? What sort of information would you like to have to help you decide whether it's likely to be correct? Imagine you had a very efficient research assistant. What sorts of information would you like your assistant to gather?

DISTINGUISHING BETWEEN EXPLANATIONS 49

You've suggested several pieces of information that might help you evaluate this explanation.

When people make suggestions, they often suggest some things that will be very useful and some that will be less useful.

Please rate how useful you think each piece of information will be.

Please also tell me why you chose each rating.

1	2	3	4	5	6	7
Not at all Useful			As Useful As Not			Completely Useful

DISTINGUISHING BETWEEN EXPLANATIONS 50

One Explanation Provided (1Exp) Condition

Event to be explained:

The platypus, an animal that lives in Australian waters, is a mammal that nevertheless lays eggs. In recent years, the platypus population has suffered a sharp decline. What might account for the decrease in the platypus population?

Possible explanation:

Environmental pollutants have introduced toxins into the water, killing many platypuses.

DISTINGUISHING BETWEEN EXPLANATIONS 52

Sometimes possible explanations for an event are plausible and sometimes they are not. Using this scale, please rate how plausible you think this explanation is.

Why did you choose this rating?

1	2	3	4	5	6	7
Not at all plausible			As plausible as not			Completely plausible

DISTINGUISHING BETWEEN EXPLANATIONS 53

Once any explanation is proposed, it needs to be evaluated to see how likely it is to be correct. How would you evaluate this explanation? What sort of information would you like to have to help you decide whether it's likely to be correct? Imagine you had a very efficient research assistant. What sorts of information would you like your assistant to gather?

DISTINGUISHING BETWEEN EXPLANATIONS 54

You've suggested several pieces of information that might help you evaluate this explanation.

When people make suggestions, they often suggest some things that will be very useful and some that will be less useful.

Please rate how useful you think each piece of information will be.

Please also tell me why you chose each rating.

1	2	3	4	5	6	7
Not at all Useful			As Useful As Not			Completely Useful

DISTINGUISHING BETWEEN EXPLANATIONS 55

Two Explanations Presented (2Exp) Condition

Event to be explained:

The platypus, an animal that lives in Australian waters, is a mammal that nevertheless lays eggs. In recent years, the platypus population has suffered a sharp decline. What might account for the decrease in the platypus population?

One possible explanation:

Environmental pollutants have introduced toxins into the water, killing many platypuses.

DISTINGUISHING BETWEEN EXPLANATIONS 57

Another possible explanation:

There has been an increase in the population of lizards that prey on platypus eggs.

How plausible do you find the second explanation?

Why did you choose this rating?

1	2	3	4	5	6	7
Not at all plausible			As plausible as not			Completely plausible

Once any explanation is proposed, it needs to be evaluated to see how likely it is to be correct. How would you evaluate the first explanation? What sort of information would you like to have to help you decide whether it's likely to be correct? Imagine you had a very efficient research assistant. What sorts of information would you like your assistant to gather?

DISTINGUISHING BETWEEN EXPLANATIONS 59

You've suggested several pieces of information that might help you evaluate this explanation.

When people make suggestions, they often suggest some things that will be very useful and some that will be less useful.

Please rate how useful you think each piece of information will be.

Please also tell me why you chose each rating.

1	2	3	4	5	6	7
Not at all Useful			As Useful As Not			Completely Useful

Appendix D

Coding and Entering Schematic

Column Name	Code	Instructions
Subject#	#	Subject Number
Gender	1	Female
	2	Male
Condition	#	Condition number (1, 2, 3, 4)
Targ AorB	1	Target is Explanation A
	2	Target is Explanation B
Story#	#	Story problem number, 1-6
T Plaus	#	Target plausibility rating
EmoRise	#	Y-axis valence on emotion grid
EmoRun	#	X-axis valence on emotion grid
A Plaus	#	Alternate plausibility rating
Evidence#	#	Order in which they state evidence
Usefulness	#	Evidence usefulness rating
Alternate ¹	0	Does not include alternative
	1	Genuine alternative
	2	Modification/combination
	3	Complimentary (general) alternative
	4	Explanation proposed becomes target (Condition 1)
CM	5	Includes A (Condition 4 only)
	0	Does not address causal mechanism
CorEvent	1	Addresses causal mechanism
	0	Not corollary event
CV Test ²	1	Involves corollary event
	0	Does not propose a contrast /covariation test
	1	Proposes contrast/covariation test of T
Distinguish ³	2	Proposes contrast/covariation test of any alternative
	0	Does not distinguish
	1	Distinguishes btw. T and a genuine alternative
	2	Distinguishes btw. T and a mod./comb. of T
	3	Distinguishes btw. T and a compliment of T
	4	No alternative specified (implied compliment)
	5	Distinguishes between T and A (Condition 4 only)

¹ “Genuine alternative,” “Modification/combination,” and “Includes A” were combined for this study as “Genuine alternative.”

² For this study, “Proposes a contrast/covariation test of T” and “Proposes a contrast/covariation test of any alternative” were combined into “Proposes a contrast/covariation test.”

³ “Distinguish btw. T and a genuine alternative,” “Distinguishes btw. T and a mod./comb. of T,” and “Distinguishes btw. T and A” were combined for this study into “Distinguishes between T and a genuine alternative.” “Distinguishes btw. T and a compliment of T” and “No alternative specified (implied compliment)” were combined for this study into “Distinguishes btw. T and a compliment of T.”

Author Note

I'd like to thank Dr. Barbara Koslowski for guiding and supporting me not only through the completion of this study but also in the course of my research in her lab; working with Dr. Koslowski for three years has been a defining experience at Cornell, and I could not have asked for a better advisor. I thank Dr. Marianella Casasola, who has been instrumental in keeping me calm and on track during the honors program process. Also, I greatly appreciate Dr. Koslowski, Dr. Casasola, and Dr. Matthew Belmonte for serving on my thesis defense committee and helping me develop as a researcher.

Of course, thank you to my parents for always urging me to challenge myself, and for recognizing my strengths before I did. Finally, thanks to Kimberly Rush, fellow research assistant and honors student, for collaborating with me on this colossal project and making monotonous randomization and coding sessions astoundingly entertaining.