Explaining December 4, 2015: Cognitive Science Ripped from the Headlines

Samuel G. B. Johnson
(samuel.johnson@yale.edu)
Department of Psychology, Yale University, New Haven, CT 06520 USA

Abstract

Do the discoveries of cognitive science generalize beyond artificial lab experiments? Or do they have little hope of helping us to understand real-world events? Fretting on this question, I bought a copy of the *Wall Street Journal* and found that the three front page headlines each connect to my own research on explanatory reasoning. I report tests of the phenomena of *inferred evidence*, *belief digitization*, and *revealed truth* in real-world contexts derived from the headlines. If my own corner of cognitive science has such explanatory relevance to the real world, then cognitive science as a whole must be in far better shape yet.

Keywords: Explanatory reasoning; ecological validity; everyday thinking; causal reasoning; theory of mind.

The Fretful Voice

Lately, I've been losing sleep.

Cognitive science excites us in part because it helps to explain broad swathes of human experience. Categories guide our stereotypes about social groups and our choices about which toothpaste to buy. Analogies help politicians to learn from history when deciding foreign policy and children to learn from examples when first encountering scientific ideas. Probability judgments determine our willingness to risk our lives and to play the lottery.

Yet, I suspect I am not the only cognitive scientist with a certain existential fear—a fretful inner voice that wonders whether our discoveries really have the generality I confidently boast to my undergrads. When a student asks about the significance of some principle of naïve physics, we can easily point to implications for science education—an important domain to be sure, but one almost custom-tailored to the scientific findings. When justifying the importance of attention research, the applications for traffic safety stand out as critical—but to what extent are various discoveries about attention generalizable across everyday experience, beyond cherry-picked case studies? It is not difficult to find real-world examples explained by cognitive theories, yet one wonders at the degrees of freedom.

I do not believe that scientific research must have direct practical implications, nor do I deny that theory-driven research can reveal genuine scientific truths. But the world is filled with truths: Isn't it our job to find the important ones—the ones that are both deep and general? To use Dennett's (2006) example, there is something undeniably elegant about the game of chess and the results in mathematics and computer science that it has inspired. But what about the (made-up) game of *chmess*, where the king moves two squares instead of one? There are just as many facts to discover about chmess as there are about chess—and they are just as true—yet chmess problems have an air of triviality that chess problems do not suffer. The insecure voice asks: Is my research more like *chess*, or more like *chmess*? Dennett quotes Donald Hebb: "If it isn't worth doing, it isn't worth doing well."

One Friday morning, I listened to the voice. I walked down Whalley Ave. to a supermarket, where I bought a copy of the only available national newspaper—the *Wall Street Journal*. The date was December 4, 2015. I was to give a talk the following week on three lines of research, each on a phenomenon of explanatory reasoning, aiming to use real-life examples from the paper to illustrate each part of the talk—to convince my audience (and myself) that my research resembles chess rather than chmess. The front page featured three principal headlines (one on a shooting, one on a central bank decision, one on military policy). Hence, there were no degrees of freedom in choosing headlines. This paper reports tests of these three phenomena in the context of these real-world events.

Explanatory Logic in Everyday Thinking

Our mental experiences consist largely of understanding observed data in terms of unobserved explanations. We make sense of events in terms of causes, features in terms of categories, behavior in terms of mental states, and retinal data in terms of 3-D organizations of the world.

To what extent do these explanatory inferences qualify as a natural kind? Do they merely share a common informational structure, or does the mind use similar mechanisms for solving these inference problems across very different types of psychological processes? I have argued that the very same mechanisms apply across these processes, via a set of heuristics I refer to as *explanatory logic*. For instance, people use an explanation's simplicity to estimate its probability, in a manner that is similar across causal reasoning (Lombrozo, 2007), categorization (Johnson, Kim, & Keil, 2016), and some visual tasks (Johnson, Jin, & Keil, 2014). Similar empirical cases have been made for several other explanatory strategies (e.g., Johnson, Merchant, & Keil, 2015; Johnson, Rajeev-Kumar, & Keil, 2014, 2015a, 2015b; Murphy & Ross, 1994; Sussman, Khemlani, & Oppenheimer, 2014).

However, if these strategies are really so general across cognition, they should also show up in everyday behavior. Is explanation not a dominant theme in our mental lives? The current studies used newspaper headlines to generate stimuli to demonstrate the wide applicability and ecological validity of three of these explanatory strategies.

General Method

Participants (N = 299) were recruited and compensated via Amazon Mechanical Turk. Participants completed three experiments in a random order, and were randomly assigned to one of three between-subjects conditions for each experiment (see Methods below). Afterwards, participants answered 12 true/false check questions. Participants incorrectly answering 33% or more of these questions were excluded from analysis (N = 9).

"California Shooters Leave Clues, but No Clear Motive"

The banner headline referred to the shooting in San Bernardino—an event that had occurred two days earlier. It was unclear at the time whether the motive was terrorism (as ultimately proved true) or an interpersonal feud. The available "clues" were stockpiles of weapons, which would be equally consistent with either motive. It would be more helpful to know whether a terrorist organization would claim responsibility (likely under the terrorism explanation, but unlikely under the interpersonal explanation); however, at the time of printing, it was too early to know. How do people think about such potentially diagnostic information when it is unavailable?

It turns out that people try to 'fill in' such information, using erroneous strategies to do so—a tendency known as the *inferred evidence heuristic* (Johnson, Rajeev-Kumar, & Keil, 2014, 2015a). People use the base rate of the *evidence* to infer whether the evidence would likely be observed, if available, even if the prior probabilities of the *hypotheses* are known, leading people to make illusory inferences (Khemlani, Sussman, & Oppenheimer, 2011). This is essentially the opposite of base rate neglect (Kahneman & Tversky, 1973)—people use *irrelevant* base rates that should be ignored.

For example, in one experiment with artificial stimuli (Khemlani et al., 2011), participants were told that magic spell *A* led to symptoms such as lumps, spots, and bumps, whereas spell *B* led only to lumps and spots. Given that Daryl has lumps and spots, but that it is unknown whether Daryl has bumps, participants believed that spell *A* was likelier. Subsequent work revealed that this bias occurs because people know that *most* people do not have bumps, and reason erroneously that Daryl must not have bumps either. This strategy explains the bias toward explanations making fewer predictions, over-and-above mechanisms such as biased prior probabilities, beliefs about the non-independence of evidence, and pragmatic inference (Johnson, Rajeev-Kumar, & Keil, 2015a).

In the shooting case, let's suppose that investigators have narrowed down the motive to two possibilities (terrorism or interpersonal feud), which have equal prior probabilities. People may nonetheless try to guess what percentage of shootings have responsibility claimed by a terrorist organization (an irrelevant piece of information once the prior of each hypothesis is known). This number is small (say, 10%), so participants might reason that there is a small chance that responsibility would be claimed in the San Bernardino case. If people then hold that *inferred* negative evidence against the terrorism motive, people would infer that the interpersonal motive is more probable than the terrorism motive—incorrectly, because this inference contradicts the prior probabilities without any new information.

To test this prediction and mechanism, participants were oriented to an anonymized version of the case:

> Imagine that a shooting occurred in the United States. Investigators have narrowed the suspect's motivation down to two possible motivations. Suppose that each motivation accounts for <u>about 2%</u> of shootings in the United States:
>
> The motivation could have been <u>interpersonal problems</u> between the suspect and one of the victims. In such cases, weapons stockpiles are typical.
>
> The motivation could have been <u>terrorist intentions</u>. In such cases, weapons stockpiles are typical, and a terrorist organization usually claims responsibility.
>
> The suspect had stockpiled weapons, but it is too early to tell whether any terrorist organization will claim responsibility.

In the *Neutral* condition, participants were not given a base rate for terrorist organizations claiming responsibility. We would expect these participants to use their tacit base rate, which would be low, and therefore to think the interpersonal motive is more likely. In the *Low Base Rate* condition, participants were explicitly given a low base rate:

> Suppose that for the vast majority of shootings, no terrorist organization claims responsibility.

Conversely, in the *High Base Rate* condition, participants were explicitly given a high base rate:

> Suppose that for the vast majority of shootings, a terrorist organization claims responsibility (regardless of whether or not they are actually responsible).

This parenthetical remark was included only in the latter condition, so that the effect base rate did not contradict the cause base rates given earlier in the problem (in the High condition), but also did not introduce a pragmatic violation (in the Low condition). If the mechanism at work here is inferred evidence, then we would expect participants to favor the interpersonal motive *more strongly* in the Low condition, and *less strongly* (or even favor the terrorism motive) in the High condition.

Method

Participants read either the text of the Neutral, Low Base Rate, or High Base Rate condition (see above). After reading this information, participants were asked "Which explanation do you think is most probable in this case?" Responses could range from -5 ("Very likely interpersonal") to 5 ("Very likely terrorism"). The order of the two explanations was randomized, and the orientation of the scale was adjusted to match this order.

On a separate page, participants in the Neutral condition were asked to report their tacit base rate: "Of all the shootings in the United States, for what percent do you think a terrorist organization claims responsibility?"

Results and Discussion

Scales were coded so that negative scores show a preference for the explanation making fewer predictions (interpersonal feud) and positive scores favor the explanation making more predictions (terrorism).

In the Neutral condition, participants preferred the interpersonal explanation [M = -0.20, SD = 0.72; $t(95)$ = 2.73, p = .008]. This prediction was predicated on participants having tacit base rates of less than 50% for terrorists claiming responsibility for shootings: Indeed, participants reported a mean 11.1% (SD = 17.1%) base rate. This base rate is normatively irrelevant, because the prior probabilities of the motives were set as equal (2%).

Further, participants strongly preferred the interpersonal motive in the Low Base Rate condition [M = -0.37, SD = 0.77; $t(96)$ = 4.74, p < .001], but not in the High Base Rate condition [M = 0.02, SD = 0.74; $t(86)$ = 0.32, p = .75]—a significant difference [$t(182)$ = 3.64, p < .001].

These results show that inferred evidence mechanisms apply not only to artificial stimuli, but also to realistic stimuli "ripped from the headlines." In addition, insofar as participants were inferring the mental states of the San Bernardino shooters, this finding suggests that people may use explanatory heuristics, such as inferred evidence, in mentalizing. Future research should address this question more fully (but see Johnson & Rips, 2014 for other explanatory heuristics used in mentalizing).

One initially surprising aspect of these results is that participants did not prefer the terrorism explanation in the High Base Rate condition. However, this is consistent with other findings in the literature (Johnson, Rajeev-Kumar, & Keil, 2015a). People's dislike of explanations making unverified predictions is multiply determined, and several other mechanisms make it difficult (though not impossible) to find a preference favoring explanations that do make such predictions.

"ECB Move Crushes Hopeful Markets"

The previous day, there had been a downturn in European markets because the European Central Bank (ECB) had not increased quantitative easing (QE), an inflationary monetary policy, as much as markets had anticipated.

Although seemingly of a very different flavor from the San Bernardino headline, the ECB story also involves an explanatory inference. Investors made inferences about the ECB's intentions based on statements from the ECB chairman. Such explanatory inferences must necessarily be uncertain (interpreting central bank statements relies on many of the same skills as tea-leaf reading). Normatively, then, this uncertainty about the correct interpretation of ECB statements should also propagate to any predictions made on the basis of such inferences.

It turns out, however, that people often *digitize* their beliefs reached through diagnostic reasoning (Johnson, Merchant, & Keil, 2015; Murphy & Ross, 1994). That is, even though people are happy to say that (for example) there is a 60% chance that an object is a skunk or that there is an 80% chance of rain, people do not treat these propositions as having graded truth; instead, they treat them as though they are certainly true or certainly false, when making inferences based on these propositions. Thus, when judging the implications of uncertain evidence (e.g., a very skunk-like and somewhat rabbit-like object), people treat the evidence as pointing to an explanation with certainty (e.g., treating it as though it is certainly a skunk) when thinking about the explanation's implications (e.g., judging whether it is likely to smell).

This tendency could partly explain why markets often react strongly to disconfirmed expectations—if the expectations are formed based on uncertain information treated as certain, the market would be overconfident. For example, suppose the bank's cryptic statement indicates a 70% chance of an aggressive monetary policy and a 30% chance of a modest monetary policy. Suppose further than there is an 80% chance of a major QE expansion, conditional on aggressive intentions, but a 20% chance of major QE expansion, conditional on modest intentions. Then, the probability of a major QE expansion is 0.8*0.7 + 0.2*0.3 = 0.62. But suppose that instead of treating the central bank's intention as uncertain, investors instead treated it as definite—then the probability of a major QE expansion would be 0.8*1 + 0.2*0 = 0.80. Hence, a failure of QE expansion would be more surprising given the 'digital' computation, leading to a bigger adjustment.

Although an experimental study cannot determine what was going through the minds of European investors this past December, the current study tested whether belief digitization occurs in stimuli relevant to such situations.

Method

The method was based on Johnson, Merchant, and Keil (2015, Exp. 2). Participants were assigned to either the *high/low*, the *low/low*, or the *low/high* condition. In the *high/low* condition, the good explanation (aggressive monetary policy) led to an event with high probability and the bad explanation (modest monetary policy) led to an event (introducing a ZT initiative) with low probability:

> Imagine that the central bank of the United States is deciding what policies to adopt.
>
> If they intend to adopt an aggressive monetary policy, they are *likely* to introduce a ZT initiative.
>
> If they intend to adopt a modest monetary policy, they are *unlikely* to introduce a ZT initiative.
>
> Suppose that the central bank chair says that the bank is concerned about the economy and considering a more aggressive monetary policy.

This last statement was intended to lead participants to think that an aggressive policy was more likely than a modest policy—an aggressive intention would be a better explanation for such a statement than a modest intention.

The *low/low* and *low/high* conditions differed only in the conditional probability of a ZT initiative given each explanation. In the *low/low* condition, the bank was *unlikely* to introduce a ZT initiative under either explanation, and in the *low/high* condition, the bank was *unlikely* to introduce a ZT initiative under an aggressive monetary policy but *likely* to do so under a modest monetary policy. The unfamiliar term "ZT initiative" was used in place of QE in order to make the three conditions equally plausible. The order of listing the good and bad explanations was randomized for each participant.

Participants were then asked a *diagnosis* question and a *prediction* question (in that order, on separate pages).

First, the *diagnosis* question asked "What do you think are the central bank's intentions?" Ratings were made independently for the options "Bank intends to adopt an aggressive monetary policy" and "Bank intends to adopt a modest monetary policy" as percentages. This question was intended to encourage participants to use graded beliefs (working against our hypothesis).

Second, the key dependent measure—the *prediction* question—asked "What do you think is the probability that the bank will introduce a ZT initiative?"

Results and Discussion

First, the results of the diagnosis question indicated that participants thought that an aggressive policy was most probable [M = 73.9%, SD = 18.9%]. However, a modest policy was nonetheless assigned a reasonably high probability [M = 28.4%, SD = 20.3%]. Thus, a failure to account for the low probability explanation could not be due to the explanation having extremely low probability.

As predicted, participants 'digitized', ignoring the low probability explanation when making predictions. There was a large difference between the high/low and low/low conditions [M = 75.4%, SD = 13.8% vs. M = 32.5%, SD = 30.7%; $t(191) = 12.89$, $p < .001$]; that is, participants changed their predictions based on the conditional probability of a ZT initiative, given the high-probability explanation. However, there was no difference at all between the low/high and the low/low conditions [M = 34.3%, SD = 26.8% vs. M = 32.5%, SD = 30.7%; $t(181) = 0.42$, $p < .001$]. Thus, participants did not change their predictions based on the conditional probability of a ZT initiative, given the low-probability explanation. This shows that participants were making inferences as though the high-probability explanation were certainly true.

As in the case of inferred evidence, these results affirm the digitization effect previously found using more artificial stimuli. And also like the inferred evidence case, the context (reading the intention of a central banker) involved mental-state inference. Future work might explore digitization effects more fully in mentalizing.

"U.S. Opening All Military Combat Roles to Women"

The final story concerned a new development in the U.S. military. The military ended a longtime policy of barring women from some combat roles, due to new evidence that women and men were equally capable in these roles.

Once again, this situation involves explanatory inference, and potentially relies on a heuristic studied in previous research (Johnson, Rajeev-Kumar, & Keil, 2015b). Our decisions depend on both the utilities of potential outcomes and our beliefs about those outcomes, which are often reached through inference (Jeffrey, 1965; Johnson, Zhang, & Keil, 2016). Sometimes situations are ambiguous, but it is nonetheless prudent to act as though a "high-stakes" hypothesis were true even if it is uncertain. In such situations, people are subject to a *revealed truth* bias—they not only *act* as though such high-stakes hypotheses are true, but they come to *believe* that they are true. When the evidence is neutral, but one action is more prudent than another, people tend to believe the corresponding hypothesis is likelier to be true. Similarly, evidence favoring a more prudent action is seen as more diagnostic than evidence favoring a less prudent action.

In the current case, the military no doubt believes that it is more problematic to make a Type II error (allowing women to serve when in fact women are less able than men) than a Type I error (forbidding women to serve when in fact women are equally able). In the former "high-stakes" case, there is a potential risk of fatalities, whereas in the latter "low-stakes" case, the risks are more intangible (e.g., discrimination, inefficiency). We would thus expect that if the military waited such a long time to open these roles up to women, it is because they made this trade-off and required overwhelming evidence that they were not making a Type II error in order to allow women to serve in these roles. Would this tendency toward conservative action—acting *as if* the hypothesis were true, that women were less capable in these roles—also make people think that women really *were* less capable?

Method

Participants were assigned to the *Neutral*, the *High-Stakes*, or the *Low-Stakes* condition. All participants were told about a disagreement between two think tanks about the abilities of a particular social group to serve in combat roles. One favors the high-stakes explanation (an error would involve fatalities) and one favors the low-stakes explanation (an error would involve inconvenience):

> One think tank argues that the members of this group are less able to engage in combat and should not be allowed to serve in combat roles. They argue that there will be a serious risk of combat fatalities if they are allowed to serve in combat.

> One think tank argues that the members of this group are equally able to engage in combat and should be allowed to serve in combat roles. They argue that there will be a minor inconvenience to this group if

they are not allowed to serve in combat.

In past cases where the think tanks have disagreed on similar issues, the two think tanks have each been proven right about half the time by objective measures.

This last statement was included to equate the prior probabilities of each explanation. The order of listing the explanations was randomized for each participant, and the orientation of each scale was adjusted to match this order.

Next, participants were given evidence concerning this group's combat abilities. In the *Neutral* condition, the evidence was ambiguous between the two explanations:

In this particular case, the evidence is unclear as to which think tank's view is right.

In the *High-Stakes* condition, the evidence favored the high-stakes explanation:

In this particular case, the evidence favors the view that members of this group are less able to engage in combat.

In the *Low-Stakes* condition, the evidence favored the low-stakes explanation:

In this particular case, the evidence favors the view that members of this group are equally able to engage in combat.

Participants then completed measures of *action* and *belief*. For the *action* question, participants were asked "Would you allow members of this social group to engage in combat?" on a scale from -5 ("Definitely no") to 5 ("Definitely yes"). For the *belief* question, participants were asked "Do you think members of this group are equally able to engage in combat or less able to engage in combat?" on a scale from -5 ("Definitely less able") to 5 ("Definitely equally able"). Thus, negative scores correspond to the high-stakes explanation and positive scores correspond to the low-stakes explanation. The order of these two questions was counterbalanced.

Results and Discussion

In the Neutral condition, participants should favor the high-stakes option in choice (i.e., not allowing women to serve), even though the evidence is ambiguous and favors neither hypothesis. Indeed, our predictions about the belief question are predicated on this assumption about the choice question. Unfortunately, this manipulation check failed: Participants were *more* likely to allow members of the group to serve, even when the evidence was ambiguous [M = 0.56, SD = 1.98; $t(97)$ = 2.80, p = .006]. In retrospect, it makes sense that many participants would not share the military's priorities, and would view the (certain) social costs of forbidding a social group from participating in the military as potentially more serious than the (potential) risk of combat fatalities. However, the large variance reveals that there were considerable individual differences in their action choices. Thus, although the High Stakes and Low Stakes conditions are uninterpretable because we cannot determine which participants viewed the Type I or Type II error risk as greater, we can analyze the Neutral condition by splitting the sample into those who chose to intervene as though the high-stakes or low-stakes explanation were true.

We first looked at participants whose choices matched our assumptions (N = 32), favoring the high-stakes over the low-stakes explanation in their actions [M = -1.78, SD = 1.23; $t(31)$ = -8.19, p < .001]. Even though the evidence was ambiguous between the two hypotheses, these participants nonetheless believed the high-stakes explanation was (marginally) more likely to be *true* [M = -0.63, SD = 1.90; $t(31)$ = -1.96, p = .059]. Thus, these participants seem to have used their decisions to infer the truth, even though their decisions were the result of prudential, rather than probabilistic, thinking.

The story is similar for those participants whose choices were opposite to our assumptions (N = 55), favoring the low-stakes explanation in their actions [M = 2.23, SD = 1.32; $t(54)$ = 12.52, p < .001]. These participants also believed the low-stakes explanation more likely to be true [M = 1.39, SD = 1.77; $t(54)$ = 5.82, p < .001].

These results support the idea that revealed truth is at work in everyday situations such as those covered by the newspaper. A shortcoming of this study was the failure of the manipulation to induce participants to consistently favor one course of action due to prudential concerns—participants appear to differ in which kind of error they deem more problematic. Hence, future research should look at naturalistic cases where the prudential concerns are more clear-cut. (Of course, the original revealed truth effect was found using artificial stimuli where prudential concerns were clear, in order to avoid this problem.)

The individual difference analysis helps to buttress our account, but has two limitations. First, since participants were selected based on their responses to the action question, it is possible that some participants ignored our insistence that the prior probabilities of the hypotheses were equal, and then based their action choices on their own antecedent beliefs. This seems unlikely given the magnitude of the effects (much more extreme for the action question than for the belief question, consistent with previous findings; Johnson, Rajeev-Kumar, & Keil, 2015b), but cannot be ruled out entirely. Second, it was not possible to test the asymmetry in evidence diagnosticity (intended to be tested with the Low-Stakes and High-Stakes conditions). These limitations should be addressed in future work with other naturalistic stimuli.

Despite these limitations, this study is encouraging for the generality of the revealed truth hypothesis. This is so not merely because the results as consistent with that hypothesis as they could be (given the failed manipulation check), but because the sort of *situation* in which the revealed truth phenomenon occurs was highlighted on the front page of the Wall Street Journal—on an arbitrary day. If situations are ecologically frequent where beliefs can be informed by choices, then the laboratory findings are likely to generalize to many real-world situations.

General Discussion

In the trenches, we forget how ubiquitous the principles of cognitive science really are. For me, this project has been an encouraging confirmation of the ecological validity of at least one corner of cognitive science research.

In previous work, I've looked at whether there are a common set of cognitive mechanisms—*explanatory logic*—at play in various diagnostic reasoning tasks, such as causal inference and categorization. To draw theoretically strong inferences requires high internal validity, so that these studies often rely on stimuli that are isolated from participants' background knowledge, such as fake diseases and magical transformations.

The current results show that people also use the same mechanisms to contemplate issues found in front page headlines. Inferences about criminal cases depend on both observed and *inferred evidence*—using irrelevant base rates to fill gaps in knowledge. Predictions in economic contexts involve *belief digitization*—treating uncertain propositions as being a sure thing. And beliefs about the capabilities of social groups may turn on *revealed truth*—choosing based on the riskiness of the options, and using that choice to infer what the truth must have been.

I do not claim that these results tell us how often these patterns of inference arise naturally in day-to-day life. Instead, this exercise demonstrates that (1) diagnostic reasoning problems are common in one naturalistic corpus; and (2) the same reasoning mechanisms found in artificial contexts apply to these types of natural problems. Future research might measure spontaneous explanatory behavior directly, to better estimate the frequency of such fallacious thinking (see Weiner, 1985, for a related effort in the domain of attribution theory).

Empirical studies often involve a trade-off between internal and external validity. Whereas cognitive science approaches (including investigations of explanatory reasoning) typically aim to maximize internal validity at virtually any cost, the current work plots a new point on the trade-off curve, increasing external validity at the expense of some experimental control. Nonetheless, I did draw some lines in the sand: I insisted on an experimental approach, where very similar stimuli could be tested across all conditions. This necessarily meant some artificiality in isolating real-world knowledge from these effects, in order to be sure that causal conclusions could be drawn from the results. Future investigations might swing even further toward external validity, perhaps using a larger variety of items drawn from real corpuses (such as newspaper articles), where the theoretically relevant dimensions (such as effect base rates) naturally vary.

If one research program has this degree of real-world relevance, I am far more hopeful for our science as a whole. This conference features hundreds of talks and posters, each reporting a discovery. This project has reinvigorated my hope that many of these discoveries can contribute toward our understanding of cognition in a broad sense. It's a relief—now I can sleep again.

Acknowledgments

I thank Frank Keil and an audience at the University of Chicago, Booth School of Business for helpful discussion. I dedicate this work to the students who keep us honest.

References

Dennett, D.C. (2006). Higher-order truths about chmess. *Topoi*, *25*, 39–41.

Jeffrey, R.C. (1965). *The logic of decision*. New York, NY: McGraw-Hill.

Johnson, S.G.B., Jin, A., & Keil, F.C. (2014). Simplicity and goodness-of-fit in explanation: The case of intuitive curve-fitting. *Proceedings of the 36th Conference of the Cognitive Science Society*.

Johnson, S.G.B., Kim, H.S., & Keil, F.C. (2016). Explanatory biases in social categorization. *Proceedings of the 38th Conference of the Cognitive Science Society*.

Johnson, S.G.B., Merchant, T., & Keil, F.C. (2015). Predictions from uncertain beliefs. *Proceedings of the 37th Conference of the Cognitive Science Society*.

Johnson, S.G.B., Rajeev-Kumar, G., & Keil, F.C. (2015). Inferred evidence in latent scope explanations. *Proceedings of the 36th Conference of the Cognitive Science Society*.

Johnson, S.G.B., Rajeev-Kumar, G., & Keil, F.C. (2015a). *Sense-making under ignorance*. Under review.

Johnson, S.G.B., Rajeev-Kumar, G., & Keil, F.C. (2015b). Belief utility as an explanatory virtue. *Proceedings of the 37th Conference of the Cognitive Science Society*.

Johnson, S.G.B., & Rips, L.J. (2014). Predicting behavior from the world: Naïve behaviorism in lay decision theory. *Proceedings of the 36th Conference of the Cognitive Science Society*.

Johnson, S.G.B., Zhang, M., & Keil, F.C. (2016). Decision-making and biases in causal-explanatory reasoning. *Proceedings of the 38th Conference of the Cognitive Science Society*.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–51.

Khemlani, S.S., Sussman, A.B., & Oppenheimer, D.M. (2011). *Harry Potter* and the sorcerer's scope: Latent scope biases in explanatory reasoning. *Memory & Cognition*, *39*, 527–35.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*, 232–57.

Murphy, G.L., & Ross, B.H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology*, *27*, 148–93.

Sussman, A.B., Khemlani, S.S., & Oppenheimer, D.M. (2014). Latent scope bias in categorization. *Journal of Experimental Social Psychology*, *52*, 1–8.

Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin*, *97*, 74–84.