Journal of Experimental Psychology:
Learning, Memory, and Cognition
2014, Vol. 40, No. 1, 261–274

© 2013 American Psychological Association
0278-7393/14/$12.00 DOI: 10.1037/a0033778

Communicating Uncertain Experimental Evidence

Alexander L. Davis and Baruch Fischhoff
Carnegie Mellon University

Four experiments examined when laypeople attribute unexpected experimental outcomes to error, in foresight and in hindsight, along with their judgments of whether the data should be published. Participants read vignettes describing hypothetical experiments, along with the result of the initial observation, considered as either a possibility (foresight) or a reality (hindsight). Experiment 1 found that the initial observation seemed more likely to be replicated when viewed in hindsight than in foresight. The remaining experiments contrasted responses to an initial observation from 1 of the 4 studies that was either expected or unexpected (based on the predictions of participants in Experiment 1). Experiments 2A–C and Experiment 3 found that unexpected results were more likely to be attributed to methodological problems than were expected ones—but to the same degree in foresight and in hindsight. Participants in Experiment 4 had more confidence in an explanation for an unexpected outcome when it was mentioned before that outcome was revealed than when it was suggested only after the surprise was known. In all the experiments, most participants recommended collecting more data before publishing the results, especially when they attributed the results to multiple causes. The results suggest that considering the causes of unexpected experimental results in foresight may improve the evaluation and communication of those results in hindsight.

Keywords: hindsight bias, confirmation bias, file-drawer problem, data sharing, causal reasoning

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Every experiment has the potential for unexpected results—otherwise it would not be worth conducting.[1] When surprises arise, scientists need to account for them, either by suggesting new theories or by raising questions about the soundness of its design—and the auxiliary assumptions needed to interpret its results (Lakatos, Worrall, & Currie, 1980). In psychology, such assumptions might include whether research participants understood the instructions and stimuli as intended, whether the experimental setup conveyed unintended clues or incentives, and whether mistakes were made during data entry or statistical analysis.

The need to make such inferences acknowledges that every study requires an assessment of construct validity, as researchers simultaneously evaluate their substantive theories and their methodological assumptions (Shadish, Cook, & Campbell, 2002). The weaker the empirical or theoretical support for those assumptions, the more the interpretation of unexpected results must rely on scientific judgment (Fischhoff, Welch, & Frederick, 1999). Surprising results should weaken researchers' confidence in their ability to identify whether problems lie in their theory, their methods, or both. Unless researchers realize that unexpected results raise questions about both theory and method, they risk allowing flawed methods to undermine valuable theories and claims of flawed methods to protect inaccurate theories from inconvenient results.

The history of physics provides a famous example of the former risk. While attempting to measure the charge of the electron, Nobel laureate R. A. Millikan discarded multiple unexpected observations, confidently attributing them to error in his experimental apparatus. Most of those instances occurred during an ambiguously defined "warm-up period," where he "gradually refined his apparatus and technique in order to make the best measurements" (Goodstein, 2000, p. 35). However, Millikan also rejected later (post-warm-up) observations where "there were no obvious experimental difficulties that could explain the anomaly," attributing the observations to nothing more explicit than "something wrong with the thermometer" (Franklin, 1997, p. 13). Later research found that Millikan's intuitions were generally right, even though he did not articulate reasons for them—and, indeed, could not have known why the anomalies occurred, given scientific knowledge at the time. (His apparatus was unreliable with charges greater than 30*e*.) Had Millikan pursued those problems, he would have delayed

This article was published Online First July 29, 2013.

Alexander L. Davis, Department of Social and Decision Sciences, Carnegie Mellon University; Baruch Fischhoff, Department of Engineering and Public Policy and Department of Social and Decision Sciences, Carnegie Mellon University.

We thank the National Science Foundation for the doctoral dissertation enhancement grant (SES-1061315), John Sperger and Terence Einhorn for help collecting the data and valuable feedback on the experiments, and Michael Gorman for his comments and suggestions. All materials and data, including reproducible statistical analyses using R and Sweave, can be obtained from the first author's Dataverse at http://hdl.handle.net/1902.1/14819. Open lab notebook can be obtained at OpenWetWare (http://openwetware.org/wiki/User:Alexander_L._Davis/Notebook/Error_Models_and_Data_Sharing_in_Hindsight).

Correspondence concerning this article should be addressed to Alexander L. Davis, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: alexander.l.davis1@gmail.com

[1] We thank the late Robyn Dawes for reminding us of this principle.

studies that made important contributions to physics, despite their flaws.

In a less happy physics example, Rene Blondlot's purported discovery of a new type of electromagnetic radiation, called n-rays, "touched off . . . a wave of self-deception that took years to subside" (Klotz, 1980, p. 170). His supporters included respected physicists who uncritically reported expected effects when they placed n-ray sources (e.g., gas burners, heated silver, sheet iron) in front of electric spark generators, while accusing scientists who failed to observe them of poor training (Wood, 1904). Thus, unexpected results, inconsistent with n-ray theory, were attributed to error, whereas expected ones were counted as support. To settle the controversy, the French physics community agreed on a definitive experiment and proposed it to Blondlot, whose refusal to participate eventually undermined faith in n-rays (Klotz, 1980; Mellers, Hertwig, & Kahneman, 2001).

Error Models

Surprising results sometimes increase an observer's need for better explanations, thereby prompting a deeper search for causes (Lau, 1984; Lau & Russell, 1980; Nisbett & Ross, 1980; Pyszczynski & Greenberg, 1981; Risen, Gilovich, & Dunning, 2007; Roese & Olson, 1996; Weiner, 1985; Wong & Weiner, 1981). One possible outcome of that search, seen in the Millikan and Blondlot examples, is *error model explanations,* which conclude that the experiment was flawed and the data should be treated as errors, thereby preserving the theory (Chinn & Brewer, 1993).[2] Such explanations have been studied under various names, including blaming the method (Dunbar, 2001), biased assimilation (Lord, Ross, & Lepper, 1979), confirmation bias (Klayman & Ha, 1989), and belief perseverance (Nickerson, 1998). In Millikan's case, error models kept science moving by "explaining away . . . odd results" in order to avoid having research "instantly degenerate into a wild-goose chase after imaginary fundamental novelties" (Michael Polanyi quoted by Gorman, 1992, p. 63). In the case of Blondlot and his supporters, error models immunized their hypotheses against valid challenges from disconfirming data (Gorman, 1989, 1992; Gorman, Tweney, Gooding, & Kincannon, 2005; Penner & Klahr, 1996).

Although researchers try to address all possible (or at least plausible) error models when designing an experiment, they must eventually decide that the experiment is good enough to conduct. Once this judgment is made, researchers may naturally set aside the uncertainty of not having considered every error model, so that their predictions are based on the theory they consider most probable (Murphy & Ross, 1994, 2010; B. Ross & Murphy, 1996). When expected results occur, this uncertainty is neglected because plausible error models never come to mind, perhaps even when writing the obligatory "limitations" section of their experimental write-up. However, when unexpected results do occur, researchers may search harder for error models. If that search is successful, they may feel as though the experimental result was not just an error but a predictable one, as would be expected from the finding that hindsight bias is greater for surprising results than for more expected ones (Fischhoff, 1975; Nestler, Blank, & von Collani, 2008; Pezzo, 2003).

The error model account is also consistent with the finding that, when asked why they do not try to publish null results, psychologists typically say that such results are more likely to be caused by flawed methods than are statistically significant ones (Greenwald, 1975). A survey of National Institutes of Health–funded scientists from various disciplines found that 15.3% reported having dropped observations based on a "gut" feeling of their being in error (Martinson, Anderson, & De Vries, 2005). An observer commented that physicists were "always doing experiments or making observations that disappoint them. They look for some phenomenon or relationship and they do not find it. Most of these negative experiments are forgotten and the results consigned to the file drawer" (Collins, 2003, pp. 661–662).

Thus, an *error model* account predicts that surprising results are more likely to be attributed to error than are expected ones, one consequence of which is making them seem less worthy of publication. That tendency represents a kind of *foresight bias* in causal attribution, whereby people neglect explanations for unexpected events until they occur. If so, researchers may benefit from making foresight more like hindsight, by producing the explanations of potential surprising results before they occur. Doing so may improve how studies are interpreted, including when researchers decide to share their results. It might even improve how studies are designed, if thinking about surprising results in foresight leads to eliminating possible sources of error.

Entropy

The error model account holds that surprising results evoke new explanations involving methodological problems. An alternative possibility is that surprising results make explanation seem less possible altogether, by creating the feeling that "anything can happen" (in the words of a participant in Experiment 2B below). Expressing ignorance this way follows a line of reasoning sometimes associated with Laplace, called the principle of insufficient reason, stating that all possibilities are equally likely absent information to the contrary (Falk & Lann, 2008). In behavioral research, this expression of ignorance may be seen in respondents' tendency to say 50% (in the sense of "fifty–fifty"), when asked for the probability that unfamiliar events will happen (Bruine de Bruin, Fischhoff, Millstein, & Halpern-Felsher, 2000; Fischhoff & Bruine de Bruin, 1999), treating the two states of the world (happen/not happen) as equally likely.

We propose that unexpected experimental results evoke such thinking. Researchers believe in a theory, then conduct an experiment whose expected results it will explain. When confronted with evidence to the contrary, they feel that anything is possible and settle for whatever error model accounts come to mind. Thus, Greenwald (1975) found that researchers feel that there are many ways to get null results but only a few ways to produce statistically significant ones, thereby reducing the value of surprises—and increasing the chances of discarding them. A similar analysis suggests that researchers who are uncertain about the causes of an experimental result will expect others to learn nothing from it (Hilton, 1990); in effect, projecting their feelings of ignorance onto others (Nickerson, 1999).

[2] Explanations invoke mechanisms (Keil, 2006). As randomness or chance is not a deterministic (Nestler et al., 2008; Wasserman, Lempert, & Hastie, 1991) or pseudo-indeterministic (Spirtes, Glymour, & Scheines, 2000) causal explanation, it is not an error model.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

We formalize this second account as an *entropy theory,* whereby an unexpected result increases uncertainty about its causes and, hence, uncertainty about its being replicated and the apparent appropriateness of sharing it. At the extreme, a surprising result may make all outcomes seem equally possible (i.e., a maximum entropy state) rather than make the expected one seem less likely.

We examine the entropy account in three ways. One is to elicit judgments of the predictability of future results, expecting higher entropy after unexpected ones (i.e., more uniform probabilities assigned to the potential outcomes). The second is to elicit explanations of past results, expecting less discrimination among possible explanations for unexpected ones. One of these explanations, that the results were due to chance, should be positively associated with the entropy of participants' predictions. Third, the higher the entropy that results evoke, the less people should recommend publishing them and the more they should require additional data before sharing the story of a study with readers.

The error model and entropy theory make complementary predictions about the effects of an unexpected result. According to the former, the substantive (nonerror model) cause leading to the expected result now seems less likely, with some of its probability now assigned to error model causes. According to the latter, a flatter distribution means that more probable causes now seem less likely and less probable causes now seem more likely, without distinguishing between substantive and error model causes.

Hindsight Bias

We examine these effects in both foresight and hindsight. Judgments of causality have been remarkably absent from the voluminous literature on hindsight bias (Blank, Musch, & Pohl, 2007; Christensen-Szalanski & Willham, 1991; Guilbault, Bryant, Brockway, & Posavac, 2004), even though the bias is typically explained in terms of individuals' propensity to make sense of reported outcomes (Blank & Nestler, 2007; Fischhoff, 1975; Nestler, Blank, & Egloff, 2010; Roese & Vohs, 2012). Experiment 1 replicates an earlier study, calibrating our stimuli. When designing the experiments that followed it, we originally predicted hindsight/foresight differences. However, in the course of conducting those experiments, as a product of our own error model thinking, we realized that our foresight and hindsight tasks differed from those used previously. Namely, we required explicit causal reasoning rather than leaving it implicit, thereby making the two perspectives more similar than they were in prior research—and might be under normal circumstances. Indeed, our growing appreciation of that forced similarity led to our positing the foresight bias addressed by Experiment 4.

Overview of Experiments

In all studies, participants received vignettes introduced by Slovic and Fischhoff (1977), describing experiments, along with an initial observation that might occur (foresight) or already had (hindsight). Depending on the experiment, participants performed some combination of the following tasks: (a) explaining that initial result, either by attributing it to specified causes or by completing an open-ended question; (b) predicting the results of a replication study; and (c) recommending publication of the results or additional data collection.

Experiment 1 repeated the original study, comparing predictions of replication in foresight and hindsight. The remaining experiments used one of the four hypothetical experiments, chosen because it had the greatest disparity between the probabilities of its possible outcomes—and, hence, had the most unexpected result. Experiments 2A, B, and C and Experiment 3 elicited judgments about possible causes of one of the two outcomes in hindsight or in foresight, as well as judgments about whether to publish the results. Experiment 2A elicited causal judgments with a structured set of possible causes. Experiment 2B refined the design of Experiment 2A with a revised set of possible causes, open-ended questions eliciting explanations, and additional measures of data sharing. Experiment 2C was a constructive replication of Experiment 2B, based on responses to its open-ended questions. Experiment 3 used Experiment 2C's design and manipulated the availability of alternative explanations, testing whether equal availability in the preceding experiments could account for the lack of hindsight/foresight differences in causal judgments. Experiment 4 tests whether explanations of unexpected results seem more credible when they are first considered before observing the experimental outcome or only afterward.

Experiment 1

Participants assessed the probability of replicating the initial observation for the four hypothetical studies presented in Experiment 1 of Slovic and Fischhoff (1977), using the identical stimulus materials. These studies described fictitious experiments testing whether (a) a virgin rat would exhibit maternal behavior following a blood transfusion from a mother rat, (b) seeding a hurricane with silver-iodide crystals would diminish its wind velocity, (c) goslings would be imprinted on a duck if exposed to its quacking before hatching, and (d) children could take another person's perspective when judging the position of a dot on a large Y.

For each scenario, foresight participants first judged the probability of two outcomes, such as A = the rat exhibited maternal behavior and B = the rat did not exhibit maternal behavior. They then judged the probability that each outcome would be replicated on all, some, or none of 10 additional observations, were it the initial observation. Participants in the hindsight condition were told that a specific outcome had occurred (either A or B); then, they assessed its probability of being replicated in all, some, or none of 10 additional observations. The design was 4 (study: rat, hurricane, duck, Y test) by 2 (time: foresight vs. hindsight) by 2 (outcome: A or B), with repeated measures on the first factor in all conditions and repeated measures on the last factor in the foresight condition, whose participants gave probabilities of replication for both outcomes.

Participants

All 268 participants were paid volunteers who responded to an Amazon Mechanical Turk (MTurk) ad offering them $1 for

participation in a 7-min study.[3] A two-part attention filter (Downs, Holbrook, Sheng, & Cranor, 2010; Oppenheimer, Meyvis, & Davidenko, 2009; Paolacci, Chandler, & Ipeirotis, 2010) at the beginning of the experiment assessed whether participants were paying attention. Only the 173 participants who passed both its parts (one easier, one harder) were included in the analysis. According to participants' reports, their average age was 32 years old (range = 18–81); 56.6% were women. Participants reported a variety of occupations, most frequently student, followed by homemaker, engineer, and miscellaneous ones such as accountants and caregivers, similar to the demographic diversity observed in previous MTurk research (Ipeirotis, 2010).

Results

For the first hypothetical study, foresight participants said that if the first virgin rat demonstrated maternal behavior after receiving a blood transfusion from a mother rat (Outcome A), then there was a (median) .20 chance of that happening on all 10 subsequent cases ($M = .28$). Hindsight participants told that the initial case had turned out that way gave a median probability of .50 for consistent replication ($M = .49$). The difference in medians was significant, $t(121) = 2.99$, $p < .01$, $d = 0.27$, and consistent with hindsight bias in predictions.[4] The corresponding means in Slovic and Fischhoff (1977) were .30 and .44, respectively. A similar hindsight/foresight contrast emerged when participants considered the possibility of Outcome B, the first virgin rat not demonstrating maternal behavior, $t(109) = 3.48$, $p < .01$, $d = 0.33$. As seen in Table 1, the same pattern held for the other three hypothetical studies.

Discussion

Slovic and Fischhoff (1977) found that people see the results of the first observation of a study as more likely to be replicated in hindsight than in foresight. In this exact replication of their Experiment 1 (except for subject population and administration mode), that result held true. In their Experiment 2, Slovic and Fischhoff found similar results when foresight participants considered only one of the two possible outcomes, rather than both (as in Experiment 1), indicating that their lower confidence in replication was not due to focusing less on each outcome.

These results are consistent with a stronger tendency in hindsight than in foresight to generate causal explanations only for the reported outcome and then find replication more likely given those explanations. Experiments 2A, B, and C ask participants to make such explanations explicit, predict future outcomes, and make recommendations regarding publication and further data collection. In order to focus participants' attention, we use a single hypothetical experiment, the Y test, which produced the most and least expected outcomes (A and B, respectively), as seen in Table 2's summary of the probabilities assigned to replication in foresight.

Table 1
Hindsight Bias Effects for the Four Hypothetical Studies

Study	Outcome A	Outcome B
Virgin rat	$t(121) = 2.99$, $d = 0.27$	$t(109) = 3.48$, $d = 0.33$
Hurricane	$t(121) = 2.33$, $d = 0.21$	$t(109) = 2.10$, $d = 0.20$
Gosling	$t(121) = 2.76$, $d = 0.25$	$t(109) = 4.56$, $d = 0.44$
Y test	$t(121) = 2.63$, $d = 0.24$	$t(109) = 4.57$, $d = 0.44$

Note. t tests compare estimated median probabilities assigned to the observed outcome replicating in all subsequent observations between those in the foresight and hindsight groups.

Experiments 2A, B, and C

Experiments 2A, B, and C contrast judgments of expected and unexpected observations, in foresight and hindsight, using variations of Experiment 1's methodology. The critical difference between these experiments and Experiment 1 is that participants were asked to evaluate possible causes of an outcome, not its probability of replication. All three experiments have a 2 (foresight vs. hindsight) by 2 (expected vs. unexpected outcome) between-subjects design. On the basis of the results of Experiment 1 (and Slovic & Fischhoff, 1977), we treat Area A as the expected outcome and Area B as the unexpected one. Note that these conditions differ from the hindsight and foresight conditions in Experiment 1 and previous research, in that they present possible explanations for each outcome rather than rely on participants to produce explanations on their own.

Participants

Participants were paid volunteers who responded to an Amazon MTurk ad offering them $1 for participation in a 7-min experiment. For Experiment 2A, 468 of 664 individuals (70%) passed both attention filters. Their average age was 31 years old (range = 18–63); 50% were women. For Experiment 2B, 359 of 448 individuals (80%) passed both attention filters. Their average age was 32 years old (range = 18–68); 151 were women (44%). For Experiment 2C, 312 of 465 individuals (67%) passed both attention filters. Their average age was 31 years old (range = 18–67); 135 were women (43%).

Experiment 2A

Experiment 2A replicated Experiment 1, with three differences: (a) Participants considered just one hypothetical study, the Y test

[3] Horton, Rand, and Zeckhauser (2011) found that MTurk participants replicated results from several classic judgments studies originally conducted with traditional (e.g., student) samples. MTurk participants are demographically diverse compared to typical Internet or college undergraduate samples (Buhrmester, Kwang, & Gosling, 2011) and tend to be similar in gender, but older and more likely to be non-White and non-American, than those in typical Internet samples. When given validated psychometric measures, MTurk participants give responses that are internally reliable, and this reliability does not vary significantly by payment amount (Buhrmester et al., 2011). Mason and Watts (2010) found that, when paid more, MTurk participants work longer but do not perform better (in terms of accuracy).

[4] Throughout the paper we use median regression (Koenker, 2009; Wooldridge, 2009) with nonparametric bootstrap to estimate standard errors (Efron & Tibshirani, 1993) for causal attributions and predictions, and linear probability models (least squares with a binary dependent variable) for discrete data-sharing judgments to publish or make a different recommendation (collect additional data or not publish).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Mean Foresight Probability and Fraction of Responses With Probabilities Greater Than .5 for Outcomes A and B

Study	P(A) *M* (*SD*)	P(B) *M* (*SD*)	One-sample *t* test	P(A) > .50	P(B) > .50
Virgin rat	.40 (.25)	.59 (.26)	$t(60) = 2.66$, $p = .01$	14/61	32/61
Hurricane	.52 (.27)	.40 (.26)	$t(60) = 2.90$, $p < .01$	29/61	12/61
Gosling	.51 (.28)	.52 (.27)	$t(60) = 0.65$, $p = .52$	28/61	25/61
Y test	.65 (.28)	.24 (.22)	$t(60) = 4.14$, $p < .01$	39/61	5/61

Note. P(A) is the average foresight probability assigned to Outcome A. P(B) is the average foresight probability assigned to Outcome B. One-sample *t* test compares P(A) to .50. P(A) > .50 indicates the number of participants who stated that Outcome A has higher than .50 foresight probability, and P(B) > .50 indicates the number of participants who stated that Outcome B has higher than .50 foresight probability.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

(see Figure 1).[5] (b) Participants assessed the probability that each of four causes accounted for the result, one of which was that the experiment was flawed. (c) Participants did not assess the probability of replicating the initial observation. The error model theory predicts that participants will invoke error more for unexpected results than for expected ones. The entropy theory predicts a more uniform distribution of probabilities assigned to the four causes for the unexpected outcome than for the expected one.

Materials

In Experiment 2A all participants received the same introductory instructions as in Slovic and Fischhoff (1977), followed by their description of the Y-test study:

> In the pretest of an experiment that she intends to run in the future, an experimenter will place a 4-year-old child in front of an easel with a large Y on it, with a dot in the lower left-hand third of the letter. The child will then be taken around to the back of the easel where he will see another Y. He will be asked to draw a dot in the "same position" on that Y as the one he had just seen.
>
> The possible outcomes are (a) the child places a dot in Area A (the lower left-hand third), (b) the child places a dot in Area B (the upper third), or (c) the child places a dot in Area C (the lower right-hand third).

Participants then completed the causal attribution task shown below, using Area A and foresight condition as an example. The parentheses contain our term for each cause.

> If the child places a dot in Area A, what is the probability that:
>
> (Note: These four probabilities should total 100%.)
>
> • (Valid) The child's understanding of the experimenter's instructions caused the child to place the dot in Area A.
>
> • (Invalid) Some error in the experiment caused the child to place the dot in Area A.
>
> • (Chance) Random chance caused the child to place the dot in Area A.
>
> • (Other) There was some other cause not already mentioned above.

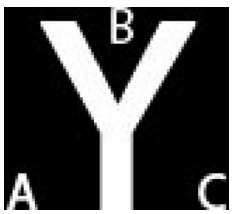

Figure 1. Image of Y.

Hindsight. As in Experiment 1, the instructions for hindsight participants differed in reporting the first observation:

> Result: The child placed a dot in Area A (the lower left-hand third).

Results

Table 3 shows median probabilities assigned to the causal explanations for Experiments 2A, B, and C.

Error model theory. As predicted, participants assigned a higher probability to the error explanation (Invalid) for the unexpected observation than for the expected one, although that main effect was not statistically significant, $t(464) = 1.68$, $p = .09$, $d = 0.08$. We also found support for the null hypothesis that attributions would be similar in foresight and hindsight, with no interaction, $t(464) = 0.00$ and $t(464) = 0.00$, for both the main effect and the interaction.

Shannon entropy. We quantify entropy as Shannon entropy (MacKay, 2003), which measures the diffuseness (spread) of a probability distribution:

$$\text{Shannon entropy}\,(A, B, C, D) = -P(A) \times \log_2[P(A)] - P(B) \times \log_2[P(B)] - P(C) \times \log_2[P(C)] - P(D) \times \log_2[P(D)],$$

where P(A) is the probability assigned to outcome (or cause) A, and so on. With four response categories, the measure ranges from 0 (all in one category; hence no uncertainty) to 2 (a uniform distribution; hence maximal uncertainty).

As predicted, participants' causal attributions were more diffuse after an unexpected observation. However, here, too, the main

[5] Experiment 2A originally included all four studies from Experiment 1. However, for the sake of simplicity, we decided to focus on the Y-test results; these used the expected and unexpected outcomes and hence best fit our research interests. We also asked exploratory questions not reported here, regarding participants' overall judgments of the strength of the experimental design and how the results should be treated. The full materials and data for all experiments can be found in Appendices A, B, C, and D or online (http://hdl.handle.net/1902.1/14819).

Table 3
*Estimated Median Probabilities (*Md*) and Bootstrapped Standard Errors of the Estimate (*SE*) Assigned to Each Cause for Experiments 2A, B, and C*

		2A		2B		2C	
Possible cause	Condition	Foresight *Md* (*SE*)	Hindsight *Md* (*SE*)	Foresight *Md* (*SE*)	Hindsight *Md* (*SE*)	Foresight *Md* (*SE*)	Hindsight *Md* (*SE*)
Valid	Expected (A)	**.60$_a$** (.08)	**.60$_a$** (.08)				
	Unexpected (B)	**.30$_b$** (.05)	**.50$_b$** (.04)				
Rotate	Expected (A)			**.58$_a$** (.07)	**.67a** (.05)	**.40$_a$** (.07)	**.35$_a$** (.06)
	Unexpected (B)			**.20$_b$** (.04)	**.25$_b$** (.03)	**.10$_b$** (.03)	**.10$_b$** (.04)
Invalid	Expected (A)	.05 (.03)	.05 (.02)	.05 (.02)	.01 (.01)		
	Unexpected (B)	.10 (.02)	.10 (.01)	.10 (.03)	.10 (.03)		
Faulty child	Expected (A)					**.10$_a$** (.02)	**.10$_a$** (.02)
	Unexpected (B)					**.25$_b$** (.03)	**.20$_b$** (.02)
Faulty task	Expected (A)					**.10$_a$** (.03)	**.10$_a$** (.02)
	Unexpected (B)					**.20$_b$** (.03)	**.25$_b$** (.05)
Chance	Expected (A)	**.20$_a$** (.02)	**.10$_b$** (.02)	**.20$_a$** (.03)	**.10$_b$** (.02)	.10 (.03)	.10 (.01)
	Unexpected (B)	**.20$_a$** (.02)	**.20$_a$** (.02)	**.25$_a$** (.03)	**.20$_a$** (.02)	.10 (.03)	.10 (.01)
Other	Expected (A)	**.09$_a$** (.02)	**.10$_a$** (.02)	**.10$_a$** (.02)	**.05$_a$** (.02)	.06 (.03)	.05 (.02)
	Unexpected (B)	**.20$_b$** (.04)	**.13$_b$** (.04)	**.20$_b$** (.03)	**.20$_b$** (.03)	.10 (.01)	.10 (.01)

Note. Causal explanations are as follows: *Valid* is "The child's understanding of the experimenter's instructions caused the child to place the dot." *Rotate* is "The child's ability to mentally rotate the image caused the child to place the dot." *Invalid* is "Some error in the experiment caused the child to place the dot." *Chance* is "Random chance caused the child to place the dot." *Faulty child* is "The child was not paying attention, and this caused the child to place the dot." *Faulty task* is "The task was confusing, and this caused the child to place the dot." Bold numbers indicate statistically significant differences between items with different subscripts.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

effect was not statistically significant, $t(462) = 1.06$, $p = .29$, $d = 0.05$. We also found support for the null hypothesis of no main effect for hindsight versus foresight, $t(462) = -1.54$, $p = .12$, $d = -0.07$, and no interaction between hindsight/foresight and outcome, $t(462) = 1.26$, $p = .21$, $d = 0.06$.

Experiment 2B

Experiment 2B changes the methodology of Experiment 2A in four ways, reflecting our own error models, prompted by the weak support for both theories (see Appendix A for materials used in Experiment 2B). (a) We reintroduced the probability of replication measure from Experiment 1, thinking that replicating a known result would be a simple way to check the validity of our modified design. This replication attempt differed from previous hindsight research, as the measure came after participants made their causal attributions, not before. (b) We sought to make the Valid option clearer by mentioning the child's mental rotation ability explicitly (with the revised wording "The child's ability to mentally rotate the image caused the child to place the dot in Area A"), thinking that some participants might have interpreted its previous wording ("The child's understanding of the experimenter's instructions . . .") as implying confusion (i.e., misunderstanding). (c) We elicited open-ended explanations before introducing the structured causal options, thinking that our options might not capture participants' intuitive ones. (We used these responses as a source for structured options in subsequent experiments, rather than analyzing them formally).

We also added a task asking participants whether they would advise the scientist to publish the research results. We expected participants who assigned a higher probability to error and provided more diffuse probabilities to the set of causes to be more reluctant to publish. The publication question was:

> If the replication of this experiment with 10 additional children comes out the way you expect, which of the following actions would you recommend that the scientist take?
> - Collect more data before publishing
> - Publish without collecting more data
> - Do not publish any of the data

Results

Hindsight bias. Hindsight participants told that the first child had placed the dot in the expected area (A) gave higher probabilities to that happening on the next 10 observations than did foresight participants asked to consider that outcome as a possibility (.50 vs. .30), $t(351) = 1.90$, $p = .06$, $d = 0.10$. Consistent replication of the unexpected result (B) was judged equally likely in hindsight and foresight (.10 vs. .10). The interaction between outcome and hindsight/foresight was marginally significant, $t(351) = -1.79$, $p = .07$, $d = -0.10$.[6]

Error model theory. As in Experiment 2A and consistent with the error model account, participants assigned a higher probability to the Invalid explanation after an unexpected observation than after an expected one, although, again, the main effect was not significant, $t(352) = 1.52$, $p = .13$, $d = 0.08$. There was again support for the null hypothesis of no interaction between whether the perspective was foresight or hindsight and whether the outcome was expected or unexpected, $t(352) = 0.90$, $p = .37$, $d = 0.05$. There was also an unexpected hindsight/foresight main ef-

[6] Note that these probability judgments were made after participants considered possible explanations, unlike in Experiment 1 (and Slovic & Fischhoff, 1977), when participants made judgments immediately after considering the initial observation.

fect, whereby participants assigned lower probability to error in hindsight than in foresight, $t(352) = -2.00$, $p = .05$, $d = -0.11$.

The higher the probability that participants assigned to error, the less likely they were to recommend publishing the data, $t(352) = 2.16$, $p = .03$, $d = -0.11$, reflecting more cautious behavior among those with the least confidence in the data. Although few participants in any condition recommended publishing (rather than collecting more data), that rate was twice as high with an expected result than with an unexpected one (0.17 vs. 0.09), $t(353) = 2.00$, $p = .05$, $d = -0.11$, equally so in hindsight and foresight (with no significant interaction). These results support the construct validity of these measures, assuming that participants wish to protect others from research that they perceive to be untrustworthy.

Entropy. Participants assigned more diffuse probabilities to potential causes after the unexpected observation than after the expected one, $t(352) = 2.45$, $p < .01$, $d = 0.13$, consistent with their reduced certainty about what would happen next. They assigned less diffuse probabilities to causes in hindsight than in foresight, $t(352) = -1.74$, $p = .08$, $d = -0.09$, with no interaction, $t(352) = 1.14$, $p = .25$, $d = 0.06$. Participants with flatter distributions (higher entropy) were less likely to recommend publication, $t(352) = -2.17$, $p = .03$, $d = -0.12$, also consistent with more cautious behavior among those who were less confident in the data.

Experiment 2C

Experiment 2C builds on the experience of Experiment 2B with two additional modifications designed to strengthen the hypothesis tests (see Appendix B for materials used in Experiment 2C). (a) Based on the open-ended explanations from Experiment 2B, we divided the Invalid method category into "the task was confusing" (Confusion) and "the child was not paying attention" (Inattention). (b) Because so many participants in Experiment 2B wanted a much larger sample before publication, Experiment 2C has participants predict the outcomes for 100 additional trials rather just 10. (c) Participants assigned probabilities to all three areas (A, B, and C), allowing us to calculate the entropy measure.

Results

Error model theory. Participants assigned higher probabilities to both Invalid method explanations when the result was unexpected, compared to when it was expected, with significant main effects for Inattention, $t(306) = 4.27$, $p < .01$, $d = 0.24$, and Confusing, $t(306) = 3.42$, $p < .01$, $d = 0.20$. When the probabilities for these two error model explanations were pooled, there was again no hindsight/foresight difference, $t(306) = 0.00$, or interaction, $t(306) = 0.60$, $p = .55$, $d = 0.03$. As in Experiment 2B, participants who assigned higher probabilities to error were less likely to recommend publication, although this time the relationship was not statistically significant, $t(308) = -1.43$, $p = .15$, $d = -0.08$. Participants recommended publication as often after an unexpected outcome (.28) as after an expected one (.31), $t(308) = -0.62$, $p = .53$, $d = -0.04$.

Entropy. As predicted, after an unexpected observation, participants assigned more diffuse probabilities to causes, $t(305) = 3.13$, $p < .01$, $d = 0.18$, and offered more diffuse predictions for future outcomes, $t(306) = 2.28$, $p = .02$, $d = 0.13$. These two entropy measures were positively correlated with one another, $t(307) = 8.54$, $p < .01$, $d = 0.49$, and negatively correlated with the probability of recommending publication, both for causal attributions, $t(307) = -1.77$, $p = .08$, $d = -0.10$, and for predictions of replication, $t(308) = -3.93$, $p < .01$, $d = -0.22$.

Participants who assigned a higher probability to chance as a cause also made predictions with greater entropy, $t(308) = 4.11$, $p < .01$, $d = 0.23$, meaning that an experiment perceived to have produced chance results once is also expected to produce unpredictable ones in the future.

Discussion

These three experiments examined whether hindsight/foresight differences, widely observed in predictions, are also observed in explanations, looking specifically at the role played by error models and entropy in judgments of causality. Consistent with the error model account, the structured attribution tasks in Experiments 2A and 2B found respondents more likely to invoke "something wrong with the experiment" when the result was unexpected. Experiment 2C found the same result more strongly, with what we believe to be a stronger design, using causal options revised in light of the open-ended explanations provided by participants in Experiment 2B. Conversely, observing the expected result evoked stronger attributions to a substantive theory and weaker attributions to methodological problems. Although all these results are consistent with the error model account, they reached statistical significance only in the final experiment, whose design took advantage of our own (error model) learning from its predecessors. The results were also consistent with the entropy account, with unexpected outcomes evoking flatter distributions across both predictions of replication and causal attributions—as might happen if unexpected outcomes led to the feeling that "anything can happen." There were no consistent differences between responses to foresight and hindsight versions of these tasks, which differed from those tasks in previous studies by requiring explanations (rather than leaving them implicit). Participants' publication recommendations were consistent with both their attributions and their predictions, affirming the construct validity of the measures—assuming that people want to protect others from the dissemination of untrustworthy scientific results.

Experiment 3

Experiment 3 uses the design of Experiment 2C to examine two open questions in its results. One is the possibility that participants' publication recommendations do not reflect deliberate decisions to avoid publishing weak data but, rather, unfamiliarity with scientific publication practices, leading them to use the "collect more data" response option as a substitute for "I don't know." That response strategy would be akin to saying "50–50" in order to avoid giving a numeric probability—unless an explicit "I don't know" alternative is offered (Fischhoff & Bruine de Bruin, 1999; Gärdenfors & Sahlin, 1982). Experiment 3 adds an "I don't know" response option in the publication recommendation task (see Appendix C for materials used in Experiment 3).

The second open question is the source of the similarity of foresight and hindsight judgments. We propose that it is due to our tasks making alternative explanations equally salient in the two

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

perspectives—and, indeed, unusually salient, compared to previous studies, which typically left explanations implicit. Slovic and Fischhoff (1977) reduced hindsight bias (in predictions of replication) by asking participants in hindsight conditions to explain how the outcome that was not observed might have occurred. However, they found that foresight participants made similar predictions for replication, regardless of whether they explained one outcome or two. In combination, these results suggest that foresight naturally evokes alternative explanations, whereas hindsight does not. Thus, the structured options that we used in Experiment 2 may have forced hindsight participants to do an unnatural act: consider explanations for unreported outcomes. As a result, because foresight and hindsight participants considered the same explanations, they made similar attributions.

Experiment 3 evaluates this proposal by manipulating the salience of the explanation that is consistent with the expected outcome (Rotate), by either including or removing it from the set that participants evaluate. When the expected outcome is reported, hindsight participants should generate that explanation spontaneously, so that the manipulation should have no effect. However, when the unexpected outcome is reported, the manipulation should affect hindsight attributions, if that explanation would not naturally come to mind. If, as we propose, foresight participants naturally consider alternative explanations, manipulating its salience should have little effect on that condition, whichever outcome they consider. We did not include a condition manipulating the salience of an explanation for the unexpected outcome because that would have removed our main dependent measure (the probability of error), forcing us to infer its probability indirectly from attribution to the "other" category.

Method

Design. Participants completed the attribution task of Experiment C, with or without the Rotate explanation in the structured option set. Thus, the causal options were either (Rotate, Faulty Task, Chance, Other; for *Rotate included*) or (Faulty Task, Chance, Other; for *Rotate removed*). The experiment was a 2 (foresight vs. hindsight) by 2 (Area A vs. Area B) by 2 (Rotate included, Rotate removed) between-subjects design. In addition, the publication recommendation task included an "I don't know" category. The Rotate included condition was identical to that in Experiment 2C.

Participants. Participants were paid volunteers who responded to an Amazon MTurk ad offering them $1 for participation in a 7-min experiment. Of the respondents, 1,360 of 1,838 individuals (74%) passed both attention filters. Their average age was 31 years old (range = 18–74); 49% were women.

Results

Error model theory. Responses to the Rotate included condition were similar to those in the Identical conditions in Experiment 2C. As before, participants assigned higher probabilities to Confusion when they received the unexpected outcome, $t(604) = 4.55$, $p < .01$, $d = 0.19$, with no significant hindsight/foresight interaction, $t(604) = -1.80$, $p = .07$, $d = -0.07$, or main effect, $t(604) = 0.00$.

In general, judgments were similar whether or not the Rotate explanation was mentioned explicitly, indicating that most participants needed no help to produce an explanation for the expected outcome, whether in hindsight or foresight. The only significant difference was that, as predicted, omitting the Rotate explanation from the set that participants initially considered led to higher probabilities for the Confusion explanation, $t(1235) = 3.18$, similarly in hindsight and foresight, interaction: $t(631) = 0.00$.

Entropy. As predicted (and as before), participants were less likely to recommend publishing the greater the entropy of their predictions, $t(1235) = -8.47$, $p < .01$, $d = -0.24$, and their attributions, $t(1233) = -6.31$, $p < .01$, $d = -0.18$. Removing the 156 participants who responded "I don't know" from the analysis (~13% of the sample), made little difference, $t(1079) = -8.15$, $d = -0.25$; $t(1077) = -5.92$, $d = -0.18$, respectively. Thus, it does not seem that participants gave flat distributions of causal attributions and predictions in order to avoid answering those questions.

Although the hindsight/foresight and Rotate manipulations were not expected to affect the entropy of participants' judgments, we report them for completeness. As before, participants' predictions of replication had higher entropy after an unexpected outcome, main effect: $t(1238) = 2.21$, $p = .03$, $d = 0.06$. Causal attributions had higher entropy after the unexpected result when the Rotate category was omitted, main effect: $t(632) = 2.91$, $p < .01$, $d = 0.12$, but, unlike the previous studies, not when the Rotate category was included, main effect: $t(604) = -0.48$, $p = .63$, $d = -0.02$. Participants who assigned higher probability to chance as a cause also made predictions with higher entropy, main effect: $t(1237) = 6.69$, $p < .01$, $d = 0.19$.

Data sharing. Unlike in Experiment 2B, but consistent with Experiment 2C, participants were equally likely to recommend publication, regardless of whether they received the expected outcome (.24) or the unexpected one (.22), $t(1236) = -0.88$. Nor were they less likely to recommend publication when they assigned a greater role to error, $t(1236) = -0.45$.

Discussion

Experiment 3 replicated the results of the previous experiments and addressed two issues raised by them. First, an "I don't know" option was added to the publication recommendations, in order to see whether recommendations to collect more data were an attempt to say "I don't know enough about scientific practices to know what to recommend." This was not the case, though, as the "I don't know" option was rarely used, and removing the few participants who gave this response had no effect on the results. Thus, publication recommendations followed participants' confidence in the data and not merely their familiarity with scientific publishing.

Second, we compared conditions that did and did not offer the Rotate explanation explicitly, in order to see whether bringing it to hindsight participants' attention led to more systematic search. However, this explanation for the expected outcome appeared to be salient regardless of whether we mentioned it. As discussed, the current design precluded similarly manipulating the salience of an explanation for the unexpected outcome. Experiment 4 addresses that question in a different way.

Experiment 4

The intuition motivating the present studies is that people typically do not generate error models until observing an unexpected

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

outcome motivates them to think even harder about what might have gone wrong (Weiner, 1985) and then to believe in whatever (error) explanations they generate. As a result, their evaluation of those explanations is biased by the very results that bring them to mind. If such error models are considered before observing results, then they should be evaluated more fairly, leading, in turn, to sounder decisions about publication and replication. In the preceding experiments, the similarity of causal judgments in hindsight and foresight suggests that our tasks created such conditions, by requiring participants to consider a full set of explanations. Experiment 4 tests this hypothesis more directly, by comparing participants who evaluate more and less complete sets of possible explanations before considering an outcome. The Complete condition has three possible explanations, along with an "other" category. Each of three Incomplete conditions omits one of those explanations, leaving participants to generate it on their own. Following Fischhoff, Slovic, and Lichtenstein (1978) and Tversky and Koehler (1994), we expect "other" to be a weak prompt for evoking missing explanations.

We propose that when people evaluate an explanation after learning of a result, they first ask how plausible that explanation would have been beforehand (Chihara, 1987; Howson & Urbach, 1989; Suzuki, 2005). If people first consider an explanation after they learn that the results fit it, then hindsight bias should increase that assessment of plausibility, compared to considering the explanation first in foresight. As a result, we predict that participants will assign higher probability to an explanation when they encounter it after an outcome that it can explain, compared to when they also evaluate it before such an outcome.

Method

Design. Experiment 4 had a 2 by 4 design, crossing two possible outcomes (A, B) with four possible sets of explanations mentioned before participants were told of the outcome. Each participant read the explanations and then checked the ones that he or she felt applied to the experiment before receiving an outcome. Participants then completed the causal attribution task, which included all of the explanations; made predictions of replication; and provided publication recommendations (without an "I don't know" category).

Participants. Participants were paid volunteers who responded to an Amazon MTurk ad offering them $1 for participation in a 7-min experiment. Use of the same attention filter as before left 969 of 1,628 individuals (60%). Their average age was 30 years old (range = 18–70); 408 (42%) were women.

Materials. The instructions were the same as in Experiment 3 (and 2C), the experiments that we considered to have the strongest design, except that before considering an initial observation, participants were asked to consider a set of explanations. In the *complete* condition, this was done by asking the following question:

> Which of the following do you think could possibly affect the experimental results (check all that apply)?
>
> 1. (Left-handed) The children selected for the study are left-handed.
> 2. (Symmetry) Children like putting things in the middle, to maintain symmetry.
> 3. (Confusing) The task is confusing.
> 4. Some other cause.

In the three other *incomplete* conditions, one of the three alternative explanations was omitted. (See Appendix D for materials used in Experiment 4).

Results

Error models theory. As before, the unexpected outcomes produced stronger error model attributions (Confusing), $t(967) = 7.74$. As a manipulation check, each of the two substantive explanations, Left-handed and Symmetry, received stronger attributions when the outcome was consistent with them (Areas A and B, respectively), $t(967) = -5.66$; $t(967) = 7.97$.

Contrary to our prediction, though, participants were not more confident when an explanation was first mentioned after observing an experimental outcome consistent with it. Left-handed was assigned a higher probability when it was mentioned before the outcome was observed, rather than just afterward, $t(484) = 3.05$, $p < .01$, $d = 0.08$,[7] and to the same degree whether or not it was consistent with the outcome, interaction: $t(484) = 0.96$, $p = .34$, $d = 0.04$. Symmetry was also assigned a higher probability when it was mentioned before the outcome that it could explain (area B) than when it was mentioned after that outcome, $t(481) = 2.34$, $p = .02$, $d = 0.16$, but there was no difference when the outcome was inconsistent, $t(481) = 0.74$, $p = .46$, $d = 0.04$, with a significant interaction, $t(481) = 2.03$, $p = .04$, $d = 0.09$. The Confusing explanation received the same probability when it was mentioned before or after the initial observation, $t(474) = 0.00$, with no interaction, $t(474) = 0.00$.

Entropy. As before, after an unexpected observation, participants assigned more diffuse causal attribution probabilities, $t(967) = 3.55$, $p < .01$, $d = 0.11$, and offered more diffuse predictions of replication, $t(967) = 7.24$, $p < .01$, $d = 0.23$. Also as before, these two entropy measures were positively correlated with one another, $t(967) = 17.46$, $p < .01$, $d = 0.56$. Participants who assigned higher probability to chance as a cause also made predictions with higher entropy, $t(967) = 8.69$, $p < .01$, $d = 0.28$. However, the entropy of causal attributions was the same whether the focal explanation was first mentioned before or after the relevant outcome was reported (all ts < 1.08).

Data sharing. Participants recommended publication less for an unexpected outcome than for an expected one, $t(967) = 1.47$, $p = .14$, $d = 0.05$, and when assigning higher probability to confusion as the cause, $t(967) = 1.53$, $p = .13$, $d = 0.05$, although neither result was significant. As before, participants also recommended publication less when there was greater entropy in their causal attributions, $t(967) = 3.55$, $p < .01$, $d = 0.11$, and in their predictions of replication, $t(967) = 7.24$, $p < .01$, $d = 0.23$.

Discussion

Experiment 4 unexpectedly found that participants had more confidence in an explanation when they were asked to consider it

[7] This analysis uses means rather than medians, as there was no resampling variance of the median in the bootstrap estimate of the model intercept.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

before observing a result consistent with it, compared to when they were asked to consider it afterward, regardless of whether that result was expected or unexpected. Thus, these observers find post hoc explanations less convincing—in a design where those explanations were missing from the set of possible causes of the result that they considered a priori. Participants assigned a similar role to confusion, regardless of whether the result was expected, in a design where substantive explanations for the possible results were always presented.

General Discussion

We present four experiments examining how the evaluation and communication of scientific evidence differs when results are expected or unexpected and when considered in foresight or hindsight, extending research on hindsight bias predictions to explanations. Experiment 1 repeats Experiment 1 of Slovic and Fischhoff (1977), 35 years later with an online (MTurk) sample, and finds similar results: An initial observation seems more likely to be replicated when considered in hindsight, compared to foresight (see Table 1). Subsequent experiments focused on one of the stimulus studies used in Experiment 1, chosen because its two outcomes had the most and least expected results (see Table 2). These experiments evaluated two accounts of how people explained unexpected outcomes, error model and entropy, finding some support for each.

The error model account holds that surprising experimental results prompt a search for causal explanations, which is often satisfied by invoking hitherto overlooked methodological problems. These error models allow observers to dismiss the experiment as having failed to provide a fair test of the hypothesis. Error models are less salient with expected results because the explanations that motivated the experiment arise naturally, but the possibility of error does not.

We found consistent support for this prediction, with results that strengthened as we refined our experimental design, drawing on our own error models, prompted by the weak support found in the initial experiments. Unexpected results evoked error models more strongly than did expected ones. The construct validity of these attributions was seen in participants' greater reluctance to publish and stronger desire for additional data, when they invoked error models. Previous studies have found that investigators are more likely to attribute unexpected results to error (Gilovich, 1983; Lord et al., 1979; Mahoney, 1977; Munro & Ditto, 1997; L. Ross, Lepper, & Hubbard, 1975; Wyer & Frey, 1983). Here, we found a similar pattern with participants who, unlike actual investigators, had no stake in the experiment's outcome—although they did have expectations about what it would be.

Although Experiment 1 replicated the familiar hindsight bias in predictions of replication, these patterns of causal attributions were the same in hindsight or foresight (for Experiments 2 and 3). We attribute this similarity to our using tasks that required considering alternative explanations in both perspectives, unlike previous studies of hindsight bias, where causal inferences were implicit and, likely, shaped by which outcomes were reported. As seen in Experiment 3, when an explanation is naturally salient, whether it is mentioned explicitly makes no difference in foresight or hindsight. However, when it is not naturally salient, as with an error model, people are as likely to produce it in hindsight or foresight, when challenged to explain the unexpected.

Experiment 4 suggested that people are uncomfortable with such post hoc explanations, expressing less confidence in explanations that were suggested only after learning about the result. Unlike the normal reporting of research results, where post hoc explanations may not be communicated separately from prior ones (Kerr, 1998), Experiment 4 explicitly distinguished explanations that were added to the set of possibilities only after considering a outcome that they could explain.

Consistent with the entropy account, an unexpected experimental outcome led to greater entropy in attributions and predictions, as though it produced a feeling of uncertainty, such that "anything can happen," rather than a certain feeling that there was an experimental error. Indeed, participants with greater entropy in their attributions also had greater entropy in their predictions, were less likely to recommend publication, and were more likely to recommend additional data collection. In sum, participants wanted more data when they felt that the evidence supported several alternative explanations, as would be expected from research showing that people are sensitive to ruling out alternative hypotheses (Koslowski, 2012; Koslowski, Marasia, Chelenza, & Dublin, 2008).

These results support both the error-model and the entropy accounts of how people respond to unexpected observations in their predictions, attributions, and recommendations for data sharing. As mentioned, these two accounts are complementary. When individuals have strong expectations, they make similar (although not identical) predictions about how people will respond to unexpected results. According to the error model account, observers will increase their attributions to error while decreasing their attributions to the previously favored explanation, with the result being greater entropy, especially when the process adds previously neglected possible explanations. According to the entropy account, observers will level the differences between all explanations, thereby increasing the importance of previously marginal explanations, including neglected error models, and decreasing the importance of previously focal explanations, including the one that favored the expected result. The error model theory is mute regarding the treatment of alternative substantive explanations. The entropy account ignores the substance of the alternative explanations, error model or otherwise. Future research is needed to explicate their converging and diverging predictions.

In many ways, participants in these studies were models of circumspection. They took unexpected results seriously, as reflected in their attributions and predictions. Although they were quick to produce error models when considering an unexpected initial observation, they were reluctant to publish without additional data. They appeared skeptical of post hoc explanations. Moreover, their publication recommendations were generally consistent with their attributions and predictions—for individuals wary of sharing uncertain results. In all these ways, research participants drawn from the general (MTurk) population had orderly, reasonable judgments.

One limit of the present studies is their reliance on structured explanations, in order to control the salience of explanations. Although not formally analyzed or reported, the open-ended explanations of Experiment 2B revealed some of the diversity in how people intuitively formulate error model explanations—and improved the structured options used in subsequent studies. In future

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

work, we envision using concurrent verbal protocols (Ericsson & Simon, 1985) eliciting participants' intuitive explanations, before, during, and after their processing of experimental outcomes, seeking to learn more about natural explanatory processes. A second limit is eliciting judgments only for a single experiment and not seeing how they would respond to receiving the additional data that most wanted to see (e.g., Would they want even data? Would the same amount of data have different impact if delivered all at once or in successive units?). A third limit is not seeing how participants would design their own additional experiments, given their explanations of the observed results, and how such active involvement would affect their response to unexpected results. Last, we used nonscientists, who are relatively naive to how research is conducted and communicated.

Would scientists exhibit similar behavior, if questioned about their explanations before and after observing unexpected results? With respect to error models, the answer seems likely to be yes. Previous studies have found that both laypeople (Penner & Klahr, 1996) and elite scientists (Dunbar, 1997) tend to attribute unexpected results to error. What they do after they generate these explanations, however, may differ from the behavior observed here. For scientists, unexpected results suggest taking a step back and checking their instruments, perhaps comparing the results to known standards, used by other scientists (Baker & Dunbar, 2000). Unlike lay observers, scientists have control over the research process and knowledge about what controls have been tried in the past and might be tried in the future. Lay observers may have little choice but to recommend repeating a study, hoping that additional data reveal a clearer picture. On the other hand, they should be freer of the motivated cognition that might make it harder to generate alternative explanations in foresight.

Conclusions

We sought to study scientific reasoning in ways that captured its inherent uncertainty (Giner-Sorolla, 2012; Kerr, 1998), as revealed in our own learning process, and by following open science practices for documenting our work (Bradley, 2007; Nosek & Bar-Anan, 2012; Pashler & Wagenmakers, 2012; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012), so that interested readers can reconstruct our process. As much as we sought to generate all possible error explanations prior to data collection, our confrontation with the evidence prompted us to identify new error models and substantive explanations. We hope to have been more like Millikan than Blondlot in second-guessing our data.

Kuhn (1996) asked, "How do scientists proceed when aware only that something has gone fundamentally wrong at a level with which their training has not equipped them to deal?" (p. 86). He answered, in effect, that they naturally attribute unexpected results to flawed experimental method, while attributing expected ones to the theory that guides them. It takes an accumulation of unexpected results, along with a deep insight, to prompt a scientific revolution. Here, we found that lay participants also saw unexpected results as due to experimental error. Moreover, they expressed cautious data-sharing policies, generally wanting more observations before publishing, especially when they felt uncertain about their predictions and explanations. One possible contributor to these noteworthy patterns is that our tasks created one of the conditions recommended for normal scientific practice: Think hard, in advance, about how you will explain unexpected results should you observe them. If so, the task eliminated a kind of foresight bias, by evoking in foresight the same need to explain that causes bias in hindsight. Hindsight bias can be reduced by explaining how what did not happen might have. Foresight bias might be reduced by explaining how what is not expected to happen still might.

References

Baker, L. M., & Dunbar, K. (2000). Experimental design heuristics for scientific discovery: The use of "baseline" and "known standard" controls. *International Journal of Human-Computer Studies, 53,* 335–349. doi:10.1006/ijhc.2000.0393

Blank, H., Musch, J., & Pohl, R. (2007). Hindsight bias: On being wise after the event. *Social Cognition, 25,* 1–9. doi:10.1521/soco.2007.25.1.1

Blank, H., & Nestler, S. (2007). Cognitive process models of hindsight bias. *Social Cognition, 25,* 132–146. doi:10.1521/soco.2007.25.1.132

Bradley, J. (2007). Open notebook science using blogs and wikis. *Nature Precedings*. Retrieved from http://dx.doi.org/10.1038/npre.2007.39.1

Bruine de Bruin, W., Fischhoff, B., Millstein, S., & Halpern-Felsher, B. (2000). Verbal and numerical expressions of probability: "It's a fifty–fifty chance". *Organizational Behavior and Human Decision Processes, 81,* 115–131. doi:10.1006/obhd.1999.2868

Buhrmester, M., Kwang, T., & Gosling, S. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. doi:10.1177/1745691610393980

Chihara, C. (1987). Some problems for Bayesian confirmation theory. *British Journal for the Philosophy of Science, 38,* 551–560. doi:10.1093/bjps/38.4.551

Chinn, C., & Brewer, W. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1–49. doi:10.3102/00346543063001001

Christensen-Szalanski, J., & Willham, C. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes, 48,* 147–168. doi:10.1016/0749-5978(91)90010-Q

Collins, H. (2003). Lead into gold: The science of finding nothing. *Studies in History and Philosophy of Science Part A, 34*(4), 661–691. doi:10.1016/j.shpsa.2003.09.002

Downs, J., Holbrook, M., Sheng, S., & Cranor, L. (2010). Are your participants gaming the system? Screening Mechanical Turk workers. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems* (pp. 2399–2402). Retrieved from http://dmrussell.net/CHI2010/docs/p2399.pdf

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. Ward, S. Smith, & J. Vafid (Eds.), *Conceptual structures and processes: Emergence, discovery, and change* (pp. 461–493). Washington, DC: American Psychological Association. doi:10.1037/10227-017

Dunbar, K. (2001). What scientific thinking reveals about the nature of cognition. In K. Crowley, C. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 115–140). Mahwah, NJ: Erlbaum.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.

Ericsson, K., & Simon, H. (1985). *Protocol analysis*. Cambridge, MA: MIT Press.

Falk, R., & Lann, A. (2008). The allure of equality: Uniformity in probabilistic and statistical judgment. *Cognitive Psychology, 57,* 293–334. doi:10.1016/j.cogpsych.2008.02.002

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychol-*

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ogy: Human Perception and Performance, 1, 288–299. doi:10.1037/0096-1523.1.3.288

Fischhoff, B., & Bruine de Bruin, W. (1999). Fifty-fifty = 50%? *Journal of Behavioral Decision Making, 12,* 149–163. doi:10.1002/(SICI)1099-0771(199906)12:2<149::AID-BDM314>3.0.CO;2-J

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 330–344. doi:10.1037/0096-1523.4.2.330

Fischhoff, B., Welch, N., & Frederick, S. (1999). Construal processes in preference assessment. *Journal of Risk and Uncertainty, 19,* 139–164. doi:10.1023/A:1007823326511

Franklin, A. (1997). Millikan's oil-drop experiments. *Chemical Educator, 2*(1), 1–14. doi:10.1007/s00897970102a

Gärdenfors, P., & Sahlin, N. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese, 53,* 361–386. doi:10.1007/BF00486156

Gilovich, T. (1983). Biased evaluation and persistence in gambling. *Journal of Personality and Social Psychology, 44,* 1110–1126. doi:10.1037/0022-3514.44.6.1110

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science, 7,* 562–571. doi:10.1177/1745691612457576

Goodstein, D. (2000). In defense of Robert Andrews Millikan. *Engineering and Science, 63*(4), 30–38.

Gorman, M. (1989). Error, falsification and scientific inference: An experimental investigation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 41*(A), 385–412.

Gorman, M. (1992). *Simulating science: Heuristics, mental models, and technoscientific thinking*. Bloomington: Indiana University Press.

Gorman, M., Tweney, R., Gooding, D., & Kincannon, A. (2005). *Scientific and technological thinking*. Mahwah, NJ: Erlbaum.

Greenwald, A. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82,* 1–20. doi:10.1037/h0076157

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology, 26,* 103–117.

Hilton, D. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 107,* 65–81. doi:10.1037/0033-2909.107.1.65

Horton, J., Rand, D., & Zeckhauser, R. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14,* 399–425. doi:10.1007/s10683-011-9273-9

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*. Chicago, IL: Open Court.

Ipeirotis, P. (2010). *Demographics of Mechanical Turk*. Working paper, Center for Digital Economy Research, NYU Stern School of Business.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57,* 227–254. doi:10.1146/annurev.psych.57.102904.190100

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2,* 196–217. doi:10.1207/s15327957pspr0203_4

Klayman, J., & Ha, Y. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 596–604. doi:10.1037/0278-7393.15.4.596

Klotz, I. (1980). The n-ray affair. *Scientific American, 242,* 168–175. doi:10.1038/scientificamerican0580-168

Koenker, R. (2009). Quantreg: Quantile regression [R package version 4]. Retrieved from http://CRAN.R-project.org/package=quantreg

Koslowski, B. (2012). Inference to the best explanation (IBE) and the causal and scientific reasoning of nonscientists. In R. Proctor & E. Capaldi (Eds.), *Psychology of science: Implicit and explicit processes* (pp. 112–136). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199753628.003.0006

Koslowski, B., Marasia, J., Chelenza, M., & Dublin, R. (2008). Information becomes evidence when an explanation can incorporate it into a causal framework. *Cognitive Development, 23,* 472–487. doi:10.1016/j.cogdev.2008.09.007

Kuhn, T. (1996). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press. doi:10.7208/chicago/9780226458106.001.0001

Lakatos, I., Worrall, J., & Currie, G. (1980). *The methodology of scientific research programmes*. New York, NY: Cambridge University Press.

Lau, R. (1984). Dynamics of the attribution process. *Journal of Personality and Social Psychology, 46,* 1017–1028. doi:10.1037/0022-3514.46.5.1017

Lau, R., & Russell, D. (1980). Attributions in the sports pages. *Journal of Personality and Social Psychology, 39,* 29–38. doi:10.1037/0022-3514.39.1.29

Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098–2109. doi:10.1037/0022-3514.37.11.2098

MacKay, D. (2003). *Information theory, inference and learning algorithms*. New York, NY: Cambridge University Press.

Mahoney, M. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1,* 161–175. doi:10.1007/BF01173636

Martinson, B. C., Anderson, M. S., & De Vries, R. (2005, June 9). Scientists behaving badly. *Nature, 435,* 737–738. doi:10.1038/435737a

Mason, W., & Watts, D. (2010). Financial incentives and the performance of crowds. *SIGKDD Explorations Newsletter, 11,* 100–108. doi:10.1145/1809400.1809422

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12,* 269–275. doi:10.1111/1467-9280.00350

Munro, G., & Ditto, P. (1997). Biased assimilation, attitude polarization, and affect in reactions to stereotype-relevant scientific information. *Personality and Social Psychology Bulletin, 23,* 636–653. doi:10.1177/0146167297236007

Murphy, G. L., & Ross, B. (1994). Predictions from uncertain categorizations. *Cognitive Psychology, 27,* 148–193. doi:10.1006/cogp.1994.1015

Murphy, G. L., & Ross, B. (2010). Uncertainty in category-based induction: When do people integrate across categories? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 263–276. doi:10.1037/a0018685

Nestler, S., Blank, H., & Egloff, B. (2010). Hindsight ≠ hindsight: Experimentally induced dissociations between hindsight components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1399–1413. doi:10.1037/a0020449

Nestler, S., Blank, H., & von Collani, G. (2008). Hindsight bias and causal attribution: A causal model theory of creeping determinism. *Social Psychology, 39,* 182–188. doi:10.1027/1864-9335.39.3.182

Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220. doi:10.1037/1089-2680.2.2.175

Nickerson, R. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125,* 737–759. doi:10.1037/0033-2909.125.6.737

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Upper Saddle River, NJ: Prentice Hall.

Nosek, B., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry, 23,* 217–243. doi:10.1080/1047840X.2012.692215

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Oppenheimer, D., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45,* 867–872. doi:10.1016/j.jesp.2009.03.009

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411–419.

Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7,* 528–530. doi:10.1177/1745691612465253

Penner, D. E., & Klahr, D. (1996). When to trust the data: Further investigations of system error in a scientific reasoning task. *Memory & Cognition, 24,* 655–668. doi:10.3758/BF03201090

Pezzo, M. V. (2003). Surprise, defence, or making sense: What removes hindsight bias? *Memory, 11,* 421–441. doi:10.1080/09658210244000603

Pyszczynski, T., & Greenberg, J. (1981). Role of disconfirmed expectancies in the instigation of attributional processing. *Journal of Personality and Social Psychology, 40,* 31–38. doi:10.1037/0022-3514.40.1.31

Risen, J. L., Gilovich, T., & Dunning, D. (2007). One-shot illusory correlations and stereotype formation. *Personality and Social Psychology Bulletin, 33,* 1492–1502. doi:10.1006/jesp.1996.0010

Roese, N., & Olson, J. (1996). Counterfactuals, causal attributions, and the hindsight bias: A conceptual integration. *Journal of Experimental Social Psychology, 32,* 197–227. doi:10.1006/jesp.1996.0010

Roese, N., & Vohs, K. (2012). Hindsight bias. *Perspectives on Psychological Science, 7,* 411–426. doi:10.1177/1745691612454303

Ross, B. H., & Murphy, G. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 736–753. doi:10.1037/0278-7393.22.3.736

Ross, L., Lepper, M., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32,* 880–892. doi:10.1037/0022-3514.32.5.880

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 544–551. doi:10.1037/0096-1523.3.4.544

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, MA: MIT Press.

Suzuki, S. (2005). The old evidence problem and AGM theory. *Annals of the Japan Association for Philosophy of Science, 13,* 105–126.

Tversky, A., & Koehler, D. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547–567. doi:10.1037/0033-295X.101.4.547

Wagenmakers, E., Wetzels, R., Borsboom, D., van der Maas, H., & Kievit, R. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7,* 632–638. doi:10.1177/1745691612463078

Wasserman, D., Lempert, R., & Hastie, R. (1991). Hindsight and causality. *Personality and Social Psychology Bulletin, 17,* 30–35.

Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin, 97,* 74–84. doi:10.1037/0033-2909.97.1.74

Wong, P., & Weiner, B. (1981). When people ask "why" questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology, 40,* 650–663.

Wood, R. (1904). The n-rays. *Nature, 70,* 530–531.

Wooldridge, J. (2009). *Introductory econometrics: A modern approach*. Nashville, TN: South-Western.

Wyer, R., & Frey, D. (1983). The effects of feedback about self and others on the recall and judgments of feedback-relevant information. *Journal of Experimental Social Psychology, 19,* 540–559.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix A

Experiment 2B Materials

> Please explain why you think the child could place the dot in Area A [or B]. (open-ended)

This was followed by modified causal attributions, as shown below, using Area A foresight condition as an example:

> If the child places a dot in Area A, what is the probability that: (Note: These four probabilities should total 100%).

- (Rotate) The child's ability to mentally rotate the image caused the child to place the dot in Area A.
- (Invalid) Some error in the experiment caused the child to place the dot in Area A.
- (Chance) Random chance caused the child to place the dot in Area A.
- (Other) There was some other cause not already mentioned above.

> If the replication of this experiment with 10 additional children comes out the way you expect, which of the following actions would you recommend that the scientist take:

- Collect more data before publishing
- Publish without collecting more data
- Do not publish any of the data

(*Appendices continue*)

Appendix B

Experiment 2C Materials

If the child places a dot in Area A, what is the probability that: (Note: These five probabilities should total 100%).

- (Rotate) The child's ability to mentally rotate the image caused the child to place the dot in Area A.
- (Faulty Child) The child was not paying attention, and this caused the child to place the dot in Area A.
- (Faulty Task) The task was confusing, and this caused the child to place the dot in Area A.
- (Chance) Random chance caused the child to place the dot in Area A.
- (Other) There was some other cause not already mentioned above.

In a replication of this experiment with 100 additional children, how many children will place the dot in the following areas:

- Area A
- Area B
- Area C

If the replication of this experiment with 100 additional children comes out the way you expect, which of the following actions would you recommend that the scientist take:

- Collect more data before publishing
- Publish without collecting more data
- Do not publish any of the data

Appendix C

Experiment 3 Materials

If the child places a dot in Area A, what is the probability that: (Note: These four probabilities should total 100%).

- (Rotate) The child's ability to mentally rotate the image caused the child to place the dot in Area A.
- (Faulty Task) The task was confusing, and this caused the child to place the dot in Area A.
- (Chance) Random chance caused the child to place the dot in Area A.
- (Other) There was some other cause not already mentioned.

Participants then completed the posterior prediction and the modified data sharing judgment:

If the replication of this experiment with 100 additional children comes out the way you expect, which of the following actions would you recommend that the scientist take:

- Collect more data before publishing
- Publish without collecting more data
- Do not publish any of the data
- I don't know

Appendix D

Experiment 4 Materials

Participants were then told the first child placed the dot in either area A or area B and were asked to attribute the cause:
(Note: These six probabilities should total 100%.)

1. (Rotate) The child's ability to mentally rotate the image caused the child to place the dot in Area A.
2. (Confusing) The task was confusing, and this caused the child to place the dot in Area A.
3. (Left-handed) The child was left-handed, and this caused the child to place the dot in Area A.
4. (Symmetry) The child likes putting things in the middle to maintain symmetry, and this caused the child to place the dot in Area A.
5. (Chance) Random chance caused the child to place the dot in Area A.
6. (Other) There was some other cause not already mentioned.

Participants then predicted the next 100 observations and made data sharing judgments, as in Experiment 3.

Received November 13, 2012
Revision received May 19, 2013
Accepted May 20, 2013 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.