Psychological Science
http://pss.sagepub.com/

Shared Visual Attention Reduces Hindsight Bias

Daw-An Wu, Shinsuke Shimojo, Stephanie W. Wang and Colin F. Camerer

Psychological Science 2012 23: 1524 originally published online 18 October 2012
DOI: 10.1177/0956797612447817

The online version of this article can be found at:
http://pss.sagepub.com/content/23/12/1524

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Association for Psychological Science

Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
23(12) 1524–1533

DOI: 10.1177/0956797612447817
http://pss.sagepub.com
SAGE

Shared Visual Attention Reduces Hindsight Bias

Daw-An Wu[1,2], Shinsuke Shimojo[2,3], Stephanie W. Wang[1,4], and Colin F. Camerer[1,3]

[1]Division of Humanities and Social Sciences, California Institute of Technology; [2]Division of Biology, California Institute of Technology; [3]Computational and Neural Systems, California Institute of Technology; and [4]Department of Economics, University of Pittsburgh

Abstract

Hindsight bias is the tendency to retrospectively think of outcomes as being more foreseeable than they actually were. It is a robust judgment bias and is difficult to correct (or "debias"). In the experiments reported here, we used a visual paradigm in which performers decided whether blurred photos contained humans. Evaluators, who saw the photos unblurred and thus knew whether a human was present, estimated the proportion of participants who guessed whether a human was present. The evaluators exhibited visual hindsight bias in a way that matched earlier data from judgments of historical events surprisingly closely. Using eye tracking, we showed that a higher correlation between the gaze patterns of performers and evaluators (shared attention) is associated with lower hindsight bias. This association was validated by a causal method for debiasing: Showing the gaze patterns of the performers to the evaluators as they viewed the stimuli reduced the extent of hindsight bias.

Keywords

visual hindsight bias, group decision making, visual priming, heuristics, theory of mind, eye movements, attention, attribution

Received 9/14/11; Revision accepted 3/22/12

In the aftermath of a surprising extreme event, such as a natural disaster, botched military operation, or last-second upset in sports, people naturally look back and ask: "Were these outcomes foreseeable? Were warning signs ignored?" Such questions are important because proper, balanced scrutiny of past decisions is necessary to learn from history and make good adjustments in procedures and leadership.

However, views of history are often dramatically distorted by hindsight. Armed with knowledge of the actual outcome, people are biased to view it as more obviously foreseeable than it really was. This tendency is a chronic challenge in legal decision making, in which standards such as foreseeability are often set as criteria for determining liability or guilt (Harley, 2007). Hindsight bias seems to cause people and societies to overreact to genuinely unpredictable surprises, harshly judging the decision makers involved. In response, decision makers who anticipate hindsight-biased second-guessing of their actions might take costly defensive measures. In this way, hindsight bias can encourage bureaucratic and legal overregulation, defensive medicine, or a refusal to deal with high-risk cases (Harley, 2007; Madarasz, 2011; Studdert et al., 2005; Terry, 2011).

Experimental studies of hindsight bias began with Fischhoff and Beyth (1975), who focused on possible public events during then-President Nixon's overseas trips in the mid-1970s to China and the Union of Soviet Socialist Republics. Subjects assigned probabilities of possible outcomes before the trip occurred and were asked to recall those probabilities after the trip occurred. Subjects recalled having predicted events more accurately than they actually did. This finding was replicated using historical scenarios (Fischhoff, 1975) and almanac questions (Fischhoff, 1977).

The hindsight effect has been shown to be large and pervasive across subject populations. Guilbault, Bryant, Brockway, and Posavac (2004) did a formal meta-analysis of 252 effects from 95 studies, reporting a mean effect size (d) of 0.39, with a 95% confidence interval (CI) of [0.36, 0.42]. Hindsight has been studied extensively in a variety of settings both inside and outside of the lab (Hawkins and Hastie, 1990; Pohl, 2007).

Corresponding Author:
Daw-An Wu, Division of Biology, California Institute of Technology, M/C 114-96, Pasadena, CA 91125
E-mail: daw-an@caltech.edu

A comparative study across four continents (Pohl, Bender, & Lachmann, 2002) found robust evidence of hindsight for all subjects except Germans and Dutch (perhaps because those nationalities were more familiar with the materials used in the study).

Using a visual paradigm, Bernstein, Erdfelder, Meltzoff, Peria, and Loftus (2011) found that hindsight bias was greatest in young children (aged 3–5 years) and in seniors (aged 61–95 years). This U-shaped pattern corresponds to the cognitive development and decline of many mechanisms over the life span. In particular, because of differing types of errors in young and old subjects, the study implicates memory as a likely mechanism for hindsight bias. Other studies have shown separate components of hindsight bias, such as the overprojection of one's own information onto the likely cognitions of others who know less (Camerer, Loewenstein, & Weber, 1989; Loewenstein, Moore, & Weber, 2006; Madarasz, 2011; Nestler, Blank, & Egloff, 2010).

Hindsight bias appears to be one of the most robust judgment errors that emerged from the "heuristics and biases" approach (Gilovich, Griffin, & Kahneman, 2002) to information processing. As with research on other types of errors, research on hindsight bias has been directed toward finding methods to reduce, or "debias," the systematic statistical biases. Effective debiasing could be of practical use and contribute to a causal understanding of the underlying mechanism generating the bias.

Unfortunately, hindsight-debiasing manipulations do not seem to work reliably. Guilbault et al.'s (2004) meta-analysis found no significant effects among studies that included manipulations designed to decrease hindsight bias. Furthermore, other studies have shown that informing subjects of the bias and asking them to avoid it does not reduce the bias (Fischhoff, 1975, 1977; Harley, Carlsen, & Loftus, 2004). Indeed, subjects do not significantly reduce their bias even after repeated testing and performance feedback (Pohl & Hell, 1996).

Prompting subjects to think through alternative outcomes has been effective in reducing other kinds of judgment biases (Koriat, Lichtenstein, & Fischhoff, 1980; Lord, Lepper, & Preston, 1984). However, this strategy has yielded decidedly mixed results for hindsight bias. Across five studies posing a total of 18 judgment questions, significant decreases were found for only 10 of the questions (Arkes, Faust, Guilmette, & Hart, 1988; Davies, 1987; Slovic & Fischhoff, 1977; Stallard & Worthington, 1998), and significant *increases* were found for 2 of them (Sanna, Schwarz, & Stocker, 2002; Slovic & Fischhoff, 1977). Sanna et al. (2002) hypothesized that the debiasing method backfired whenever subjects found it difficult to implement the strategy. The metacognitive impression that it is difficult to think of alternatives can strengthen the perceived inevitability of the actual outcome.

The only intervention that reduces hindsight bias somewhat reliably is to tell subjects that the original outcome information was false (Erdfelder & Buchner, 1998; Hasher, Attig, & Alba, 1981, Experiment 3). However, this method has limited use in practical cases (e.g., in law and medicine) in which an event has actually occurred, because it requires convincing deception about what happened. A small study found that presenting subjects with a written record of their prejudgment thoughts during recollection did reduce hindsight bias (Davies, 1987, Experiment 1); this study can be seen as a precursor to our Experiment 3.

In the experiments reported here, we used a visual task adapted from Harley et al. (2004), who showed subjects blurred celebrity faces that slowly become clearer second by second and asked subjects to identify the celebrity as early as possible. Afterward, subjects viewed the unblurred image and were asked either to identify the level of blur at which they made the identification or to predict the level of blur at which another subject could make an accurate identification. Hindsight bias was observed in both conditions: Subjects chose higher levels of blur than was present at the actual time points of identification and at which other naive subjects could not identify the celebrity in the photo.

Though visual judgments may seem to have little in common with the evaluation of historical or legal scenarios, the way that hindsight operates across such low- and high-level materials seems worthy of investigation. An influential cognitive-process model of hindsight (called Selective Activation and Reconstructive Anchoring, or SARA; Pohl, Eisenhauer, & Hardt, 2003) posits that hindsight arises from biased activation in the mental sampling process. This model may be applicable to the visual domain because biases in *perceptual* sampling (spatial attention, eye movements) are known to occur when the subject has been primed by previous exposure or semantic information (reviewed in Hannula et al., 2010). If such biases in front-end sampling contribute to visual hindsight bias, this would help explain why the bias is so cognitively impenetrable and would also suggest a point at which debiasing strategies could be directed.

Hindsight bias regarding the foreseeability of events has also been attributed to the subjective ease with which mental representations are generated (Sanna et al., 2002). Harley et al. (2004) propose a similar account for visual hindsight bias, which they term fluency misattribution. Exposure to a visual outcome increases the ease with which an ambiguous stimulus is processed, which leads to a misattribution of that ease to the predictability of the outcome. To the extent that these mechanisms are similar or overlapping, they might also produce similar effects across domains of probability judgments.

We used a visual task for three reasons. First, the post hoc evaluation of visual materials is central to many decisions of consequence. For example, medical images are often evaluated during litigation, and video recordings of public behavior are becoming ubiquitous. Second, there may be common or analogous mechanisms operating in different domains of judgment; comparing our visual results with those from higher-level cognitive paradigms will test how general hindsight bias is across different domains. Finally, if visual hindsight results are similar in size and reliability to those in cognitive judgments, then the visual task might serve as a useful model system because it has more direct tools than in other experimental domains: Visual-task difficulty can be controlled quantitatively, manipulations of visual attention are well-developed, and eye tracking can be used as a window into the mind to see how hindsight bias emerges from patterns of visual attention.

We conducted three experiments that established visual hindsight bias, showed how to predict its occurrence from eye tracking, and showed how to "debias" hindsight in a simple way. Further, we connected visual hindsight bias to the findings of a seminal cognitive study. The perceptual task in all three experiments was to decide whether blurred photos contained human beings. Subjects were undergraduate and graduate students at the California Institute of Technology and the University of California, Los Angeles. All protocols were approved by the institutional review boards of both schools. (Full methods for the three experiments are presented in the Supplemental Material available online.)

Experiment 1: Establishing Visual Hindsight Bias

Method

One hundred fifty-seven subjects participated in Experiment 1. One group of subjects ("performers"; n = 97) had to decide whether blurred photos contained human beings (50% of the photos did picture humans). Three other groups of subjects ("evaluators"; n = 20 each) estimated the proportion of performers who indicated the presence of a human in each photo. Performers were paid $10 for participation, plus a performance bonus of up to $10 based on a quadratic scoring rule that incentivized accurate reports of beliefs. Evaluators also earned $10 for participation and up to $10 for accurate performance, with a linear scoring rule determining their bonus. Subjects performed the task in groups of 8 to 20 in a computer lab setting, with dividers obscuring views between neighbors.

For all subjects, the main stimulus in each trial was a highly blurred photo that became clearer over time but stopped at a set point short of full clarity (Fig. 1). Stopping points were varied between photos to achieve a distribution of task difficulties (presentation times varied between 2 s and 26 s).

Performers saw the main stimulus at the beginning of each trial, and then were asked to rate the probability that a human was present in it. They were then given visual and verbal feedback in the form of a clear photo and an on-screen statement indicating whether a human was present, respectively.

For evaluators in the visual- and verbal-priming conditions, the main stimulus was preceded by an unblurred photo or an on-screen statement indicating whether a human was present in the photo, respectively (Fig. 1). Evaluators had to successfully perform a task to verify that they understood this prime in order to see the main stimulus. After seeing the main stimulus, they were asked to estimate the proportion of performers who reported that a human was present. Evaluators in the control condition did not receive a prime; they made their estimates after seeing the main stimulus alone.

Results and discussion

Data from evaluators' judgments in the visual-priming condition are shown in Figure 2a. The separation between the regression lines for stimuli with and without humans clearly indicates hindsight bias: When evaluators knew that a stimulus contained humans, they tended to overestimate the probability that performers would judge the stimulus as containing humans; the opposite was true for stimuli known to contain no humans. Results from an analysis of covariance (ANCOVA) indicate that slopes for the two types of stimuli were not significantly different (p = .42) but that the regression intercept for parallel fits was much higher for photos containing humans than for photos that did not contain humans ($\Delta = 0.35$), $F(1, 77) = 175.35$, $p < .0001$, $\omega^2 = 0.62$.

In the verbal-priming condition (Fig. 2b), regressions for the two types of stimuli also had slopes that were not significantly different (p = .52). The separation between intercepts ($\Delta = 0.18$) was about half the magnitude of the same separation in the visual-priming condition and was still highly significant, $F(1, 73) = 74.73$, $p < .0001$, $\omega^2 = 0.32$.

In the control condition (Fig. 2c), evaluators had no idea whether the picture contained a human, so we did not expect a difference in judgments about the two types of stimuli. Indeed, the regressions were very similar: There was no significant difference between slopes (p = .24) and a small but significant difference in the intercept ($\Delta = 0.08$), $F(1, 77) = 11.34$, $p = .0012$, $\omega^2 = 0.06$.

The results from the control condition show that without outcome knowledge, evaluators do not show strong biases. They treated photos with and without humans the same when translating their internal sense of task difficulty into an estimation of performers' choices. However, the outcome knowledge provided in the other conditions caused a large bias in estimates toward the known outcome.

Visual priming induced more hindsight than did verbal priming. An ANCOVA comparing regressions across those conditions showed that this was the case for photos with humans, $\Delta = 0.13$, $F(1, 75) = 29.25$, $p < .0001$, $\omega^2 = 0.27$, and for photos without humans, $\Delta = 0.04$, $F(1, 75) = 5.15$, $p = .03$, $\omega^2 = 0.05$. These differences indicate that a large portion of the effect was attributable to the visual processing initiated by the outcome prime, beyond the abstract outcome knowledge provided by a verbal prime.

This finding is consistent with the fluency-misattribution account of hindsight bias, according to which increased processing fluency for imagining an outcome leads one to consider it to be more foreseeable (Bernstein & Harley, 2007; Sanna et al., 2002; Whittlesea & Leboe, 2003). As compared with verbal information, visual experience has been found to be especially powerful in increasing processing fluency (Hawkins & Hastie, 1990; Roese, Fessel, Summerville,

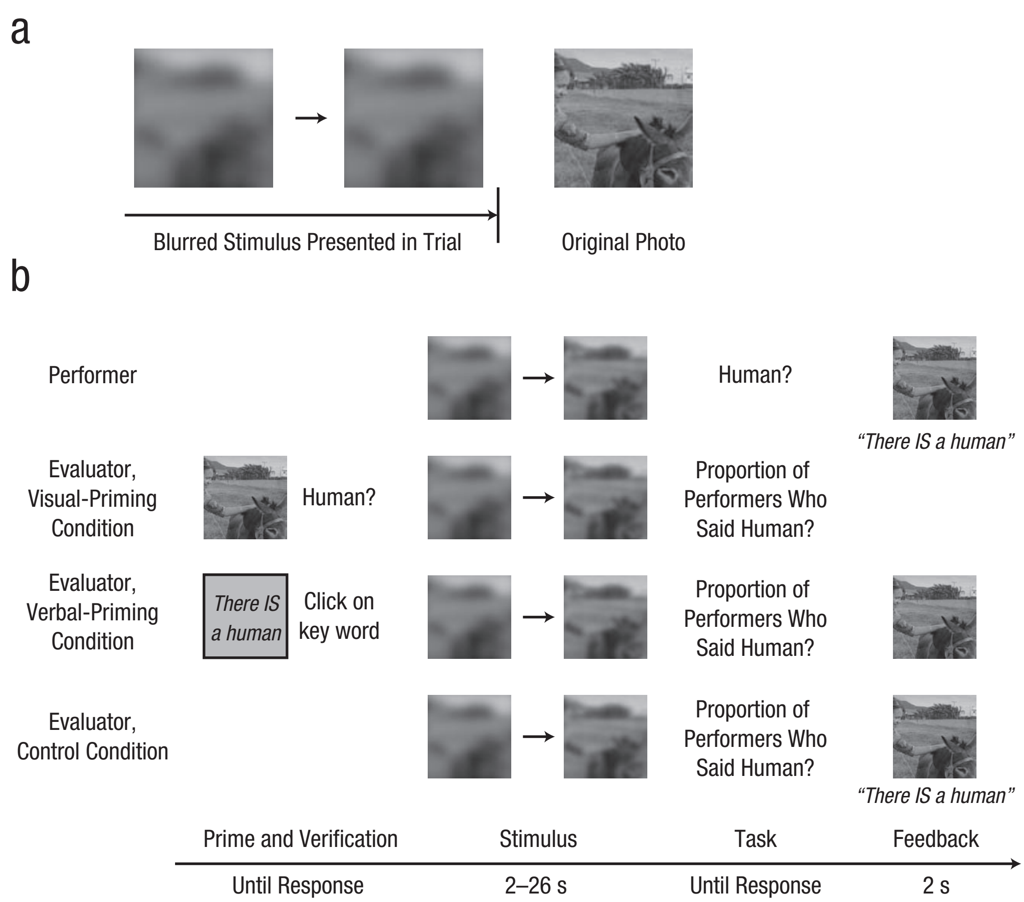

Fig. 1. Sample stimulus and trial sequences for Experiment 1. Each stimulus (a) was a highly blurred photo that clarified over time, terminating at a set point short of full clarity. Task design and trial sequences for the four groups are shown in (b). In each trial, performers were prompted to rate the probability that a human was present in the blurry stimulus. They were given visual and verbal feedback afterward in the form of the unblurred photo and a statement indicating whether a human was present. For evaluators in the visual- and verbal-priming conditions, the visual or verbal portion of the feedback, respectively, was moved to the beginning of the trial, thereby becoming a prime. Evaluators in these conditions had to successfully perform a task to verify that they understood this prime in order to advance to the main portion of the trial, in which they saw the same stimulus that was presented to performers. They then estimated the proportion of performers who reported that a human was present. Evaluators in the control condition did not receive a prime but made the same estimates based on the main stimulus alone.

Kruger, & Dilich, 2006). In our case specifically, differences in processing fluency could be generated by the effects of perceptual interference (Bruner & Potter, 1964): Visually primed subjects know where the target is throughout stimulus inspection; however, verbally primed subjects most likely generate a series of incorrect hypotheses about the blurry stimulus. Accordingly, visually primed subjects experience less uncertainty, more processing fluency, and more hindsight bias than verbally primed subjects do.

There is also a clear asymmetry to this visual component of the priming effect, which primarily shifts the estimates for the stimuli containing humans but leaves the estimates for the stimuli not containing humans largely unchanged. This is consistent with findings from recent studies of metacognition in detection tasks, which show an asymmetry in the ability of a subjects' perceptual confidence to predict accuracy in target-present and target-absent situations (Fleming & Dolan, 2010; Kanai, Walsh, & Tseng, 2010).

One coarse way to compare our results with the findings of other hindsight studies is on the basis of effect size. Guillbault's meta-analysis found that Cohen's d averaged 0.39 across studies (95% confidence interval = [0.36, 0.42]). Across our 80 images, we found effect sizes averaging 0.52 (SEM = 0.07), which would seem somewhat higher than average. However, Figure 2 shows that hindsight bias varied greatly as a function of stimulus difficulty (as indexed by performers' naive response probabilities). A deeper comparison would incorporate this interaction by including the variation across the range of stimuli; therefore, we compared our hindsight data with data from Fischhoff and Beyth's (1975) study of

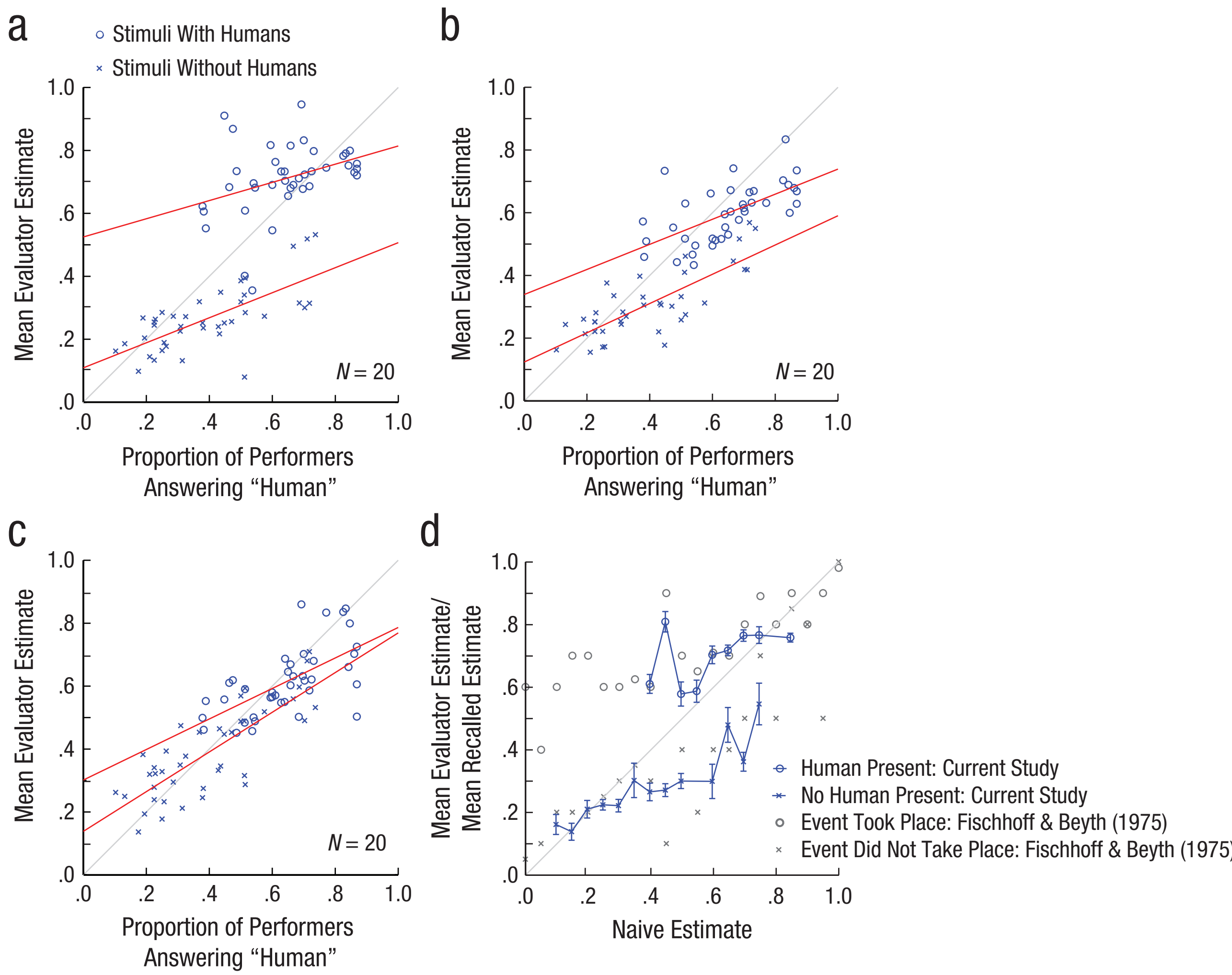

Fig. 2. Results for Experiment 1. In (a) through (c), the scatter plots (with best-fitting regression lines) show mean evaluator estimates of the proportion of performers who indicated the presence of a human in each stimulus as a function of the actual proportion of performers' responses and stimulus type. Results are shown separately for the (a) visual-priming condition, (b) verbal-priming condition, and (c) control condition. Accurate estimates would lie along the diagonal. Hindsight bias would result in a separation between regressions for the two stimulus types. In (d), the visual-priming data from (a) are overlaid on a replotting of data from Fischhoff and Beyth's (1975) study of hindsight on historical events. Subjects in that study estimated the likelihood of future events and then attempted to recall their estimates after event outcomes were known. Mean recalled estimates from Fischhoff and Beyth's study are presented as a function of actual estimates and whether or not an event occurred. Data from the current study were binned to be consistent with Fischhoff and Beyth's data. Error bars indicate ±1 *SEM*.

hindsight on historical events (described previously in the introduction). Their study tapped a higher-level domain of judgment and required subjects to make estimates about their own past beliefs (rather than the beliefs of other people, as in our study). Nevertheless, direct comparisons of the quantitative results are sensible because, in both studies, probabilities regarding binary outcomes were measured and probability estimates made by informed subjects were analyzed as a function of corresponding probabilities generated by naive subjects.

Data from our visual-priming condition were binned and averaged following the procedures used by Fischhoff and Beyth (1975) so that the two sets of data could be directly compared (Fig. 2d). The mutual fit was quite good ($R^2 = .87$), $t(20) = 11.67$, $p < .0001$. This reflects a correspondence that is as strong as two linear regressions fitted directly on the historical data ($R^2 = .87$), $t(20) = 11.62$, $p < .0001$. This quantitative correspondence suggests that hindsight bias might be parametrically robust across a wide range of paradigms and judgment domains. The numerical comparability of our results with those of the earliest and the meta-analyzed studies suggests that visual priming could be used as a model system to study hindsight in a general sense. This would allow researchers to take advantage of experimental methods available in the vision sciences. The next two experiments go in this direction.

Experiment 2: Eye Tracking

As mentioned previously, hindsight bias in the evaluation of verbal scenarios and questions has been attributed to sampling biases in the exploration of possible outcomes (Pohl et al., 2003). Subjects with outcome knowledge may have an increased tendency to evaluate the materials in terms of their relation to the known outcome while neglecting alternatives that would be considered by subjects without outcome knowledge. It is difficult to monitor such evaluation processes in the case of verbal scenarios, but in our visual task, we monitored the distribution of subjects' attention by tracking the location of their eye gaze. We could therefore directly test the possibility that visual sampling bias underlies hindsight bias.

Method

For Experiment 2, we recruited new sets of performers (n = 31) and evaluators (n = 19), who performed the same task with the same stimuli as in Experiment 1. However, evaluators were tested solely in the visual-priming condition. Furthermore, we tracked the eye movements of all subjects using an EyeLink II (SR Research, Mississauga, Canada) head-mounted eye tracker. Incentives for performance were the same as in Experiment 1, except that the base participation payment was raised to $20 to compensate for any discomfort from the eye tracking. Subjects were tested individually and were different from those in Experiment 1. For each image, we calculated two spatial maps of subject gaze distributions: one for performers as a group and one for evaluators. The degree of spatial correspondence between the distributions (i.e., their similarity) was calculated for each stimulus by summing the zero-phase two-dimensional cross-correlation of the two probability maps (illustrated in Fig. 3a).

If the attention-based hypothesis for hindsight is valid, then we would expect that stimuli that were viewed most differently by the evaluators and performers would generate more hindsight bias. To avoid spurious connections between a given subject's gaze patterns and his or her own reports, we evaluated this hypothesis using subject reports from Experiment 1, which involved a different set of subjects.

Results and discussion

Regression analysis of hindsight against the similarity metric showed significant results but included a few outlying data points with excessive influence, which possibly exaggerated the correlation (see Results in the Supplemental Material). Thus, we used a more conservative nonparametric analysis, dividing the images into quartiles according to the gaze-similarity metric and comparing median hindsight bias elicited by images in the top and bottom quartiles (Fig. 3). To test the significance of these differences, we ran a bootstrapping procedure to determine the empirical distributions of the medians. We resampled the individual eye-tracking trials to generate 5,000 sets of gaze data; for each synthetic data set, we recalculated the gaze maps, similarity metrics, quartiles, and medians.

For images containing humans, we found that the bottom quartile (least similarity between gaze patterns) elicited significantly more hindsight (95% CI = [0.097, 0.169]) than did the top quartile (most similarity between gaze patterns; 95% CI = [0.018, 0.104]; p = .011). Meanwhile, for images without humans, the difference between quartiles did not reach significance (bottom quartile: 95% CI = [0.062, 0.156], top quartile: 95% CI = [0.112, 0.175]; p = .095).

These results are consistent with the hypothesis that differences in the distribution of attention are a cause of hindsight bias. That the correlation was only significant in stimuli containing humans is consistent with the patterns of hindsight seen in Experiment 1; although the visual component of hindsight (visual vs. verbal forms of priming) was found to be quite large for stimuli with humans, less difference was found for stimuli without humans. An intuitive rationale is that target-present primes have obvious locations toward which attention might be biased, whereas it is unclear how the location of attention would be systematically affected when the prime lacks a target. It is also possible that more sophisticated measures of eye movements and attention would reveal further hindsight-relevant dimensions along which performers and evaluators differ.

Experiment 3: Hindsight Debiasing

The results of Experiment 2 connect hindsight bias to a mismatch between gaze patterns of performers and evaluators. This front-end sampling bias is analogous to the SARA model proposed by Pohl et al. (2003). Although top-down cognitive strategies aimed at changing these mental sampling patterns have not been shown to reliably decrease hindsight bias, the eye-tracking data in our study may provide a bottom-up approach to debiasing. The performers' gaze patterns are a clear record of attention patterns in the naive state. Presenting these patterns to evaluators may provide them with information needed to alter their own attention patterns, take note of neglected regions, and reduce misattributions of perceptual fluency.

The transfer of gaze patterns across subjects of differing expertise has been tested in cooperative puzzles (Velichkovsky, 1995) and practical training situations in which novices could benefit from knowledge of how an expert would view the subject materials. Novices benefited from this "gaze transfer" in examining chest x-rays, aircraft fuselages, and computer programs (Litchfield, Ball, Donovan, Manning, & Crawford, 2008; Mehta, Sadasivan, Greenstein, Gramopadhye, & Duchowski, 2005; Stein & Brennan, 2004). Could this method be useful in the opposite direction, when transferring gaze from novice to expert?

Method

In Experiment 3, we modified the visual-priming condition of Experiment 1 by splitting the experimental session into two halves. In the first half (baseline hindsight treatment), the

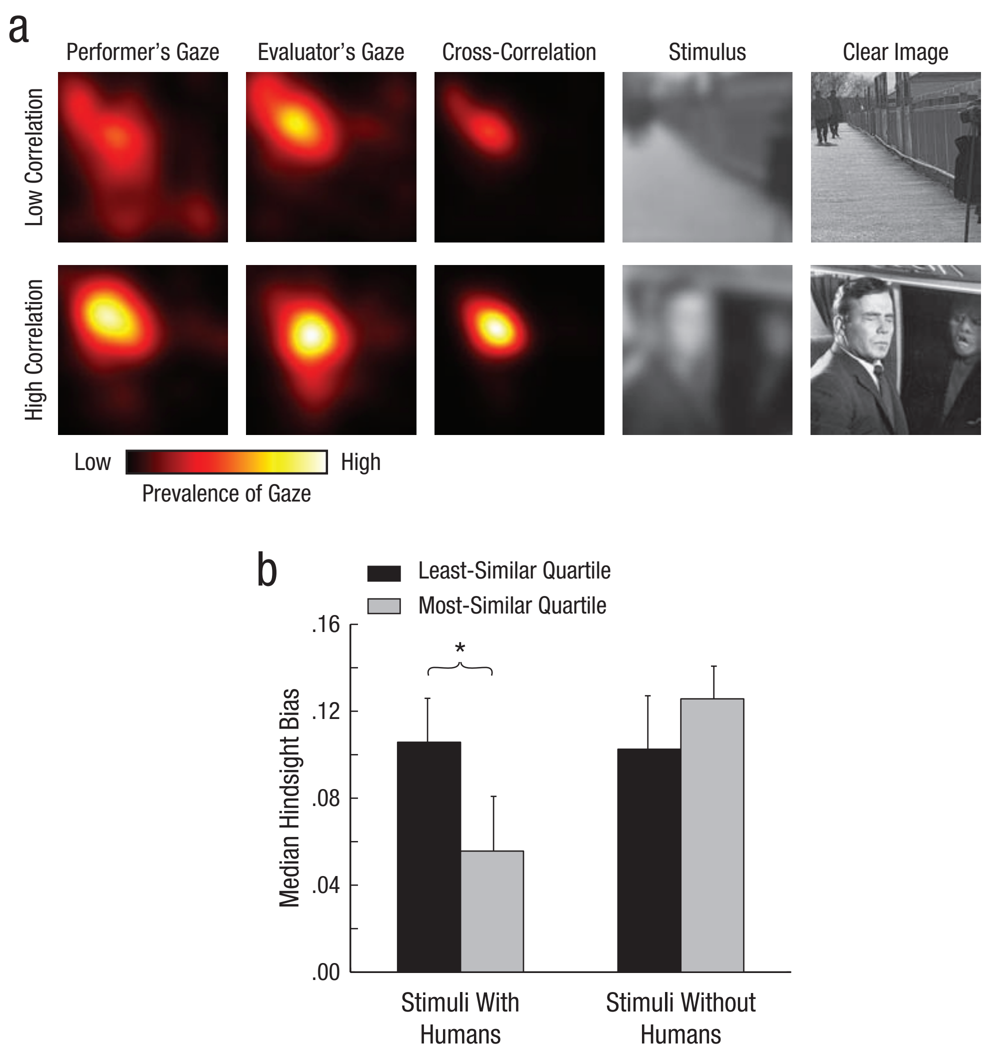

Fig. 3. Analysis of hindsight-bias data from Experiment 1 based on gaze distributions from Experiment 2. The examples in (a) illustrate how stimuli were sorted into quartiles according to performers' and evaluators' gaze similarity. The example in the top row is from the least-similar quartile, and the example in the bottom row is from the most-similar quartile. These particular stimuli elicited mean hindsight biases of 0.098 and 0.024, respectively, in Experiment 1. For each example, the figure shows the spatial distribution of performers' gazes, the distribution of evaluators' gazes, and the correlation between the two, as well as the stimulus and clear image from which these data were taken. The graph in (b) shows median hindsight bias elicited in Experiment 1 as a function of stimulus type and quartile of gaze similarity. Error bars represent standard deviations of the bootstrapped empirical distribution. The asterisk indicates a significant difference between quartiles ($p < .05$).

stimuli were the same as in Experiment 1. In the second half (gaze overlay treatment), performers' gaze patterns were depicted as dots moving across the surface of the stimulus. To prevent evaluators from using spurious connections between a given performer's eye movement and responses, we used performers' gaze data from Experiment 2 and scored evaluator estimates based on performer response data from Experiment 1. Subjects were informed of the source of the gaze data but were not given any explicit strategy for using it. Subjects were tested in groups and were different from those in the previous experiments.

We calculated within-subjects changes (n = 34) in root-mean-square error (RMSE) and in mean bias from the first half to the second half of the experiment. One-tailed statistical tests for significant decreases were based on weighted means across subjects, with subject weights computed from the inverse of the variance of their trial-by-trial errors. As a control, we ran the same analysis on subjects (n = 37) who performed the ordinary visual-priming task in both halves of the experiment. Given the pattern of results found in the preceding experiments, we analyzed stimuli with humans and stimuli without humans separately (p values are Bonferroni-corrected).

Results and discussion

Gaze-transfer treatment effects showed the same pattern seen in the preceding experiments, with large and significant effects on evaluations with stimuli that contained humans and nonsignificant effects with stimuli that did not contain humans (Fig. 4a). For stimuli with humans, RMSE was reduced by 0.039, $t(29.1) = 3.69$, $p < .001$, $d = 0.69$, and hindsight bias was reduced by 0.035, $t(29.1) = 2.08$, $p < .05$, $d = 0.32$. Reductions in RMSE and hindsight bias for the pictures without humans did not reach significance (both ps > .2). In control subjects, no significant reductions were found (all ps > .2; Fig. 4b).

Direct comparisons showed that the effect of gaze transfer on RMSE was significantly greater in the gaze-debiasing condition than in the control condition, Welch's $t(56.25) = 2.58$, $p < .01$, $d = 0.66$. Meanwhile, the reduction in hindsight bias was not significantly greater in the gaze-debiasing condition than in the control condition, $t(60.42) = 0.84$, $p = .20$, $d = 0.23$.

The treatment effects were substantial, as measured against the bias magnitudes exhibited in the initial half of the experiment. Hindsight bias for subjects in the gaze-transfer condition declined by a little more than half, from 0.067 to 0.032. Control subjects showed a declining trend, from 0.079 to 0.063. A notable difference between these declines is that the bias shift in subjects in the gaze-transfer condition was accompanied by an equivalent improvement in accuracy, as measured by RMSE. The small bias reduction in control subjects did not improve their RMSE accuracy.

It is apparent that some control subjects may have figured out our rather blatant priming manipulation and attempted to compensate. However, such blind bias shifts do not necessarily improve the accuracy of evaluations. As shown in Figure 2a, the magnitude of overestimation in evaluator estimates varies across stimuli of different difficulties. Shifts in estimates can therefore be subject to both under- and overshooting for stimuli in different locations in the range. The gaze-transfer treatment may be helpful in providing a basis for estimating performer reliability, allowing evaluators to calibrate and selectively apply shifts in their estimates.

General Discussion

Experiment 1 established between-subjects hindsight bias using a visual paradigm. We found that visual priming led to more hindsight bias than verbal priming did, which suggests that visual-processing fluency enhanced the strength of the effect. A strong quantitative similarity between our data and those of Fischhoff and Beyth (1975), despite the use of very different methods, underscores the appropriateness of this visual paradigm for studying hindsight as a general phenomenon.

In Experiment 2, we tracked the eye movements of both the performers and the evaluators in the visual-priming task and found evidence that an attentional-sampling bias contributes to hindsight bias. The magnitude of hindsight bias varied with the degree of difference between performer and evaluator attention patterns. In Experiment 3, we were able to partially debias evaluators by showing them the gaze patterns of the performers as the evaluators looked at the pictures. Putting the knowledgeable evaluators in the eyes of the naive performers in this way reduced evaluators' hindsight bias by half, objectively improving performance (cf. Davies, 1987, Experiment 1).

Our results suggest several future directions for hindsight research. Eye-tracking data is a rich source of attention-related

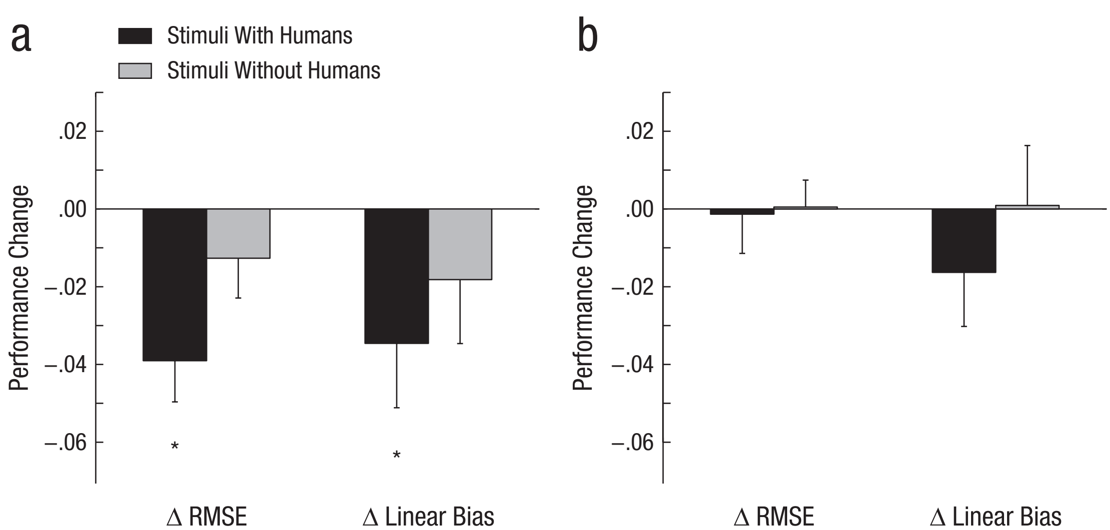

Fig. 4. Weighted mean changes in performance for (a) subjects who received the gaze-debiasing treatment and (b) subjects who received no treatment. Performance changes between the first and second halves of Experiment 3 were measured as the change in root-mean-square error (RMSE) and linear bias. Results are shown separately for each type of stimulus. Error bars indicate 1 *SEM*. Negative values indicate improvement in performance, with significant improvement indicated by an asterisk ($p < .05$).

information and can be used to test a range of existing attention-based hypotheses. Regarding the debiasing technique, more work needs to be done to understand the impact of being observed on the performers' attention allocation and which types of observations are most effective for reducing hindsight bias. Further, it remains to be seen how widely the technique can be applied across different kinds of ambiguity or image degradation.

Practically speaking, the relevance of visual hindsight and its debiasing increases as technology progresses. More and more, decision-making processes produce audiovisual records available for post hoc review (e.g., from e-mail and Web browsing, video conferences, and operation of military drones). Recordings are also made proactively to increase accountability (e.g., in police work). Naive and informed viewing patterns on such materials could be acquired using eye-tracking technology that is increasingly affordable and unobtrusive. Given the pressures on decision makers to protect themselves from second guessing, even small amounts of protection offered by such data could motivate the integration of such systems into professional practice.

Acknowledgments

We wish to thank Daniel M. Bernstein, Daniel J. Simons, and an anonymous reviewer for their helpful comments on the manuscript.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was funded by grants from Trilience Research (D.-A. W., C. F. C., and S. S.), the Betty and Gordon Moore Foundation (C. F. C. and S. W. W.), Tamagawa-Caltech Global COE program (C. F. C. and S. S.), and the CREST program of the Japan Science and Technology Agency (S. S.).

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

References

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, *2*, 305–307.

Bernstein, D. M., Erdfelder, E., Meltzoff, A. N., Peria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of age. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 378–391.

Bernstein, D. M., & Harley, E. M. (2007). Fluency misattribution and visual hindsight bias. *Memory*, *15*, 548–560.

Bruner, J. S., & Potter, M. C. (1964). Interference in visual recognition. *Science*, *144*, 424–425.

Camerer, C. F., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*, 1232–1254.

Davies, M. F. (1987). Reduction of hindsight bias by restoration of foresight perspective: Effectiveness of foresight-encoding and hindsight-retrieval strategies. *Organizational Behavior and Human Decision Processes*, *40*, 50–68.

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 387–414.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288–299.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 349–358.

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen": Remembered probabilities of once-future things. *Organizational Behavior and Human Performance*, *13*, 1–16.

Fleming, S. M., & Dolan, R. J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Consciousness and Cognition*, *19*, 352–363.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York, NY: Cambridge University Press.

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

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4*, 1–16.

Harley, E. M. (2007). Hindsight bias in legal decision making. *Social Cognition*, *25*, 48–63.

Harley, E. M., Carlsen, K. A., & Loftus, G. R. (2004). The "saw-it-all-along" effect: Demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 960–968.

Hasher, L., Attig, M. S., & Alba, J. W. (1981). I knew it all along: Or, did I? *Journal of Verbal Learning and Verbal Behavior*, *20*, 86–96.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311–327.

Kanai, R., Walsh, V., & Tseng, C.-H. (2010). Subjective discriminability of invisibility: A framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition*, *19*, 1045–1057.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 107–118.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2008). Learning from others: Effects of viewing another person's eye movements while searching for chest nodules. In B. Sahiner & D. J. Manning (Eds.), *Medical imaging 2008: Image perception, observer performance, and technology assessment* (Proceedings of SPIE, Vol. 6917). Bellingham, WA: SPIE.

Loewenstein, G., Moore, D. A., & Weber, R. A. (2006). Misperceiving the value of information in predicting the performance of others. *Experimental Economics*, *9*, 281–295.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*, 1231–1243.

Madarasz, K. (2011). Information projection: Model and applications. *The Review of Economic Studies*. Advance online publication. doi:10.1093/restud/rdr044

Mehta, P., Sadasivan, S., Greenstein, J., Gramopadhye, K., & Duchowski, T. (2005, September). *Evaluating different display techniques for communicating search strategy training in a collaborative virtual aircraft inspection environment*. Paper presented at the Human Factors and Ergonomics Society Annual Meeting, Orlando, FL.

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.

Pohl, R. F. (2007). Ways to assess hindsight bias. *Social Cognition*, *25*, 14–31.

Pohl, R. F., Bender, M., & Lachmann, G. (2002). Hindsight bias around the world. *Experimental Psychology*, *49*, 270–282.

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory*, *11*, 337–356.

Pohl, R. F., & Hell, W. (1996). No reduction of hindsight bias with complete information and repeated testing. *Organizational Behavior and Human Decision Processes*, *67*, 49–58.

Roese, N. J., Fessel, F., Summerville, A., Kruger, J., & Dilich, M. A. (2006). The propensity effect: When foresight trumps hindsight. *Psychological Science*, *17*, 305–310.

Sanna, L. J., Schwarz, N., & Stocker, S. L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 497–502.

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 544–551.

Stallard, M. J., & Worthington, D. L. (1998). Reducing the hindsight bias utilizing attorney closing arguments. *Law and Human Behavior*, *22*, 671–683.

Stein, R., & Brennan, S. E. (2004, October). *Another person's eye gaze as a cue in solving programming problems.* Paper presented at the 6th International Conference on Multimodal Interfaces, State College, PA.

Studdert, D. M., Mello, M. M., Sage, W. M., DesRoches, C. M., Peugh, J., Zapert, K., & Brennan, T. A. (2005). Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *Journal of the American Medical Association*, *293*, 2609–2617.

Terry, K. (2011). What is smart sex offender policy? *Criminology & Public Policy*, *10*, 275–282.

Velichkovsky, B. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition*, *3*, 199–224.

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, *49*, 62–79.