The “Saw-it-All-Along” Effect: Demonstrations of Visual Hindsight Bias*

Erin M. Harley
University of California, Los Angeles

Keri A. Carlsen and Geoffrey R. Loftus
University of Washington

*The writing of this article and the laboratory described in it were supported by NIMH Grant MH41637 to G. Loftus. We thank Daniel Bernstein for valuable comments on previous versions of this manuscript. We also thank Chris Dessert, Eunice Poon, Sarah Hayes, and Erik Gust for assistance in data collection.
Abstract

This article addresses whether a hindsight bias exists for visual perception tasks. In three experiments observers equipped with visual target information predicted the performance of a naïve self or peer. In phase 1 of all experiments, observers identified degraded images of celebrity faces as they resolved to full clarity. Following phase 1, Experiment-1 and 2 observers were given a surprise memory test in which they attempted to adjust the clarity of each face until it looked as it did when recognized in phase 1. Experiment-3 observers viewed phase-1 faces plus a new set of faces in full clarity at the outset, and then stopped the resolution process when they thought a naive peer would identify the face. In all experiments, observers overestimated identification performance of naïve observers. This visual hindsight bias was greater for more familiar faces—those shown in both phases of the experiment—and was not reduced following education and instructions to observers to avoid the bias. We conclude that once the content of a visual image is known, observers are unable to discount this information. As a result, they overestimate their ability, and the ability of others, to identify the image at a more degraded state. We propose a fluency misattribution theory to account for the bias. Implications for medical malpractice litigation and eyewitness testimony are discussed.
Imagine the following scenario: A patient at a local hospital undergoes routine chest radiography at time 1, and Radiologist 1 interprets the films as normal. Three years later, the patient experiences discomfort and undergoes chest radiography a second time, whereupon Radiologist 2 discovers a large tumor. Despite treatment, the patient dies. The patient’s family files a medical malpractice lawsuit against Radiologist 1, claiming the tumor should have been detected at time 1. Radiologist 2, testifying on the family’s behalf, views the original time-1 radiographs and, seeing the tumor missed by Radiologist 1, claims that it was visible at time 1.

Such a scenario is typical; the vast majority of medical malpractice litigation in Radiology over the past 20 years has involved diagnostic, perceptual, or decision-making errors (Berlin, 1996a, 1996b; Berlin & Hendrix, 1998). In such cases, the defendant radiologist being accused of negligence for a decision made prospectively (without the benefit of outcome knowledge) is often judged by radiology experts who have full knowledge of what future radiographs revealed (Berlin, 2000). Are such experts biased by this knowledge; i.e., can a radiologist who views a tumor at time 2 make an appropriate prediction about whether another radiologist should have detected the tumor when it was smaller and less visible at time 1? This is a question of visual hindsight, which is akin to verbal hindsight reported in the past by many investigators.

First reported by Fischhoff (1975), hindsight bias, or the knew-it-all-along effect (Wood, 1978), is the tendency for individuals with outcome knowledge to claim that they would have estimated a higher probability of occurrence for the reported outcome than was estimated in foresight. In other words, it is the after-the-fact feeling that some outcome was very likely to happen, or was predictable, even though it was not predicted to happen beforehand. Hindsight bias is thought to result from cognitive reconstruction processes that occur after outcome information is received (for a review see Hawkins & Hastie, 1990). Judges reanalyze the event so that the beginning and the middle connect causally to the end. During this process evidence consistent with the reported outcome is elaborated and evidence inconsistent with the outcome is minimized or discounted. The result of this “rejudgment” process is that the given outcome seems inevitable, or at least more plausible than alternative outcomes.

The verbal hindsight bias effect has been shown to be quite robust, occurring in both within- and between-subject designs, in spite of explicit instructions to subjects to avoid the bias, and across a range of time intervals between initial judgments, outcome feedback, and second judgments (see Christensen-Szalanski & Willham, 1991; Hawkins & Hastie, 1990; Fischhoff, 1982; for reviews). The bias is sensitive to task difficulty—it is greater for difficult than easy items, and greater for events initially judged to be
least plausible (Fischhoff, 1977; Wood, 1978; Arkes, Wortmann, Saville, & Harkness, 1981)—and is greater when evidence supporting the given outcome is more easily brought to mind (Sanna, Schwarz, & Small, 2002). Hindsight bias has been studied extensively in the cognitive domain, with broadly ranging types of events, e.g., outcomes of historical events, psychiatric cases, scientific experiments, consumer purchases, sporting events, economic decisions, election outcomes, medical and legal cases, and answers to almanac trivia questions.

In contrast, there has been remarkably little research on the kind of visual hindsight bias exemplified by the radiology example sketched earlier. The main purpose of this article is to begin to correct this deficit. We first describe how hindsight bias might apply to visual perception, and in this context, introduce the broader topic of metaperception—people’s insights into their perceptual abilities. Next, we present two experiments that show evidence of visual hindsight bias, and propose a theory of fluency misattribution to account for the bias. Finally, we present a third experiment in which a prediction of the fluency-misattribution theory is tested and confirmed.

**Visual Hindsight Bias**

To illustrate how hindsight bias might apply to visual perception we return to the medical scenario described earlier. In the field of Radiology, it is common in medical malpractice litigation for a physician with outcome knowledge to judge whether another physician, who did not have the benefit of such knowledge, should have detected an abnormality on a medical image (Berlin, 2000). When making this judgment, it may upon initial inspection seem reasonable for the judging physician to simply assess his or her own ability to see the missed abnormality. However, the physician cannot use this method if the missed abnormality is more easily detected in hindsight.

There is evidence to suggest that medical abnormalities are in fact more visible when the physician has outcome knowledge. A striking example of tumors only evident in hindsight comes from Muhm, Miller, Fontana, Sanderson, and Uhlenhopp (1983), who conducted a screening program at the Mayo Clinic for men at high risk of lung cancer. The 4,618 members of the study group obtained chest radiographs every four months, each of which was read by 2-3 radiologists or chest physicians. Over the course of the 6-year study, 92 tumors were detected in the study group. Of these, 75 (82%) were, as the authors termed it, “visible in retrospect (p. 611).” This means that when the physicians looked at the previous sets of radiographs, in 82% of the cases they found they could detect the tumor on at least one,
and often multiple, prior radiographs (ranging from 4 to 53 months prior) that had initially been interpreted as normal.

When speculating as to why the 75 tumors visible in hindsight were not detected on initial readings, Muhm et al. (1983) made the following comment: “The fact that some of the cancers had been overlooked was usually due to perceptual errors [italics added] by the observers” (p. 612). Does it make sense that 2-3 experienced radiologists made errors on 82% of the tumor-containing radiographs? We argue that the error may well lie not in the physicians’ initial interpretations of the radiographs, but rather in their retrospective interpretations of the radiographs. The “overlooked” tumors became visible in hindsight not because they were more detectable to begin with, but because the physicians had the benefit of outcome knowledge.

Multiple factors contribute to allow a radiologist with outcome knowledge, or hindsight, to detect abnormalities that previously could not be seen. Since Bartlett’s (1932) pioneering work, and Neisser’s landmark *Cognitive Psychology* (1967), visual perception has been construed as an active and creative process involving far more than a direct translation of images projected onto the retina. What we perceive is influenced by top-down processes that include prior knowledge, expectations, context, and a large number of assumptions about how objects in the world behave (see, e.g., Palmer, 1999). The radiologists who initially read the films in the Mayo Clinic’s screening program had relatively low expectations for finding a tumor, and if a tumor were present, they had no knowledge of what type of tumor to search for, or where the tumor might be located. In contrast, the post-detection viewing conditions were quite different. During retrospective analysis of the radiographs the physicians had high expectations for finding a tumor, had full knowledge of both tumor type and tumor location, and because the outcome was known, no longer had to be cautious about avoiding false positives. These factors combined to make visible many of the abnormalities that had originally gone undetected.

There is some research indicating that observers who know what they are looking for will have higher detection rates in a difficult visual search task. Bruner and Potter (1964) reported that observers claim to be able to perceive a visual target at a more degraded state when viewed in a clear-to-blurry progression than in a blurry-to-clear progression. Given that target information can improve an observer’s ability to detect or identify a visual target, an observer asked to estimate the performance of a naïve peer is faced with a difficult challenge. The observer cannot simply assess his or her current detection ability—it has

1. Even though the patients were at high risk for lung cancer, over the 6-year screening only 92 tumors were detected in a group of 4,618 patients (roughly 2% of the sample).
been enhanced by outcome information— but instead must discount target information, and imagine what a naïve observer would perceive.

Metaperception

When asked to estimate the visual performance of a naïve self or peer, an observer must attempt to gain insight into the strengths and limitations of their perceptual abilities, as well as how these abilities are affected by experience and knowledge. We will term this insight *metaperception*. Although a large amount of research has been dedicated to the investigation of metacognition, and most specifically, metamemory (see Metcalfe & Shimamura, 1994; Mazzoni & Nelson, 1998; Chambres, Izaute & Marescaux, 2002, for reviews), very few researchers have explored metaperception.

Recent work by Levin, Momen, Drivdahl, & Simons (2000), and Levin (2002) has investigated metaperception in relation to change blindness, the counterintuitive finding that observers commonly fail to detect large visual changes in their environment (see, e.g., Blackmore, Brelstaff, Nelson, & Troscianko, 1995; Henderson, 1997; Pashler, 1988; Phillips, 1974; Rensink, O'Regan, & Clark, 1997; for a review see Simons, 2000). The existence of change blindness suggests that observers do not retain many visual details in memory from one view to the next, and that focusing attention on the changing item is critical for successful change detection. Most relevant to the present work is that change blindness is a startling finding in that it contradicts people’s intuitions about their perceptual abilities. Prior to experience with change-detection tasks, most observers wrongly assume that they will have no trouble detecting changes in visual scenes (Levin, Drivdahl, Momen, & Beck, 2002). For example, Levin et al (2002) found that 90% of participants believed they would notice a scarf disappear on an actor across a movie cut that 0% of observers in the original experiment noticed. This metacognitive error has been termed *change blindness*, and is evidence that under some conditions naïve observers have grossly inaccurate insights into their own and others’ perceptual abilities. This conflict between actual visual performance and estimated visual performance warrants further investigation of metaperception. Specifically, under what conditions will observers make inaccurate estimates of their perceptual abilities? Judging in hindsight may be one such case.

Only two studies that we are aware of have examined hindsight bias in the visual domain. Winman, Juslin, & Bjorkman (1998) found a reverse visual hindsight bias, i.e., an underestimate of performance, when observers with outcome information estimated prior, naïve performance in a two-alternative-forced-choice line-length discrimination task. In contrast, Harley, Carlsen, & Loftus (2001) found that observers
show positive hindsight bias when predicting the performance of their peers in a digit-identification task, but only when the task was most difficult. The conflicting results of these two studies suggest that we cannot assume a priori that hindsight bias will exist for all perceptual tasks, nor can we assume that if it does it will be a positive bias, like the traditional verbal hindsight bias, rather than a negative bias, like the one found by Winman et al (1998).

Another reason we cannot assume a priori that visual hindsight bias will mirror verbal hindsight bias is that confidence in foresight knowledge judgments appears to trend differently in the intellectual and sensory domains. People tend to be overconfident when making intellectual judgments, e.g., “Who was the third president of the United States”, but underconfident when making sensory judgments, e.g., “What color are your colleague’s eyes?” (e.g., Adams, 1957; Bjorkman, Juslin, & Winman, 1993; Dawes, 1980; Keren, 1988; Olsson & Winman, 1996; Winman and Juslin, 1993). If hindsight confidence ratings are based on the same information as foresight confidence ratings—Winman, Juslin, & Bjorkman (1998) suggest that the two are both based on an assessment of task difficulty—then a reverse hindsight bias may prove more prevalent in the visual domain. If, on the other hand, hindsight bias is a general phenomenon that affects decisions made in multiple domains, including visual, then we expect to find evidence for a positive visual hindsight bias, i.e., an overestimation of naïve performance after target identity is known.

In a series of experiments reported here we tested the hypothesis that hindsight bias is a general phenomenon—an overestimation of foresight knowledge following the receipt of outcome knowledge—that is not restricted to the intellectual domain, but occurs in the visual domain as well. To test this hypothesis we examined whether an observer with knowledge of target identity, akin to outcome knowledge, could accurately predict the level of visual degradation at which a naïve self or peer would be able to identify a celebrity face. In all experiments observers exhibited visual hindsight bias. The bias was exhibited for judgments made about both self and others, as well as in spite of education and warnings to avoid the bias. We propose a fluency-misattribution theory to account for visual hindsight bias and provide confirmatory evidence of a prediction of the theory in Experiment 3.

**Experiment 1**

In Experiment 1, observers identified degraded pictures of celebrity faces as they gradually clarified. Later, in a surprise memory test, observers recalled the degree of degradation present at time of original identification.
Method

Observers

Forty-two University of Washington undergraduates, all with normal or corrected-to-normal vision, participated in exchange for course credit.

Apparatus

Data-collection took place in a room equipped with four Macintosh eMac computers—each equipped with a G4 processor and a 17" monitor—allowing for up to four observers to participate in each data-collection session. Curtains were hung between computers to prevent observers from viewing other monitors during the session. Each data collection session lasted approximately 30 minutes. The experiment was written and executed in MATLAB using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

Stimuli

Stimuli were 36 grayscale pictures of celebrity faces. The set included well-known actors, musicians, politicians, and sports figures, e.g., Jerry Seinfeld, Harrison Ford, Madonna, Hillary Clinton, and Michael Jordan (see Figure 1). Each face measured 500 pixels from the bottom of the chin to the top of the head, and subtended a visual angle of about 21.8° vertically. The display monitor’s background luminance measured 2.94 cd/m².

Thirty successively more blurred versions of each face were created. Each blurred version was accomplished by Fourier-transforming the image from pixel space into spatial-frequency space, multiplying the resulting frequency amplitude spectrum by a Gaussian low-pass filter and inverse-Fourier transforming the result back into pixel space. The low-pass filter was designed such that for each blur level it passed frequencies perfectly, i.e., had a value of 1.0 up to some value \( f_0 \) cycles per face height, and then fell parabolically, reaching zero at the cutoff frequency value of \( f_c = f_0 \times 3 \) cycles per face height. As \( f_0 \) and \( f_1 \) are made smaller, the filter cuts off more spatial frequencies, and the resulting face becomes blurrier (see Loftus, 2001 for additional explanation). See Figure 1, row 2, for samples of a filtered face. In each of the three experiments reported here, \( f_1 \), the filter cutoff frequency, was used as the dependent variable—a measure of the degree of blur present in the image when the observer was able to (or believed they would be able to) identify the celebrity. We reemphasize that a blurrier picture is implied by a smaller \( f_1 \) value.
Design and Procedure

Phase 1 of the experiment, baseline identification, was a simple identification test. For each face, the 30 blur images were displayed, in order, from most to least blurred, at a rate of 500 ms per image. To the observer it appeared as if the celebrity’s face was slowly clarifying over time. Observers were read aloud the following instructions:

Each celebrity will start very blurry and slowly become clear. Press the space bar as soon as you recognize who the celebrity is. After you press the space bar type in this guess and hit RETURN. After you enter your guess the picture will continue to become clear. If your guess changes hit the space bar again and type in a new guess. You can guess as many times as you wish until you are certain you have identified the celebrity correctly. If you get all the way to the clearest image and you don’t know who the celebrity is just type in ‘don’t know’ or ‘?’. Observers were allowed to identify a celebrity in a number of ways, including, any portion of the celebrity’s name, the name of a character they play on TV, the name of a movie in which they have starred, etc. Anything that would indicate to the investigators that the celebrity was recognized was scored.

Figure 1. Samples of celebrity faces used in Experiments 1-3. Row two contains 4 of the 30 low-pass filtered versions of Jennifer Aniston, with corresponding $f_1$ (cycles per face height) filter cutoff values.
as an accurate response. At the completion of each trial the observer was asked to verify his or her final identity guess. For each celebrity face all 30 blur images were displayed regardless of whether or when the face was identified. This was done to equate as best as possible the total time an observer viewed each face. The order in which the 36 celebrities appeared was randomized for each observer.

In phase 2 of the experiment, observers were given a surprise memory test. The same celebrities were shown again in a different, random order. At the start of each trial the face was shown in its most degraded form and the observer’s response from phase 1 was printed on the screen below the face so that the celebrity’s identity would be known despite beginning in a degraded state. Observers were read aloud the following instructions: “Now you are going to perform a memory test. …Use the arrow keys to adjust the blurriness of the celebrity until it matches what the celebrity looked like when you correctly identified him/her in the first half of the experiment.” Observers were allowed to range back and forth among the 30 filters until they were satisfied with their decisions; no time limits were imposed.

Observers completed two practice trials prior to each of the two tasks: baseline-ID and memory test. Celebrities shown in practice trials did not appear in the experiment proper.

Results and Discussion

Evidence for visual hindsight bias was found in Experiment 1. When asked to recall how blurry each face looked when identified in Phase 1 of the experiment, observers systematically overestimated the degree of blur. Only trials for which an observer correctly identified the celebrity during baseline identification were included in the analysis. The mean baseline-ID point was \( f_1 = 25.68 \), whereas the mean estimated identification point from the memory test was \( f_1 = 20.73 \), implying that the observers remembered identifying the celebrity in a blurrier state (\( f_1 = 20.73 \)) than was actually the case (\( f_1 = 25.68 \)). The hindsight ratio (HR) can be quantified as the ratio of these two numbers: baseline-ID \( f_1 \) over memory-test \( f_1 \). To the extent that observers show hindsight bias, HR will be greater than 1, as is the case here: \( HR = 1.28 \pm 0.07 \) (see Figure 2A). The bias was not driven by a few outlying observers; 37 of the 42 observers (88%) showed visual hindsight bias.

As mentioned in the Introduction, there is evidence that verbal hindsight bias is greater for difficult than easy items (Fischhoff, 1977; Wood, 1978). To assess the influence of task difficulty on the

\[ \text{The notation “x ± y” refers to a mean plus or minus a 95% confidence interval.} \]
magnitude of visual hindsight bias, celebrities were divided into four difficulty quartiles based on baseline-ID point. Faces with lower \( f_1 \) identification values, i.e., identified at a more degraded state, were considered easier than those with higher \( f_1 \) identification values, i.e., identified at a less degraded state. This was done separately for each observer, and the means for each quartile were then averaged across observers.

Hindsight ratios for the four difficulty quartiles are shown in Figure 2A. No bias was found for the easiest faces, HR = 0.99 ± 0.05. For the remaining three difficulty quartiles, hindsight ratios increase as difficulty increases, ranging from a small bias for faces in difficulty quartile 2, HR = 1.18 ± 0.06, to a large bias for the most difficult faces, HR = 1.72 ± 0.19.

Experiment 2

In the verbal domain, many researchers have tried to eliminate or at least reduce the hindsight bias effect, and most attempts to do this have been unsuccessful. Fischhoff (1977) found that neither educating participants about hindsight bias nor warning them to do everything they could to avoid the bias were successful in reducing the effect. In Experiment 2 we used a similar approach to test whether visual hindsight bias is cognitively impenetrable. A mental function is said to be cognitively impenetrable if it cannot be influenced by purely cognitive factors such as goals, beliefs, and inferences (Pylyshyn, 1980).
Experiment 2 was a replication of Experiment 1 with one addition: Immediately prior to the memory test, observers were educated about visual hindsight bias, and were warned to avoid it and perform as accurately as possible.

**Method**

**Observers**

Fifty-four University of Washington undergraduates, all with normal or corrected-to-normal vision, participated in exchange for course credit.

**Apparatus and Stimuli**

All equipment and stimuli were identical to that used in Experiment 1.

**Design and Procedure**

The design of Experiment 2 was identical to that of Experiment 1 with one difference. Following the baseline-identification task and instructions on how to perform the memory test, observers were read aloud the following:

> When remembering how blurry each celebrity was when you first recognized him or her, I would like you to be aware of hindsight bias. Hindsight bias is when someone who knows the outcome of an event thinks they would have predicted that outcome before it happened. In previous versions of this experiment, your peers tended to believe that they recognized celebrities earlier than they did in phase 1 of the experiment. We believe that already knowing who the celebrity is before viewing the clarification is what causes this effect. The result is that your peers think they recognized celebrities earlier, at a blurrier point, than they actually did. Please try to avoid this bias and be as accurate as possible when performing the memory test.

**Results and Discussion**

The instructions and warning given to observers in Experiment 2 did not reduce the size of the hindsight bias. As was the case with Experiment-1 observers, Experiment-2 observers systematically overestimated the degree of blur when asked to recall baseline-ID performance.

Data, averaged across 54 observers, are shown in Figure 2B. Only trials for which an observer correctly identified the celebrity during the baseline-ID task were included in the analysis. The mean baseline-ID point was $f_1 = 29.13$, whereas the mean estimated identification point from the memory test was $f_1 = 23.36$. Recall that the hindsight ratio (HR) can be quantified as the ratio of these two numbers: baseline-ID $f_1$ over memory-test $f_1$; $HR = 1.31 \pm 0.10$. 

To assess the influence of task difficulty on the magnitude of the bias, celebrities were divided into four difficulty quartiles using the procedure used for the Experiment-1 data. The difficulty effect found in Experiment 1 was replicated in Experiment 2 (see Figure 2B). Identical to the pattern observed in the Experiment-1 data, no bias was found for the easiest faces, $HR = 0.92 \pm 0.05$, and for the remaining three difficulty quartiles, hindsight ratios increase as difficulty increases, ranging from a small bias for faces in difficulty quartile 2, $HR = 1.10 \pm 0.06$, to a large bias for the most difficult faces, $HR = 2.09 \pm 0.25$.

Note that there was no reduction in the hindsight bias effect observed in Experiment 2 when compared to the original Experiment-1 data. Educating observers about hindsight bias, and warning them to avoid the bias and perform as accurately as possible were not effective in reducing bias. Experiment-2 data suggest that visual hindsight bias is similar to verbal hindsight bias in that they are both automatic, unconscious processes that observers are not easily able to control.

A Fluency-Misattribution Theory of Visual Hindsight Bias

When an observer views a visual stimulus, it is processed with some degree of perceptual fluency that reflects processing speed, effort, and accuracy. Stimulus variables such as clarity, long exposure duration, familiarity and semantic relatedness, can serve to enhance perceptual fluency. Jacoby and Whitehouse (1989) demonstrated that if participants are unaware of why fluency has been enhanced, they might misattribute the fluency to something else, e.g., in their study, recent prior exposure to the stimulus. Since this work, research has shown that enhanced fluency can be misattributed to numerous factors in addition to prior exposure, such as, stimulus duration, clarity, truth, and liking (for a review see Winkielman, Schwarz, Reber, & Fazendeiro, 2003).

We propose that visual hindsight bias results from misattribution of enhanced perceptual fluency following the receipt of target-identity information. In the present experiments, processing of a degraded face after target identity is known (during the memory test) is enhanced compared to when it is not known (during the baseline-ID task). Observers must ignore the enhanced fluency if they are to accurately estimate when a naïve self or peer will make a correct identification. If observers fail to discount the enhanced fluency, or do not discount enough, they may mistakenly misattribute some or all of it to the predictability of the given outcome resulting in an overestimation of naïve performance, i.e., hindsight bias (see Bernstein, Whittlesea & Loftus, 2002; Whittlesea & Williams, 2000; for similar accounts).
Experiment 3

The fluency-misattribution theory makes the following prediction: The more fluently a target is processed following the receipt of identity information, the larger should be the size of the hindsight effect. Experiment 3 was designed to test this prediction. Phase 1 was identical to the baseline-ID task used in Experiments 1 and 2: Observers viewed celebrity faces as they clarified over time, and stopped the process when identification of the face was possible. Phase 2 differed from Experiments 1 and 2. Instead of a memory test, Experiment 3 used what is referred to as a hypothetical hindsight design in which observers with outcome information predict the performance of a naïve peer. Briefly, in the hindsight task observers viewed an outcome stimulus—an unfiltered version of the face—followed by the same clarification process used during the baseline-ID task. Observers stopped the clarification when a naïve peer, someone who did not see the outcome stimulus, would be able to identify the celebrity.

Critically, two sets of faces were shown in phase 2: the faces shown during phase 1 (old faces), and new, previously unseen faces. We expect observers to show hindsight bias for both types of faces because processing fluency during the hindsight task will be enhanced as a result of viewing the outcome stimuli. However, we expect a larger bias for old faces compared to new because processing fluency for old faces will receive additional enhancement from viewing those faces during the baseline-ID task.

Experiment 3 also allows for a test of the cognitive reconstruction theories proposed to account for verbal hindsight bias. Note that for the new faces shown in the hindsight task observers will view the outcome stimulus prior to any exposure to the degraded face. Cognitive reconstruction theories cannot account for bias in an outcome-first presentation because at the time outcome information is received there is no prior evidence to be reworked or rejudged. If hindsight bias is observed for the new faces, rejudgment processes cannot be necessary to produce visual hindsight bias.

Method

Observers

Fifty-three University of Washington undergraduates, all with normal or corrected-to-normal vision, participated in exchange for course credit.

Apparatus and Stimuli

All equipment and stimuli were identical to those used in Experiment 1.
Design and Procedure

Each observer participated in a baseline-identification task followed by a hindsight task. The baseline-identification task was identical to that used in Experiments 1 and 2. Briefly, each face progressed from highly degraded to full clarity over 15 seconds, and observers stopped the resolution process as soon as they recognized the face.

Prior to completing the hindsight task observers were read aloud the following instructions:

… In the second phase of the experiment you are going to see the same celebrities again plus some new ones. Instead of indicating at what point you recognize the celebrity, your task this time will be to estimate at what point one of your peers would recognize the celebrity. So that you know the correct answer we will show you a clear version of the celebrity at the beginning of each trial. If you know who they are type that in. If you do not know who the celebrity is, type in a question mark and it will skip to the next trial. After you identify the clear picture of the celebrity, the face will go blurry and slowly resolve—just like in phase 1. Now imagine that a same-age peer is seeing the face for the first time, meaning they did not see the clear picture first. Press the space bar when you think your peer would recognize who the person is.

When the observer stopped the clarification process, the following question appeared on the screen, “Is this the point at which you think your peer would recognize this person?” If the observer entered “y” for yes, the trial ended. If the observer answered “n” for no, the clarification process continued. Observers were allowed to stop the process as many times as necessary until they were satisfied with the degree of blur present in the image.

Half, 18 of 36, of the celebrities were shown in the baseline-ID task. The 18 baseline faces plus 18 new faces were mixed and shown in the hindsight task. The choice of which celebrities were shown twice (i.e., shown in both the baseline-ID and hindsight tasks), and the order of the 64 celebrities across trials, were counterbalanced across observers. Observers completed two practice trials prior to each of the two tasks. Celebrities shown in practice trials did not appear in the experiment proper.

Results and Discussion

Experiment-3 data are shown in Figure 3. Only trials for which an observer could correctly identify the celebrity were included in the analysis. Recall that the average $f_i$ value for the baseline-ID condition represents the average degree of blur present at time of correct identification, whereas average $f_i$ values for the two hindsight conditions represent the degree of blur present when observers predicted a naïve peer would make an accurate identification. As predicted by the fluency-misattribution theory, average $f_i$
identification point was greatest in the baseline-ID condition ($M = 29.8$), followed by the new faces in hindsight ($M = 21.4$), followed by old faces in hindsight ($M = 18.4$).

Two orthogonal planned comparisons (C1 and C2) were tested to compare the three conditions: Baseline-ID performance was compared to the two hindsight conditions (C1), and the two hindsight conditions were compared to one another (C2). For each observer a ratio of the weighted conditions was computed to test each comparison (sum of positive-weighted conditions over sum of negative-weighted conditions), and the mean of the ratios was computed. Compared to baseline-ID, observers demonstrated hindsight bias for both the old and new faces, $M_{C1} = 1.45 \pm 0.11$. Additionally, as predicted by the fluency-misattribution theory, the hindsight effect was larger for old than for new faces, $M_{C2} = 1.08 \pm 0.02$.

![Figure 3. Experiment-3 data (n=53). Performance in the three conditions is plotted as a function of $f_1$, filter cutoff frequency expressed as cycles per face height. Error bars represent standard errors.](image-url)
General Discussion and Implications

Data from three experiments provide evidence for visual hindsight bias; observers operating with the benefit of target identity information may be biased when asked to estimate the performance of a naïve peer or self. It was found for judgments made about both self (Experiments 1 and 2) and others (Experiment 3), and despite education and explicit instructions to observers to avoid the bias (Experiment 2). The bias was larger when post-outcome processing was more fluent, and when pre-outcome identification was more difficult. We now discuss these last two findings in more detail.

Fluency-Misattribution Theory

We have proposed a theory of fluency misattribution to account for visual hindsight bias. Exposure to outcome information increases the perceptual fluency with which a degraded image is processed. This fluency must be discounted if an observer is to accurately predict identification performance of a naïve self or peer. If the observer fails to fully discount the enhanced fluency, it is misattributed to the predictability of the outcome. The theory predicts larger hindsight bias for targets processed more fluently following the receipt of identity information. This prediction was confirmed in Experiment 3 in which a larger bias was found for old faces—those shown both in the baseline-ID task and the hindsight task—compared to new faces shown only in hindsight.

Jacoby and Whitehouse (1989) found that participants who are aware of why fluency has been enhanced were able to discount the fluency. Only when participants were made unaware of the source of the enhanced fluency did they misattribute it to something else. In the experiments reported here, observers may have had some awareness that processing fluency was enhanced via learning the identities of the celebrities during the baseline-identification task. Although the data clearly indicate that observers did not successfully discount all of the enhanced fluency, it is possible that observers discounted some of the enhanced fluency. If the source of increased fluency were made more obscure, a larger bias might be found. In fact, decreased awareness of the source of enhanced fluency may have contributed to the larger bias observed for old faces compared to new faces in Experiment 3.

If hindsight bias is a general metacognitive error to which all modalities are vulnerable, as we believe it is, then the fluency-misattribution theory should account for verbal hindsight bias in addition to visual hindsight bias. There is some evidence that this may be the case. Using different terminology, Sanna, Schwarz, & Small (2002) suggest that fluency plays a large role in verbal hindsight bias. The authors demonstrated that hindsight bias is reduced when evidence supporting alternative outcomes is easy to
bring to mind, but it is increased when such evidence is difficult to bring to mind. They term the ease or difficulty with which these thoughts are brought to mind, subjective accessibility, which we argue, is really just another way of describing fluency. The outcome, and evidence for it, that is processed more fluently is seen as more likely and more predictable. In the case of verbal hindsight bias, fluency is conceptual rather than perceptual. This work suggests that a fluency-misattribution theory may account for verbal hindsight bias as well as visual. Further studies are warranted to examine this possibility.

**Task Difficulty**

As mentioned in the Introduction, there is evidence that verbal hindsight bias is greater for difficult than easy items, and greater for events initially judged to be least plausible (Fischhoff, 1977; Wood, 1978; Arkes, Wortmann, Saville, & Harkness, 1981). In the visual domain, Harley, Carlsen, and Loftus (2001) found that visual hindsight bias only occurred for the most difficult-to-detect targets. The results reported here are consistent with these findings. For Experiments 1 and 2, no hindsight effect was observed for the easiest faces, and the effect then increased monotonically as the degree of foresight identification difficulty increased.

The influence of task difficulty on the size of visual hindsight bias can be accounted for by the fluency-misattribution theory if we assume that outcome information is more beneficial to the processing of difficult targets. To illustrate, prior to the receipt of target identity information, some degraded targets will be more difficult to identify than others. Following the receipt of target identity information, all of the images become identifiable at a more degraded state. If target identity information is more beneficial for an item that was originally difficult to identify compared to one that was not difficult, then the discrepancy between baseline processing fluency and hindsight processing fluency will be larger for more difficult items. The greater this discrepancy, the more fluency an observer must discount to make an accurate prediction about a naïve observer’s ability. As evidenced by Experiment-1 and Experiment-2 data, observers do adjust appropriately for the easiest faces, those for which the outcome information presumably provides the least benefit, but the adjustment falls shorter and shorter as difficulty increases.

One could argue that the difficulty effect observed in Experiments 1 and 2 resulted from observers’ failure to stop the clarification process at the point of recognition due to an inability to remember a celebrity’s name. Recall that faces were divided into difficulty quartiles based on the point at which they were identified during the baseline-ID task; faces identified later, i.e., at a clearer state, were categorized as more difficult than those identified earlier. It is possible that faces identified later (coded as more difficult) were faces for which observers recognized the celebrity but had trouble recalling a name or
other identifying remark. In such cases, observers may have let the clarification process continue until a name could be recalled. Although this account is plausible, we have two reasons to believe that observers were not performing the task in this manner. First, observers were instructed to, “press the space bar as soon as you recognize who the celebrity is.” They were also told that if their guess changed they could stop the clarification process again, and that there would be no penalty for guessing multiple times. Second, in watching observers perform the baseline-ID task they appeared to stop the clarification process as soon as the face was recognizable, and then wait until a name or identifying remark could be generated before continuing. Given, however, that our second piece of evidence is anecdotal, and that we cannot know if all observers followed the directions, the alternative account of the difficulty effect cannot be ruled out.

**Legal Implications**

Visual hindsight bias has a number of important legal implications. We return once again to the medical malpractice scenario described at the outset of the paper. When asked to estimate whether Radiologist 1 should have detected a tumor at time 1, Radiologist 2 should proceed with extreme caution. Not only will the tumor be more visible to an observer operating with the benefit of outcome knowledge (e.g., Muhm, et al., 1983), but also, based on the results reported here, it is highly likely Radiologist 2 will not fully discount this benefit, and as a result, will overestimate the detection ability of a naïve observer.

Physicians judging the visibility of missed tumors are not the only legal players who may be susceptible to visual hindsight bias. Eyewitnesses to a crime are often asked to judge their perceptual abilities, and in doing so, may overestimate their ability to detect or identify a visual target under poor viewing conditions. For example, if Warren Witness views someone fleeing from the scene of a crime 30 feet away, on a dark and rainy night, his memory of that person’s face is likely to be poor. If Warren is later shown a clear picture of Sam Suspect in a police photo lineup (akin to outcome knowledge), Warren may integrate the clear picture of Sam with the original degraded image he has stored in memory. The result will be a “cleaned up” memory representation that more closely resembles Sam Suspect. This is clearly problematic, as Sam may or may not be the criminal Warren saw, but the trouble does not end here. When asked to testify against Sam Suspect, Warren Witness may overestimate his ability to have identified Sam under the original, poor viewing conditions.

The notion that feedback can influence an eyewitness’s estimate of original viewing conditions is not new. Verbal confirmatory post-identification feedback has been shown not only to inflate witnesses’ confidence in the accuracy of their identification, but also to distort their estimates of the original
witnessing conditions (Wells & Bradfield, 1998; Bradfield, Wells, & Olson, 2002). Goodness of view, speed of identification, amount of attention paid to the suspect’s face, and clarity of memory for the suspect, are given inflated ratings by eyewitnesses who received confirmatory feedback following identification of a suspect. The data reported here add to the mounting evidence that outcome information, be it verbal or visual, can distort an eyewitness’s beliefs about the original viewing conditions.

**Concluding Remarks**

We have provided evidence that, under certain conditions, observers operating with the benefit of hindsight do not have accurate insights into the strengths and limitations of their perceptual abilities. This is evidenced by their failure to accurately predict the performance of a naïve peer or self in visual identification tasks. Like verbal hindsight bias, visual hindsight bias appears to be moderated by task difficulty, with a greater bias occurring for more difficult (i.e., ambiguous) images, and appears to be cognitively impenetrable. A fluency-misattribution theory was proposed to account for the bias. The theory posits that exposure to target identity information results in enhanced processing fluency of the degraded image. When asked to judge the performance of a naive observer, the increased fluency is not fully discounted, and instead may be misattributed to the predictability of the given outcome. The theory predicts a larger hindsight effect for more fluently processed targets. This prediction was confirmed in Experiment 3 in which a larger bias was observed for targets made more familiar via repeated exposures.
References


