Auditory Hindsight Bias:

Fluency Misattribution versus Memory Reconstruction

Philip A. Higham[1], Greg J. Neil[1], & Daniel M. Bernstein[2]

[1] University of Southampton, UK

[2] Kwantlen Polytechnic University, Canada

Word count: 13,461

Author Notes

Philip A. Higham, Department of Psychology, University of Southampton; Greg J. Neil, Department of Psychology, University of Southampton; Daniel M. Bernstein, Department of Psychology, Kwantlen Polytechnic University.

The authors would like to thank Charlotte Heathcote, Cherise Yau, and Andrew Huebert for assistance collecting data. The Canada Research Chairs Program (950-228407) helped to fund this work. Portions of this research were presented at the 11th biennial meeting of the Society for Applied Research in Memory and Cognition (SARMAC), Victoria, BC, Canada and at the 52nd Annual Meeting of the Psychonomic Society, Seattle, WA, USA.

Correspondence concerning this article should be directed to Philip A. Higham, Department of Psychology, University of Southampton, Highfield, Southampton, SO17 1BJ. Email: higham@soton.ac.uk, phone: +44 (0)23 8059 5942.

Abstract

We report 4 experiments investigating auditory hindsight bias – the tendency to overestimate the intelligibility of distorted auditory stimuli after learning their identity. An associative priming manipulation was used to vary the amount of processing fluency independently of prior target knowledge. For hypothetical designs, in which hindsight judgments are made for peers in foresight, we predicted that judgments would be based on processing fluency and that hindsight bias would be greater in the unrelated- compared to related-prime context (differential-fluency hypothesis). Conversely, for memory designs, in which foresight judgments are remembered in hindsight, we predicted that judgments would be based on memory reconstruction and that there would be independent effects of prime relatedness and prior target knowledge (recollection hypothesis). These predictions were confirmed. Specifically, we found support for the differential-fluency hypothesis when a hypothetical design was used in Experiments 1 and 2 (hypothetical group). Conversely, when a memory design was used in Experiments 2 (memory group), 3A and 3B, we found support for the recollection hypothesis. Together, the results suggest that qualitatively different mechanisms create hindsight bias in the two designs. The results are discussed in terms of fluency misattributions, memory reconstruction, anchoring-and-adjustment, sense making, and a multi-component model of hindsight bias.

Keywords: auditory hindsight bias, fluency attributions, hypothetical design, memory design, associative priming

Statement of the Public Significance of the Work

This study was aimed at uncovering the basic cognitive mechanisms of auditory hindsight bias, the finding that people informed about the content of a distorted audio signal prior to hearing it subsequently overestimate how objectively intelligible the signal actually is. The study suggests that the bias is not produced by a unitary cognitive mechanism, but is based on different mechanisms depending on whether one focuses on the past (i.e., trying to remember how intelligible a signal seemed earlier before learning the content) or on a hypothetical scenario (i.e., estimating how intelligible the signal would be for naïve others who are unaware of the content). Specifically, in the context of remembering, knowledge of content can distort how memories are reconstructed. Alternatively, in a hypothetical scenario, knowledge of content enhances perception of the signal which is not properly discounted when estimating for naïve others.

Auditory Hindsight Bias:
Fluency Misattributions versus Memory Reconstruction

Hindsight bias is a common judgment error that occurs when people who are knowledgeable of an outcome overestimate its probability. For example, the attacks on the World Trade Center in 2001 might seem predictable to most of us now even though, prior to the attacks, many people did not expect them. Hindsight bias, first documented by Fischhoff in the 1970s (e.g., Fischhoff, 1975, 1977; Fischhoff & Beyth, 1975), occurs across the lifespan and in different cultures (e.g., Bayen, Erdfelder, Bearden, & Lozito, 2006; Bernstein, Erdfelder, Meltzoff, Peria, & Loftus, 2011; Pohl, Bayen, & Martin, 2010; Pohl, Bender, & Lackman, 2002). It affects many situations, including medical and legal decisions, business, consumer satisfaction, sporting events, and election outcomes (Arkes, Wortman, Saville, & Harkness, 1981; Blank, Fischer, & Erdfelder, 2003; Giroux, Coburn, Harley, Connolly, & Bernstein, 2016; Harley, 2007; Leary, 1981; Zwick, Pieters, & Baumgartner, 1995).

Investigating Hindsight Bias: Hypothetical versus Memory Designs

There are two common designs used to investigate the hindsight effect: *hypothetical* and *memory*. With the typical hypothetical design, participants rate the likelihood of an event having a particular outcome (e.g., “How likely is it that the British won the British - Gurkha war of 1814?”; Fischhoff & Beyth, 1975). Participants in *foresight* make this rating with no knowledge of the event’s actual outcome. Participants in *hindsight* learn the outcome to the event (e.g., “The British won the war.”) and then rate how likely the outcome would seem to participants in foresight who do not share this knowledge. The hypothetical judgments of the hindsight groups are then compared to the actual judgments made by participants in foresight.

With the typical memory design, participants first make a judgment about the likelihood of an outcome in foresight. Then, after a delay, they learn the outcome and try to remember their previous foresight ratings (e.g., Wood, 1978). Regardless of the design used, there is usually evidence of a failure to discount outcome knowledge because likelihood estimates made in hindsight exceed those made in foresight. For example, participants told that the British won the war overestimate how likely foresight participants rate that outcome (hypothetical design) and overestimate their own previous foresight ratings (memory design).

Generally speaking, hindsight effects are larger with hypothetical designs than memory designs (e.g., Campbell & Tesser, 1983; Davies, 1992; Fischhoff, 1977; Wood, 1978). This difference is thought to derive from the fact that some participants in memory designs may be able to accurately recollect their earlier foresight rating, which would result in bias-free hindsight ratings. Indeed, longer retention intervals between the original foresight ratings and the later hindsight ratings have been shown to increase hindsight bias in memory designs (e.g., Blank et al., 2003; Fischhoff & Beyth, 1975; Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988; Pennington, 1981), presumably because fewer participants are able to accurately recollect their previous ratings. Conversely, because participants only make one rating in hypothetical designs, they do not have foresight ratings to recollect. Consequently, all ratings are potentially open to bias, resulting in more hindsight bias in hypothetical compared to memory designs. Thus, the larger hindsight bias in hypothetical designs compared to memory designs may be more apparent than real. To accurately compare the size of the bias in the two designs, it is necessary to carefully consider whether ratings can be recollected in memory designs and remove such cases prior to making the comparison (see Pohl, 2007 for discussion).

The fact that veridical recollection of prior ratings partly determines the size of hindsight bias in memory designs but not hypothetical designs suggests that the bias is not based on a unitary mechanism (although see Schwarz & Stahlberg, 2003). Consistent with this idea, Blank, Nestler, von Collani and Fischer (2008) have argued that there are actually three components to hindsight bias: memory distortions, impressions of foreseeability, and impressions of necessity. In support of these distinctions, they found that certain variables affected one component but had no effect, or even the opposite effect, on another. For example, in their Study 2, participants rated prior to an election the percentage of votes that various political parties would obtain. These ratings were completed either 10 weeks or 6-10 days before the election. After the election, participants in hindsight tried to remember their earlier percentage estimates and also rated how foreseeable the outcome was. Blank et al. found that there was hindsight bias as measured by memory distortion; that is, memory for earlier percentage ratings was biased toward the winner. Furthermore, hindsight bias was greater following a long retention interval (i.e., for participants who completed the foresight ratings 10 weeks prior to the election) compared to a short one (i.e., for participants who completed the foresight ratings 6-10 days prior to the election). In contrast, participants in hindsight indicated that, overall, the election outcome was *not* foreseeable. Moreover, participants who made hindsight judgments after a short retention interval rated the outcome as slightly *more* foreseeable than participants who made judgments after a long retention interval. In other words, foreseeability ratings responded to changes in retention interval in a manner that was opposite to how memory distortion responded. Blank et al. argued that dissociations such as this suggest that hindsight bias is not a unitary phenomenon and that different mechanisms support the different components.

Auditory Hindsight Bias

In contrast to the classic demonstrations involving event scenarios described above, more recently, various classes of *perceptual hindsight bias* have been reported in the literature. These cases, which include visual, auditory, and gustatory judgments, typically challenge the perceptual system in some way by requiring participants to identify perceptually degraded stimuli (e.g., Bernstein & Harley, 2007; Bernstein, Wilson, Pernat, & Meilleur, 2012; Calvillo & Gomes, 2011; Harley, Carlsen, & Loftus, 2004; Pohl, Schwarz, Sczesny, & Stahlberg, 2003; Wu, Shimojo, Wang, & Camerer, 2012). In this paper, we focus on one particular type of bias on perceptual judgments: *auditory hindsight bias*.

Bernstein et al. (2012) recently reported four hypothetical-design experiments in which they aurally presented participants with word targets. The words were perceptually degraded with a low-pass filter, which removes high frequencies from the signal, making the words sound muffled and difficult to identify. On foresight trials, participants attempted to identify the words. On hindsight trials, participants received knowledge of the targets' identities (i.e., prior target knowledge) before they were asked to estimate the percentage of foresight participants who would be able to identify the muffled words. The authors found that participants in hindsight overestimated identification performance in foresight. Furthermore, this overestimation occurred despite instructions to avoid it, demonstrating that it was difficult to overcome the effect that prior target knowledge had on judgments. These experiments extend other demonstrations that people have difficulty discounting "top-down" effects on the perception of degraded auditory stimuli (e.g., Epley, Keysar, Van Boven, & Gilovich, 2004; Lange, Thomas, Dana, & Dawes, 2011; Vokey & Read, 1985).

Currently, very little is known about the underlying mechanisms of auditory hindsight bias and the primary purpose of the research reported here is to fill this gap in the literature. Like Blank et al. (2008), we suspect that different mechanisms underpin auditory hindsight bias when tested in the context of remembering versus in the context of hypothetical scenarios. To understand hindsight bias in the latter scenario, it is worth considering the basic problems people experience when trying to identify degraded stimuli in foresight. In our view, there are two such problems. The first problem is that there are too many possibilities to consider (*stimulus ambiguity*). For example, a degraded auditory target such as *long* is easily confusable with *strong* or *wrong*. The second problem is that there are too few possibilities to consider (*generation difficulty*). People in foresight may hear a degraded stimulus and may not be able to generate any candidates *at all* (“What was *that*?”). Prior target knowledge solves both these problems: it disambiguates cases for which there are too many possibilities and it provides an option for cases when there is none.

We suggest that when prior target knowledge solves either the problem of stimulus ambiguity or generation difficulty, there is facilitated processing of the degraded target which participants experience as high *subjective fluency*; that is, prior target knowledge makes processing the degraded target feel relatively easier. A body of research has suggested that when participants experience subjective fluency, they (consciously or unconsciously) attribute its cause to a source (e.g., Bernstein & Harley, 2007; Higham & Vokey, 2000; Jacoby & Dallas, 1981; Jacoby, Kelley, & Dywan, 1989; Jacoby, Woloshyn & Kelley, 1989; Lloyd & Miller, 2011; Olds & Westerman, 2012; Werth & Strack, 2003; Westerman, 2008; Whittlesea, 1993, 2011). The chosen source may or may not be the true cause of subjective fluency, and if fluency is attributed to the wrong source, then a *fluency misattribution* has occurred. Fluency misattributions can result in illusions of memory as well as other types of illusions of subjective experience. For example, briefly presented old words that appeared in a list shown earlier in the experiment were judged to have longer display durations than new words, presumably because they were processed more fluently (Witherspoon & Allen, 1985). In the case of auditory hindsight bias in the hypothetical scenario, fluency resulting from prior target knowledge may be misattributed to target clarity, creating the illusion for hindsight participants that the target is objectively easier to hear than it actually would be for foresight participants.

Although fluency misattributions provide a reasonable account of auditory hindsight bias in hypothetical designs, it is less clear how fluency might play a role in a memory design. In memory designs, the focus of attention is on the past rather than current processing because participants face a recall problem (i.e., recollecting their previous foresight ratings). Although some types of memory decisions such as recognition judgments may be based on the fluency of processing the recognition stimulus (e.g., Whittlesea, 1993), subjective fluency offers no assistance toward veridical recall. Consequently, participants in hindsight may essentially ignore their current processing experience with the muffled target in the memory design and instead focus on the past in an attempt to recollect their prior ratings.

Experimental Predictions and Rationale

To help us learn more about the mechanisms of auditory hindsight bias, we examined how the bias responds to a priming manipulation that, along with prior target knowledge, likely affects the fluency of processing the degraded target. Specifically, before hearing a perceptually degraded target word and either trying to identify it or judge its clarity, participants heard a clear prime word that was either associatively related to the target (e.g., *reflect-mirror*) or not (e.g., *humid-guard*). Prime relatedness was then factorially combined with a manipulation of prior target knowledge: on half the trials, the target was presented in a clear form prior to it being played in a distorted form. On the remaining trials, a neutral stimulus was presented in place of the clear target. This procedure created four within-subjects experimental conditions: unrelated-foresight (UF), unrelated-hindsight (UH), related-foresight (RF), and related-hindsight (RH). These four conditions, schematically depicted in Figure 1, form the core of our design and they are present in all our experiments.

We assume that related primes increase the fluency with which participants process target words for reasons very similar to prior target knowledge; that is, related primes help solve the problems of stimulus ambiguity and generation difficulty. If both related primes and target knowledge affect processing in a similar manner, they should have similar effects on subjective experience (i.e., both produce qualitatively similar types of subjective fluency). This similarity allowed us to generate our main prediction for the hypothetical design: the subjective fluency that is created by enhancing participants' processing of degraded targets by providing them with target knowledge will be less on related-prime trials (when fluency for the trial is already high due to the prime) than on unrelated-prime trials (when fluency is low).[1] This hypothesis is based on a generalization of the psychophysical law relating subjective intensity and stimulus intensity: greater change in stimulus intensity is needed with high compared to low stimulus intensities to achieve the same change in subjective intensity. This basic psychological principle is found with many different types of stimuli and subjective states and it is the essence of Weber's fraction, Fechner's logarithmic law, and Steven's power law (see Stevens, 1957 for a review). Applying this principle to auditory hindsight bias would mean that subjective fluency caused by prior target knowledge would be greater on low-fluency, unrelated-prime trials than on high-fluency, related-prime trials. Empirically, we predict that this differential fluency, in turn, will lead to an interaction between prime relatedness and prior target knowledge, with greater hindsight bias for unrelated- compared to related-prime trials. Hereafter, we refer to this hypothesis as the *differential-fluency hypothesis*.

Assuming veridical recollection does not occur, judgments in the memory design may also be subject to hindsight bias. However, the nature of this bias derives from the attempt to reconstruct the prior rating rather than fluency. We believe the nature of hindsight bias in memory designs is similar to other well-documented memory distortions in the memory literature such as the *misinformation effect* (Loftus, Miller & Burns, 1978). New knowledge can affect people's schemata (i.e., cognitive frameworks), whether that knowledge is misleading details about a witnessed event or target knowledge in a perceptually demanding

[1] It is important to note that it is specifically the processing of the target in its *degraded* form that is critical here, not the processing of the target in its *clear* form that occurs only on hindsight trials. It is the target in its degraded form that presents a perceptual challenge to participants and there are variables that may facilitate processing (such as prior target knowledge) that must be discounted to achieve an unbiased rating.

task. Schemata, in turn, can bias how people reconstruct past events (e.g., Bartlett, 1932). However, if schemata alteration rather than fluency is the basis of hindsight bias in the memory design, then there is no reason to assume that it will be moderated by prime relatedness in the same way as with the hypothetical design. Thus, in contrast to the hypothetical design, our empirical prediction for the memory design is two independent main effects of prime relatedness and prior target knowledge. Hereafter, we refer to this independence pattern in the memory design as the *recollection hypothesis*.

A Word About Baselines

In experiments on auditory hindsight bias, there is a choice about which baseline should serve as a control. One obvious option is the foresight identification rates. That is, participants can be instructed to try to identify the distorted targets and then hypothetical or memory hindsight ratings can be compared to the successful identification rates (e.g., Bernstein et al., 2012). Alternatively, in a memory design, hindsight ratings in Phase 2 might be compared to foresight ratings in Phase 1 as is done in most memory-design experiments investigating non-perceptual hindsight effects.

However, in our view, there are potential problems with both these baselines. First, although it is tempting to directly compare ratings on a 0-100 scale to actual levels of performance (e.g., percentage of successful identifications) to infer whether participants were overconfident in their ratings (i.e., showed hindsight bias), this process completely ignores how participants map internal or subjective feelings of certainty onto the particular scale values, a mapping that can have profound effects on the conclusions drawn (e.g., see Higham, Zawadzka, & Hanczakowski, 2016 for details; see also Hanczakowski, Zawadzka, Pasek & Higham, 2013 and Zawadzka & Higham, 2015, 2016). A second problem is that comparing ratings between different phases potentially introduces confounding variables such as familiarity with the stimuli. For example, by comparing foresight ratings in Phase 1 against hindsight ratings in Phase 2 in memory designs, one is also comparing stimuli presented for the first time (in Phase 1) with stimuli presented twice (in Phase 2). Consequently, hindsight ratings in Phase 2 could be higher than foresight ratings in Phase 1 not because of outcome knowledge, but because of fluency from target familiarity being misattributed to target clarity.

To avoid both these problems in all the experiments that we report here, we used foresight ratings made in the same phase as hindsight ratings to serve as a control. Unlike comparison with identification performance, comparing same-phase hindsight and foresight ratings makes no assumptions about the absolute meaning of scale values; the comparison only leads to conclusions about relative differences. Furthermore, by ensuring that the ratings being compared are always made in the same phase, we also ensure that the level of exposure to the word pairs is the same between the hindsight and foresight conditions.

Power Analysis

We based our sample size on the observed effect size for auditory hindsight bias obtained in Bernstein et al. (2012) as it provided one of the few previous examples of this type of bias in the literature. For a within-subjects *F* test based on power $(1 - \beta) = .99$, alpha $= .05$, one degree of freedom in the numerator, and their observed effect size of $R^2 = .63$, the estimated sample size was seven. We tested a minimum of 20 participants in all conditions reported in this manuscript, which far exceeded the estimated number needed to detect an effect of this size.

Experiment 1

Experiment 1 was our initial test of the differential-fluency hypothesis for the hypothetical design. It consisted primarily of our core 2 (knowledge: hindsight, foresight) x 2 (prime status: related, unrelated) repeated-measures design (Figure 1) conducted within a single experimental phase. Participants listened to muffled target words that were preceded by clearly presented prime words that were either related or unrelated to the target words. Half the prime-target pairs were preceded by a clear presentation of the target, which provided prior target knowledge (hindsight condition), whereas the other half were not (foresight condition). Participants' primary task was to judge how easy it would be for foresight peers (i.e., those without prior target knowledge) to identify the target words. The foresight and hindsight ratings were then compared to determine whether there was auditory hindsight bias, and if so, whether it was moderated by the priming manipulation. As explained in detail above, the differential-fluency hypothesis for the hypothetical design is that hindsight bias will be greater on unrelated- compared to related-prime trials.

We also included some foresight trials for which participants tried to identify the target words rather than to rate them. As with the trials requiring a rating, the distorted target was preceded by either a related or unrelated prime word. We operationalized high fluency as enhanced ability to identify the primes. Consequently, successful identification rates on these trials provided an independent measure of processing fluency and allowed us to test our assumption that related primes enhance fluency relative to unrelated primes.

Method

Participants. Participants were 24 students (six male) from the University of Southampton (age: *M* = 21.08, *SEM* = 0.62). One participant was dropped because English was not their native language. They participated for course credits or £4.50 cash.

Design and materials. Two test lists of 120 word pairs each were created for Experiment 1. To create the lists, 120 related word pairs were selected from the Nelson, McEvoy, and Schreiber (2004) word-association norms. In the Nelson et al. norms, the forward association is the proportion of people that would respond with the target word when given the prime word. The 120 word pairs were randomly split into two sets of 60 highly related word pairs, with the restriction that the pairs in each set of 60 had the same mean associative strength of 0.50, with no word pair having an associative strength of less than 0.35. Two different test lists of 120 word pairs were created, one of which was used with a particular participant. To create the first test list of 120 word pairs, we assigned one set of 60 related word pairs to the list with their original pairings (e.g., *annual-yearly*), whereas the remaining set was transformed into 60 unrelated word pairs by randomly reassigning the prime words of the word set to different targets (e.g., *digit-hungry*). Re-pairing the first half of the pairs while leaving the second half intact created the second test list of 120 word pairs. Thus, each test list contained the same target and prime words, with the assignment of primes to targets defining whether a word pair was related or unrelated. Assignment of test list to participant was counterbalanced.

Audio recordings of all the word pairs in each test list were then created using an Apple Macintosh computer and Audacity software. A second, distorted recording of each target word from each pair was then created using a low-pass filter, which blocked any frequency above 48 Hz. This created a set of distorted target words that sounded muffled, as if the speaker was talking through a pillow. To homogenize volume levels across the stimulus set, all recordings were subjected to a compressor and a limiter using Logic Pro. Each sound file ended with silence for 0.5 seconds. The experiment was programmed and administered using LiveCode.

A 2 (relatedness: related, unrelated) X 2 (knowledge: foresight, hindsight) repeated-measures design was used. The design also included an additional set of items reserved for identification attempts. To achieve this design, the two test lists were split into thirds, with each third containing 20 related and 20 unrelated word pairs. One third was assigned to the hindsight condition, with the task being to provide a hypothetical rating (see below) for these word pairs (the RH and UH conditions). The second third was assigned to the foresight condition with the task again being a hypothetical rating task (RF and UF conditions). The final third was assigned to the foresight identification task during which participant were simply required to type in an identification attempt. Word pairs were rotated through conditions and task following a Latin-square design, such that all target words appeared in each condition equally often, and each test list had 20 items in each of the RH, UH, RF, UF conditions as well as 40 identification trials (20 related; 20 unrelated). All stimuli in the experiment were administered to each participant individually using an Apple iMac with the words being played through headphones and responses being entered using a keyboard and mouse.

Procedure. Experiment 1 consisted of one phase only. Participants heard the 120 word pairs played in a random order. Each test trial started with the word "ready" appearing on screen for 1.5 seconds. For 40 hindsight trials, this prompt was followed one second later by the clear (i.e., undistorted) target word being played through the headphones. After a pause of 0.25 seconds, the prime word was played.[2] After another pause of 0.25 seconds, the muffled target word was played through the headphones. Once the muffled target word had finished playing, a "rate" prompt appeared on screen, along with text asking participants to judge the number of people out of 100 who would be able to identify the target word. It was emphasized that this estimate should be based on everything that they heard on the trial *except* the clear target at the beginning of the trial. In other words, it was necessary to assume that the 100 hypothetical people heard the prime word before the muffled target, but that they did not hear the clear target presented at the beginning of the trial. They made their response by typing a number between 0% and 100% into the computer. The next trial began once participants had pressed the continue button (self-paced).

For the 40 foresight trials, the order and timing of events on each trial was the same, but instead of the clear target word playing at the beginning of the trial, participants heard a

[2] Note that the pause was actually 0.75 seconds due to the 0.5 seconds of silence at the end of each sound clip.

neutral stimulus (beep). As with hindsight trials, in response to the “rate” prompt, participants rated how many others with prime knowledge would be able to identify the target word.

The remaining 40 trials were the same as the foresight trials except that participants tried to identify the muffled target by typing a response into a text box. So that participants knew to identify rather than to rate the muffled target, an “identify” prompt was used instead of a “rate” prompt. No feedback was provided.

Results and Discussion

All tests were conducted with an alpha level of 0.05 for this and all subsequent experiments.

Identification decisions. Mean identification accuracy (%) for related- and unrelated-prime trials is shown in Table 1. Identification performance was much better for related- (M = 57, SEM = 4) than for unrelated-prime trials (M = 12, SEM = 3), $F(1,22) = 158.37$, $p < .001$, $\eta_p^2 = 0.88$. This difference is consistent with our expectation that participants processed related-prime targets more fluently than unrelated-prime targets.

Auditory hindsight bias. Mean ratings on the rate trials are shown in Figure 2. To test for an auditory hindsight effect, we conducted a 2 (knowledge: hindsight, foresight) X 2 (relatedness: related, unrelated) repeated-measures ANOVA on the ratings. The results of this analysis revealed main effects of prime relatedness, $F(1,22) = 40.69$, $p < .001$, $\eta_p^2 = 0.65$, and knowledge, $F(1,22) = 13.02$, $p = .002$, $\eta_p^2 = 0.37$. Ratings were higher for related- (M = 51, SEM = 3) than unrelated-prime trials (M = 37, SEM = 3) and they were higher in hindsight (M = 48, SEM = 3) than in foresight (M = 40, SEM = 3). The interaction was also significant, $F(1,22) = 6.22$, $p = .021$, $\eta_p^2 = 0.22$. The interaction reflected the fact that the hindsight effect was greater for unrelated- than for related-prime trials, although both effects were significant, $F(1,22) = 14.71$, $p = .001$, $\eta_p^2 = 0.40$ and $F(1,22) = 4.33$, $p = .049$, $\eta_p^2 = 0.16$, respectively.

Summary. The results of Experiment 1 are consistent with the view that fluency misattributions were the cause of auditory hindsight bias in the hypothetical design.[3] First, related primes increased both the chances of successfully identifying the muffled targets on *identify* trials, and ratings assigned on *rate* trials, suggesting that related primes enhanced the fluency of processing the degraded target. We suggest that prior target knowledge also enhanced the fluency of processing the degraded target and participants were unable to fully discount this fluency despite instructions to do so. The failure to discount fluency was particularly evident on unrelated-prime trials, which provided empirical support for the differential-fluency hypothesis. According to the differential-fluency hypothesis, subjective fluency caused by prior target knowledge was greater and harder to discount on low-fluency, unrelated-prime trials than on high-fluency, related-prime trials.

Although the data from Experiment 1 were consistent with the differential-fluency hypothesis, we considered it important to replicate the results and contrast them to an experimental group undergoing a similar procedure but given memory instructions rather than hypothetical instructions. As explained above, we did not expect prime relatedness to moderate auditory hindsight bias if judgments are based on recollections rather than fluency (recollection hypothesis). Experiment 2 tested this prediction.

Experiment 2

Experiment 2 included a group of participants that received hypothetical instructions (*hypothetical* group, similar to Experiment 1), and a group that received memory instructions (*memory* group). As noted, our aim was to directly compare performance in the two groups, but we faced a problem: there are a different number of phases between the two designs (one

[3] We consider some alternative explanations of our results in the General Discussion.

in the hypothetical design; two in the memory design). The critic could argue, therefore, that any differences observed in the results obtained between the designs were due to confounding factors introduced by the varying number of phases (e.g., more familiarity with the items, and hence more fluent processing, with two phases compared to one).

To address this issue, both the hypothetical and memory groups in Experiment 2 completed a two-phase experiment in which the first phase consisted of foresight ratings regarding the percentage of peers who would be able to identify the muffled targets with the assistance of the prime word. However, in Phase 2, the procedure for the two groups diverged. In particular, the hypothetical group was told to ignore their earlier ratings and focus on their experiences in Phase 2 to make their judgments. In contrast, the memory group was told to focus on the past and try to remember their previous ratings.

As stated above, we hypothesized that the basis of responding may be different when participants focus on the past in the memory design rather than on current processing fluency as in the hypothetical design. If so, and if our instructional manipulation is effective, then we predict independent main effects of prime relatedness and prior target knowledge in the memory group, but an interaction between these variables in the hypothetical group.

In addition to adopting a two-phase design and including two experimental groups instead of one, there were a few other methodological changes between this experiment and the last. First, to ensure that all participants were aware of which primes were related to targets and which were not, even on trials for which the muffled targets could not be identified, the word "related" or "unrelated" appeared on the screen as the prime word sounded. We considered this addition to be important because knowledge regarding the relationship between the prime and target could affect the strategy adopted (e.g., participants may only rely on the prime for assistance in identifying the target if they are aware that the two are related). Second, because our baseline was Phase-2 foresight ratings rather than foresight identification performance, and because Experiment 1 established that prime relatedness had a large effect on identification likelihoods, there were no trials in the experiment for which participants were required to explicitly identify muffled targets. Third, to simplify the auditory aspects of the task on hindsight trials, the clear target word appeared visually on the computer screen rather than being played through headphones. The visually presented target word remained on the screen until participants made a response to ensure that it was encoded prior to hearing the muffled version. Fourth, we dropped some items in this experiment, eliminating those that had identification rates near the floor. Fifth, following each rating in Phase 2, participants were asked to indicate whether they were confident in their response (Y/N). These data were used to ascertain whether the auditory hindsight effect still occurred when only high-confidence responses were analyzed. One possibility is that people only turn to fluency-based responding when more reliable bases of responding (that likely lead to higher confidence) are unavailable. For example, Johnston, Dark, & Jacoby (1985) found that participants only responded to processing fluency when making recognition memory judgements if their ability to recollect the items from a previously presented list was poor. If something similar occurs in our paradigm, then we may find that the differential-fluency hypothesis does not apply if the analysis is restricted to high-confidence responses.

Finally, as a manipulation check, we included a third phase at the end of the experiment to ensure that participants in the hypothetical and memory groups were following our instructions. Specifically, participants in each group were shown all the word pairs on the screen along with their ratings and confidence levels and were asked whether the rating for each item was based on *current impression*, *remember*, or *don't know*. If our instructional manipulation was successful, then compared to the memory group, participants in the hypothetical group should respond *current impression* more often and *remember* less often.

Method

Participants. Participants were 67 students from Kwantlen Polytechnic University who participated in exchange for course credit. One participant was removed after reporting that she closed her eyes to avoid reading the clear target word. Thirty-five (six male) of the remaining 66 participants were randomly assigned to the *hypothetical* group (age: *M* = 21.77, *SEM* = 0.54) and 31 (four male) to the *memory* group (age: *M* = 21.87, *SEM* = 0.66).

Design and materials. A 2 (relatedness: related, unrelated) X 2 (knowledge: foresight, hindsight) x 2 (instructions: hypothetical, memory) mixed design was used, with instructions as the only between-subjects factor.

Eighty related word pairs were selected from the pairs used in Experiment 1, omitting any single-syllable targets because, in a distorted form, they were virtually impossible to identify. The 80 word pairs were randomly split into four sets of 20 pairs, with the restriction that each set of 20 had a mean associative strength of 0.50 and no word pair had an associative strength of less than 0.35. To create a test list, a procedure similar to that in Experiment 1 was used. Specifically, the related pairs in two out of the four sets remained intact (e.g., *annual-yearly*), with one set assigned to the RH condition (20 pairs) and the other to the RF condition (20 pairs). The remaining two sets of 20 related pairs were transformed into unrelated word pairs (e.g., *digit-hungry*) by swapping the prime words between the sets; one set was assigned to the UF condition and the other to the UH condition. Using four test lists, word pairs were rotated through the UF, UH, RF, and RH conditions in their sets of 20 following a Latin-square design, such that all target words appeared in each condition equally often across participants. Thus, each test list contained the same target and prime words, with the assignment of prime to target defining whether a word pair was related or unrelated. Assignment of test-lists to participants was counterbalanced. Clear and muffled audio recordings were created in the same manner as Experiment 1.

Procedure. The experiment was administered to each participant individually using an Apple iMac with the words being played through headphones and responses being entered using a keyboard and mouse. Experiment 2 consisted of a familiarization phase, two main experimental phases, and a final judgment-basis selection phase. The purpose of the familiarization phase was to give the Canadian participants an opportunity to become accustomed to the English accent of the speaker. Familiarization took the form of a set of participants passively listening to a set of eight words (not used in the later phases) played one after another and then repeated.

Following familiarization, all participants read the Phase-1 instructions, which were the same for all participants. Each participant heard 80 word pairs (40 related; 40 unrelated) played in a random order. Word pairs were played in the same manner as the pairs in the identification and foresight conditions from Experiment 1, with one exception. For each pair, the phrase “The two words on this trial were related [unrelated]” appeared on screen after the muffled target was played (i.e., at the time of the judgment) so that there was no ambiguity about the nature of the prime-target relationship. Participants’ task was to judge the percentage of other people who would be able to identify the target word with the help of the prime word. Unlike Experiment 1, participants were not required to explicitly try to identify the muffled targets themselves. They made their response by typing a number between 0% and 100% into the computer. The next trial began once participants had pressed the continue button (self-paced).

Phase 2, which was also self-paced, ensued immediately after Phase 1 and realized our 2 X 2 core design (Figure 1). There were 80 test trials in Phase 2 (20 RF, 20 RH, 20 UF, and 20 UH) that were similar to “rate” trials in the single phase of Experiment 1, barring several minor changes. First, as in Phase 1, participants were informed on the screen whether the prime and target were normatively related or unrelated at the time of making their judgment. Second, instead of hearing the clear target at the beginning of the trial in the hindsight condition, the target word appeared visually on screen one second before the muffled target was played and remained there until a response was made. This change was meant to ensure that participants knew the identity of the target on hindsight trials prior to making a rating. In the foresight condition, the target word displayed on the computer screen was replaced with "XXXX." Third, after making a rating, participants were asked to indicate whether they were confident that their rating was correct with a yes/no (Y/N) response.

Participants in the hypothetical group were instructed to rate on each trial the percentage of other people who would be able to identify the muffled target word with the help of the prime word, but without any prior target knowledge. The instructions also made it clear that participants should ignore their Phase-1 ratings, and even forget what the words sounded like in Phase 1. To emphasize this point, participants were told that in previous studies, relying on Phase-1 memory resulted in 50% worse performance compared to not relying on Phase-1 memory. Conversely, participants in the memory group were instructed to try to remember their Phase-1 rating and to ignore how the words currently sounded in Phase 2. To emphasize these instructions, participants were told that relying on how the words sound now would result in 50% worse performance compared to relying on memory for Phase-1 ratings.

Finally, all participants completed an additional selection task after Phase 2. Participants were shown the word pairs on the screen along with their rating and Y/N confidence responses. For each word pair, they were asked to indicate the basis of their judgment by selecting one of the following options: *remember*, *current impression,* or *don't know*. Participants were told that *remember* indicated that they remembered their Phase-1 judgment and used that information to make the Phase-2 judgment. *Current impression* indicated that they ignored their Phase-1 judgment and instead used their current experience with the words to make their Phase-2 judgment. If the participant could not remember or did not know how they made their judgment, they could select *don't know*. No information was provided in this final phase about whether the word pairs were normatively related or unrelated, or whether the pair corresponded to a foresight or hindsight trial.

Results and Discussion

Phase-1 identification estimations. The mean Phase-1 identification estimations are presented in Table 1. These ratings were analyzed with a 2 (relatedness: related, unrelated) X 2 (instructions: hypothetical, memory) mixed ANOVA. There was only a main effect of relatedness, $F(1,64) = 209.25$, $p < .001$, $\eta_p^2 = 0.77$, with related-prime trials ($M = 63$, $SEM = 2$) having much higher ratings than unrelated-prime trials ($M = 30$, $SEM = 2$). There were no other effects, largest $F(1,64) = 1.06$, $p = .306$, $\eta_p^2 = .016$.

Auditory hindsight bias. The mean Phase-2 ratings in the hypothetical group are shown in the top panel of Figure 3. As in Experiment 1, the ratings were analyzed with a 2 (knowledge: hindsight, foresight) X 2 (relatedness: related, unrelated) repeated-measures ANOVA.[4] There were main effects of relatedness, $F(1,34) = 81.08$, $p < .001$, $\eta_p^2 = 0.70$, and knowledge, $F(1,34) = 29.81$, $p < .001$. $\eta_p^2 = 0.47$, and an interaction between relatedness and knowledge, $F(1,34) = 7.39$, $p = .010$, $\eta_p^2 = 0.18$. The interaction arose because the hindsight effect was greater for unrelated- than for related-prime trials, although both effects were significant, $F(1,34) = 31.38$, $p < .001$, $\eta_p^2 = 0.48$ and $F(1,34) = 18.14$, $p < .001$, $\eta_p^2 = 0.35$, respectively. The pattern of data in the hypothetical group (two main effects and an interaction) replicated that observed in Experiment 1 in which participants also received

[4] Conceivably, group (hypothetical, memory) could be included as a factor in this analysis. However, as will become apparent, the data file for the memory group needed to be cleaned prior to analysis to remove exact Phase-1/Phase-2 matches reflecting veridical recollection. This cleaning meant that the group means were based on different items, potentially compromising direct statistical comparisons. Consequently, we considered it more appropriate to analyze the groups separately.

hypothetical instructions (cf. Figure 2 and the top panel of Figure 3). Similar results were obtained when the analysis was restricted to high-confidence responses.

Turning to the memory group, we first cleaned the data file by removing exact matches between Phase-1 and Phase-2 ratings to eliminate cases of veridical recollection. As others have noted (e.g., Coolin, Bernstein, Thornton, & Thornton, 2014; Blank et al., 2003; Pohl, 2007; Schwarz & Stahlberg, 2003), veridical recollection of prior ratings in memory designs essentially protects participants from bias and can potentially distort the pattern of results (e.g., artificially decrease or eliminate hindsight bias; see Erdfelder & Buchner, 1998 for a multinomial model of this process). On average across participants, 35% (*SEM* = 5%) of Phase-2 ratings matched earlier Phase-1 ratings. No participants were eliminated (i.e., no participant had matching Phase-1/Phase-2 ratings on 100% of trials). The mean Phase-2 ratings after matches were removed are shown in the bottom panel of Figure 3.

A 2 (knowledge: hindsight, foresight) X 2 (relatedness: related, unrelated) repeated-measures ANOVA on the remaining Phase-2 ratings revealed main effects of both relatedness, $F(1,30) = 47.70$, $p < .001$, $\eta_p^2 = 0.61$, and knowledge, $F(1,30) = 8.00$, $p = .008$, $\eta_p^2 = 0.21$. Phase-2 ratings were higher for related- ($M = 53$, $SEM = 3$) than unrelated-prime trials ($M = 30$, $SEM = 3$) and they were higher in hindsight ($M = 46$, $SEM = 3$) than in foresight ($M = 38$, $SEM = 3$). However, unlike the previous analysis on the hypothetical group, the interaction was not significant, $F < 1$. Analogous results were obtained when the analysis was restricted to high-confidence responses.

Judgment-basis selections. The similarity of the pattern of results between Experiment 1 and the hypothetical group of this experiment, and the dissimilarity of this pattern to that observed in the memory group, suggests that our instructions were successful. Thus, the hypothetical instructions persuaded participants to focus on current processing and to ignore their previous ratings whereas the memory instructions persuaded participants to focus on the past and to ignore their current processing. To confirm this conclusion, we analyzed the judgment bases that participants selected at the end of the experiment. In particular, the proportion of *remember* and *current impression* selections were analyzed with a 2 (instructions: hypothetical, memory) x 2 (judgment basis: remember, current impression) mixed ANOVA to confirm that participants were performing the hypothetical and memory tasks as instructed. These mean proportions, along with the mean proportion of *don't know* responses, are presented in Table 2.[5] The ANOVA revealed a main effect of judgment basis, $F(1,64) = 11.47$, $p = .001$, $\eta_p^2 = 0.15$, but no effect of instructions, $F < 1$. There were more *remember* ratings ($M = 0.44$, $SEM = 0.03$) than *current impression* ratings ($M = 0.29$, $SEM = 0.02$) overall. More critically, the interaction was also significant, $F(1,64) = 14.16$, $p < .001$, $\eta_p^2 = 0.18$. The interaction reflected the fact that the selection rate of *current impression* was higher in the hypothetical group than the memory group, $F(1,64) = 6.97$, $p = .010$, $\eta_p^2 = 0.10$, whereas the opposite was true with the selection rate of *remember*, $F(1,64) = 13.96$, $p < .001$, $\eta_p^2 = 0.18$. Hence, the pattern of judgment-basis selections indicates that participants in the two experimental groups followed their respective instructions reasonably well.

Summary. Instructing participants in the hypothetical group to focus on their current processing when making ratings for others produced a pattern of results which closely resembled that obtained in Experiment 1. That is, we observed a hindsight effect, but it was larger in the unrelated-prime condition than in the related-prime condition, which confirmed the differential-fluency hypothesis. This replication occurred despite several methodological changes between the experiments, which attests to the robustness of the results.

Participants in the memory group also demonstrated a hindsight effect when they were instructed to try to remember their Phase-1 foresight ratings rather than to focus on their current Phase-2 experience. However, consistent with the recollection hypothesis, the

[5] *Don't know* responses were equivalent across the different instructions and excluded from the analysis.

hindsight effect observed in the memory group was independent of the effect of prime relatedness. Together, the two experimental groups in Experiment 2 have demonstrated that a simple instructional manipulation regarding what to base judgments on can modify the pattern of results obtained. Furthermore, the different data patterns that we observed could not be attributable to differential exposure to the stimuli between the groups (i.e., two exposures to the stimuli in the memory design but only one exposure in the hypothetical design). As noted above, differential exposure to the stimuli could be particularly problematic if participants are basing judgments on fluency.

Because cases of veridical recollection can confound the results if they are left to contaminate the data (e.g., Erdfelder & Bucher, 1998; Pohl, 2007), we removed perfect rating matches between Phases 1 and 2 prior to analyzing the data. However, there was a significant proportion of such cases (35%), most likely due to the fact that Phase 2 followed immediately after Phase 1, which would have allowed veridical recall of prior ratings. The necessary elimination of so many items raises concerns that the different pattern of results obtained in the memory and hypothetical groups was attributable to an item-selection artifact. The question remains, therefore, as to whether an auditory hindsight effect (with or without an interaction with prime relatedness) would be obtained with a memory focus if remembering the Phase-1 ratings was made more difficult, thereby allowing more items to be included in the analysis. We tested this possibility in Experiments 3A by including a two-day retention interval between Phases 1 and 2.

Experiment 3A

Experiment 3A consisted of two phases as in Experiment 2, although participants received only memory instructions. Other aspects of the procedure were also similar to the memory group in Experiment 2, except that the second phase ensued two days after Phase 1 instead of immediately afterwards. Our primary interest was to determine whether we would observe hindsight bias with a memory focus when the rate of veridical recall was less (due to the longer retention interval) and the data set was more complete, compared to the memory group in Experiment 2. Our secondary interest was to determine whether prime relatedness and hindsight bias, if found, would be independent as we observed in the memory group of Experiment 2. Finding independence in a more complete data set would provide stronger support for the recollection hypothesis. Furthermore, it would suggest that the different pattern of results obtained in the memory and hypothetical groups of Experiment 2 was not attributable to item differences but rather that the mechanisms underlying hindsight bias in the two designs are qualitatively different.

Method

Participants. Participants were 20 students (nine male) from the University of Southampton (age: $M = 20.8$, $SEM = 0.19$). They participated for course credits or £4.50 cash.

Design and materials. As in previous experiments, a 2 (relatedness: related, unrelated) X 2 (knowledge: foresight, hindsight) repeated-measures design was used. However, unlike Experiment 2 which had both memory and hypothetical groups, only one group with memory instructions was tested in Experiment 3A. The materials were the same, and counterbalanced in the same manner, as in Experiment 2. Counterbalanced assignment of test list to participants was also the same as in previous experiments.

Procedure. The procedure was similar to the memory group of Experiment 2 except for the following. The most important difference was that Phase 2 ensued two days after Phase 1 rather than immediately afterwards. Second, there was no familiarization phase (as the participants were English) or final judgment-basis selection phase. Third, because the results of the hypothetical groups of Experiments 1 and 2 were similar, explicit indication of whether the prime and target words were related was considered superfluous and removed from both Phases 1 and 2. Fourth, participants were instructed to read the visually-presented target aloud before making their memory judgment. This aspect of the procedure ensured that participants encoded the clear, visually-presented version of the target word (rather than, for example, closing their eyes to avoid being biased by it, a strategy adopted by at least one participant in Experiment 2). For the foresight trials, the order of events on each trial was the same, but instead of the target word appearing onscreen, *XXXX* appeared, and the participants were instructed to say “Xs” before making their judgment. Finally, although the instructions still emphasized the importance of remembering previous ratings rather than relying on current processing, the wording was slightly different from the memory group of Experiment 2. In particular, participants were told, “You may be tempted to write down what you think now rather than how you responded during part 1. However, the correct response is to write down your response from part 1 because it shows that your memory is accurate.” All other aspects of the procedure were the same as for the memory group in Experiment 2.

Results and Discussion

Phase-1 identification estimations. Mean identification estimations, shown in Table 1, were much higher for related-prime trials ($M = 61$, $SEM = 3$) than for unrelated-prime trials ($M = 31$, $SEM = 3$), $F(1,19) = 168.29$, $p < .001$, $\eta_p^2 = 0.90$.

Auditory hindsight bias. Mean Phase-2 memory judgments are shown in Figure 4. As in Experiment 2, the data file was cleaned prior to analysis to remove exact Phase-1/Phase-2 matches, a large portion of which would have represented veridical recollection. As expected, there were far fewer such cases in this experiment ($M = 14\%$, $SEM = 2\%$ versus $M = 35\%$, $SEM = 5\%$ in the memory group of Experiment 2), most likely due to the longer retention interval (two-days versus immediate). To test for auditory hindsight bias, we conducted a 2 (knowledge: hindsight, foresight) X 2 (relatedness: related, unrelated) repeated-measures ANOVA on the identification-rating memories in Phase 2. The results of this analysis revealed main effects of prime relatedness, $F(1,19) = 110.58$, $p < .001$, $\eta_p^2 = 0.85$, and knowledge, $F(1,19) = 29.13$, $p < .001$, $\eta_p^2 = 0.61$. Ratings were higher for related-prime trials ($M = 59$, $SEM = 3$) than unrelated-prime trials ($M = 32$, $SEM = 4$) and they were higher in hindsight ($M = 51$, $SEM = 3$) than in foresight ($M = 40$, $SEM = 3$). The interaction was not significant, $F < 1$, indicating that prime relatedness did not moderate the effect of prior target knowledge. As in Experiment 2, similar results were obtained when the analysis was restricted to high-confidence responses.

Summary. Participants in Experiment 3A showed a bias toward higher memory ratings in Phase 2 (regarding their Phase-1 likelihood ratings about identification success for others) if they were armed with prior target knowledge (hindsight) than if they were not (foresight). In other words, an auditory hindsight effect occurred. Moreover, in contrast to the interactive patterns observed with hypothetical instructions in Experiments 1 and 2, but consistent with the recollection hypothesis, hindsight bias in Phase 2 of this experiment was not moderated by prime relatedness. This independence pattern was obtained despite the fact that, compared to the memory group in Experiment 2, fewer than half the number of trials needed to be removed because Phase-1/Phase-2 matched ratings. The fact that we observed the independence pattern in this experiment on 85% of items suggests that the analogous pattern observed in the memory group of Experiment 2 was not attributable to an item-selection effect.

Experiment 3B

Although the implementation of a two-day retention interval in Experiment 3A meant that fewer trials needed to be removed prior to analysis (due to Phase-1/Phase-2 rating matches), it would be preferable if no trials were removed. If an independence pattern was obtained with memory instructions on the full set of items rather than a subset of them, we could be more confident that different mechanisms are causing auditory hindsight bias in hypothetical versus memory designs and that the dissociation is not due to item differences.

To achieve this end in Experiment 3B, we again used a memory design, but we altered the information that participants needed to remember in Phase 2. In the memory groups of both Experiment 2 and 3A, participants in Phase 2 were attempting to remember previous identification *estimations for others* (i.e., numbers on a scale). This methodology allowed for exact Phase-1/Phase-2 rating matches because participants sometimes recollected the particular rating value they assigned on an earlier trial. To sidestep this issue in Experiment 3B, we instead had participants attempt to identify the muffled targets in Phase 1, but no feedback was provided. Later in Phase 2, participants rated the probability that their Phase-1 identification attempt was correct. Because no rating was provided in Phase 1, participants could no longer recollect the particular scale value assigned to any Phase-1 trial. This alteration to the methodology allowed us to test whether auditory hindsight bias and the recollection hypothesis still hold in the memory design if all trials are included in the analysis.

Method

Participants. Participants were 20 students (six male) from the University of Southampton (age: $M = 20.65$, $SEM = 0.15$). They participated for course credits or £4.50 cash.

Materials and design. The materials and design used in Experiment 3B were identical to Experiment 3A.

Procedure. The procedure was identical to Experiment 3A in all regards except for the judgments that participants made. In Phase 1, rather than rating identification success for others, participants attempted to identify the stimuli themselves with or without the assistance of a related prime word. They made their response by typing their identification attempt into a textbox on the computer. No feedback was provided. In Phase 2, after reading the target aloud (hindsight) or saying "Xs" (foresight), participants were instructed to judge the likelihood that they correctly identified the muffled target word in Phase 1. They did so by entering a number between 0% and 100%. This rating was thus a memory-based probability judgment because participants were required to remember whether their previous identification attempt was correct or not.

Results and Discussion

The analyses conducted on the Experiment 3B data were the same as for Experiment 3A except that there was no analysis of Phase-1/Phase-2 correspondence. Because the data in Phase 1 were binary (yes/no identification success) whereas the Phase-2 judgments were made on a 0-100% likelihood scale, there was no simple method of identifying exact matches.

Phase-1 identification. As in Experiment 1, the presence of a related prime substantially improved identification performance, $F(1,19) = 165.48$, $p < .001$, $\eta_p^2 = 0.90$ (see Table 1). This result again confirms that prime relatedness enhances processing fluency.

Auditory hindsight bias. Mean ratings in Phase 2 are shown in Figure 5. To test for an auditory hindsight effect, we conducted a 2 (knowledge: hindsight, foresight) X 2 (relatedness: related, unrelated) repeated-measures ANOVA on the memory judgments made in Phase 2. It revealed main effects of prime relatedness, $F(1,19) = 118.42$, $p < .001$, $\eta_p^2 = 0.86$, and knowledge, $F(1,19) = 25.17$, $p < .001$, $\eta_p^2 = 0.57$. Ratings were higher for related-prime trials (*M* = 69, *SEM* = 3) than unrelated-prime trials (*M* = 31, *SEM* = 3) and they were higher in hindsight (*M* = 56, *SEM* = 3) than in foresight (*M* = 44, *SEM* = 3). Prime relatedness did not moderate the effect of prior target knowledge, $F < 1$ for the interaction. These results replicate almost exactly those found in Experiment 3A. As in previous experiments, similar results were obtained when the analysis was performed on only those trials assigned high-confidence.

Summary. Experiment 3B established that prior target knowledge biased Phase-2 memory judgments even though those judgments were about previous identification performance rather than ratings for others. Hence, the results of Experiment 3B attest to the robustness of the biasing effect that prior target knowledge has on memory. Additionally, prime relatedness had a large effect on identification success in Phase 1 and, analogous to Experiment 3A, on the ratings in Phase 2. However, it did not moderate the size of the hindsight effect, confirming the recollection hypothesis. This independence pattern is consistent with those obtained in the groups given memory instructions in Experiment 2 (memory group) and 3A and it stands in contrast to the interactive pattern found with hypothetical instructions in Experiments 1 and 2 (hypothetical group). Moreover, it was obtained on the full set of items rather than a subset, thereby eliminating potential item-selection problems. As noted above, the different data patterns associated with memory and hypothetical designs suggest different underlying mechanisms, a point we elaborate on in the General Discussion.

General Discussion

In four experiments, we investigated the underlying mechanisms of auditory hindsight bias in both hypothetical and memory designs. We hypothesized that judgments for foresight peers in the hypothetical design would be based on processing fluency. Furthermore, we hypothesized that the fluency derived from prior target knowledge would be greater and harder to discount in the context of a low-fluency trial than a high-fluency trial (the differential-fluency hypothesis). To test this prediction, we independently manipulated the fluency context by introducing an associative priming manipulation. As expected, in the hypothetical groups of Experiments 1 and 2, auditory hindsight bias was greater in the context of unrelated primes (low fluency) than related primes (high fluency). This finding resembles analogous results obtained in psychophysical experiments that have shown that a change to a stimulus' intensity brings about smaller subjective effects if the stimulus intensity is already high versus low (see Stevens, 1957 for a review). This principle applied to auditory hindsight bias suggests that the same facilitation to processing brought about by possessing target knowledge has less of a subjective impact, and is therefore more easily discounted, when fluency on the trial is already high (because of a related prime) than when it is low (because of an unrelated prime).

In contrast to the results obtained with the hypothetical design, those for the memory design used in Experiments 2, 3A, and 3B revealed no evidence of prime relatedness moderating hindsight bias, which supported the recollection hypothesis. Instead, similar-sized hindsight biases were obtained between all the related- and unrelated-prime conditions in those experiments. In Experiment 2, for which the retention interval between the foresight ratings and the subsequent memory task was short, hindsight bias was observed after eliminating cases of exact veridical recall (i.e., when memory for ratings in Phase 2 matched the actual ratings in Phase 1). However, the proportion of cases of veridical recall was high (35%), which potentially meant that the independence pattern observed in that experiment was due to item-selection artifacts. To reduce the rate of veridical recollection in Experiments 3A, we implemented a two-day retention interval. The longer retention interval reduced veridical recollection to 14% and we observed the same independence pattern again. Finally, in Experiment 3B, using a methodology that did not allow veridical recollection to play a major role, we again observed hindsight bias and an independence pattern in the complete data set. Together, the results of Experiments 2, 3A, and 3B suggest that the independence pattern is real and the recollection hypothesis has some validity rather than being based on item selection.

Alternative Accounts of Auditory Hindsight Bias

In this section, we consider some alternative explanations of auditory hindsight bias and demonstrate that they are insufficient to account for our full set of results. In particular, we first focus on two scaling accounts (the ceiling-effect and anchoring-and-adjustment accounts). We then consider sense making and, finally, a multi-component model of hindsight bias.

Scaling accounts. The results from the memory groups eliminate a host of potential "scaling" accounts of the interactive pattern observed in the hypothetical groups. For example, one account might be that the high baseline performance in the related-prime conditions caused the interaction. Stated simply, there was less "room" for prior target knowledge to increase ratings in the related-prime conditions compared to the unrelated-prime conditions because baseline (foresight) performance was already near the top of the judgment scale. However, counter to this hypothesis, examination of Figures 2-5 indicates that no condition had near-ceiling performance. Moreover, the ratings were comparable between the memory and hypothetical designs (e.g., baseline differences were present in both designs). This account fails to explain, therefore, why the difference in baseline performance between the related- and unrelated-prime conditions produced an interaction in the hypothetical groups but not in the memory groups.

Another scaling account of hindsight bias is *anchoring and adjustment*. Several theorists have suggested that outcome knowledge provides a judgment anchor reflecting certainty (100%). In an attempt to discount outcome knowledge, participants in hindsight must adjust their rating downward to match foresight ratings that are more uncertain. However, in most cases, the adjustment is insufficient to reach the level of uncertainty that exists in foresight, which results in hindsight bias (e.g., Hawkins & Hastie, 1990; Epley & Gilovich, 2006). To extend this logic to our methodology incorporating related- and unrelated-prime trials, one might propose that the interactive pattern observed in the hypothetical design was due to the vastly different foresight baselines. That is, the adjustment necessary to achieve an unbiased judgment was far less for related- compared to unrelated-prime trials, which resulted in less hindsight bias.

However, just as with the ceiling-effect account, it is not clear why we found independent main effects of prime relatedness and prior target knowledge in the memory design in Experiments 2, 3A, and 3B if anchoring-and-adjustment was the primary mechanism. The foresight baselines in those conditions were also drastically different between the related- and unrelated-prime conditions and so a similar interactive pattern would be expected there as well. The question remains, therefore, as to why participants use the anchoring-and-adjustment heuristic in hypothetical designs but not in memory designs because the heuristic should be equally applicable to both designs.

Sense making. Pezzo (2003) proposed a sense-making model of hindsight bias in which outcomes that provide high initial surprise invoke a process of trying to make sense of how the outcome could have happened. If that process is successful, then the outcome seems obvious and hindsight bias results. If it is not, then hindsight bias can be reduced or even reversed ("No one would have predicted *that*!"). Unsurprising outcomes do not invoke the sense-making process and therefore do not produce much hindsight bias either. More recently, Calvillo and Gomes (2011) successfully applied Pezzo's sense-making model to a perceptual hindsight-bias paradigm involving predictions about the inevitability of animated automobile accidents.

Although the distorted word identification task like that used in our experiments is not as conducive to sense-making processes as more traditional hindsight-bias paradigms involving event scenarios, one might still reason that the related prime helps participants make sense of prior target knowledge. That is, on related-prime trials, participants could have crosschecked target candidates against the prime. For example, after being told that the muffled target is *doctor* (hindsight), participants might make better sense of this knowledge if the prime word was *nurse* (related) than if it was *table* (unrelated). However, if such a sense-making process is occurring, then it makes the prediction that hindsight bias will be larger with related- compared to unrelated-prime trials. This result was not confirmed in any of our experiments with either the hypothetical or memory design suggesting that this form of sense making was not producing our effects.

This post hoc application of Pezzo's (2003) sense-making model to our auditory hindsight paradigm is speculative at best. The model was designed to explain hindsight bias in scenarios involving narratives or general knowledge that might be made sense of using pre-existing schematic knowledge, semantic memory, and reasoning processes rather than perceptual processes. Moreover, even if the model can explain the interactive results pattern we observed in the hypothetical design, like the other accounts we have considered, it is not clear why sense making did not also occur in the memory design to produce an analogous interactive pattern.

Blank et al.'s (2008) three-component model of hindsight bias. As discussed in the Introduction, Blank et al. (2008) have argued that hindsight bias is not a unitary construct but consists of three components: memory distortion, impressions of foreseeability, and impressions of necessity. We agree wholeheartedly with making a distinction between hindsight judgments that are memory based versus those based on the subjective aspects of current processing. Indeed, that distinction is the crux of the contrasting predictions of the differential-fluency and recollection hypotheses. However, it is not clear to us how the distinction between impressions of foreseeability and impressions of necessity might apply to our auditory hindsight bias paradigm. According to Blank et al., impressions of foreseeability pertain to subjective predictions and they are influenced by metacognition. For example, a person may judge that that they foresaw a particular election result because they knew the results of certain opinion polls prior to the election. In contrast, impressions of necessity pertain more to objective states of affairs and are based on causal attributions. For example, knowing that a pre-election debate went well for one candidate in particular, her winning the election may seem inevitable. While this distinction is useful in the context event-based scenarios such as election outcomes, there does not seem to be any straightforward way of making this distinction when it comes to identifying distorted words with or without prior target knowledge. However, the model may be useful in accounting for perceptual hindsight bias in other paradigms (e.g., the inevitability versus foreseeability of simulated car accidents; Calvillo & Gomes, 2011).

Conclusions

Together, the results from the current set of experiments suggest strongly that different mechanisms were driving auditory hindsight bias in the hypothetical and memory designs. Consequently, we believe that no account involving only a unitary mechanism will suffice. In our view, the most parsimonious explanation of our results as a whole is that hypothetical designs engender fluency attributions whereas memory designs engender reconstructive recall processes. Since its inception, the fluency-attribution framework has emphasized flexibility of processes; that is, the framework assumes that there are multiple bases of judgments and that processing fluency may be utilized or ignored by participants depending on the stimuli and context. For example, Johnston et al. (1985) observed that participants were less likely to use fluency as a judgment basis for recognition judgments if recollection was made easier by having participants study words instead of nonwords. In the same way, compared to hypothetical designs, we assume that memory designs cause participants to rely less on fluency and more on recollection. However, by switching their focus to the past, our participants faced a new problem: bias in memory reconstruction that worsened as the retrieval conditions worsened (i.e., longer retention intervals). [6]

The fact that auditory hindsight bias was evident in two different designs and driven by two completely different mechanisms demonstrates its robustness. Auditory hindsight bias remains a fascinating but largely understudied phenomenon. Our hope is that our current research provides a modest step toward filling a gap in the literature by providing a theoretical foundation. Future research might concentrate on topics such as debiasing and discovering how widespread the bias is. For example, do singers who are familiar with their own lyrics overestimate how easy it is for others to hear the words in their songs? Do police investigators who view suspect interviews with transcripts overestimate how intelligible the voices would be for a jury viewing the interview without the transcript (e.g., Lange et al., 2011)? In our view, these and other topics related to auditory hindsight bias are worthy of more attention so that the true scope of the bias can be determined.

[6] One potential criticism of "flexible" theories is that they are irrefutable and therefore not of much value to scientific progress. We recognize that the fluency hypothesis is potentially open to this criticism and hope that future research might better identify the parameters that determine whether fluency is or is not the primary basis of decisions in auditory-hindsight paradigms.

References

Arkes, H. R., Wortmann, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnosis. *Journal of Applied Psychology*, *66,* 252–254. doi: 10.1037/0021-9010.66.2.252

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.

Bayen, U. J., Erdfelder, E., Bearden, J. N., & Lozito, J. P. (2006). The interplay of memory and judgment processes in effects of aging on hindsight bias. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*, 1003-1018. doi: 10.1037/0278-7393.32.5.1003

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, & Cognition, 37*, 378-391. doi: 10.1037/a0021971

Bernstein, D. M., & Harley, E. M. (2007). Fluency misattribution and visual hindsight bias. *Memory*, *15*, 548-560. doi: 10.1080/09658210701390701

Bernstein, D. M., Wilson, A. M., Pernat, N., & Meilleur, L. (2012). Auditory hindsight bias. *Psychonomic Bulletin & Review*, *19,* 588–593. doi: 10.1080/09658210701390701

Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory, 11*, 491-504. doi:10.1080/09658210244000513

Blank, H., Nestler, S, von Collani, G, & Fischer, V. (2008). How many hindsight biases are there? *Cognition, 106*, 1408-1440. doi: 10.1016/j.cognition.2007.07.007

Calvillo, D. P., & Gomes, D. M. (2011). Surprise influences hindsight–foresight differences in temporal judgments of animated automobile accidents. *Psychonomic Bulletin & Review*, *18,* 385–391. doi: 10.3758/s13423-011-0062-4

Campbell, J. D., & Tesser, A. (1983). Motivational interpretations of hindsight bias: An individual difference analysis. *Journal of Personality*, *51,* 605. doi: 10.1111/j.1467-6494.1983.tb00868.x

Coolin, A., Bernstein, D. M., Thornton, A. E., & Thornton, W. L. (2014). Age differences in hindsight bias: The role of episodic memory and inhibition. *Experimental Aging Research*, *40*, 357-374. doi: 10.1080/0361073X.2014.896667

Davies, M. E (1992). Field dependence and hindsight bias: Cognitive restructuring and the generation of reasons. *Journal of Research in Personality, 26*, 58-74. doi: 10.1016/0092-6566(92)90059-D.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17*, 311-318. doi: 10.1111/j.1467-9280.2006.01704.x

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327-339. doi: 10.1037/0022-3514.87.3.327

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. doi: 10.1037/0278-7393.24.2.387

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. doi: 10.1037/0096-1523.1.3.288

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 349-358. doi: 10.1037/0096-1523.3.2.349

Fischhoff, B., & Beyth, R. (1975). 'I knew it would happen': Remembered probabilities of once-future things. *Organizational Behavior & Human Performance, 13*(1), 1-16. doi:10.1016/0030-5073(75)90002-1

Giroux, M. E., Coburn, P. I., Harley, E. M., Connolly, D. A. & Bernstein, D. M. (2016). Hindsight bias and law. *Zeitschrift für Psychologie, 224*, 190–203. doi: 10.1027/2151-2604/a000253

Hanczakowski, M., Zawadzka, K., Pasek, T. & Higham, P. A. (2013). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. *Journal of Memory and Language, 69*, 429-444. doi: 10.1016/j.jml.2013.05.003.

Harley, E. M. (2007). Hindsight bias in legal decision making. *Social Cognition*, *25,* 48–63. doi: 10.1521/soco.2007.25.1.48

Harley, E. M., Carlsen, K., & Loftus, G. R. (2004). The ‘saw it all along’ effect: Demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30,* 960–968. doi: 10.1037/0278-7393.30.5.960.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311-327. doi: 10.1037/0033-2909.107.3.311

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, *16,* 533–538. doi: 10.3758/BF03197054

Higham, P. A., & Vokey, J. R. (2000). Judgment heuristics and recognition memory: Prime identification and target processing fluency. *Memory & Cognition, 28*, 574-584. doi: 10.3758/BF03201248

Higham, P.A., Zawadzka, K., & Hanczakowski, M. (2016). Internal mapping and its impact on measures of absolute and relative metacognitive accuracy. In J. Dunlosky and S. Tauber

(Eds.) *Oxford Handbook of Metamemory* (pp. 39-61). doi: http://dx.doi.org/10.1093/oxfordhb/9780199336746.013.15

Jacoby, L. L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*, 306-340. doi: 10.1037/0096-3445.110.3.306

Jacoby, L. L., Kelley, C. M. & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., Woloshyn, V. & Kelley, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118*, 115-125. doi: 10.1037/0096-3445.118.2.115

Johnston, W. A., Dark, V. A., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 3-11. doi: 10.1037/0278-7393.11.1.3

Lange, N. D., Thomas, R. P., Dana, J., & Dawes, R. M. (2011). Contextual biases in the interpretation of auditory evidence. *Law and Human Behavior*, *35*, 178-187. doi: 10.1007/s10979-010-9226-4

Leary, M. R. (1981). The distorted nature of hindsight. *Journal of Social Psychology*, *115,* 25–29. doi: 10.1080/00224545.1981.9711984

Lloyd, M. E., & Miller, J., K. (2011). Are two heuristics better than one? The fluency and distinctiveness heuristics in recognition memory. *Memory & cognition 39*, 1264-1274. doi: 10.3758/s13421-011-0093-0

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 19-31. doi: 10.1037/0278-7393.4.1.19

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, and Computers, 36*, 402-407. doi: 10.3758/BF03195588

Olds, J. M., & Westerman, D. L. (2012). Can fluency be interpreted as novelty? Retraining the interpretation of fluency in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 653-664. doi: 10.1037/a0026784

Pennington, D. C. (1981). On being wise after the event: An investigation of hindsight bias. *Current Psychological Research, 1*, 271-282. doi: 10.1007/BF03186737

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

Pohl, R. F. (2007). Ways to assess hindsight bias. *Social Cognition, 25*, 14-31. doi: 10.1521/soco.2007.25.1.14

Pohl, R. F., Bayen, U. J, & Martin, C. (2010). A multiprocess account of hindsight bias in children. *Developmental Psychology*, *46*, 1268-1282. doi: 10.1037/a0020209

Pohl, R.F., Bender, M. and Lachmann, G. (2002). Hindsight bias around the world. *Experimental Psychology*, *49*, 270-282. doi: 10.1026//1618-3169.49.4.270

Pohl, R. F., Schwarz, S., Sczesny, S., & Stahlberg, D. (2003). Hindsight bias in gustatory judgments. *Experimental Psychology*, *50*, 107-115. doi: 10.1026//1618-3169.50.2.107

Schwarz, S., & Stahlberg, D. (2003). Strength of hindsight bias as a consequence of meta-cognitions. *Memory*, *11,* 395–410. doi: 10.1080/09658210244000496

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181. doi: 10.1037/h0046162

Vokey, J. R., & Read, J. D. (1985). Subliminal messages: Between the devil and the media. *American Psychologist, 40*, 1231-1239. doi:10.1037/0003-066X.40.11.1231

Werth, L., & Strack, F. (2003). An inferential approach to the knew-it-all-along phenomenon. *Memory*, *11,* 411–419. doi: 10.1080/09658210244000586

Westerman, D. L. (2008). Relative fluency and illusions of recognition memory. *Psychonomic Bulletin & Review, 15*, 1196–1200. doi: 10.3758/ PBR.15.6.1196

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1235–1253. doi: 10.1037/0278-7393.19.6.1235

Whittlesea, B. W., A. (2011). Remembering under the influence of unconscious expectations. In P. A. Higham and J. Leboe (Eds.). *Constructions of remembering and metacognition* (pp. 225-236). *Essays in honor of Bruce Whittlesea*, Palgrave MacMillan

Witherspoon, D., & Allan, L. G. (1985). The effect of a prior presentation on temporal judgments in a perceptual identification task. *Memory & Cognition, 13,* 101-111. doi: 10.3758/BF03197003

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 345-353. 10.1037/0096-1523.4.2.345

Wu, D. A., Shimojo, S., Wang, S. W., & Camerer, C. F. (2012). Shared visual attention reduces hindsight bias. *Psychological Science*, *23*, 1524–1533. doi: 10.1177/0956797612447817

Zawadzka, K., & Higham, P. A. (2015). Judgments of learning index relative confidence, not subjective probability. *Memory & Cognition, 43*, 1168-1179. doi: 10.3758/s13421-015-0532-4

Zawadzka, K., & Higham, P.A. (2016). Recalibration effects in judgments of learning: A signal detection analysis. *Journal of Memory and Language, 90,* 161-176. doi: http://dx.doi.org/10.1016/j.jml.2016.04.005

Zwick, R., Pieters, R, & Baumgartner, H. (1995). On the practical significance of hindsight bias: The case of the expectancy-disconfirmation model of consumer satisfaction.

Organizational Behavior and Human Decision Processes, *64,* 103–117. doi: 10.1006/obhd.1995.1093

Table 1

Mean Actual (Experiments 1 and 3B) and Mean Estimated (Experiments 2 and 3A) Percentage of Correctly Identified Muffled Targets of in Experiments 1-3 as a Function of Prime Relatedness. Standard Errors Are Shown in Parentheses.

	Prime Type	
Experiment and Group	Unrelated	Related
Experiment 1 (Identification trials)	11 (3)	57 (4)
Experiment 2		
Hypothetical Group (Phase 1)	31 (3)	65 (3)
Memory Group (Phase 1)	29 (3)	60 (3)
Experiment 3A (Phase 1)	31 (3)	61 (3)
Experiment 3B (Phase 1)	24 (2)	69 (4)

Note. Both identification attempts and identification estimates for peers were performed without prior target knowledge

Table 2

Mean Rates of Selecting Each Judgment Basis as a Function of Experimental Group in Experiment 2. Standard Errors Are Shown in Parentheses.

	Judgment Basis		
Group	Current Impression	Remember	Don't Know
Hypothetical	.37 (.04)	.35 (.03)	.25 (.03)
Memory	.22 (.04)	.53 (.03)	.25 (.04)

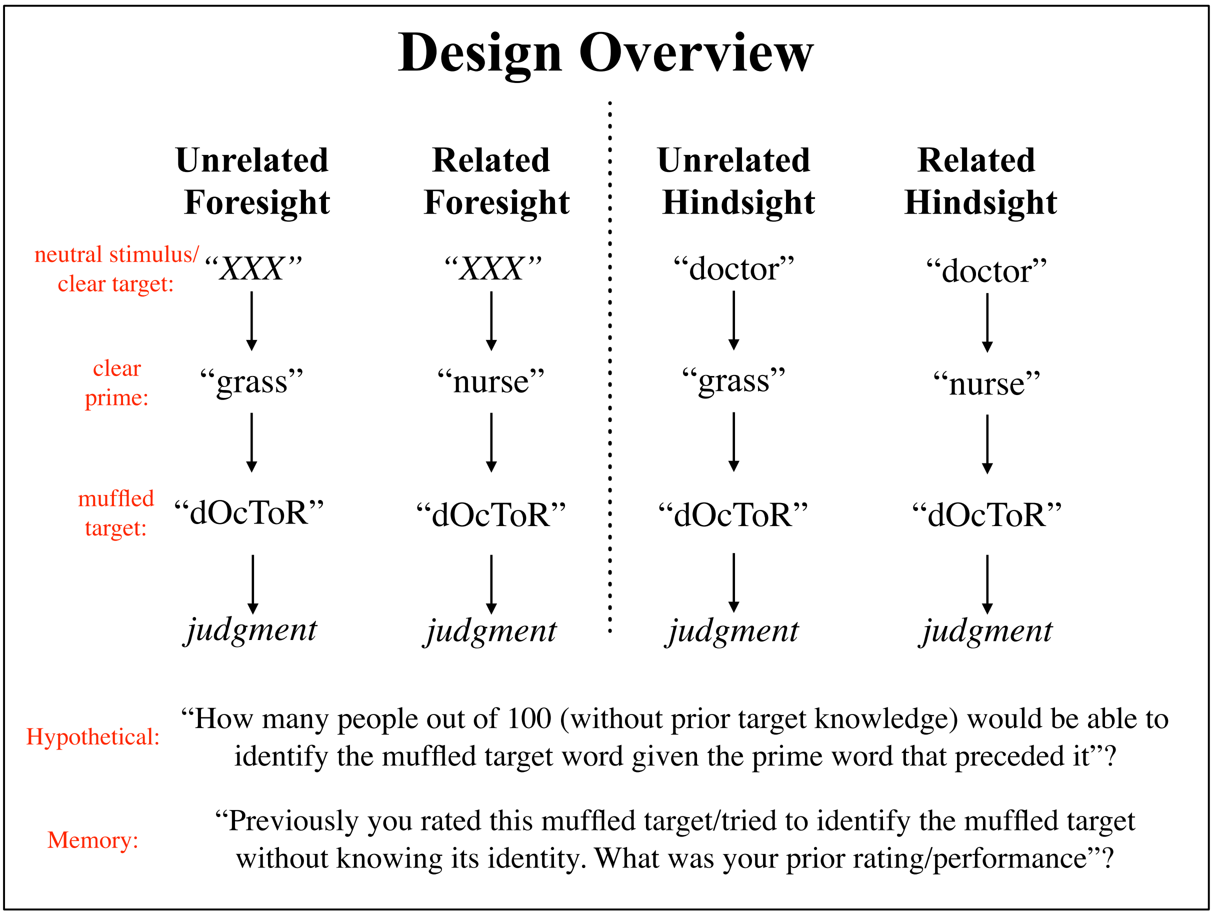

Figure 1: The core 2 (knowledge: foresight, hindsight) x 2 (prime relatedness: related, unrelated) within-subjects design used in all experiments. Details regarding how the different stimuli in the design were presented to participants across experiments vary, but the important design feature is that participants made either a hypothetical or memory judgment about each muffled target either after being made aware of its identity (hindsight) or not (foresight) and after it was preceded by either a related or unrelated prime word. “Hypothetical” and “memory” refer to the judgment that participants were required to make about the muffled target. The alternating upper and lower case letters for the muffled target is meant as a visual depiction of auditory target distortion.

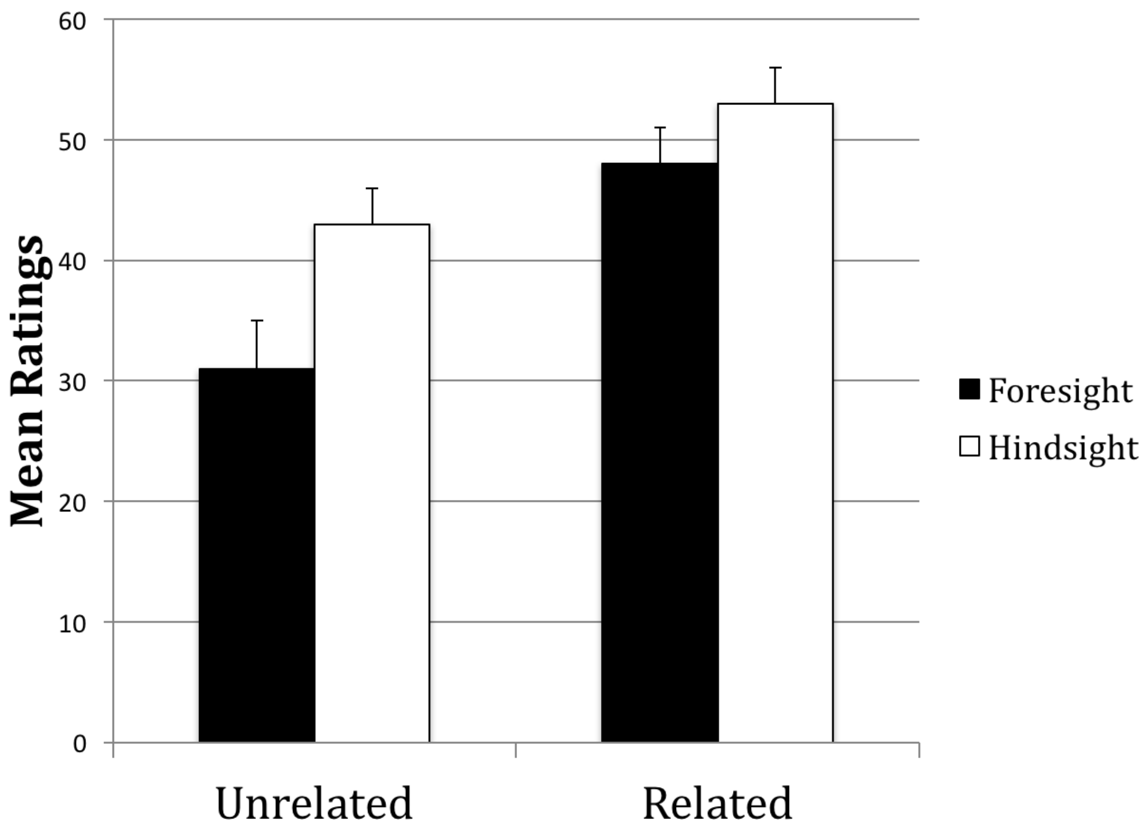

Figure 2. Mean ratings of the percentage of people who would be able to correctly identify the muffled target without prior target knowledge, but with the help of the associative prime on "Rate" trials in the hypothetical design of Experiment 1. The data are plotted as a function of prime relatedness (related, unrelated) and prior target knowledge (foresight, hindsight). Error bars represent standard error of the mean.

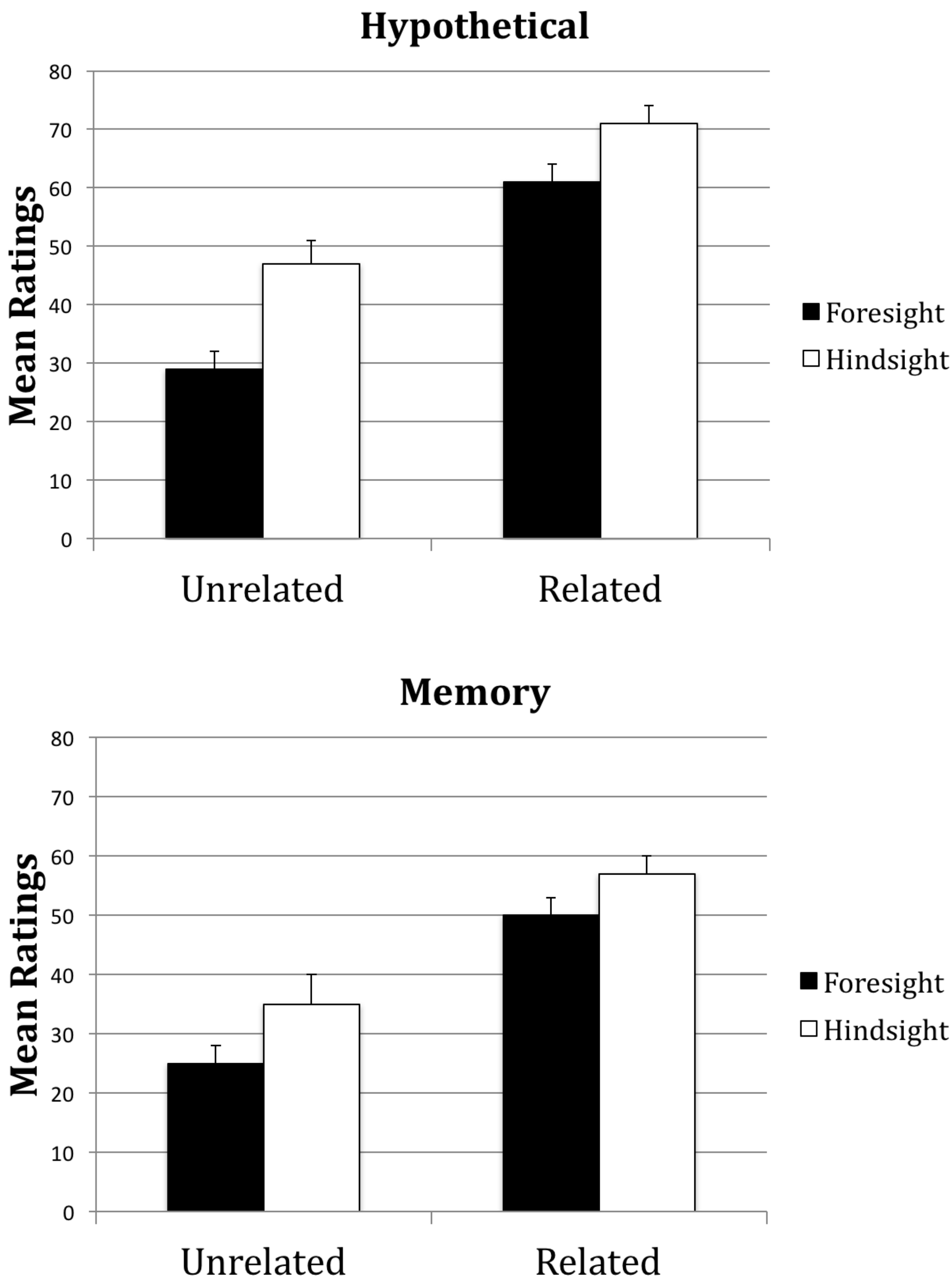

Figure 3. Mean ratings in Phase 2 in the "Hypothetical" (top panel) and "Memory" (bottom panel) groups of Experiment 2. Participants in the Hypothetical group rated the percentage of their peers who would be able to correctly identify the muffled target without prior target knowledge, but with the help of the associative prime. Participants in the Memory group tried to remember analogous judgments made previously in Phase 1. The data are plotted as a function of prime relatedness (related, unrelated) and prior target knowledge (foresight, hindsight). Error bars represent standard error of the mean.

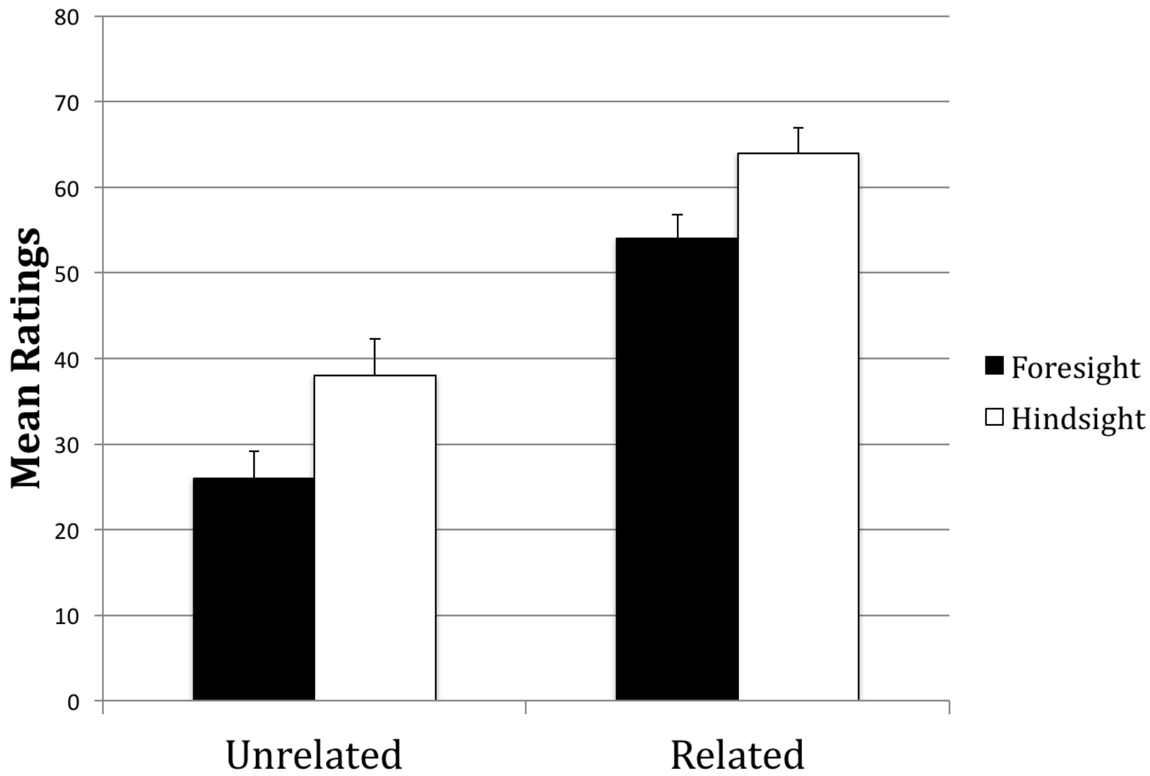

Figure 4. Mean Phase-2 memory estimates for Phase-1 muffled-target identification estimates in Phase 2 of Experiment 3A as a function of prime relatedness and prior target knowledge. Error bars represent standard error of the mean.

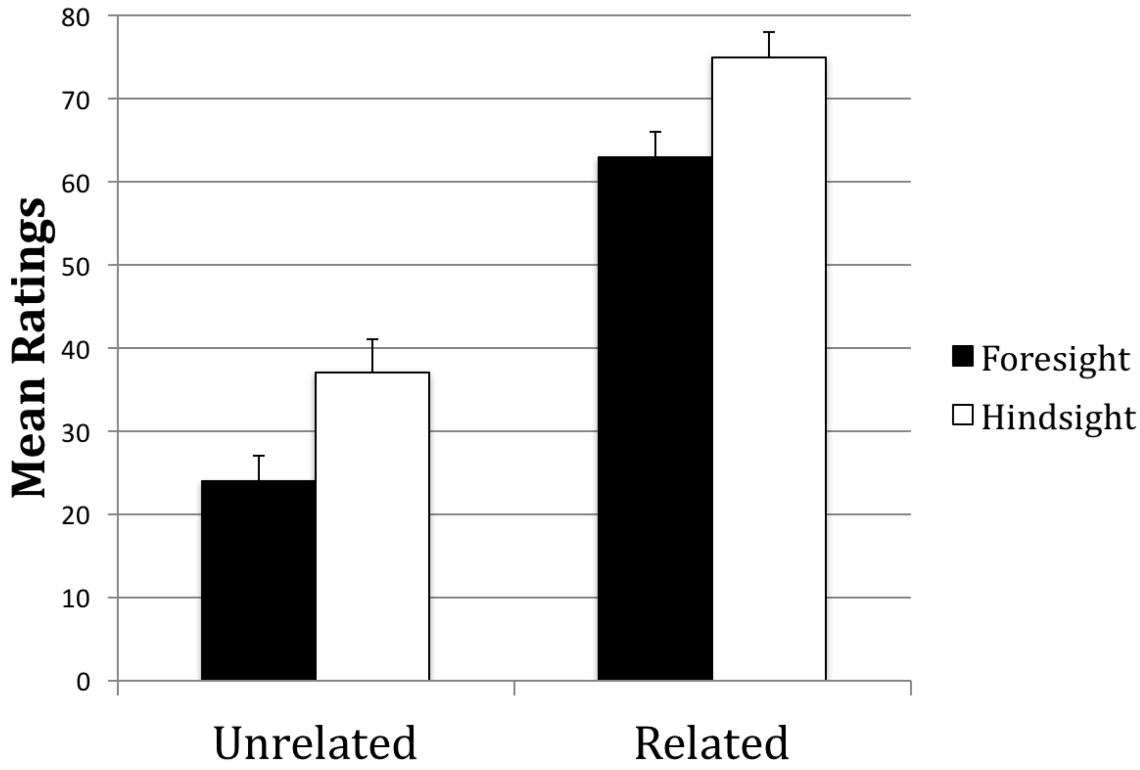

Figure 5. Mean Phase-2 memory estimates for Phase-1 muffled-target identification attempts in Experiment 3B as a function of prime relatedness and prior target knowledge. Error bars represent standard error of the mean.