"I Knew That Answer Before You Told Me...Didn't I?": Subjective Experience Versus Objective Measures of the Knew-it-all-along Effect

by

Michelle Marie Arnold
B.A., University of Lethbridge, 1997
M.Sc., University of Victoria, 2001

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Psychology

Supervisor: Dr. D. Stephen Lindsay

Abstract

The knew-it-all-along (KIA) effect occurs when individuals report that they had previously known something that they learned only recently. Participants in a traditional KIA experiment first rate on a number scale the likelihood of one or more given responses being the correct answer for trivia-like questions (Phase 1); in the feedback phase they are shown the correct answers for a portion of the questions; and in the final phase they are asked to ignore the feedback and give the same number rating for each question that they had given in the first phase. Although several studies have shown that people often have difficulty retrospectively determining the level of knowledge they had prior to the occurrence of feedback, there is no research exploring the subjective experience of the effect. We incorporated a Remember/Just Know/Guess judgment in a traditional (Experiment 1) and a modified-traditional (Experiment 2: 2-alternative-forced-choice) KIA paradigm. In the modified paradigm the number scale was eliminated, and participants simply chose which of two response alternatives they believed to be the correct answer for each trivia question. Experiments 3 - 5 were similar in format to Experiments 1 and 2, but the trivia stimuli were replaced with word puzzles, which were expected to be better suited to inducing a feeling of having known it all along because answers to trivia questions typically seem arbitrary, whereas solutions to word puzzles give rise to ah-ha experiences. A typical KIA effect was observed in all five experiments, but evidence for an accompanying subjective *feeling* of knew-it-all-along was found only with word puzzle stimuli.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I never know how adequately to complete this section – so many people deserve thanks for the various contributions that they have made to my progress toward the completion of my degree. There does not seem to be the right words to express my appreciation, but I will try my best.

I thank my supervisor, Steve Lindsay, for his guidance and assistance throughout this whole process, and for putting up with me over the years. I also thank my committee members, Mike Masson, Helena Kadlec, and John Anderson, for their valued input that has helped to shape this dissertation.

Many friends and colleagues have made this process so much more enjoyable than it would have been in their absence. Thanks to Iris van Rooij and Denise Broekman for being amazing friends, and for always knowing the best things to say to me when things got tough. I also wish to thank Dana Braseth, who has been an incredible supporter from the beginning of this wild ride. A special thanks to my co-conspirators in the lab, Joshua Mira Goldberg and Leora Dahl, who have seen both my highs and lows and have stuck by me through both. Thanks are also in order for those who went before me and led by example; Vincenza Gruppuso, Tanya Berry, and Anna-Lisa Cohen.

Finally, I wish to thank my family and friends back home in Alberta, who always tried to support me in the best ways they could, and who were always smart enough to never let me forget my roots.

Dedication

To my grandmother, Jean Arnold, who has always been my greatest and most beloved teacher.

"I Knew That Answer Before You Told Me...Didn't I?": Subjective Experience Versus Objective Measures of the Knew-it-all-along Effect

Several studies have shown that people often have difficulty retrospectively determining the level of knowledge they had prior to the occurrence of an event (e.g., Fischhoff, 1975; Hasher, Attig, & Alba, 1981; Wood, 1978). Fischhoff (1977) coined the phrase the *knew-it-all-along* (KIA) effect to describe this result, although it also is commonly referred to as *hindsight bias*. In a traditional KIA paradigm participants respond to a set of questions, after which they are given feedback (i.e., the correct answers) for a portion of the questions; participants are then given the questions and instructed to respond with the same answers that they had given prior to being exposed to the feedback. The KIA effect occurs when participants correctly answer significantly more feedback questions (in comparison to non-feedback questions), indicating an overestimation in the amount of knowledge they believe they previously possessed (Fischhoff, 1977).

Most studies of the KIA effect have used one of two paradigms. In a *memory design,* the effects of feedback are determined by comparing the foresight and hindsight judgments within-subjects (e.g., Dehn & Erdfelder, 1998; Fischhoff & Beyth, 1975). Specifically, each participant completes the same set of judgments twice, once before and once after exposure to correct feedback. Instructions for the second test of the memory condition usually involve telling participants to complete the judgments with the exact same answers that they had given to the items prior to receiving the feedback (Hasher et al., 1981), or as someone who had not been exposed to the answers (Wood, 1978). A KIA effect is found for the memory condition when hindsight judgments are significantly

closer to the correct answers for feedback items in comparison to items for which feedback was not presented. For example, Wood (1978; Experiment 1) gave participants 40 true/false statements to rate on a 7-point scale (from *definitely false* to *definitely true*); in the second part of the experiment participants were asked to study 20 of the statements, with each statement marked as true or false. In the final part of the experiment the participants were given all of the true/false statements, and they were instructed to ignore the feedback information that they had received and to rate the items as they had in the first part of the experiment. The results demonstrated a typical KIA effect: The ratings in the final part of the experiment for the feedback items showed a greater shift toward the correct answers (for both true and false statements) than the ratings for the items that had not received any feedback information.

In a *hypothetical design*, foresight and hindsight judgments are compared between-subjects (e.g., Fischhoff, 1975; Mazursky & Ofir, 1990). Judgments made in the presence of feedback for one group (hindsight group) are compared to judgments of the same stimuli for a second group of participants not exposed to the answers (foresight group). The participants in the hindsight group are normally instructed to complete the judgments as they would have had they not been provided with the answers (Fischhoff, 1975). In the hypothetical design, a hindsight bias is said to exist if the effect of feedback leads to a significant difference between the judgments in the foresight and hindsight groups, in that the hindsight group rates the items more similar to the solutions than the foresight group. It is important to emphasize that, although a KIA effect has been demonstrated with both a memory and hypothetical design, there is no guarantee that both paradigms tap the same mechanism. A comprehensive analysis of hypothetical

versus memory designs is beyond the scope and interest of the present paper, and because the focus of the present work is on memory (i.e., the experimental work presented below is all based on a memory design), this paper concentrates on memory designs. However, the issue of hypothetical versus memory designs will be addressed in the General Discussion, as this issue impacts the more detailed examination of the theoretical explanation of the KIA effect.

Hindsight bias has garnered a large volume of research since the mid-1970s, and it has been demonstrated in a wide variety of settings, such as relationship satisfaction (Halford & Griffith, 2002), forensic psychology (Williams, 1992), gustatory judgments (Pohl, Schwarz, Sczesny, & Stahlberg, 2003), and sporting events (Bonds-Raacke, Fryer, Nicks, & Durr, 2001). Exploration of the KIA effect has taken two somewhat differing paths: 1) a focus on the applied impact of the bias, with an emphasis on manipulations that may reduce/eliminate the effect, and 2) exploring the breadth and moderation of the effect, but with an emphasis on theoretical explanation rather than finding applications specific to fixing the problem in real-world settings. In terms of the applied impact of the effect, a relatively recent focus of hindsight bias research has involved its legal ramifications (e.g., Lieberman & Arndt, 2000). For example, Stallard and Worthington (1998) found that participants were much more likely to assign blame in a litigation case when they were fully informed of the plaintiff's complaint (hindsight condition), in comparison to participants who were provided with all of the details of the case except for the outcome of the event (foresight condition).[1] Although the applied aspect of the

[1] Stallard and Worthington (1998) also included a hindsight debiasing condition that mirrored the hindsight condition except that it included instructions that were intended to focus participants on using only the information up to the outcome of the event (e.g., instructing them to use only the information the

KIA effect is an interesting and important issue, the focus of the research and discussion in this paper is on the effect itself and the potential theoretical explanations of the hindsight bias data.

In terms of a theoretical explanation of the effect, Fischhoff (1975, 1977; Fischhoff & Beyth, 1975) proposed that the KIA effect may result from an automatic assimilation of the correct feedback with pre-existing knowledge (i.e., a memory impairment approach). A feeling of "knew it all along" occurs because the assimilation of new information with prior knowledge effectively eradicates the original knowledge state, making it impossible for an individual to "recapture" his/her previous level of knowledge. Because this process is automatic and immediate it is difficult for people to comprehend the impact post-event information has on their perception of past knowledge, even when they are warned about the phenomenon.[2]

The automatic assimilation hypothesis does not explicitly make a distinction between semantic and episodic memory systems, such that the assimilation of feedback occurs within a semantic memory system. However, the hypothesis implicitly denotes a

---

defendants would have had available at the time of their actions that led to the case, and not to be swayed by outcome information). These debiasing instructions appeared to moderate the effect, as the researchers found that participants in this hindsight debiasing condition performed more like the participants in the foresight condition than those in the hindsight condition (although there was still a hindsight bias present).

[2] The idea of feedback "overwriting" memory in a KIA paradigm is very similar to one prominent explanation of the misinformation effect. Specifically, Loftus and colleagues (Loftus, 1979; Loftus, Miller, & Burns, 1978) have found that misleading post-event information can lead participants to report details of an event that were only suggested. This outcome is known as the misinformation effect. In the classic misinformation paradigm, participants view a series of slides of an automobile accident, with the critical slide containing either a stop sign or yield sign. After viewing the slides, participants are exposed to consistent or misleading information. For example, in the misleading condition participants who saw a slide containing a stop sign might be asked "How fast was the car going when it went through the yield sign?" Not only will some participants come to report the suggested information, but in some cases this misleading information is also rated high on a confidence scale for having occurred (Loftus et al., 1978). Loftus (1979) proposed that misleading information can become integrated into recollection and alter a person's memory for that event. Further, this alteration of memory can make it impossible for a person to "recapture" the memory they had for an event prior to receiving misleading suggestions.

separation between semantic and episodic memory: Feedback is said to alter semantic knowledge, but the automatic assimilation theory does not explicitly claim that it overwrites the prior episodic event itself. For example, giving participants the correct answers to general-knowledge questions in a KIA experiment hinders their ability to remember accurately the answers they gave prior to the feedback, but it likely does not leave them unable to recognize the fact that they had previously answered the same questions. Further, this hypothesis implies that the updating effect of feedback on existing knowledge leaves no trace of its occurrence in the semantic system; rather, the relevant knowledge and beliefs are simply altered in accord with the new information. Fischhoff (1977) pointed to the failure of his debiasing instructions (i.e., informing participants of the effect and cautioning them to avoid overestimating their previous knowledge) to reduce the hindsight bias as supporting the notion that participants lack awareness that the post-event information influenced their hindsight judgments: Even when warned of the bias, participants did not adjust adequately for the effects of being exposed to feedback information.

Following Fischhoff's (1975, 1977; Fischhoff & Beyth, 1977) seminal work on the phenomenon, a large majority of the research exploring the KIA effect has been designed to explore the theoretical implications of the effect, and specifically, the validity of this memory impairment account of the phenomenon (e.g., Davies, 1987; Hasher et al., 1981; Hoffrage, Herwig, & Gigerenzer, 2000). As an alternative to the assimilation account, Jacoby and Kelley (1987) proposed an attributional approach to the KIA effect, in which they argued that giving participants feedback "spoils" their subjective experience, thereby contaminating the chief basis upon which they would judge an

answer. For example, it is likely that feedback information is more accessible at test (e.g., due to recency) than prior knowledge, and this accessibility leads to more fluent processing of the feedback information. High accessibility of the feedback information would not, in itself, lead to a KIA effect, but rather individuals must erroneously attribute the fluently generated feedback information to prior knowledge. It is important to note that there is a critical distinction between Jacoby and Kelley's (1987) attributional approach and the automatic assimilation hypothesis. For the attributional approach, no claims of a destruction of original memory are made and the operation of unconscious processes is not constrained to the time of feedback, in that the unaware influences of memory could occur at the time of retrieval/reconstruction of the hindsight judgments. Unlike the automatic assimilation theory, an attributional approach to the KIA effect does not distinguish between remembering an original knowledge state and reconstructing it.

Although it does not address the issue directly (i.e., with direct experimental work), the attributional approach to the KIA effect does raise the important question of subjective phenomenology: How do participants subjectively experience the KIA effect? Many researchers discuss the effect in terms that imply that participants have a *feeling* of knowing the newly-acquired knowledge in foresight (e.g., Mazursky & Ofir, 1990; Sanna, N. Schwarz, & Small, 2002; Stahlberg & Maas, 1998), but to date there has been no published research that has concretely measured subjective experience in a typical KIA paradigm. The issue of subjective phenomenology is important because increased ratings for feedback items in hindsight do not necessarily reflect anything about participants' beliefs regarding the nature of their recollective experience for previously answering those test items. For example, imagine that a participant who is given the

statement "Absinthe is a liqueur" (Fischhoff, 1977) claims that s/he is 70% sure that the statement is true, but after receiving confirmatory feedback states that s/he had given a rating of 85% in foresight. This increase in the rating demonstrates the typical hindsight bias, but it reveals nothing about how the participant feels about her/his prior state of knowledge. That is, the hindsight rating cannot be taken to mean that the individual is now 85% sure that s/he had known that particular answer in foresight (because the task only asks for the original number, which is not the same as asking for how confident someone was that they had possessed the knowledge prior to receiving feedback). It is possible that an increase in rating for KIA items is accompanied by a belief that this knowledge was known prior to the feedback phase, but the increase in the rating on its own cannot be generalized to a participants' level of confidence or quality of memory experience for these items.

Establishing a distinction between participants' subjective and objective experience for the KIA effect is important because it could be argued that the nature of most KIA paradigms in fact works against the creation of illusory remembering/knowing for feedback items. That is, many previous studies have required participants to respond to a large number of items (e.g., Hell, Gigerenzer, Gauggel, Mall, & Muller, 1988; Sharpe & Adair, 1993), and they often involve collecting responses with large number scales (e.g., having participants respond to each item with any number between .00 and 1.00) or numerous alternatives (e.g., Fischhoff, 1975; Goethals & Reckman, 1973; Hardt & Pohl, 2003). For example, Goethals and Reckman (1973) had their participants complete agree/disagree ratings for 30 statements (e.g., the use of bussing to achieve ethnic balance in schools); each of these ratings was performed on a 31-point scale.

Further, the participants had to give their degree of confidence for each statement rating on a 17-point scale. In the second part of the experiment the participants took part in a discussion of one of the 30 statements, and during this discussion a confederate attempted to change the participants' attitudes on that issue by presenting persuasive reasons for the opposite belief (e.g., pro-bussing participants heard anti-bussing reasons). Finally, participants were required to re-rate 8 of the 30 original statements (including the discussed statement), and they were specifically asked to remember how they had rated each of the statements in the first part of the experiment.

Goethals and Reckman (1973) found that participants were significantly more likely to move their ratings toward the opposite belief for the statement that had been used during the discussion (i.e., participants who had originally been pro-bussing subsequently claimed to have been closer to the anti-bussing side of the scale) than for the statements that had not been discussed. The researchers claimed that the participants altered their past attitudes so that they matched their current attitudes because "this allows them to *feel* [italics added] that the position they hold now is the one they have always held" (p. 498). However, there are no data that demonstrate that the participants did feel that the re-ratings they provided in fact did match their original ratings; that is, participants had to reconstruct the ratings they gave on a 31-point scale for numerous items (even though they only had to re-rate a portion of the original statements), and it is possible that they had little confidence in the re-ratings that they were forced to provide. Therefore, although the objective measures demonstrated a hindsight bias in these types of paradigms (whether it be consistency in attitudes or a KIA effect), it is unlikely that participants believe in hindsight (i.e., experience a feeling of remembering or knowing)

that they gave those specific responses for the feedback items because of the inherent difficulty of the final test (e.g., reconstructing exact numbers on a large scale for a large set of items).

One of the main motivations for the present set of experiments was to measure separately both the objective and subjective characteristics of the KIA effect. To gauge the recollective experience of the KIA effect we chose to implement a "Remember/Know" judgment, which has typically been defined in the following terms:

> Often, when *remembering* a previous event or occurrence, we consciously recollect and become aware of aspects of the previous experience. At other times, we simply *know* that something has occurred before, but without being able consciously to recollect anything about its occurrence or what we experienced at the time. (Gardiner & Java, 1990, p. 25)

Because participants are required to give a response in the final test of the KIA paradigm (i.e., they must respond with the value they believe they gave in Test 1, even when unsure of their Test 1 responses) we added a "Guess" category to the judgment (cf. Gardiner, Ramponi, & Richardson-Klavehn 2002). Consequently, for the KIA test items, if participants truly have the belief or feeling that they knew the answers in foresight then they often should give these items a rating of "know" (or perhaps "remember"). However, if participants do not have an accompanying subjective feeling of knew-it-all-along, then the pattern of results should show a higher frequency of "guess" responses for the judgment task.

A key issue surrounding the use of Remember/Know judgments is that the interpretation of these judgment data depends on the underlying theoretical model of

memory. More specifically, as Gardiner et al. (2002) noted, there are two general types of Remember/Know theories; quantitative versus qualitative. The majority of Remember/Know models fall under the qualitative approach, and these types of theories emphasize the idea that remembering is the result of two distinct processes that give rise to different types of subjective experience; namely, recollection and familiarity. However, the qualitative approaches differ in how they define the nature of the underlying structures responsible for recollection and familiarity. For example, in a standard Remember/Know paradigm, some researchers interpret the "remember" option as a measure of recollection and the "know" option as an index of familiarity (e.g., Gardiner, 1988; Gardiner, Kaminska, Dixon, & Java, 1996). Conversely, Jacoby and colleagues (e.g., Jacoby, Jones, & Dolan, 1998; Jacoby, Yonelinas, & Jennings, 1997) have argued that the "know" option should not be taken as a straightforward measure of familiarity because "remember" responses displace "know" responses when recollection and familiarity co-occur: An individual who believes that an event is old will only choose "know" if s/he is unable to recollect specific details of this prior event. Additionally, the equations for estimating recollection and familiarity in Jacoby's (1991) dual process model rest upon the assumption that conscious (recollection) and unconscious (familiarity) processing are independent of one another; that is, conscious and unconscious processing can occur either in isolation or together. Jacoby (1991; Kelley & Jacoby, 1998, 2000) argued that this independence assumption is able to incorporate the experimental results better than either the assumption that the two types of processing never occur together (exclusivity) or the assumption that conscious processing can never occur without unconscious processing (redundancy).

Quantitative approaches to Remember/Know data specify that the difference between remembering and knowing is dependent on the decisional processes; both judgments are based on the same memory traces (i.e., the same information), and they simply reflect differences such as trace strength (e.g., Donaldson, 1996; cf. Dunn, 2004).[3] Similar to qualitative models of Remember/Know judgments, the quantitative approaches differ in how they define the decisional processes that lead to a "remember" or "know" response. For example, a classic quantitative interpretation of Remember/Know data is that "know" responses in a recognition task represent the divide between judging items to be "old/new," whereas the "remember" responses correspond to the high confidence "old" judgments (Donaldson, 1996). Conversely, Rotello, Macmillan, and Reeder (2004) argued that, although recollection and familiarity are not independent processes, two dimensions are required to model recognition data; one dimension is responsible for producing the overall "old/new" recognition judgments, and the second dimension distinguishes between "remember/know" experiences.

The goal of the present research is to measure relative differences in subjective phenomenology, rather than compare and contrast quantitative and qualitative models (e.g., focus is on whether participants always claim to be guessing that they possessed the feedback information in foresight, or whether they at least sometimes claim to remember/know they previously knew the information). Nonetheless, the related issue of Remember/Know theoretical models will be examined further in the General Discussion

[3] Gruppuso, Lindsay, and Kelley (1997; see also Bodner & Lindsay, 2003) proposed an explanation of the Remember/Know distinction that combined aspects of both the qualitative and quantitative approaches. They suggested that, "rather than arising from two distinct and a priori memory processes, recollection and familiarity are ad hoc categories of memory influences, with the constitution of the two categories dependent on the specifics of the situation" (p. 273). This presentation of the Remember/Know distinction and underlying theories is meant as a simple introduction to the topic, and Gruppuso et al.'s theoretical account, along with some of the other approaches, will be explored further in the General Discussion.

because interpreting "know" responses relies on assumptions about the underlying relationship between remembering and knowing.

Although a Remember/Know judgment was chosen over a confidence measure (a more straightforward judgment of subjective experience) for use in the present experiments, there is the important question of how well participants understand the distinction between the "remember" and "know" options (e.g., recollection vs. familiarity in the absence of recollection), and how effectively they use them. For example, researchers have found that the testing procedure, such as the inclusion of a "guess" option or a one-step versus two-step judgment, can alter how the "remember" and "know" categories are used (e.g., Eldridge, Sarfatti, & Knowlton, 2002; Gardiner et al., 2002). Although the potential difficulty of understanding typical Remember/Know instructions is an important topic to consider, it should be noted that great care was taken with the Remember/Know/Guess instructions in all five of the experiments presented in this paper. That is, participants were given extensive instructions on how to use the three options (always with the emphasis that they should be highly confident for both the "remember" and "know" options), given examples of each of the three judgment types, required to answer practice trials (and subsequently questioned about their judgment choices on the practice trials), and required at the end of the experiment to describe in general terms why they chose each type of judgment (i.e., how they used each of the three options in the final test). Additionally, the data from any participant who did not meet the requirements for understanding the distinction between the options (e.g., a participant could not explain how s/he used the options during the experiment) was excluded from the analyses.

The issues raised in the above paragraph surrounding the (potential) difficulty of implementing Remember/Know procedures should not be taken lightly, and they should factor into whether a Remember/Know judgment is used instead of a confidence judgment to measure subjective performance. The Remember/Know judgment is a more complicated measure of subjective phenomenology than a confidence rating (e.g., requires more instructions than a confidence rating, it may seem less intuitive to participants than confidence, etc.), but one of the reasons that it was chosen over a confidence rating to investigate the subjective KIA effect component is that the Remember/Know judgment attempts to move beyond the simple level of confidence in a response; that is, the definition of the judgment allows an individual to be just as confident for a "know" as a "remember" response, with the main distinction between the two categories being the quality of the recollective experience (i.e., presence vs. absence of accompanying details). Indeed, several studies (e.g., Gardiner & Conway, 1999; Gardiner & Java, 1990; Holmes, Waters, & Rajaram, 1998) have shown that, although Remember/Know judgments sometimes are correlated with confidence measures (e.g., high confidence co-occurring with "remember" responses), the two measures do not necessarily converge; a high level of confidence does not automatically equate into an individual being able to recollect consciously details of a prior event (Gardiner & Java, 1990). Further, although it is to be expected that these two measures are correlated in many situations (e.g., remembering specific details leads you to be more confident than if an item just "feels" old), there is evidence to suggest that the two measures are not interchangeable. For example, Rajaram, Hamilton, and Bolton (2002) administered a confidence measure and a Remember/Know measure to both control and amnesic

participants. The researchers argued that if the two judgments quantify the same information then amnesic participants should be impaired on both measures, in comparison to the control participants.

To test this hypothesis, Rajaram et al. (2002) modified Gardiner and Java's (1990) word-nonword paradigm. Specifically, the control and amnesic participants each studied two separate lists that contained both words and nonwords (with a 1-week interval between lists). Participants were required to make studied/new judgments for both lists, and for any items given a "studied" response they had to complete either a Remember/Know judgment (list 1) or a confidence judgment (list 2). Rajaram et al. found that control participants demonstrated a cross-over effect: On the Remember/Know measure there were more R judgments for words and more K responses for nonwords, but "sure" responses on the confidence judgment were higher than "unsure" responses for both the words and nonwords. Amnesic participants did not show the same pattern as the control participants on the Remember/Know judgment (i.e., no effect or interaction of item type [word/nonword] or response type [remember/know]), but like the controls the amnesics did show more "sure" than "unsure" judgments for both words and nonwords. The researchers argued that this pattern of results – amnesic showing impairment on one judgment but not the other, relative to controls – "showed that states of awareness that accompany memory performance and levels of confidence that accompany memory performance are sensitive to independent variables in very different ways" (p. 234).

The level of overlap between Remember/Know judgments and confidence ratings is important to consider because, as mentioned earlier, the Remember/Know task is more difficult to administer to participants, as well as more difficult for participants to perform

(e.g., prior to participating in a Remember/Know experiment, participants likely never consciously attempted to distinguish between "remembering" details of a past event and just "knowing" that it occurred). Therefore, if these two different judgments are not significantly different from each other (i.e., the Remember/Know measure is not adding any distinct information over a confidence measure) then it would be more beneficial and straightforward to use a confidence rating to measure the subjective phenomenology of hindsight bias. Evidence against an absolute correspondence between the two measures was provided in the previous paragraph, but there also is compelling data against a strict correspondence of the two judgments from some of the Signal Detection Theory (SDT) research in the area (e.g., Rotello et al., 2004; Wixted & Stretch, 2004).[4]

In their examination of the relationship between Remember/Know and confidence judgments, Rotello et al. (2004) argued that the Receiver Operating Characteristic (ROC) curves that typically are observed for recognition data (i.e., data in the form of "old/new" judgments made with a rating scale ranging from low to high confidence) should also be observed for the Remember/Know data. Specifically, they emphasized that the well-known research findings that demonstrate that recognition *z*-transformed ROCs (*z*ROCs) typically have a slope around 0.80 should also be found for the *z*ROCs that are constructed for Remember/Know judgments.[5] Rotello et al. conducted a meta-analysis that looked at *z*ROCs for both recognition (confidence measure) and Remember/Know data, and they found that the mean slopes for the Remember/Know data did not match the typical pattern; the mean *z*ROC slope for the Remember/Know measure ($M = 1.01$)

[4] This discussion assumes that the reader has some background knowledge of SDT (e.g., Green and Swets, 1966).

[5] That is, *z*ROCs calculated from the *z* scores for the probability of hits to false alarms for each point on the confidence scale. For Remember/Know judgments, the (transformed) curve provides a two-point *z*ROC.

appeared to be much greater than the mean recognition *z*ROC slope ($M = .77$). The researchers concluded from these (and additional) data that "remember responses are not simply high-confidence old decisions..." (p.606). Relatedly, Wixted and Stretch (2004) found that the confidence ratings associated with Remember/Know responses (i.e., looking at the "old/new" data when both types of judgments are collected in an experiment) show variability. That is, "remember" and "know" responses can overlap to varying degrees on a confidence scale, and therefore not all "remember" responses are made with high confidence and not all "know" responses are made with lower confidence.

The issue of Remember/Know versus confidence measures of subjective phenomenology is related strongly to the types of theoretical models used to account for Remember/Know data (e.g., single-component approaches typically argue that the high correlation between Remember/Know and confidence judgments provides evidence against the idea of separate processes), and therefore further discussion on the matter is left to the Remember/Know theoretical approaches section of the General Discussion. Nonetheless, as discussed in the preceding paragraphs, there is strong evidence to support the argument that the Remember/Know judgment does not simply capture the same information as a confidence rating, and therefore it is valuable for the present research. For example, one of the reasons we favoured a Remember/Know measure over a confidence scale is that the claims from previous researchers regarding the KIA effect have centred around the idea of the feeling of knowing, (i.e., participants come to believe they possessed the information in foresight, but without necessarily remembering details of having given that information). Consequently, we were interested not only in whether

participants would come to believe that they had possessed the feedback information in hindsight (e.g., always choosing the "guess" option vs. sometimes choosing "know" or "remember"), but also how they would classify these beliefs (e.g., illusory recollection vs. feelings of knowing/familiarity).

The five experiments reported below were designed to explore the subjective experience component of the KIA effect in both a "typical" hindsight experimental design (i.e., with procedures/materials that have often been used to test for the effect) and modified designs. The first two experiments explored subjective phenomenology using standard KIA materials (trivia questions) within a traditional design (respond to questions using a number scale; Experiment 1) and a modified-traditional design (respond to questions by choosing one of two alternatives; Experiment 2). Experiment 3 and 4 matched the traditional and modified-traditional designs of Experiment 1 and 2, respectively, but the trivia questions were replaced with word puzzles to test if differences in the type of stimuli used would lead to differences in the subjective measure of the effect. Finally, Experiment 5 was designed to rule out any impact that the timing of the feedback may have on a word puzzle modified-traditional paradigm (i.e., replication of Experiment 4 with a feedback-timing manipulation).

Experiment 1

The first experiment was designed to replicate the typical KIA effect, with the addition of a measure exploring participants' subjective experience of the effect. In the first test, participants were required to answer a set of trivia questions; half of the questions were difficult to answer (critical items) and the other half of the items were relatively easy to answer (filler items). In the feedback phase participants were shown the

correct answers to half of the critical items, and in the final test participants were given the same trivia items as in the initial test and asked to respond with the exact same answer that they had given to each item in Test 1. Additionally, the final test required participants to make judgments regarding whether they remembered, just knew, or were guessing that they had given those answers in Test 1.

Method

Participants. Nineteen University of Victoria students participated in exchange for optional extra credit in an introductory psychology course. The data from three participants were excluded from the analyses because these participants failed to understand the instructions of the tasks and/or R-JK-G judgment.

Materials. A set of 100 trivia questions was constructed from various sources (e.g., Nelson & Narens, 1980). Half of the questions were critical items that were constructed to be difficult to answer (e.g., "*What do you call a baby echidna?*"), whereas the other fifty questions were designed to be easier to answer and were included as filler items (e.g., "*Which precious gem is red?*"). There were two responses assigned to each question; the correct answer and a plausible foil (e.g., "*puggle*" and "*chuttle*," respectively, for "What do you call a baby echidna?"). Two feedback lists were constructed (*feedback-list* factor) to counterbalance between participants which critical items were shown with feedback (i.e., participants either received feedback for arbitrarily numbered critical items 1-25 or 26-50). A re-worded trivia question was constructed from each critical item for the feedback phase, and these re-worded questions always contained the answer to the critical item (e.g., "*For what animal is a baby called a puggle?*"). Additionally, 15 new filler trivia questions were created for the feedback phase.

Procedure. All of the participants were tested individually on an IBM-compatible personal computer using Schneider's Micro-Experimental Laboratory Professional software package (Schneider, 1988). Participants were seated directly in front of the computer, with the experimenter off to the side. In each phase, the experimenter read the instructions aloud. Participants were instructed that, in Test 1, for each trial a trivia question would appear on the screen and their task was to read the question aloud. After participants completed this task on each trial, both the correct answer and foil for that question were displayed on the screen. Participants were told that the correct and incorrect responses would be separated vertically by a number scale ranging from 1 to 10, and that they must choose the number that they believed best corresponded to the correct answer (see Figure 1 for an example of a complete test trial). Specifically, participants were instructed to use the number scale to indicate their confidence that one of the two responses was the correct answer; a response of 1 or 10 was an indication that they were absolutely sure that the response on that end of the scale was the correct answer, whereas a response of 5 or 6 meant that they were only guessing that the response on that end of the scale was the correct response to the question. For example, if "puggle" was at the 1-endpoint of the scale and "chuttle" was at the 10-endpoint of the scale, and the participant was sure that "puggle" was the correct answer, then s/he was told that s/he should respond by saying "one." Conversely, if the participant was completely guessing that "puggle" was the answer, then s/he was instructed that s/he should respond by saying "five." Further, participants were instructed that it was important that they used the full range of the number scale; for example, the number 3 could be used to indicate that they had more confidence that the response on the 1-endpoint of the scale was the correct

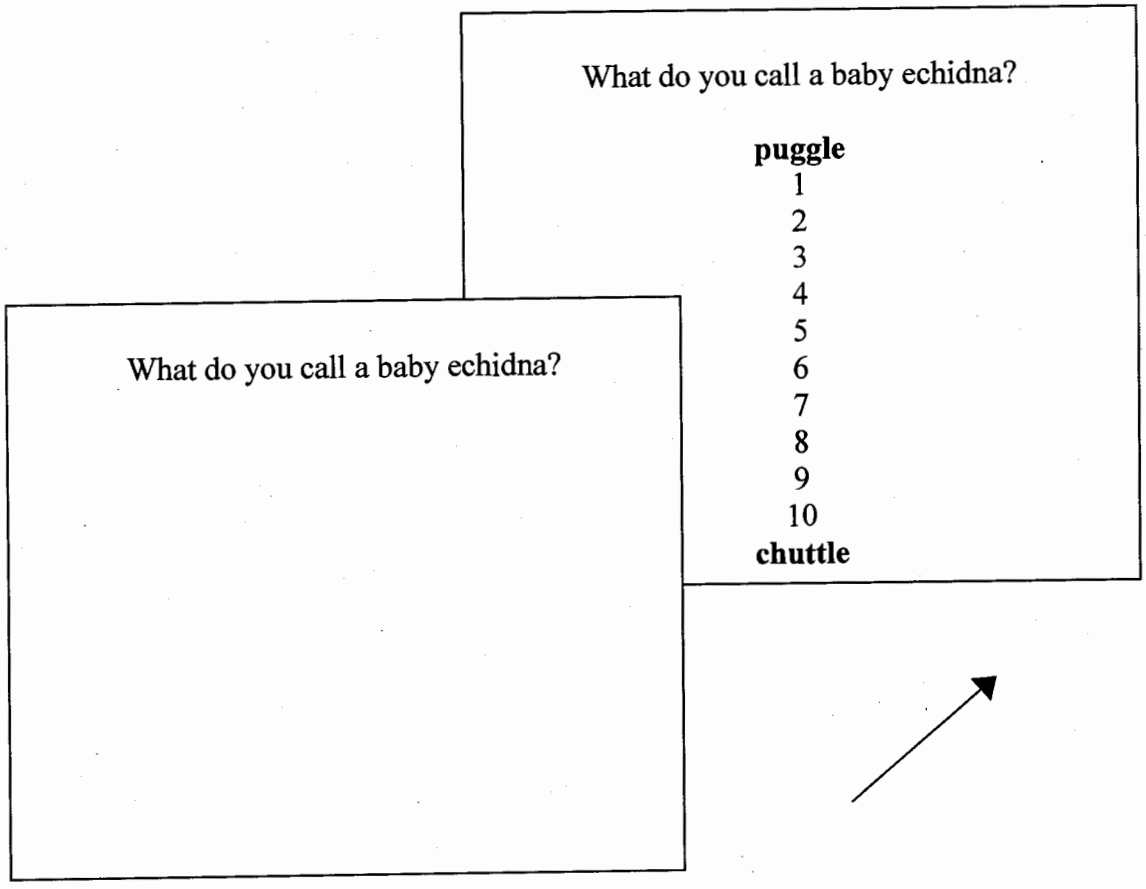


Figure 1. Example of a full trial in Test 1 of Experiment 1.

answer than if they chose the number 5.

To ensure that participants fully understood how to use the number scale, the experimenter walked them through an example prior to starting Test 1. During this example, the experimenter emphasized to participants that there was no midpoint to the scale (i.e., no neutral response), in that the numbers 1-5 always corresponded to the response at the 1-endpoint of the scale and the numbers 6-10 always corresponded to the response at the 10-endpoint of the scale. Additionally, the experimenter accurately informed participants that the correct answers to the trivia questions had been randomly assigned to either the 1- or 10-endpoint of the scale. Finally, participants were told that some of the questions would be very difficult to answer, and therefore it was acceptable if they had to guess at the correct answer (and that they should not become discouraged if they found the test difficult). After finishing Test 1, as a delay activity, participants completed a 20-min, unrelated filler task in which they were shown a series of Snodgrass and Vanderwart (1980) fragmented pictures for 20 different items. Participants were instructed to identify each item as quickly as possible (i.e., as soon as the fragmentation was low enough that they could recognize each item), and they were required to continue with an item until they could correctly identify the picture.

The feedback phase occurred immediately after the filler task. In an attempt to make the feedback less obvious, participants were informed that they were going to complete a two-part Speeded Reading Task (SRT). The feedback phase was disguised as a SRT because pilot testing demonstrated that, because the trivia items themselves were quite memorable, it would be difficult to find a large KIA effect if the feedback was presented in a manner that allowed participants to focus on actively recalling their Test 1

responses (i.e., if participants are given ample time to look at the correct answers presented with the trivia items). In the first part of the SRT, 40 trivia questions (25 re-worded critical items and 15 new filler items) were presented. For each trial, the question was presented near the bottom of the screen, and the correct answer appeared above the question. Participants were told that their first task was to read the answer aloud into a hand-held microphone (i.e., before even looking down at the question); as soon as the microphone picked up the participants' responses, the answer disappeared and participants then were required to read the question aloud. The experimenter explained to participants that their goal in the first part of the SRT was to associate the answer that they had just read aloud with its question. More specifically, participants were informed that it was important that they try their best to associate the correct answers with the questions because doing so would help them in the second part of the SRT (i.e., that reaction time would be measured in Part 2, and that associating the correct responses to the questions would help them reduce their reaction times). Prior to starting the first part of the SRT, participants were told that many of the trivia questions were similar to the questions from Test 1, but that none were the exact same questions from the first test.

The 40 trivia questions from the first part of the SRT were presented two times in the second part of the SRT. Participants were instructed that on each trial a question would appear near the top of the computer screen and that they were to read the question to themselves; they were told that once they had identified the question they were to push a button (which caused the question to disappear), and the answer to the question would be presented in the center of the screen. Participants were instructed to say the answer as quickly as possible into the microphone, and they were informed that once they had

finished saying the response their reaction time would be displayed on the screen. The experimenter emphasized to participants that their goal was to improve their reaction time (i.e., respond faster) as they moved through the 80 trials, and therefore it was important for them to push themselves to respond both accurately and quickly.

The final test (Test 2) occurred immediately after the SRT. Participants were informed that they would be presented with 50 of the 100 trivia questions from Test 1 (to reduce the length of the testing session, only the critical items were presented in Test 2), and that their task was to choose the same number that they had given to each question in Test 1. Further, the experimenter stressed that the researchers were interested in whether participants could consistently select the same numbers that they had chosen in Test 1, and therefore that it was important for participants to ignore the SRT and concentrate on remembering the original number that they had given for each question in Test 1. After choosing their Test 1 response, participants were required to complete two separate Remember-Just Know-Guess (R-JK-G) judgments; the first judgment referred to the side of the scale they had been on in Test 1 (e.g., whether they had chosen a number on the "puggle" or "chuttle" side of the scale) and the second judgment pertained to the specific number they had chosen as their answer. Participants were told to say "Remember" (R) if they could recollect something about having chosen that particular response/number in Test 1, and to say "Just Know" (JK) if they knew that they had chosen that response/number in Test 1 but could not recall anything specific about choosing that response/number for the question.[6] Finally, participants were instructed to say "Guess"

[6] We changed the traditional "know" judgment option to "just know" because we believed participants would better understand the task with this alteration (e.g., "Even though I don't remember any specific details, I *just know* that I gave that response in the first test!"). Therefore, the terms "know" (K) and "just

(G) if they were unsure whether they had chosen that response/number for the trivia question in Test 1. The three R-JK-G judgment options were displayed on the screen during both judgment tasks, and participants always completed the response judgment before the number judgment; to ensure that participants kept on task, the R-JK-G judgment screen was constructed to remind participants whether they were currently completing the response or number judgment task. Finally, to ensure that participants were correctly using the R-JK-G scale, at the end of the experiment they were required to describe the three judgment options in their own words.

Results and Discussion

Initial omnibus within-subject analyses of variance (ANOVAs) showed no reliable effects of the counterbalancing factor of feedback list (all Fs ≤ 1.01, ps $\geq .33$), and therefore the data were collapsed across this variable.

Objective measures of the KIA effect. There was a reliable difference between feedback and control items in the average absolute change in number on the 10-point scale from Test 1 to Test 2; the overall average absolute change in number was greater for feedback items ($M = 1.53$, $SEM = .08$) than control items ($M = 1.34$, $SEM = .07$), $F(1, 15) = 6.00$, $MSE = .05$, $\eta_p^2 = .29$, $p < .03$.

Although the average change in number from Test 1 to 2 was higher for feedback items, this result does not demonstrate a KIA effect because it does not establish the direction of change, and therefore it is important to split the data into items that moved toward versus away from the correct answer on Test 2. In terms of the overall proportion of items given a different number on Test 2 that moved toward the correct answer, there

---

know" (JK) will be used interchangeably throughout the remainder of this paper.

was no reliable difference between feedback items ($M = .50$, $SEM = .04$) and control items ($M = .49$, $SEM = .04$), $F < 1$. However, even though there was no reliable difference between the conditions in the proportion of items moving toward the correct answer on Test 2, there was a significant effect of feedback on the average change in number. Specifically, the average change in number for items moving toward the correct answer on Test 2 was greater for the feedback items ($M = 1.63$, $SEM = .11$), than control items ($M = 1.28$, $SEM = .06$), $F(1, 15) = 16.77$, $MSE = .06$, $\eta_p^2 = .53$, $p = .001$. Conversely, there was no difference in the average change in number for items moving away from the correct answer on Test 2 for the feedback items ($M = 1.41$, $SEM = .07$) and control items ($M = 1.38$, $SEM = .11$), $F < 1$.

Because the number scale did not contain a mid-point (i.e., participants had to choose one of the two responses as the correct answer), the data also can be broken down to look at the number of items switching from the correct to incorrect response or from the incorrect to correct response between Test 1 to Test 2. A KIA effect was found for switching to the correct answer: Participants were more likely to switch from the incorrect answer on Test 1 to the correct answer on Test 2 in the feedback condition ($M = .16$, $SEM = .03$) than the control condition ($M = .07$, $SEM = .02$), $F(1, 15) = 9.54$, $MSE = .01$, $\eta_p^2 = .39$, $p < .01$. The proportion of items switching from the correct answer on Test 1 to the incorrect answer on Test 2 was slightly higher for the feedback ($M = .15$, $SEM = .04$) than control items ($M = .10$, $SEM = .03$), but this difference was not statistically significant, $F(1, 15) = 2.15$, $MSE = .01$, $\eta_p^2 = .13$, $p = .16$.

Subjective measures of the KIA effect. Quantitatively analyzing the R-JK-G judgment was not as straightforward as analyzing the objective data because the

subjective measure of the effect has some data cells with very few observations per participant (e.g., few R responses are given for critical items switched from one side of the scale to the other). To alleviate this problem we transformed the data by taking the natural log of the proportions, which resulted in more normal distributions of the data.[7] The inferential tests reported below – and for the subjective measures of the KIA effect in the following four experiments – are based on the transformed data, but to foster clarity the accompanying means and standard error of the means are reported for the raw proportions. However, the means of the transformed data for the R, JK, and G ratings of the response (i.e., side of the scale) and number judgments for each experiment are shown in Appendix A (Table A1 and A2, respectively).

An additional issue surrounding the analyses of the subjective experience data involves the interpretation of the judgment options. As mentioned in the introduction, certain researchers (e.g., Gardiner et al., 1996) interpret the R option as a measure of recollection and the JK option as a measure of familiarity (F), whereas other researchers (e.g., Jacoby et al., 1998) have claimed that the nature of the R and JK options (i.e., the instructions for when to label something R vs. when to label something JK) leads to an underestimation of F. In general, the main implication of an event that has been labelled as R is that specific details of that event can be brought to mind (e.g., details of when it took place, where something occurred, who was present, etc.), but it often also implies that there is an accompanying feeling of familiarity. Further, events that both are

---

[7] We also added a constant of .50 to both the numerator and denominator of the proportion equation prior to transforming the data to deal with the issue of empty R, JK, or G data cells. Additionally, any participants who did not have data for both the feedback and control measures of interest were removed prior to the analyses (e.g., participants who had switched items from the incorrect answer on Test 1 to the correct answer on Test 2 for feedback items but had no such switches for control items were dropped from the analyses). I thank Michael A. Hunter for suggesting this transformation approach to analyzing the subjective data.

recollected and feel familiar will be grouped under the R response rather than the JK response, and therefore the JK responses overall will be underrepresented. To correct for this problem, "familiarity under independence is conditionalized on the opportunity to have a [JK] judgment" (Jacoby et al., 1998, p. 706), and thus F is calculated by dividing the JK judgments by (1 - R). This measure of familiarity from Jacoby's (1991) independence R/K procedure (IRK) will be included in the results section, where relevant (i.e., in cases where the JK and F data differ or F is reliably different across the feedback and control conditions), for all five experiments.

The overall proportions of items given an R, JK, or G rating for the response and number judgments are shown in Figure 2 and 3, respectively. The transformed proportions of R-JK-G designations for the response judgment and the number judgment trivia items that moved toward the correct answer on Test 2 (but did not switch sides of the scale) were analyzed in separate 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. The main effects of item type and judgment option are not informative (i.e., because, in terms of the raw proportions, these measures sum to 1.00) and therefore only the interaction and subsequent planned comparisons are reported, which holds true for all omnibus ANOVAs reported for the R-JK-G data of the five experiments reported in this paper. One participant was dropped from these analyses for having no feedback items move toward the correct answer on Test 2. Overall, there was no significant interaction between item type and judgment option for the response judgment, $F(2, 28) = 1.90$, $MSE = .37$, $\eta_p^2 = .12$, $p = .17$. However, planned follow-up comparisons showed one trend; that is, there was a tendency for more R response judgments to be given to feedback items ($M = .58$,

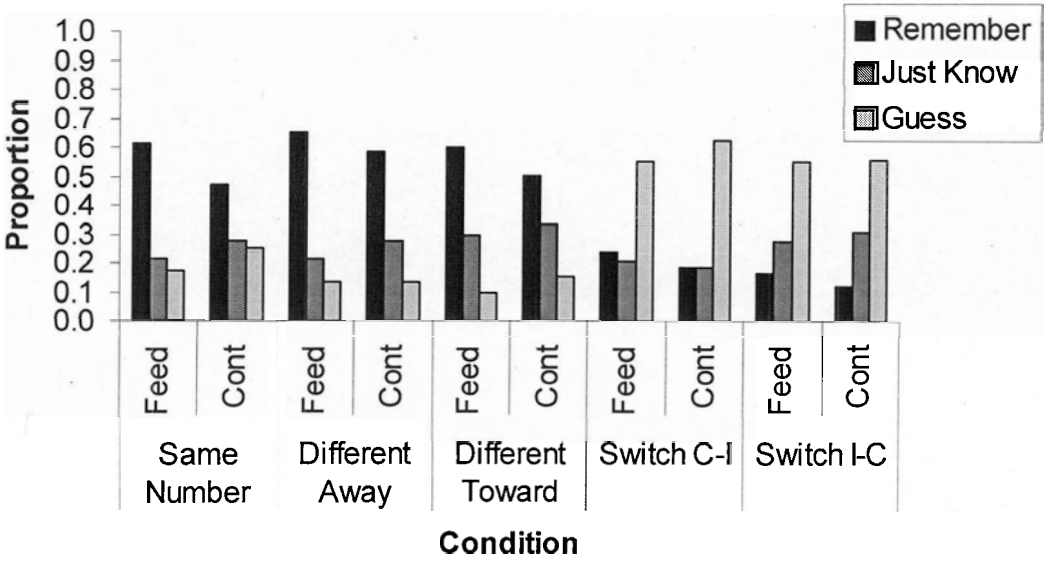


Figure 2. R-JK-G ratings for the response judgment (i.e., side of scale) in Experiment 1 for the feedback and control conditions (collapsed across participants). The judgments are separated by Test 1 and Test 2 responses; a) items given the same number on Test 1 and Test 2 (*same number*), b) items given a number that moves away from the correct answer on Test 2, but does not switch sides of the number scale (*different away*), c) items given a number that moves toward the correct answer on Test 2, but does not switch sides of the number scale (*different toward*), d) items that switch from the correct response on Test 1 to the incorrect response on Test 2 (*switch C-I*), and e) items that switch from the incorrect response on Test 1 to the correct response on Test 2 (*switch I-C*).

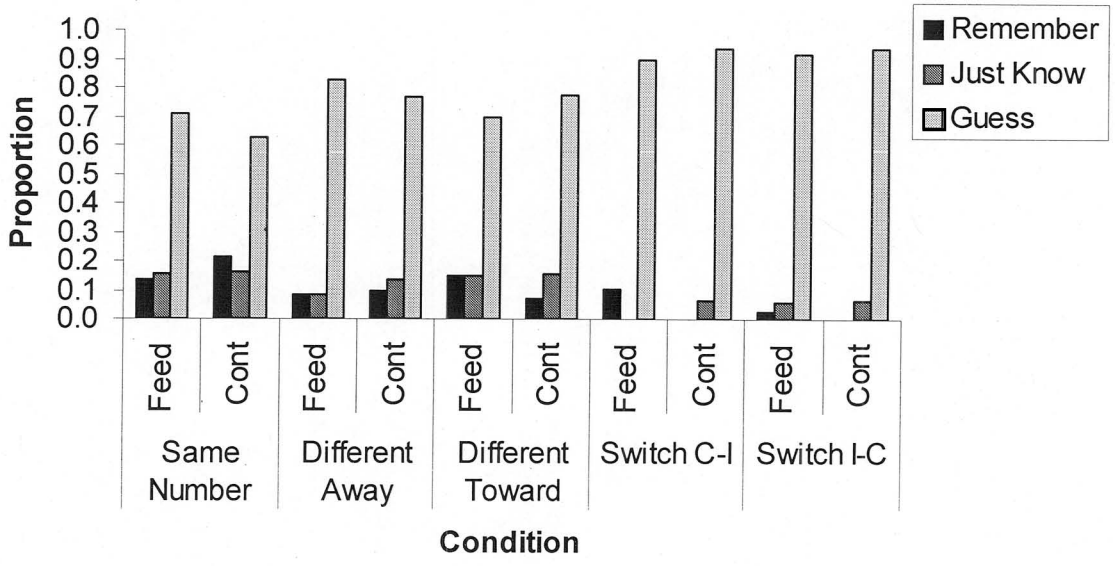


Figure 3. R-JK-G ratings for the number judgment (i.e., side of scale) in Experiment 1 for the feedback and control conditions (collapsed across participants). The judgments are separated by Test 1 and Test 2 responses; a) items given the same number on Test 1 and Test 2 (*same number*), b) items given a number that moves away from the correct answer on Test 2, but does not switch sides of the number scale (*different away*), c) items given a number that moves toward the correct answer on Test 2, but does not switch sides of the number scale (*different toward*), d) items that switch from the correct response on Test 1 to the incorrect response on Test 2 (*switch C-I*), and e) items that switch from the incorrect response on Test 1 to the correct response on Test 2 (*switch I-C*).

SEM = .08) than control items (*M* = .43, *SEM* = .07), *t* (14) = 2.03, *p* = .06. As for the number judgment, there was no interaction between the feedback and control items for the R-JK-G judgment, $F < 1$.

The transformed proportion of R-JK-G designations for the response judgment and the number judgment trivia items that switched from the incorrect answer on Test 1 to the correct answer on Test 2 were also analyzed in separate 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. Seven participants were excluded from the analyses for having zero feedback and/or control items switch to the correct answer on Test 2. For the response judgment, there was no interaction between item type and R-JK-G choices, $F < 1$. However, there was a significant interaction for the number judgment, $F(2, 16) = 6.80$, $MSE = .06$, $\eta_p^2 = .46$, $p < .01$. The planned comparisons showed a reliable difference between the feedback (*M* = .03, *SEM* = .03) and control items (*M* = .00, *SEM* = .00) for the JK judgment option, *t* (8) = 2.70, *p* = .03. It is important to note though that this difference arises because a participant used the JK option one time for the number judgment of his/her feedback items. Additionally, under the IRK model there was no significant difference for F between the feedback (*M* = .03, *SEM* = .03) and control (*M* = .00, *SEM* = .00) conditions of the number judgment, *t* (8) = 1.33, *p* = .22.

The purpose of the current set of experiments is to explore the subjective phenomenology of the KIA effect, but it also may be informative to look at the overall R, JK, and G judgments for the trivia items that participants accurately chose their Test 1 responses for on the final test (i.e., items that are either given the same number, or a number that moves toward/away from the correct answer but does not switch sides of the

scale). The following analyses were conducted on the raw proportions because, unlike for the KIA items, there was a suitable number of observations in each condition (i.e., across the R, JK, and G categories). The proportion of R-JK-G judgments for the trivia questions that were given the same responses on both Test 1 and Test 2 were analyzed in a 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. There was an interaction between item type and judgment, $F(2, 28) = 5.18$, $MSE = .02$, $\eta_p^2 = .27$, $p = .01$. The planned comparisons showed that participants were more likely to judge at Test 2 that they remembered choosing the same responses in Test 1 (i.e., choosing that side of the number scale) for the feedback items ($M = .63$, $SEM = .06$) than for the control items ($M = .52$, $SEM = .06$), $t(14) = 2.58$, $p = .02$. Additionally, there was a marginal trend for a higher level of JK responses in the control condition ($M = .29$, $SEM = .05$) than the feedback condition ($M = .23$, $SEM = .06$), $t(14) = 1.90$, $p = .08$, and this same trend was also found for the G responses of the control ($M = .19$, $SEM = .03$) and feedback ($M = .14$, $SEM = .02$) conditions, $t(14) = 1.92$, $p = .08$. Finally, although there was a marginal trend for the JK judgments, there was no difference in the IRK estimate of F between the feedback ($M = .54$, $SEM = .07$) and control items ($M = .54$, $SEM = .07$), $t < 1$.

The results of Experiment 1 clearly demonstrated a typical hindsight bias, and this pattern was found both in the number scale and in the proportion of items switching from the incorrect side of the scale on Test 1 to the correct side of the scale on Test 2. However, the R-JK-G measure did not produce any concrete evidence that the KIA effect was accompanied by a subjective feeling of knowing the feedback information in foresight. Interestingly, the failure to find evidence of a subjective component to the KIA

effect cannot be due simply to an overall lack of qualitative discriminability between feedback and control items (e.g., that, overall, the feedback and control items just "feel" the same to participants, regardless of any type of manipulation). That is, the analyses for the subjective measure of the trivia questions for which participants stayed on the same side of the scale on both tests revealed that participants were more likely to claim that they were remembering their Test 1 responses for the feedback items.

One potential explanation for this failure to find an accompanying subjective phenomenology for the KIA items is that the structure of responding to the stimuli (i.e., the number scale) interfered with producing a difference in phenomenology between the feedback and control conditions. Experiment 2 was designed to explore whether eliminating the number scale as the objective measure of the bias would lead to a reliable difference between feedback and control items for the subjective component of hindsight bias.

Experiment 2

Because the majority of KIA paradigms require participants to give a numerical response to test items (and because participants are often required to make these numerical responses to a number of test items in succession) it is not surprising that participants have a difficult time trying to reconstruct their Test 1 responses (e.g., participants may often feel that they recollect what particular response they chose, but not the specific number they assigned to it). Given this, it is also not surprising that participants in Experiment 1 rarely reported illusory feelings of remembering or knowing which number they had selected on the first test because – in the context of dozens of ratings – scale responses for any particular trial are unlikely to be memorable.

One of the main goals of Experiment 2 was to produce a KIA effect without using a number scale; participants simply had to choose one of two responses as the correct answer for each trivia question. Additionally, of interest was whether this change in how participants were required to respond to each item would lead to changes in participants' subjective experience of the effect. Therefore, participants were required (as in the first experiment) to complete an R-JK-G judgment in the final test.

Method

Participants. Thirty three University of Victoria students participated in exchange for optional extra credit in an introductory psychology course. The data from five participants were excluded from the analyses because these participants failed to understand the instructions of the tasks and/or R-JK-G judgment.

Materials. The trivia questions from Experiment 1 were used, with the only change to the set of stimuli being that, to make the final test more difficult, a second plausible foil was created for each critical trivia question (i.e., both foils were presented on the final test, along with the correct response). The two foils for each question were counterbalanced (*Test 1-Test 2 foil* factor) so that each foil occurred equally often in Test 1 across participants (e.g., if one foil was presented with the correct answer in Test 1, then for another participant it was only presented in Test 2, and vice versa). Finally, as in Experiment 1, two feedback lists were constructed (*feedback-list* factor) to counterbalance between subjects which critical items were presented during the feedback phase.

Procedure. The basic procedure from Experiment 1 was implemented, but with two major modifications. The first modification involved changing the format of the tests

from a number scale to 2-alternative-forced-choice (2AFC). In Test 1, participants were told that, after reading each question aloud, they would be shown two possible responses and that their task was to choose the response that they believed was the correct answer to the question. As in Experiment 1, the experimenter warned participants that some of the questions were difficult, and that it was fine if they had to guess which of the two responses was the correct answer to the question. In Test 2, participants were instructed that they would be shown three possible responses for each question; (a) the correct response that had been presented in Test 1, (b) the incorrect response that had been presented in Test 1, and (c) an incorrect response that had not been shown in Test 1. The second foil was added to Test 2 to make the task more difficult, as well as to provide a measure of consistency (i.e., if participants were not simply randomly selecting a response on Test 1 without really looking at the two responses, then for the majority of trials they should be able to rule out the new foil as their Test 1 response).[8] All of the instructions for Test 1 and 2 were modified to reflect the 2AFC format, and the R-JK-G judgment in Test 2 was changed from a 2-part task to a single judgment (by eliminating the number R-JK-G judgment task).

The second major modification to the procedure was to conduct the experiment over two days; each participant completed the SRT and Test 2 24 hrs after Test 1. This change was added to make Test 2 more difficult for participants. Specifically, moving to a 2AFC format made Test 1 very memorable for participants, and pilot testing indicated that it was necessary to add a delay to the procedure to reduce the overall hit rate (i.e., correctly choosing the Test 1 response) on Test 2.

---

[8] Participants rarely chose the new foil on the final test, and therefore this feature is not discussed further.

Results and Discussion

None of the initial omnibus within-subject analyses of variance (ANOVAs) showed significant effects of the counterbalancing factors of Test 1-Test 2 foils and feedback list (all Fs < 1), and therefore the data were collapsed across these variables.

Objective measures of the KIA effect. As anticipated, a KIA effect was observed; the proportion of items switching from the incorrect response on Test 1 to the correct response on Test 2 was higher in the feedback condition (M = .23, SEM = .03) than in the control condition (M = .10, SEM = .02), $F(1, 27) = 26.77$, MSE = .01, $\eta_p^2 = .50$, $p < .001$.

There also was a significant effect for items switching from the correct answer on Test 1 to the incorrect answer on Test 2. Here the proportion of switching was higher for the control items (M = .16, SEM = .02) than for the feedback items (M = .09, SEM = .02), $F(1, 27) = 5.43$, MSE = .01, $\eta_p^2 = .17$, $p = .03$. At first glance it may appear that this result is simply an artifact: Because participants have a much higher rate of switching from incorrect to correct for feedback items, the feedback items are necessarily "underrepresented" in the switching from correct to incorrect analysis. However, the proportions of switching for both sets of analyses are not computed from the same base. That is, the proportion of items switching from incorrect on Test 1 to correct on Test 2 is calculated by dividing the number of items that changed from incorrect to correct by the number of items assigned an incorrect response on Test 1, whereas the proportion of items switching from correct on Test 1 to incorrect on Test 2 is calculated by dividing the number of items that changed from correct to incorrect by the number of items given a correct response on Test 1. Further, a comparison of Test 1 responses confirmed that there was no significant difference between feedback and control items in the number of

items given a correct or incorrect response; participants were just as likely to pick the correct response on Test 1 for feedback items ($M = 11.50$, $SEM = .40$) as they were for the control items ($M = 11.18$, $SEM = .36$), $t(27) < 1$. Therefore, it could be argued that, although feedback leads participants to judge incorrectly that they had given the correct information in Test 1, the feedback information also enhances accuracy by reducing the instances of incorrectly claiming that the correct information *had not* been given in foresight.

It is important to note that there was a significant difference between the control items for switching rates; more control items were switched from correct to incorrect than from incorrect to correct, $t(27) = 2.28$, $p = .03$. It is not clear why such differences are found in switching rates for the control condition stimuli (although it is quite possible that the results are simply a Type I error and subsequently would not be replicated), and therefore the interpretation of the feedback versus control result (i.e., showing a higher rate of switching to the incorrect answer for the control items) becomes less clear.

Subjective measures of the KIA effect. The overall proportions of items given an R, JK, or G rating for the response judgment are shown in Figure 4. The transformed proportion of R-JK-G designations for the trivia items that switched from an incorrect answer on Test 1 to a correct answer on Test 2 were analyzed in a 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. Ten participants were excluded from the analyses for having either no feedback or no control items switch to the correct answer on Test 2. There was no significant interaction between item type and judgment option, $F < 1$.

The number scale was eliminated in this experiment to make the task of

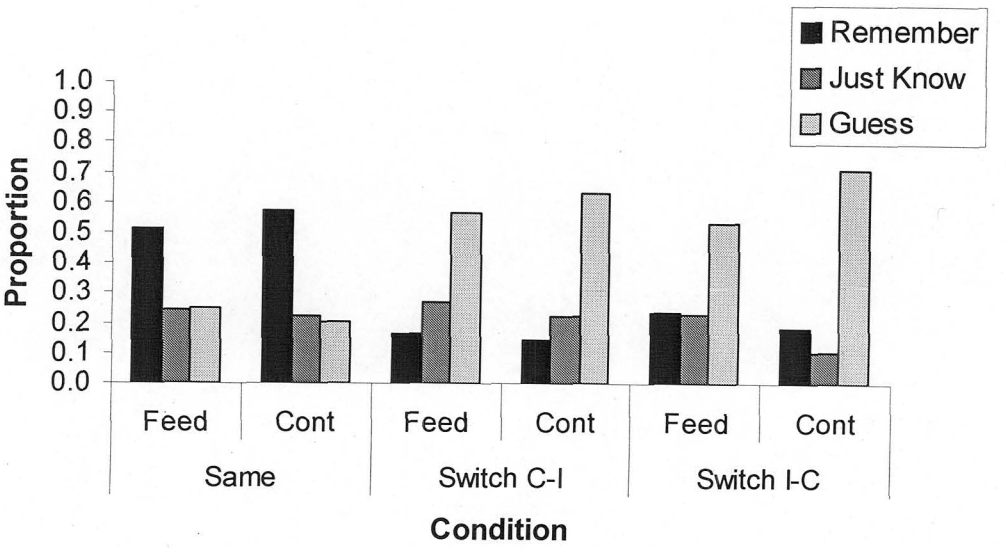


Figure 4. R-JK-G ratings for the response judgment in Experiment 2 for the feedback and control conditions (collapsed across participants). The judgments are separated by Test 1 and Test 2 responses; a) items given the same response on Test 1 and Test 2 (*same*), b) items that switch from the correct response on Test 1 to the incorrect response on Test 2 (*switch C-I*), and c) items that switch from the incorrect response on Test 1 to the correct response on Test 2 (*switch I-C*).

reconstructing Test 1 recollections less complicated for participants, and – although the goal of this manipulation was to find differences in subjective phenomenology between feedback and control KIA items – this change also may impact the subjective experience measure for the items that were given the same responses on both tests. Therefore, even though no difference in subjective experience was found for the feedback and control KIA items, it is important to look at whether this procedural change influenced the subjective component for the stimuli that were given the same responses on both Test 1 and Test 2.

The proportions of R-JK-G judgments for the trivia questions that participants chose the same responses to on both Test 1 and Test 2 were analyzed in a 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. There was only a marginal trend for the interaction, $F(1, 54) = 2.66$, $MSE = .01$, $\eta_p^2 = .09$, $p = .08$, but planned comparisons showed a significant effect for the R judgment; participants were more likely to claim that they remembered giving those same responses in Test 1 to the control items ($M = .57$, $SEM = .04$) than for the feedback items ($M = .52$, $SEM = .03$), $t(27) = 2.19$, $p < .04$. This result for the R option is opposite to the same measure in the first experiment (i.e., where it was the feedback items that received more claims of remembering), but, similar to Experiment 1, there was no reliable difference between control and feedback items for the JK and G options, $ts \leq 1.70$, $ps \geq .10$, or for the IRK estimate of F, $t < 1$.

The data from Experiment 2 replicated the main findings from the first experiment; that is, the objective measure of hindsight bias showed a significant difference between feedback and control items, but no such difference between the

conditions was evident for the subjective measure of the bias. Taken together, the results from Experiment 1 and 2 suggest that the KIA effect is not accompanied by a substantial subjective experience in hindsight that the answers had been remembered or known from foresight. Additionally, as in Experiment 1, there was a significant difference between the feedback and control items for the R judgment of the items for which participants chose the same response to on both Test 1 and 2. Unlike Experiment 1, though, this result in the second experiment demonstrated higher levels of R judgments for the control items than the feedback items. Nevertheless, the important point for the current discussion is that, as in Experiment 1, it is not the case that the phenomenological quality of the feedback and control cannot ever be differentiated (i.e., and that therefore this is the reason why no subjective component to the bias has been found).

The main difference between Experiment 1 and 2 was the methodology used to objectively measure the effect (i.e., number scale vs. 2AFC format). The goal of the next experiment was to investigate whether changing the type of stimuli used to test for the KIA effect would influence the subjective component of the bias.

Experiment 3

The lack of evidence in the first two experiments for a subjective component to the bias suggests that the KIA effect might better be called the "I guess I knew it all along" effect. Nonetheless, although no evidence for an accompanying subjective phenomenology was found in Experiment 1 and 2, it cannot be concluded from these two experiments that the KIA effect is simply (or always) the result of a simple bias toward the correct answers after-the-fact (i.e., after feedback) and that therefore the effect will never be accompanied by some feeling or belief of knowing that specific information

prior to seeing the correct answers. Indeed, it may be possible to demonstrate that, under the right circumstances, participants can have the subjective feeling of knowing or remembering these answers, even though they were unable to provide the correct information in foresight.

Experiment 3 was designed to explore the KIA effect with a set of stimuli that should be more suited to inducing a *feeling* of knew-it-all-along than the trivia items used in the first two experiments. The items used in this third experiment were word puzzles – commonly referred to as "Wordies" – which can be re-arranged or translated to form common words, phrases, or clichés (see Appendix B for examples of Wordies used in this experiment). One of the important differences between the trivia items in the previous experiments and the Wordies is that the Wordies invite a greater degree of interaction between the participants and the items: Even when participants do not immediately know the solution to a word puzzle, they can use various strategies and techniques to try to arrive at the correct solution. Moreover, whereas answers to trivia questions have an arbitrary quality, solving Wordies often gives rise to "ah-ha!" experiences. Additionally, once the answer to a Wordies puzzle is discovered it may seem obvious, as per Kelley and Jacoby's (1987) explanation regarding "spoiled" subjective experiences. Finally, unlike trivia questions, Wordies contain a wealth of information that (once known) specifies the correct answer. Therefore, the nature of the Wordies and the tasks involved in the experiment should lend themselves better to promoting a sense of "I knew that before the feedback!" than the trivia items in the previous two experiments.

Method

Participants. Twenty-four University of Victoria students participated in

exchange for optional extra credit in an introductory psychology course. The data from three participants were excluded from the analyses because these participants failed to understand the instructions of the tasks and/or R-JK-G judgment.

Materials. A pool of 92 Wordies was created from a variety of sources (e.g., Brain Teasers!, n.d.), and a plausible foil was constructed for each correct solution (e.g., "marked counter-top" as a foil for the solution "check-out counter"). Two of the Wordies were used as practice trials at the beginning of Test 1 and 2, and 10 of the Wordies were used as training items during the feedback phase. The remaining 80 items were used as test items during Test 1 and 2 (50 critical items and 30 filler items), and the designation of Wordies to the critical and filler item categories was accomplished through pilot testing of the foils/solutions to determine which items were the most difficult to answer (i.e., the most difficult items designated as critical items). No counterbalancing conditions were necessary because the computer program randomly selected which critical items were presented during the feedback phase for each participant.

Procedure. Participants were tested individually on an IBM-compatible personal computer using Schneider, Eschman, and Zuccolotto's (2002) E-prime software package. In Test 1, participants were told that for each trial they would see a word puzzle appear on the screen for 3 sec; during that period they were to look at the puzzle and come up with ideas and/or solutions for the puzzle in their heads (i.e., they were specifically instructed not to guess out loud). Further, participants were informed that after the 3 sec the puzzle would disappear and two responses separated by a vertical 10-point number scale would appear on the screen. The experimenter informed the participants that one of the responses was always the correct solution to the puzzle, and that their task was to

choose a number that they believed corresponded to the correct solution and that indicated their confidence in that answer. As in Experiment 1, participants were given in-depth instructions regarding how to use the number scale. Prior to starting the first test, they were also shown an example of a word puzzle (i.e., to ensure that the participants understood what it meant to solve a word puzzle, what the format of the puzzles would look like, etc.), and they were also shown a concrete example of how to use the number scale. After finishing Test 1, participants completed a 5-min embedded figures filler task.

As in Experiment 1, the feedback phase occurred immediately after the filler task; however, due to the nature of the Wordies, the feedback tasks deviated from those of the first experiment. The participants were informed that they were going to complete a 2-part Wordies judgment task. In Part 1, referred to as the "understanding-of-the-solution" judgment, they were shown 40 of the word puzzles from Test 1 (25 critical items, 15 filler items). For each trial the puzzle was shown for 10 sec, with the correct solution directly below the puzzle. The experimenter told participants that their task during the 10 sec was to decide if they understood *how* to arrive at that correct solution for the puzzle, and that once participants felt they understood why the solution was the answer they should push a key (and that if they failed to push the key during the 10 sec, the computer would go on to the next trial). To ensure that participants were focused on the task and not simply reading the solution before hitting the button, the experimenter gave participants a brief training period on the topic of what was meant by truly understanding how to arrive at a puzzle's solution. More specifically, the participants were shown different types of puzzles and their solutions, and the experimenter talked them through how to arrive at the solutions (e.g., "The solution is 'once upon a time' because the word

'once' has physically been placed on top of the phrase '4:56 pm,' which itself stands for the more abstract concept of 'time'"). After completing the training session, participants were given the final instruction that, by pushing the button for a particular puzzle, they would be indicating that they could explain to someone else (just as the experimenter had done with them) how to arrive at the solution to the puzzle.

The second part of the Wordies judgment task occurred immediately after Part 1, and its purpose simply was to give participants another pass at the correct solutions to the critical items. For Part 2, referred to as the "common/rare" judgment task, participants were told that they would be shown the same 40 puzzles/solutions from Part 1, but that the puzzles/solutions would only be displayed for 2 sec. After 2 sec, the scale "common - - - - rare" appeared on the screen, and at this point the participants had to decide whether the puzzle/solution represented a common word or phrase (i.e., something they thought they might encounter in a typical day) or whether it represented a rare word or phrase (i.e., something they didn't think they'd typically come across in everyday conversations). Finally, the experimenter informed participants that once the common/rare scale appeared on the screen they would only have 2 sec to respond; if they did not respond within 2 sec the computer would beep at them and go on to the next trial (and they were told it was important to avoid timing out).

Test 2 took place immediately after the Wordies judgment. Participants were told that they would be presented with the same 80 puzzles from Test 1, and that their task was to choose the exact same number that they had given to each puzzle in the first test. The format of the trials was exactly the same as in Test 1, except that the prompt "Your Wordies Puzzle Test 1 answer for this puzzle?" appeared at the bottom of the

responses/number scale screen. As in Experiment 1, after selecting their Test 1 response participants were required to complete two separate R-JK-G judgments (a response judgment and a number judgment). Instructions for both judgments were exactly the same as in Experiment 1, and participants were shown how to use the R-JK-G scale with one of the Wordies examples from Test 1. Finally, both judgment screens were the same as in Experiment 1, and participants were asked to describe the definitions for the R-JK-G options in their own words at the end of the experiment.

Results and Discussion

Objective measures of the KIA effect. There was a significant difference in the average absolute change in number on the 10-point scale from Test 1 to Test 2, with the average absolute change being greater for feedback items ($M = 2.16$, $SEM = .18$) than control items ($M = 1.79$, $SEM = .09$), $F(1, 20) = 9.19$, $MSE = .15$, $\eta_p^2 = .32$, $p < .01$.

The overall proportion of items given a different number that moved toward the correct answer on Test 2 was reliably higher for the feedback items ($M = .65$, $SEM = .03$) than the control items ($M = .44$, $SEM = .03$), $F(1, 20) = 20.77$, $MSE = .02$, $\eta_p^2 = .51$, $p < .001$. Additionally, there was a significant effect of the feedback on the average change in number for items moving toward the correct answer on Test 2: The average change in number was greater for feedback items ($M = 2.45$, $SEM = .27$) than control items ($M = 1.91$, $SEM = .15$), $F(1, 20) = 9.24$, $MSE = .34$, $\eta_p^2 = .32$, $p < .01$, which demonstrates the typical KIA effect. There was no difference in average change in number for items moving away from the correct answer on Test 2 for the feedback ($M = 1.62$, $SEM = .13$) and control items ($M = 1.72$, $SEM = .12$), $F < 1$.

Turning to the data for switching sides of the scale from Test 1 to Test 2 (i.e.,

switching from one response to the other), a KIA effect was found for switching to the correct answers, in that participants were more likely to switch to the correct answer on Test 2 for the feedback items ($M = .42$, $SEM = .08$) than the control items ($M = .19$, $SEM = .05$), $F(1, 14) = 6.52$, $MSE = .06$, $\eta_p^2 = .32$, $p = .02$. There was no significant difference in the proportion of items switching from the correct response on Test 1 to the incorrect response on Test 2 for the feedback ($M = .05$, $SEM = .01$) versus control items ($M = .07$, $SEM = .02$), $F < 1$.

Therefore, as in the previous experiments where the materials were difficult trivia items, the objective measures in Experiment 3 demonstrated that a KIA effect (both in the number scale and for switching sides of the scale) can be found when the items used to test for the effect are word puzzles.

Subjective measures of the KIA effect. The overall proportions of items given an R, JK, or G rating for the response (i.e., side of the scale) and number judgments are shown in Figure 5 and 6, respectively. The transformed proportions of R-JK-G designations for the response judgment and the number judgment items that moved toward the correct answer on Test 2 (but did not switch sides of the scale) were analyzed in separate 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. Overall, there was no significant interaction between item type and judgment option for the response judgment, $F(2, 40) = 2.15$, $MSE = .58$, $\eta_p^2 = .10$, $p = .13$. Planned follow-up comparisons showed one marginal trend; there was a tendency for more G response judgments to be given to control items ($M = .21$, $SEM = .06$) than feedback items ($M = .14$, $SEM = .05$), $t(20) = 1.83$, $p = .08$. There was no interaction between the feedback and control items for the number judgment, $F(2,$

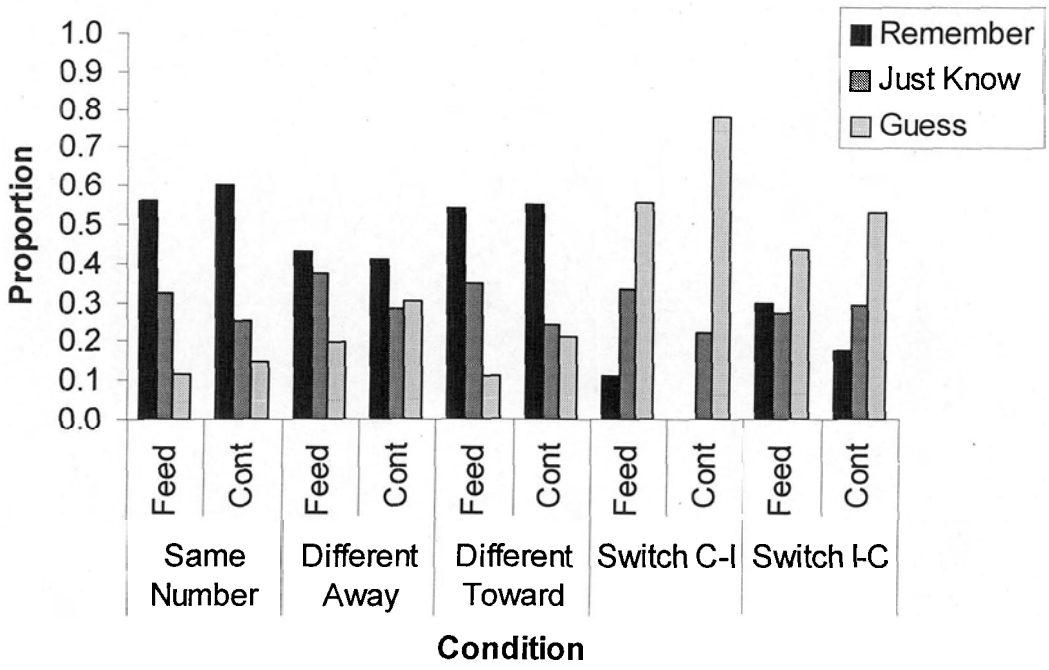


Figure 5. R-JK-G ratings for the response judgment (i.e., side of scale) in Experiment 3 for the feedback and control conditions (collapsed across participants). The judgments are separated by Test 1 and Test 2 responses; a) items given the same number on Test 1 and Test 2 (*same number*), b) items given a number that moves away from the correct answer on Test 2, but does not switch sides of the number scale (*different away*), c) items given a number that moves toward the correct answer on Test 2, but does not switch sides of the number scale (*different toward*), d) items that switch from the correct response on Test 1 to the incorrect response on Test 2 (*switch C-I*), and e) items that switch from the incorrect response on Test 1 to the correct response on Test 2 (*switch I-C*).

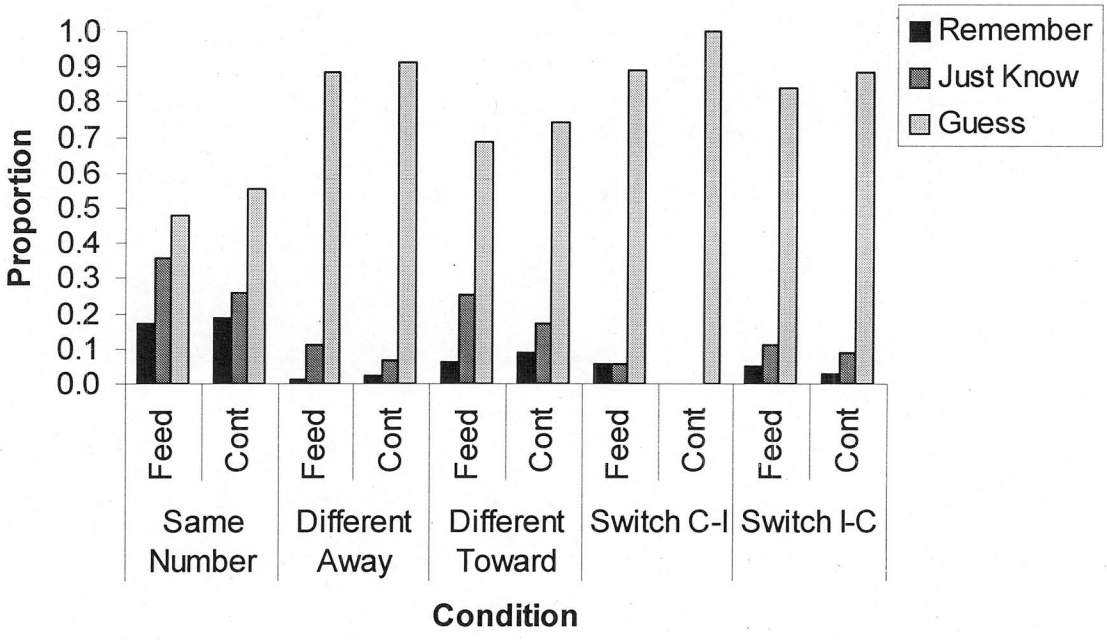


Figure 6. R-JK-G ratings for the number judgment (i.e., side of scale) in Experiment 3 for the feedback and control conditions (collapsed across participants). The judgments are separated by Test 1 and Test 2 responses; a) items given the same number on Test 1 and Test 2 (*same number*), b) items given a number that moves away from the correct answer on Test 2, but does not switch sides of the number scale (*different away*), c) items given a number that moves toward the correct answer on Test 2, but does not switch sides of the number scale (*different toward*), d) items that switch from the correct response on Test 1 to the incorrect response on Test 2 (*switch C-I*), and e) items that switch from the incorrect response on Test 1 to the correct response on Test 2 (*switch I-C*).

40) = 2.20, MSE = .39, η_p^2 = .10, p = .12, and no strong trends were found in the follow-up comparisons, all ts ≤ 1.66, ps ≥ .11.

The transformed proportions of R-JK-G selections for the response and number judgments of items that switched from the incorrect answer on Test 1 to the correct answer on Test 2 were also analyzed in separate 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVAs. Six participants were dropped from the analyses for having zero feedback and/or control items switch to the correct answer on Test 2. For the number judgment, there was no interaction between item type and R-JK-G choices, $F < 1$. The interaction between item type and judgment option also failed to reach significance for the response judgment, $F(2, 28) = 2.74$, $MSE = .27$, $\eta_p^2 = .16$, $p = .08$, although the planned comparisons showed a reliable difference for the G option. Specifically, participants were less likely to choose the G option for feedback items ($M = .38$, $SEM = .09$) than for the control items ($M = .56$, $SEM = .10$), $t\,(14) = 3.16$, $p < .01$. There was also a marginal trend for higher JK response in the control condition ($M = .33$, $SEM = .10$) than in the feedback condition ($M = .28$, $SEM = .08$), $t\,(14) = 1.90$, $p = .08$. However, the data for the IRK model shows a reversed, but reliable, trend; that is, F was higher in the feedback condition ($M = .40$, $SEM = .11$) than in the control condition ($M = .34$, $SEM = .10$), $t\,(14) = 2.29$, $p < .04$. Therefore, although the raw JK response judgment showed no difference between the conditions, the IRK data demonstrated that the correct responses for the feedback items that participants switched to in the final test appeared to feel more familiar to them (i.e., as being their Test 1 responses) than the correct responses that they switched to for the control items.

The proportions of R-JK-G response judgments for the trivia questions that were given the same responses on both Test 1 and Test 2 were analyzed in a 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. There was an interaction between item type and judgment, $F(2, 40) = 6.35$, $MSE = .01$, $\eta_p^2 = .24$, $p < .01$. Planned comparisons showed no significant difference in the R judgment between the feedback ($M = .51$, $SEM = .05$) and control items ($M = .52$, $SEM = .04$), $t < 1$. There were more JK judgments given to the feedback items ($M = .35$, $SEM = .04$) than the control items ($M = .27$, $SEM = .04$), $t\ (20) = 3.12$, $p < .01$, and this pattern was mirrored in the IRK data: A higher level of F was found for the feedback items ($M = .68$, $SEM = .05$) for which the same alternatives were chosen on both tests than for the control items ($M = .52$, $SEM = .05$), $t\ (20) = 4.01$, $p < .01$.

Differences between Experiment 3 and the earlier experiments hint that the phenomenological experience of KIA may depend to some degree on the materials/procedure used to test for the effect (i.e., Figure 5 shows some changes to the R-JK-G judgments), but there was no overwhelming evidence that word puzzles produce a substantially more vivid subjective experience of knew-it-all-along over the traditional trivia stimuli. However, modifications to the procedure of Experiment 3, such as eliminating the number scale and adding a delay between judgments, may lead to memory illusions by the participants of actually producing those KIA responses in foresight.

Experiment 4

The objective of this experiment was similar to that of Experiment 2; can a KIA effect be produced (with the Wordies stimuli) without using a number scale to collect

participants' responses on Test 1 and 2? Additionally, it was also expected that the changes to the procedure, such as the elimination of the number scale (along with the nature of the Wordies themselves) would make it more likely that the participants would falsely come to know/remember giving the correct information to KIA items in Test 1. For example, the number scale makes both the task of responding to the stimuli (objective measure) and the Remember/Just Know/Guess judgment (subjective measure) more difficult; the numbers themselves are abstract and therefore likely lead to interference (e.g., "Did I pick 3 or 4...now I'm not sure what I said on Test 1!").

The other changes to the procedure (which are outlined fully below) also should work toward producing a subjective component to the bias. For example, the 24 hr delay between Test 1 and Test 2, along with changing the feedback phase to include a recall task (i.e., of the correct solutions), should make it more difficult for participants to differentiate between how they responded to items on Test 1 and how they performed during the feedback phase. Therefore, participants may mistake their performance in the feedback phase with how they performed during Test 1, and this confusion may lead to the production of illusory feelings of knowing/remembering the feedback information from Test 1.

Method

Participants. Twenty three University of Victoria students participated in exchange for optional extra credit in an introductory psychology course. The data from three participants were excluded from the analyses because these participants failed to understand the instructions of the tasks and/or R-JK-G judgment.

Materials. The same Wordies stimuli from Experiment 3 were used and, as in the

previous experiment, no counterbalancing conditions were created because for each participant the computer program randomly selected which critical items were shown in the feedback phase.

Procedure. The basic procedure from Experiment 3 was used, but it included the two major modifications from Experiment 2; that is, the test format was changed to a 2AFC, and the experiment was conducted over two days.

The level of performance in Experiment 3 for critical items on the first test (+/- 73%) indicated that it would be difficult to produce a KIA effect in the absence of a number scale. That is, if participants are only wrong on about 25-30% of the items in Test 1, then the number of items that can be included in the calculation the KIA effect will be small (i.e., on average, only about 6 to 7 items would be included in the analyses of the KIA effect, regardless of whether participants do switch from incorrect on Test 1 to correct on Test 2). Therefore, it was necessary to reduce participants' performance on Test 1. During the test phases of Experiment 3, participants were given an unlimited period of time to choose an answer. In Test 1 of Experiment 4, participants were given 10 sec to respond after the alternatives for the puzzle appeared on the screen. Reducing response time should, at least for some of the trials, force participants to choose an alternative before they have been able to work out the correct solution in their head (and therefore increase the number of trials for which they choose the incorrect response). However, response time for Test 2 of this experiment was not limited to 10 sec because it was important that the participants have as much time as needed to try to recall their Test 1 performance. Additionally, reducing the amount of time that the word puzzles themselves are displayed on the screen (i.e., before the alternatives are shown) also could

help lower Test 1 performance. Therefore, the time that the Wordies remained on the screen in both Test 1 and Test 2 was reduced from 3 sec (in Experiment 3) to 2 sec.

It was hoped that placing the final test on a second day would not only help produce the KIA effect, but would also promote higher levels of false remembering/knowing for the KIA items. To further aide in creating these memory illusions, changes were made to the feedback phase of the experiment. First, participants completed the feedback phase on the first day of the experiment (unlike the 2-day paradigm in Experiment 2). Second, Part 2 of the Wordies judgment feedback phase (the common/rare judgment) was replaced with a forced recall task; after completing the understanding-of-the-solution judgment (Part 1), participants were shown all 40 of the puzzles again, and their task was to free-recall the solutions to each puzzle (Part 2: recall-of-the-solutions task). It was anticipated that requiring participants to free-recall the solutions for the critical feedback items on the first day of the experiment would lead them, on Day 2, to feel that they had actually chosen those responses on the first test (e.g., on the second day the free-recalled items from the previous session may feel obvious/easy to participants, and therefore they may have the illusion of remembering/knowing they had chosen those responses in Test 1).

Results and Discussion

Objective measures of the KIA effect. A typical KIA effect was observed; the proportion of items switching from the incorrect response on Test 1 to the correct response on Test 2 was higher in the feedback condition ($M = .62$, $SEM = .07$) than in the control condition ($M = .29$, $SEM = .04$), $F(1, 19) = 32.73$, $MSE = .03$, $\eta_p^2 = .63$, $p < .001$.

As in Experiment 2, there was also a significant effect for items switching from

the correct answer on Test 1 to the incorrect answer on Test 2, although again the proportion of switching was higher for the control items ($M = .08$, $SEM = .02$) than for the feedback items ($M = .04$, $SEM = .01$), $F(1, 19) = 4.70$, $MSE = .003$, $\eta_p^2 = .20$, $p = .04$. A comparison of Test 1 responses confirmed that there was no significant difference between feedback and control items in the number of items given a correct or incorrect response; participants were just as likely to pick the correct response on Test 1 for feedback items ($M = 18.20$, $SEM = .60$) as they were for the control items ($M = 17.45$, $SEM = .79$), $t(19) < 1$.

Subjective measures of the KIA effect. The overall proportions of items given an R, JK, or G rating for the response judgment are shown in Figure 7. The transformed proportions of R-JK-G designations for the trivia items that switched from an incorrect answer on Test 1 to a correct answer on Test 2 were analyzed in a 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. Five participants were excluded from the analyses for having either no feedback or no control items switch to the correct answer on Test 2. There was a reliable interaction between item type and judgment option, $F(1, 28) = 11.89$, $MSE = .37$, $\eta_p^2 = .46$, $p = .001$. Planned follow-up comparisons showed that for the items that switched from an incorrect response on Test 1 to a correct response on Test 2, participants were significantly more likely to claim that they remembered giving those correct answers on Test 1 for the feedback items ($M = .37$, $SEM = .06$) than for the control items ($M = .05$, $SEM = .04$), $t(14) = 4.95$, $p < .001$. There was no difference between the feedback and control items for the JK judgment option, $t < 1$, although there was a marginal trend for higher F in the feedback ($M = .52$, $SEM = .09$) than in the control condition ($M = .32$,

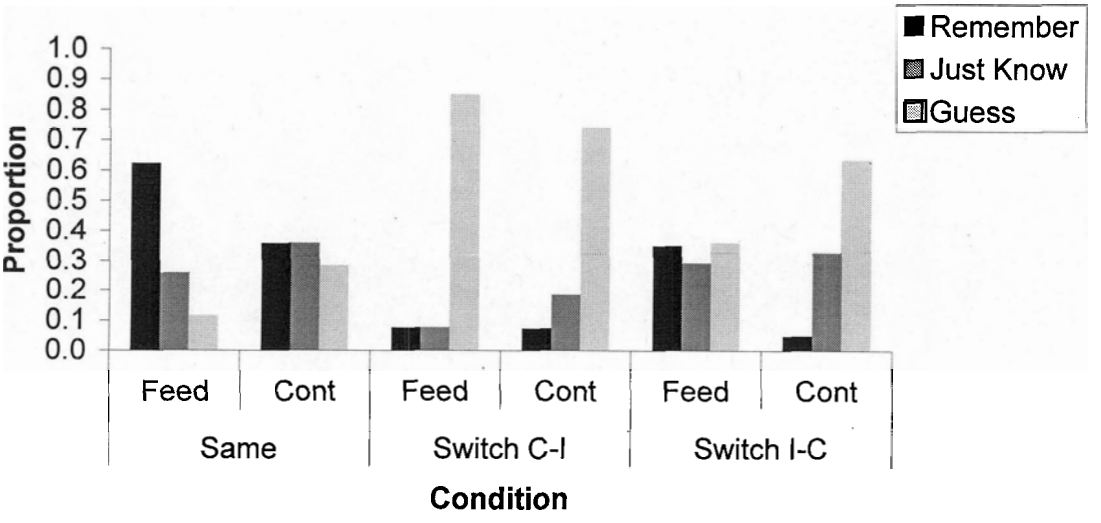


Figure 7. R-JK-G ratings for the response judgment in Experiment 4 for the feedback and control conditions (collapsed across participants). The judgments are separated by Test 1 and Test 2 responses; a) items given the same response on Test 1 and Test 2 (*same*), b) items that switch from the correct response on Test 1 to the incorrect response on Test 2 (*switch C-I*), and c) items that switch from the incorrect response on Test 1 to the correct response on Test 2 (*switch I-C*).

SEM = .07), t (14) = 1.99, p = .07. Finally, participants were less likely to use the G option for the feedback items that they had switched to the correct answer (*M* = .33, *SEM* = .07) than for the control items (*M* = .67, *SEM* = .08), t (14) = 3.85, $p < .01$.

The proportions of R-JK-G judgments that participants assigned to the items for which the same alternatives were chosen on both Test 1 and Test 2 were analyzed in a 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. There was an interaction between item type and judgment, F(2, 38) = 23.00, *MSE* = .02, η_p^2 = .55, $p < .001$. The follow-up comparisons showed that participants were much more likely to assign an R rating to the feedback items (*M* = .62, *SEM* = .05) than to the control items (*M* = .35, *SEM* = .04), t (19) = 5.94, $p < .001$. There was a trend for participants to give more JK judgments to the control (*M* = .36, *SEM* = .03) than feedback items (*M* = .26, *SEM* = .04), t (19) = 2.09, p = .05. However, the IRK data demonstrated an opposite marginal trend to the JK data, with participants showing lower levels of F for the control items (*M* = .58, *SEM* = .04) than the feedback items (*M* = .68, *SEM* = .04), t (19) = 1.97, p = .06. Finally, participants were less likely to use the G option for the feedback items (*M* = .12, *SEM* = .02) than the control items (*M* = .29, *SEM* = .04), t (19) = 6.70, $p < .001$.

The results for the objective measure of hindsight bias from the current experiment replicated Experiment 2; specifically, a KIA effect was produced without having to use a number scale as the objective measure of the bias. More importantly, the data for the subjective measure of the effect in Experiment 4 are the first in this series of studies to demonstrate clear evidence for a subjective component to hindsight bias. Further, the higher claims of remembering for the feedback KIA items seem to indicate

that it is not simply that participants experience the feedback information in Test 2 as the "more obvious" choice of what they would have given in Test 1 (i.e., otherwise you would expect significantly higher JK or F ratings). That is, the higher level of R responses for the feedback KIA items may indicate that participants are recollecting aspects of their performance in the feedback phase (e.g., remembering themselves generating the steps to the solutions during the understanding-of-the-solution task), and that they then mistakenly attribute these aspects to having occurred during Test 1. This hypothesis will be explored further in the General Discussion (e.g., source monitoring; Johnson, Hashtroudi, & Lindsay, 1993).

Experiment 5

One major methodological difference between the 2AFC trivia questions KIA experiment (Experiment 2) and the 2AFC Wordies puzzles experiment (Experiment 4) was the timing of the feedback manipulation. That is, feedback was given on Day 2 for the trivia questions paradigm, whereas it was presented on Day 1 for the Wordies version. Therefore, it is not clear if the difference found in subjective experience between the two experiments was due mainly to the difference in stimuli (as intended), or whether the timing of the feedback may have had an impact on the subjective measure. Therefore, Experiment 5 was constructed to replicate Experiment 4 but with the addition of a feedback timing manipulation.

Method

Participants. Forty University of Victoria students participated in exchange for optional extra credit in an introductory psychology course.

Materials. The same Wordies stimuli from Experiment 4 were used and, as in the

previous two experiments, no counterbalancing conditions were created because the computer program randomly selected which critical items were shown in the feedback phase.

Procedure. The same procedure from Experiment 4 was used, but it included a between-subjects manipulation of feedback presentation; that is, half of the participants completed the feedback phase on Day 1 and half of the participants completed the feedback phase on Day 2 of the experiment.

Results and Discussion

Objective measures of the KIA effect. As expected, a KIA effect was found; the proportion of items switching from the incorrect response on Test 1 to the correct response on Test 2 was higher in the feedback condition ($M = .48$, $SEM = .22$) than in the control condition ($M = .27$, $SEM = .14$), $F(1, 38) = 29.17$, $MSE = .03$, $\eta_p^2 = .43$, $p < .001$. Further, there was no effect of feedback presentation (Day 1 vs. Day 2) on the KIA effect, $F < 1$.

Unlike in Experiment 2 and 4, there was no significant main effect for the proportion of items switching from the correct answer on Test 1 to the incorrect answer on Test 2 between the control ($M = .10$, $SEM = .09$) and feedback items ($M = .11$, $SEM = .10$), $F < 1$. There was a marginal trend for the interaction, $F(1, 38) = 3.17$, $MSE = .01$, $\eta_p^2 = .08$, $p = .08$, and the pattern of switching was opposite for the two feedback conditions. That is, the pattern of results showed that switching to the incorrect answer on Test 2 was higher for feedback ($M = .12$, $SEM = .12$) than control items ($M = .09$, $SEM = .08$) when feedback was given on Day 2, but switching to the incorrect answer on Test 2 was higher for control items ($M = .12$, $SEM = .10$) than feedback items ($M = .09$, $SEM =$

.08) when feedback was given on Day 1. Although potentially interesting, neither of these patterns for switching to the incorrect answer on Test 2 was significant, $ts \leq 1.40$, $ps \geq .18$. Thus, the feedback-timing manipulation had little if any effect.

Subjective measures of the KIA effect. The overall proportions of items given an R, JK, or G response judgment are shown in Figure 8. The transformed proportions of R-JK-G designations for the Wordies items that switched from an incorrect answer on Test 1 to the correct answer on Test 2 were analyzed in a 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA, with feedback timing (day 1 vs. day 2) as the between-subjects factor. Four participants were excluded from the analyses for having either no feedback or no control items switch to the correct answer on Test 2. As anticipated, there was a reliable interaction between item type and judgment option, $F(2, 68) = 7.86$, $MSE = .42$, $\eta_p^2 = .19$, $p = .001$. There was no effect of feedback timing on this interaction, $F(2, 68) = 1.22$, $MSE = .42$, $\eta_p^2 = .04$, $p = .30$, and therefore this manipulation is dropped from the remaining analyses.

The planned follow-up comparisons for the interaction showed the same patterns as in Experiment 4. That is, for the items that switched from an incorrect response on Test 1 to a correct response on Test 2, participants were significantly more likely to claim that they remembered giving those correct answer on Test 1 for the feedback items ($M = .27$, $SEM = .05$) than for the control items ($M = .04$, $SEM = .02$), $t(35) = 3.50$, $p = .001$. Also, there was no difference between the feedback and control items for the JK judgment option, $t < 1$, but again there was a trend toward a higher F value in the feedback condition ($M = .38$, $SEM = .06$) than in the control condition ($M = .29$, $SEM = .06$), $t(35) = 1.80$, $p = .08$. Finally, participants were less likely to choose G for the

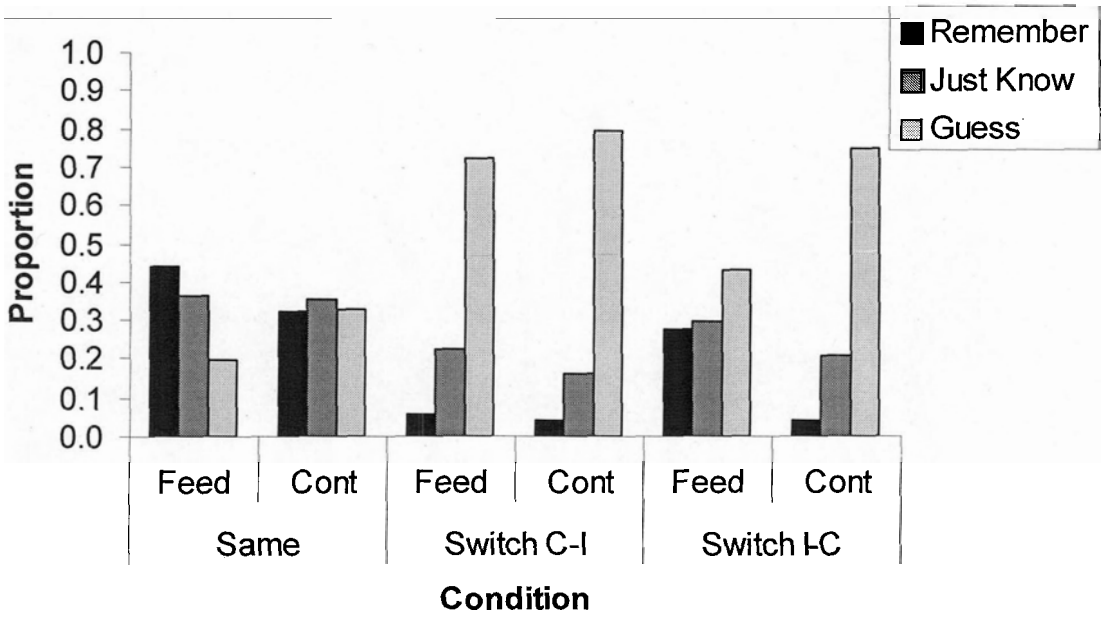


Figure 8. R-JK-G ratings for the response judgment in Experiment 5 for the feedback and control conditions (collapsed across participants and feedback condition). The judgments are separated by Test 1 and Test 2 responses; a) items given the same response on Test 1 and Test 2 (*same*), b) items that switch from the correct response on Test 1 to the incorrect response on Test 2 (*switch C-I*), and c) items that switch from the incorrect response on Test 1 to the correct response on Test 2 (*switch I-C*).

feedback items that they had switched to the correct answer ($M = .42$, $SEM = .05$) than for the control items ($M = .71$, $SEM = .06$), $t\,(35) = 3.84$, $p < .001$.

The proportions of R-JK-G judgments for the items that participants chose the same responses for on both Test 1 and Test 2 were analyzed in a 2 (Item type: feedback vs. control) x 3 (Judgment option: remember, just know, guess) within-subjects ANOVA. There was an interaction between item type and judgment, $F(2, 78) = 11.82$, $MSE = .03$, $\eta_p^2 = .23$, $p < .001$. As in the previous experiment, the planned comparisons showed that participants were more likely to assign an R judgment to the feedback ($M = .44$, $SEM = .03$) than the control items ($M = .32$, $SEM = .03$), $t\,(39) = 3.59$, $p = .001$. There was no difference between the feedback and control items for the JK option (both Ms = .36, SEMs = .02), $t < 1$, but the IRK estimate showed a higher level of F for the feedback items ($M = .65$, $SEM = .03$) than the control items ($M = .52$, $SEM = .02$), $t\,(39) = 3.73$, $p = .001$. Finally, more G judgments were given to control ($M = .33$, $SEM = .02$) than feedback items ($M = .20$, $SEM = .02$), $t\,(39) = 4.73$, $p < .001$.

The results of Experiment 5 replicated the previous experiment; a subjective component to the KIA effect was found, and participants were more likely to label feedback KIA items as "remembered" than control KIA items. More importantly, the data from the current experiment showed that this evidence for the subjective component of hindsight bias was not due to a feedback-timing manipulation. Specifically, placing the feedback phase on Day 1 (rather than on Day 2, as in the trivia KIA paradigms) was not the factor responsible for producing an accompanying subjective phenomenology to the KIA effect. Instead, it was the change in stimuli and the procedures in Experiment 4 and 5 that led to finding an effect in both the objective and subjective measures of the bias.

General Discussion

All five experiments demonstrated a KIA effect in the objective measures used to quantify the phenomenon. More interestingly, the first three experiments provided little evidence that participants had any strong accompanying subjective feeling of previous knowledge for the KIA items. Further, the results from Experiment 4 and 5 illustrated that individuals do sometimes report illusory memories of having known the feedback information in foresight.

It is important to emphasize that the goal of the present research has never been to discredit the KIA effect; even though the first three experiments did not demonstrate an accompanying subjective phenomenology for the KIA effect, a robust hindsight bias was found in all three experiments. That is, the lack of an accompanying subjective component does not make the KIA effect less "real." But, the presence or absence of a subjective component to the bias can provide us with valuable information regarding the nature of (and possible explanations for) the effect. For example, if all experimental paradigms that were used to test for subjective phenomenology never produced evidence that the hindsight bias could be accompanied by a feeling of remembering/knowing the feedback information in foresight, then one conclusion that could be drawn is that the effect is simply the result of a bias. Specifically, when an individual is unable to confidently reconstruct his/her previous performance on a task, having been exposed to the correct information between judgments systematically moves his/her hindsight judgment toward that correct information (but without an accompanying sense or feeling of having known that correct information during the original judgment).

The nature of most typical hindsight bias experiment – answering numerous

difficult general knowledge questions on a number scale – gives little reason to assume that individuals would have a feeling of knowing/remembering for possessing the feedback information in foresight. Therefore, it is not all that surprising that the traditional and modified-traditional paradigms used in Experiment 1 and 2 did not result in a belief on the part of the participants that they had known the feedback information during the first judgment task. Given that the trivia items used to test for the effect must be very difficult (i.e., to find the hindsight bias you need participants to have low confidence in their responses on Test 1, or to get a lot of the questions wrong), such stimuli tend to hold little meaning for participants. Additionally, providing feedback for the trivia stimuli allows participants to learn the correct answers to these items, but in a passive and pallid manner because the questions themselves are somewhat meaningless to them. That is, difficult general knowledge trivia items are somewhat similar to nonsense stimuli, in that, even when the correct answers are provided they tend to "feel" meaningless (e.g., participants often don't know why those specific responses are the correct answer, so they don't elicit any "ah-ha" type of reactions).

One of the principal benefits of the word puzzle stimuli used in Experiments 3 - 5 is that, not only do participants learn the answers to the puzzles in feedback, but they also learn techniques for figuring out *why* those are the correct solutions. The goal of the training phase that occurred prior to feedback in the Wordies paradigms was to provide participants with a variety of ways to think about the solutions to the word puzzles, as well as techniques to help them solve the puzzles (i.e., so that they could understand why the solutions given were the correct answers to the Wordies). In comparison to the traditional and modified-traditional paradigms, it could be argued that participants in the

Wordies KIA experiments were required to be much more engaged with both the feedback and with the stimuli (e.g., due to training, etc.). However, it is not simply the change from trivia questions to word puzzles that led to an accompanying subjective feeling of knew-it-all-along for the hindsight bias. For example, even though word puzzles were used in Experiment 3 there was no strong evidence that participants believed they had known the solutions to the feedback items in foresight, in comparison to the control items. It appears that the important ingredients for producing illusory feelings of knowing/remembering for hindsight bias found in both Experiment 4 and 5 was the combination of the 2AFC testing format (which allows for more memorable Test 1 experiences than a number scale) and the Wordies stimuli (which, once understood, seem inherently correct); this combination leads individuals to feel that they remember or know that they knew it all along.

The remainder of the General Discussion tackles the KIA effect in relation to three different topics: 1) the technique used to measure the subjective experience of the KIA effect (Remembering vs. Just Knowing vs. Guessing), 2) the two broad theoretical approaches to the effect (Memory Impairment vs. Biased Reconstruction), as well as a more comprehensive approach, and 3) a related phenomenon (the forgot-it-all-along effect).

Remembering Versus Just Knowing Versus Guessing

Interpreting the present R-JK-G data. As mentioned in the introduction, a Remember/Know judgment was adapted for the present KIA experiments to measure participants' subjective experience because we believed this technique would provide information beyond just asking for a confidence response. For example, participants are

instructed to choose G if they are unsure of their response, but to parse their confident responses into R or JK (and therefore the idea of confidence is built into the R and JK posts of the scale). Thus, a participant can go beyond simply stating that s/he is confident by providing qualitative information, such as the production of details regarding the past event, and this provides a further distinction in the type of subjective experience that accompanies the KIA effect. The focus of the experiments was on exploring the subjective phenomenology of hindsight bias (e.g., will participants almost always choose G for KIA items, which would speak to a lack of accompanying belief of possessing the knowledge in foresight, or will they at least sometimes choose R or K?). However, collecting qualitative information about the subjective experience is also useful for understanding both the objective and subjective components of the effect.

As we had expected, the results from the traditional and modified-traditional KIA paradigms (Experiment 1 and 2) did not provide evidence for the assertion that the effect is accompanied by a belief or feeling of having known the feedback information in foresight. As previously hypothesized, this pattern of results is likely due to at least two main factors; that is, that the use of a number scale (traditional paradigm) and the nature of the stimuli (traditional and modified-traditional paradigms) are not conducive to producing a subjective feeling of having known the answers prior to the presentation of feedback. The use of a number scale cannot be the main roadblock to producing an accompanying feeling of knowing for the feedback items (i.e., getting rid of the scale did not lead to subjective KIA experiences in the modified-traditional paradigm), but the results from Experiments 3 - 5 suggested that the number scale may interfere with producing a subjective experience of having known the information all along. Therefore,

even though it cannot firmly be concluded that the change from a number scale in Experiment 3 to 2AFC in Experiment 4 and 5 is what led to the difference in results for the subjective measure (i.e., because there were other important differences between the paradigms, such as the time to respond to the puzzles in Test 1), finding supporting evidence for a subjective component to the effect is likely more probable when a number scale is not used to collect the objective measure of the bias.

As mentioned earlier, one of the reasons we favoured the R-JK-G measure over a confidence scale is that the assumptions from previous researchers regarding the subjective component of the KIA effect have centred around the idea of the feeling of knowing (but without strong claims that participants explicitly remember details of giving the feedback information in foresight). In fact, our hypothesis was that, if there was an accompanying subjective feeling of knew-it-all-along it would be observed in the JK responses: Although participants may not be able to construct specific details of giving the feedback information during the first test, they may come to believe that, because they currently know the correct information they must have known it prior to the feedback (e.g., perhaps because the answers now seem obvious or the more logical choices of the given options). The data from the IRK model seemed to support this assumption because there was a trend in both Experiment 4 and 5 for higher F values for the KIA items in the feedback condition than in the control condition. However, the more convincing and interesting evidence for the existence of subjective phenomenology accompanying the KIA effect in Experiments 4 and 5 was found in the R responses, where the participants were reliably more likely to claim that they remembered some detail of giving the correct answers to the feedback than the control KIA items.

Although a significant difference between feedback and control items in the subjective measure was found only in the R responses of Experiment 4 and 5, the pattern of JK responding for items switching from the correct to incorrect responses and the incorrect to correct responses seems to be suggestive of a generic "knew it" feeling for the KIA items (see Figures 7 and 8). Specifically, the JK judgments for the KIA feedback items (i.e., switching from incorrect to correct) appear to be greater than the JK judgments for the feedback items that switched from the correct to incorrect response; this pattern is also evident for the control items. One interpretation of these patterns is that learning has occurred between tests for both the feedback and control items. The training phase that participants completed during the feedback phase may also have impacted the control items because participants could subsequently use the skills they learned to solve (i.e., come to understand) some of the control items during the final test. Further, if participants fail to recall their (incorrect) Test 1 responses to these items (and, in the case of feedback items, no diagnostic details from the feedback phase come to mind) then they may experience a feeling of knowing for the correct solutions based on their understanding of *why* they are the correct answers (e.g., "Well, I don't specifically remember giving this answer in Test 1, but I do know why it's the correct solution to the puzzle, so I'm sure I knew it was the correct answer in the first test").

If there is a generic "knew it" feeling for feedback and control KIA items in Experiment 4 and 5, then there should be a significant difference between the JK ratings for the items that switched from incorrect to correct and items that switched from correct to incorrect. Such post-hoc analyses could not be conducted on the data from Experiment 4 because there were too few observations (i.e., many participants had no switching from

the correct to incorrect response, and consequently there were not enough participants that could be included in the analyses). However, a comparison of the JK judgments in Experiment 5 showed no reliable difference in the JK judgment for the feedback items that switched from correct to incorrect and from incorrect to correct, as well as no significant differences for this same comparison of the control items.[9] Although no reliable differences were found, this notion of a generic "knew it" feeling (especially if demonstrated for both feedback and control items) is an interesting concept, and it would be useful for future research to focus specifically on investigating this concept.

Putting aside for the moment the issue of finding subjective phenomenology of the KIA effect under only certain testing conditions (i.e., trivia questions vs. word puzzles, etc.), it is informative to look at why support for the presence of subjective phenomenology was found mainly in the R rather than the JK judgment. There are likely many factors involved in producing this pattern of results for the subjective measure, but an overall key aspect is the intertwined relationship between the type of stimuli (Wordies), feedback structure (training/understanding judgment directly followed by forced recall of solutions) and, to a lesser extent, the testing procedure (2AFC).

In terms of the type of stimuli used to test for the KIA effect, overall, participants seemed to respond better to the Wordies than the trivia questions. For example, for the Wordies stimuli participants seemed to understand that if they did not immediately know

---

[9] Only the participants who had items that switched both from correct to incorrect and incorrect to correct could be included in the analyses; therefore 24 of the 40 participants from Experiment 5 were dropped from the analyses. Additionally, the analyses were conducted on the natural log transformed data. For the feedback items, there was no difference between JK judgments for items switching from correct to incorrect ($M = .25$, $SEM = .08$) and incorrect to correct ($M = .31$, $SEM = .08$), $t < 1$. The same pattern was found for control items: There was no significant difference between JK judgments for items switching from correct to incorrect ($M = .17$, $SEM = .08$) and incorrect to correct ($M = .28$, $SEM = .09$), $t < 1$. The same analyses were not conducted on the IRK estimates of F because a large number of the F values were .00 across the four conditions.

the correct solution to the puzzle that they could attempt to work it out by trying to solve the puzzle in their head using different techniques, or by working backwards from the two alternatives that they were shown for each puzzle. However, the trivia questions are not amenable to this type of interaction because the difficult questions that are needed to test for the effect are typically "all-or-nothing" in nature (i.e., participants either know the correct answer to the question or they don't, and when they didn't know the answer they typically are unable to use/find any techniques that could help them get to the correct answer). Additionally, the structure of the feedback phase likely is an important piece in producing the subjective experience effect in the R rather than JK judgment option. That is, in the trivia questions paradigms there was an attempt to mask the feedback by changing the wording of the questions and answers, but participants were never told (or given tools to help them understand) why an answer was correct. In the Wordies paradigm though, participants were explicitly given training on how to solve some of the different types of puzzles used as stimuli in the experiment in the hopes that this technique would ensure that participants learned both the answers to the feedback puzzles and *why* those answers were the correct solutions (i.e., by figuring out for themselves the steps for why a particular response was the correct solution to a puzzle).

Taken together, it could be argued that the nature of the feedback training/tasks in the Wordies paradigm (specifically Experiment 4 and 5) and the nature of the Wordies themselves led to problems in source monitoring for participants (e.g., Johnson et al., 1993). The concept of source monitoring used within the source monitoring framework (SMF) is commonly defined as the decisional processes that are involved in assessing the origin of various mental experiences (Johnson et al., 1993; Mitchell & Johnson, 2000).

The concept of source monitoring is closely related to Johnson and Raye's (1981) earlier work on reality monitoring, which was defined as the processes used by individuals to determine whether a memory is the product of an external source (perception) or an internal source (imagination). Importantly, researchers have emphasized that, although related to it, source monitoring is not simply a form of recognition memory: Variables that affect source monitoring do not necessarily affect recognition memory, and vice versa (Lindsay & Johnson, 1991). Therefore, it is not necessary to assume that participants' ability to recollect their original responses in a KIA experiment and their ability to accurately monitor the source of information used to determine this recollection (e.g., whether they believe they knew the information in Test 1 or it was supplied in the feedback phase) are tied together or influenced by the same factors. That is, the level and type of subjective phenomenology of the bias (e.g., recollection vs. knowing/familiarity) need not change systematically with any changes that may occur in the objective measure of the KIA effect.

The SMF does not define familiarity and recollection in terms of discrete categories, but rather the SMF differentiates the two concepts on the basis of the "specificity" of the activated information that contributes to each mental experience. For example, as the specificity of the memorial information that has been activated increases, the more likely it is that an individual will feel that s/he is "remembering" an event rather than "just knowing" that it happened (i.e., assuming that the individual attributes the information to some source in the past; Dodson & Johnson, 1996; Mitchell & Johnson, 2000). In the Wordies KIA paradigms, requiring participants to complete a training phase, followed immediately by the chance to use that training (i.e., the understanding-of-

the-solution task) provides an opportunity for specific details related to the correct solutions to be formed, and the recall-of-the-solution task (Experiments 4 and 5) further complicates the situation by ensuring that participants have said aloud the correct solutions (which corresponds to the response mode of Test 1).

Another important principle of the SMF is that source recollection is not an all-or-nothing process, in that it can have "intermediate qualities." For example, you may remember that a sibling knocked your tooth out when you were playing together as children, but you may not remember which sibling was responsible for the accident (e.g., "Did I get my tooth knocked out by my older brother or middle brother?," Mitchell & Johnson, 2000). Relatedly, it is highly doubtful that participants in the Wordies KIA experiment "forgot" that they had been shown the correct solutions for the feedback items, but rather they may fail to appreciate the impact of having completed a feedback phase has on their subsequent performance. For example, at Test 2 participants may remember the steps that they used to solve the puzzles during the feedback phase, but they may fail to tag those steps as having occurred *only* during the feedback phase (and not also during Test 1). Further, if participants fail to realize that their understanding of the correct solutions for the feedback KIA items originally occurred during the feedback phase, then it makes sense that they would choose the R judgment option because they would label the bringing to mind of details for solving these puzzles as occurring during Test 1.

It is important to note that a fundamental principle of the SMF view is that both veridical and illusory memories originate from the same processes: No special mechanisms need to be postulated to account for how false memories are created

(Henkel, Franklin, & Johnson, 2000; Johnson et al. 1993; Mitchell & Johnson, 2000). Johnson and colleagues have claimed that one way memory distortions occur is through factors that render accessible information non-diagnostic of source (see also Gruppuso, et al., 1997). For example, Lindsay, Johnson, and Kwon (1991) demonstrated that individuals were more likely to make source monitoring errors if the potential sources were similar than if they were dissimilar. That is, high similarity renders memories of perceptual details less diagnostic than they otherwise would be. Relatedly, the structure of Test 1 and the feedback phase in Experiment 4 and 5 could be argued as being "similar sources": In both phases the goal is to solve word puzzles using a variety of techniques and to say that solution aloud (and the correct solution appears in both parts). There are certain diagnostic features that participants can use to distinguish between Test 1 and the feedback phase (e.g., difference in time between the two phases, recollecting specific details of giving the incorrect response in Test 1), but the similarity of the tasks between the phases likely makes source monitoring more difficult. However, it is important to keep in mind that – because memory distortions are the result of the same processes as veridical memory – the production of illusory memory is constrained by such aspects as an individual's beliefs, knowledge, biases, etcetera. More specifically, a person will not judge an imagined event to be a recollected memory if it does not feel plausible when compared to his/her beliefs and self-knowledge (Johnson & Raye, 1981; Johnson & Sherman, 1990). Therefore, one prediction from the SMF would be that participants in a Wordies KIA paradigm would not claim to *remember* having given a correct solution to a feedback item (even if the objective measure demonstrates the bias) if they were unable to understand the solution to that puzzle during the feedback phase.

The principles of the SMF discussed above provide one promising way to begin thinking about the patterns of R-JK-G judgments that were observed in the five experiments presented in this paper. Although the SMF is not being proposed here as an encompassing explanation for the objective and subjective measures of hindsight bias, aspects of the framework will arise throughout the remainder of the General Discussion.

Theoretical approaches to R-JK-G data. As noted in the introduction, in general, the interpretation of R/JK data is dependent on an individual's theoretical concept of the underlying structure and nature of memory. The two broad types of R/JK theories can be defined as single- (quantitative) versus dual-component (qualitative), although not all models fit neatly into these two categories (Bodner & Lindsay, 2003; Gardiner et al., 2002; Rajaram et al., 2002; Yonelinas, 1994). Although testing the various theoretical models of R/JK data was not the goal of this paper (as discussed further at the end of this section), it is informative to touch on some of the approaches to the use of the R/JK procedure.

In brief, single-component models of R/JK judgments (often referred to as trace-strength models) largely concentrate on decisional criteria that lead to the distinction between R and JK responses, which allows for the application of SDT. Donaldson (1996) argued that the simplest way to explain R/JK data was to posit a two-fold criteria account: The first criterion distinguishes between old and new responses, and the second criterion is established to parse the old responses into R (upper end of the distribution) and JK categories. However, this type of one-dimensional SDT model has had difficulty accounting for all of the different patterns of data found using R/JK judgments (e.g., Donaldson, 1996; Gardiner et al., 2002; Yonelinas, 1994).

Rotello et al. (2004) proposed a two-dimensional SDT account – the sum-difference theory of remembering and knowing (STREAK) – and they argued that performance on both a recognition task ("old/new" decision) and a Remember/Know task could be accounted for through the combination of global and specific strength dimensions. Specifically, they claimed that every item has both an overall memory strength (global component) and associated "specific details" (specific strength); these specific details either support or challenge the global memory strength. Further, the STREAK account assumes that recollection is graded, in that, "distributions of studied and unstudied items differ in their specific strength *as well as* [italics added] global strength" (p. 592). In the STREAK model, old/new recognition judgments are based on the summed values of global and specific memory strengths, whereas the R/JK judgments are based on the difference between these two types of memory strengths (e.g., a higher level of specific strength is needed for an R response when there is a high level of global strength).

In their comparison of the STREAK model with an equal-variance and unequal-variance one-dimensional SDT model, Rotello et al. (2004) found that, in most of their test situations, STREAK fit the data as well as or better than the other two models. The researchers noted, though, that one drawback to the STREAK model is that the current (simple) version of STREAK does have difficulties explaining the combined recognition and Remember/Know data patterns when the two measures are highly correlated. However, they argued that this issue of correlation between recognition judgments (i.e., confidence ratings) and Remember/Know is not necessarily "devastating" to STREAK because such correlations are not "pervasive." For example, they stated that how the two

measures are collected (within- vs. between-subjects) impacts the rate of correlation, in that, a correlation typically is found only when the judgments are collected from the same participants. Although Rotello et al. contended that this correlation normally arises only when the measures are collected within-subjects, Wixted and Stretch (2004) have found otherwise. That is, Wixted and Stretch re-analyzed published data from experiments that had collected these two judgments between-subjects (e.g., Gardiner & Java, 1990; Rajaram, 1993; Yonelinas, 2001) and they found that, in most of the experimental conditions, the two measures were correlated. Therefore, although it seems to be a promising single-component model for R/JK judgments, it is important that the STREAK model address the issue of potential correlations between recognition and Remember/Know judgments so that such correlations do not reduce the fit of the model to the data.

The use of SDT principles is not exclusive to single-component models of Remember/Know data. For example, Yonelinas (1994; 2001) developed a dual-process SDT model for recollection and familiarity that he contended could account for the pattern of data from both recognition and Remember/Know judgments. Specifically, he argued that SDT works well for capturing the use of familiarity, but that an additional factor, such as recollection, is needed to account for the varying data patterns found using R/JK procedures. Within this dual-process SDT model, recollection is viewed as a threshold process, and it "is assumed to lead to high-confidence responses, whereas familiarity assessment is assumed to support a wider range of memory confidence responses" (Yonelinas, 2001, p. 362). In a series of experiments measuring confidence and Remember/Know judgments under full- and divided-attention, Yonelinas (2001)

found that the predictions made by the dual-process SDT model fit very closely with the observed Remember/Know data. Further, he argued that the asymmetrical ROC curves that typically are found for recognition memory (e.g., *z*ROCs with a slope around 0.80) are due to "remembering" and not "knowing;" that is, that an R response reflects qualitatively different information than a JK response, and that it is this distinct information that contributes to the asymmetry (e.g., remembering details leads to more accurate responding, which is not simply interchangeable with an increase in confidence). Finally, Yonelinas showed that his ROC curves became symmetrical when he removed the recollection component (i.e., "remembering"), but that the removal of items given a high confidence ratings was not enough to produce similar symmetrical patterns.

Interestingly, Wixted and Stretch (2004) re-analyzed Yonelinas' (2001) data and found that the *z*ROC curves for the recognition judgments (confidence measure) and Remember/Know judgments were highly correlated, which would seem to go against the dual-process SDT model's assertion that R responses reflect distinct information from high confidence. However, Yonelinas (2001) never claimed that recollection was not related to confidence, in that, the recollection threshold process does typically lead to high confidence responding (as mentioned in the previous paragraph). Rather, his claim was that the (separate) recollection process results in a "remember" experience that involves more than just high confidence (e.g., qualitative information, such as specific details of the prior event), which is one avenue for distinguishing R from K judgments (e.g., K judgments can be made with the same level of confidence as an R judgment, but without any accompanying contribution from the recollection threshold process). It is not clear, though, at what level a correlation between the two judgments goes from indicating

a relationship between the measures (i.e., a link, but still a difference between what the two measures tap into) to indicating the two measures capture the same information (e.g., Rotello et al., 2004; Wixted & Stretch, 2004).

The underlying assumption for all dual-component models of R/JK judgments is that performance on a memory task is composed of two separate processes. Perhaps one of the most widely known examples of a dual-component model of memory is Jacoby's (1991; Jacoby, Yonelinas, & Jennings, 1997) Process Dissociation Procedure (PDP). Briefly, the equations for estimating recollection and familiarity in the PDP rest upon the assumption that conscious (recollection) and unconscious (familiarity) processing are independent of one another; that is, conscious and unconscious processing can occur either in isolation or together (Jacoby, 1991). At the core of Jacoby's PDP is the idea that to produce an accurate measure of conscious and unconscious memory – that is, to separate out the two processes – performance on a task in which both processes contribute (facilitation) must be compared to a task in which the two processes work in opposition (interference).

Jacoby and colleagues' IRK model (e.g., Jacoby, 1991; Jacoby et al., 1998) that was used to analyze some of the results in this paper is part of the PDP approach to memory processes. The IRK model was proposed in response to Gardiner's (1988; Gardiner & Java, 1990) dual-process model, which regards R as a measure of explicit memory (recollection) and JK as an implicit measure of memory (familiarity). Although the IRK model alleviates the problem of the raw JK judgments not being appropriate as a pure estimate of familiarity (i.e., as discussed earlier, because recollection "wins out" if an item is both recognized and feels familiar), this model also has difficulty capturing all

of the R/JK data. For example, Bodner and Lindsay (2003) found that manipulating the levels of processing (LOP) used to study two sets of words impacted the level of R and JK judgments given at test: Items studied with a medium LOP task received higher R judgments when they were tested with items that had been studied with a shallow LOP task than when the medium-studied items were mixed with items that had been studied with a deep LOP task. As noted by Bodner and Lindsay, it is difficult for any dual-component model that posits two separate and stable memory systems to account for these results (including the IRK model).

Gruppuso et al. (1997) proposed a functionalist model of recognition memory that merged features of both single- and dual-component approaches. For example, although this functional approach emphasizes a qualitative distinction between recollection and familiarity, no claims are made that separate processes lead to experiencing either recollection or familiarity. Further, Gruppuso et al. characterized recollection as the ability to bring to mind "information that enables one to accomplish the task set by the situation...and familiarity as retrieval of memory information that gives rise to a feeling of recognition but does not enable one to accomplish the task set by the situation" (p. 264). The important feature of this functionalist model of recognition is that, because the type of recognition experience is context specific, the kind of information that contributes to an R response in one situation could lead to a JK response in a different setting (although the researchers did note that certain types of information may typically contribute to one type of response over the other, such as the recall of conceptual information/processing more often leading to R than JK experiences). In summary, in terms of the R/JK paradigm, the authors argued that the R and JK options are not factor

pure (i.e., they are not separate processes), and the type of information that contributes either to an R or JK response depends on the context of the situation.

Although the goal of the KIA research presented in this paper was not to test the various theories that have been proposed for R-JK-G data, it could be argued that the functionalist approach is better suited than either single- or dual-component models for interpreting the KIA subjective experience data. Specifically, of the three types of models described in the preceding paragraphs, the functionalist approach is the best fit with the source monitoring interpretation of the subjective phenomenology evidence found in Experiment 4 and 5. In fact, Gruppuso et al. (1997) noted that a large part of their functionalist approach was influenced by the SMF (e.g., Johnson et al., 1993). The greater flexibility inherent in the functionalist approach (i.e., compared to dual-process models that posit separate processes for R and K) would better capture the data patterns, such as explaining why, under certain testing conditions, the levels of R and JK between conditions do or do not differ from one another.

Additional research that is focused specifically on employing the R-JK-G method in a KIA experiment is warranted before strong conclusions regarding single/dual component and the functionalist approach can be made. Because the purpose of the present research was not to test the various theories for the underlying structures of R and JK responses, the paradigms used in all five of the experiments are not highly conducive to making claims regarding which of the three models discussed above would be the best fit for the data. For example, because none of the items on the final test in any of the KIA paradigms were new, it may be more difficult to interpret what a JK response means. For instance, if participants choose the JK judgment option in a KIA paradigm, it may be

because the response they chose for the test item felt familiar but they couldn't recall details of choosing that response in Test 1 (which is the classical definition of what a JK judgment is meant to imply about the type of recognition specific to that trial), or because they "just know" that the response they chose for that trial is the correct answer and they are sure that, even though they don't remember details of choosing that answer in Test 1, they would have known that was the correct answer in the first test (but without a feeling of familiarity for the response). This issue surrounding the interpretation of JK judgments within a KIA paradigm is not a concern for the goals of the research presented in this paper (i.e., because the current interest is simply whether a difference in subjective phenomenology exists between feedback and control KIA trials), but it is problematic for specifically testing models of R/JK data.

Theoretical Explanations of the Effect

Theories pertaining to the KIA effect typically fall under two general categories; memory impairment and biased reconstruction (e.g., S. Schwarz & Stahlberg, 2003). The remainder of this section provides an overview of the chief memory impairment and biased reconstruction approaches to hindsight bias. Further, the section concludes with an attempt to provide a beginning account of hindsight bias that is rooted in a comprehensive theoretical memory and cognition framework (i.e., one that has not specifically been proposed for the KIA effect).

Memory Impairment. As mentioned in the introduction, Fischhoff (1977) was the first researcher to refer to the bias as the knew-it-all-along effect, and his initial work on this topic often is credited for spawning a great deal of interest and follow-up research on testing theoretical explanations of the effect (Christensen-Szalanski, & Willham, 1991;

Dehn & Erdfelder, 1998). Fischhoff's automatic assimilation hypothesis is also the most well-known memory impairment theory of the KIA effect, and he applied this theory to both hypothetical and memory designs. That is, Fischhoff (1975, 1977; Fischhoff & Beyth, 1975) hypothesized that the KIA effect results from an automatic assimilation of the correct feedback with pre-existing knowledge, and that this integration does not necessarily mean that old information is "overwritten" by the feedback, because pre-existing knowledge may be "reinterpreted" to make sense of, and account for, the new information. Additionally, this overwriting or reinterpretation occurs regardless of whether or not initial judgments (i.e., pre-feedback) were made and therefore the same underlying processes are at work for both hypothetical and memory KIA paradigms.

Although Fischhoff (1975, 1977) never directly addressed the issue of an accompanying subjective experience component to the KIA effect in his automatic assimilation hypothesis, the language he used to describe the bias appeared to imply that there is an associated feeling of having known the feedback information in foresight. For example, he claimed that individuals do not understand the impact that feedback information can have on their judgments of previous knowledge, and "thus they *believe* [italics added] that they knew all along that the reported event was going to happen" (Fischhoff, 1977, p. 349). Consequently, it appears that the automatic assimilation approach treats objective performance (movement toward the feedback information in hindsight) and subjective performance (type of recognition for the foresight judgment; recollection vs. familiarity vs. guessing) as the same entity; that is, when hindsight bias is found it is accompanied by subjective phenomenology, and therefore both types of performance are based on the same type/application of information. Because it appears

that the assimilation assumption implies that individuals not only are prone to producing a hindsight bias after exposure to feedback, but also that they come to believe that they had possessed this information pre-feedback, it is important to look at the research that has been directed at testing the automatic assimilation approach (although, again, none of the work has focused specifically on including subjective experience as a separate variable).

In their attempts to investigate the automatic assimilation hypothesis, Hasher et al. (1981) both reduced and eliminated the hindsight bias through experimental instructions in a memory-design paradigm. Participants in their *disconfirmed: mistake* memory condition received feedback in the form of a study phase; half of the items were stated as true and half as false. After the feedback was no longer present, participants were told that the experimenter had been mistaken, and in fact the true statements were false and the false statements were true. Hasher et al. hypothesized that this manipulation would force participants to adopt a different strategy for performing the hindsight judgments, possibly directing participants to focus on memory for their previous judgments. Results indicated that, in comparison to the *standard* memory condition (no instruction of experimenter mistake), participants in the disconfirmed: mistake condition showed a much lower KIA effect. Nevertheless, even for the disconfirmed: mistake group, items originally studied as false and then called true were given a higher value in hindsight than foresight, and items originally studied as true and then called false were given a lower value than on foresight judgments. A third condition, the *disconfirmed: wrong* group, was told after the study phase (and again, after feedback was no longer present) that the items they had just studied were wrong, and therefore should be ignored. Interestingly, this set

of instructions was effective in eliminating the hindsight bias: Ratings for items whose (false) answers had been provided during the study phase did not differ from ratings for non-feedback items.

Hasher et al. (1981) contended that their results cannot be easily reconciled with an automatic assimilation approach, and they argued that the reduction/elimination of the hindsight bias demonstrated that individuals are able to remember the level of knowledge they possessed prior to feedback. In terms of eliminating the effect, a potential criticism of their conclusion is that their results do not rule out the possibility that, rather than remembering their prior judgments, participants relied on some other strategy to rate the items. More specifically, it could be that the feedback was integrated with prior knowledge, and that this assimilation would have resulted in higher probability ratings for feedback items in hindsight (standard hindsight bias). However, because participants were later told that the feedback information was false, this new information was also assimilated with prior knowledge and the probabilities for the false-feedback items were then lowered. Granted, this two-stage integration of information may appear too sophisticated to occur automatically, in that it assumes the processes responsible for the integration of new information are able to determine which items previously received feedback.

A reduction in the hindsight bias has been used by other researchers to argue against the automatic assimilation of feedback hypothesis. For example, Hell et al. (1988) found that requiring participants to give reasons for their judgments during the foresight phase of a memory design (*reason-requested* condition) reduced the level of hindsight bias, in comparison to participants who gave no reasons for their responses (*reason-not-*

requested condition). Hell et al. maintained that the reason-requested condition led to a stronger memory trace for the Phase 1 items (i.e., foresight judgments) than the reason-not-requested condition. Consequently, these stronger traces allowed participants sometimes to gain access to their original judgments, thereby resulting in a smaller KIA effect. One potentially troublesome issue, though, involves initial hindsight judgments for the two groups. More specifically, requiring participants to give reasons for their responses may change how they answer the questions, and it may influence the answer they give: Stating reasons for an answer may lead you to think longer, harder, and more elaborately about a response, which could better help you come up with the correct solution. If participants in the reasons-requested condition were more accurate in foresight than participants in the reasons-not-requested condition, then there is less room for a hindsight bias to be discovered in the reasons-requested group. Unfortunately, Hell et al. do not provide results for the foresight judgments of the reason-requested and reason-not-requested groups.

In a study related to that by Hell et al. (1988), Davies (1987) manipulated encoding by requiring one group of participants to write notes during the foresight phase. Participants were given descriptions and possible outcomes for four different experiments and asked either to read the descriptions or to read the descriptions and also take notes on their thoughts/ideas/feelings relating to the experiments. Davies found that taking notes reduced the hindsight bias, but only when participants were allowed to review their notes prior to making the hindsight judgments (Experiments 1 and 2).

Hell et al. (1988) did not allow participants to review the reasons they had written for their answers in the foresight phase yet, unlike Davies (1987), they still found an

effect of encoding manipulations. Therefore, at first glance it is unclear which set of results accurately reflects the effect of encoding manipulations, but there are some methodological differences between the two studies that should be considered. For instance, there was a one-week delay between foresight and hindsight judgments in Hell et al.'s (1988) experiment, whereas there was a two-week delay in the experiments conducted by Davies (1987). It is possible that the delay in Davies' study was too long to show the impact of the encoding manipulation of writing versus not writing notes (in the absence of reviewing the notes) on reducing hindsight bias. Also, the items participants had to judge differed between the two experiments. Hell et al. (1988) used trivia-like questions, and it could be argued that Davies (1987) used event-like items (i.e., scenarios of experiments/potential results). Could it be that the encoding manipulations influenced how participants originally answered the trivia items (as mentioned in the previous paragraph), but not how participants originally judged the event items? A comparison of foresight judgments from Hell et al. (1988) and Davies (1987) could possibly answer this question, but unfortunately the necessary numbers were not provided.

Both Hell et al. (1988) and Davies (1987) stated that although their experiments do not invalidate the automatic assimilation theory, the results do indicate that the KIA effect can be moderated, at least to some degree, by conscious processes. Problems with Hell et al.'s (1988) study have already been discussed, and until it becomes more clear whether their effects are due to the differences in encoding on hindsight judgments and not some other factor (e.g., differences in foresight judgments) the impact of their results on the automatic assimilation hypothesis is questionable. As for Davies (1987), the fact that he found a differential hindsight bias for participants who were allowed to view their

notes, compared to participants who were not allowed to reread their notes (Experiment 1), is not really surprising and can be accommodated by the automatic assimilation theory. First, it is important to note that participants allowed to review their notes did not do so until after they had been given feedback. According to the assimilation hypothesis, these participants integrated the feedback information with their prior knowledge, but this integration was followed by an assimilation of the information contained in the notes. Because participants were allowed to review their notes after feedback, they could have updated their knowledge with the information of what they had thought/believed/etc. at the time they had made foresight judgments. Therefore, one would expect a much larger KIA effect for those participants who did not have the benefit of rereading the notes that they made during their foresight judgments.

Although the idea of automatic assimilation of feedback information has been difficult to invalidate based on the majority of past findings from KIA paradigms, a major problem for the hypothesis has been that it does not readily incorporate the differential rates of hindsight bias found for hypothetical and memory conditions. As mentioned at the beginning of this section, Wood (1977) found that making pre-feedback judgments did alter the hindsight bias for the hypothetical condition: Participants in the memory condition who were told to rate the items as they had done so prior to feedback exhibited a smaller hindsight bias than the participants in the hypothetical condition. If new information is immediately assimilated with existing knowledge, then why should it make a difference whether you have made foresight judgments prior to making hindsight judgments? A lower rate of hindsight bias for the memory group may indicate that participants are able to access/reconstruct their memory for foresight judgments (original

knowledge), but that either they tend not to (e.g., alternate strategy), or have a hard time doing so.

It could be argued that the automatic assimilation hypothesis has been difficult to invalidate because, in the way that most researchers have chosen to define it, the automatic assimilation hypothesis is unfalsifiable: Even if strong evidence for mitigation of the effect by controlled processes is found, one could still argue that, due to the manipulations employed to reduce the effect, participants switch to some other strategy to recreate prior knowledge rather than "remembering" it. Specifically, the problem stems from memory for prior knowledge being treated as a "thing" that either gets overwritten or can be remembered, and often the results that show a mitigation of the bias have spurred researchers to argue that at least some of the "original knowledge state" has been left intact, allowing participants to "gain access" to it, even though "subjects must exert unusual effort to retrieve pre-updated information" (Hasher et al., 1981, p. 95). The biased reconstruction approach to the KIA effect that is described in the next sub-section was developed in part to address this issue.

A related, but perhaps larger problem for the automatic assimilation approach is that it is not clear how it would (parsimoniously) account for differences in subjective phenomenology for the hindsight bias. Specifically, an objective measure of hindsight bias was found in all five experiments reported in this paper, but strong evidence for an accompanying belief that the feedback information was known in foresight was demonstrated only in the final two experiments (i.e., 2AFC with word puzzles). It cannot be argued that, in general, participants' ability to "recognize" the responses they chose on Test 1 was compromised in the first three experiments (compared to Experiments 4 and

5) because Figures 2-8 seem to indicate comparable levels of R/JK across experiments on performance for items given the same response on Test 1 and 2. Highlighting this similar performance is not meant to imply that it is therefore odd or unexpected to find a difference in subjective experience between experiments for the KIA items, but simply that the items used in the first three experiments (and in particular the trivia stimuli) can lead to R and JK judgments (at least for the response R-JK-G judgment).

Biased Reconstruction. Unlike Fischhoff's (1975, 1977) automatic assimilation hypothesis of the KIA effect, the more recent biased reconstruction approaches (e.g., Sanna, N. Schwarz, & Small, 2002; Sanna, N. Schwarz, & Stocker, 2002; S. Schwarz & Stahlberg, 2003) emphasize the contribution of subjective experience characteristics to the creation of hindsight bias (i.e., to the level of the objective measure), although admittedly this focus has not included the issue of *accompanying* subjective experience for the effect.

Sanna, N. Schwarz, and Stocker (2002) argued that two different kinds of information become available when an individual tries to recall details from memory; *accessible content* (the details that come to mind) and *accessibility experiences* (how easy/difficult it was to bring those details to mind). Further, they stated that accessibility experiences are an important type of information because they "qualify" the inferences that an individual draws from the accessible content. To test this assertion, Sanna, N. Schwarz, and Stocker manipulated the number of participants' thoughts pertaining to different outcomes of KIA items because "the most frequent recommended remedy for debiasing the hindsight effects is to search for reasons why the event might have turned out otherwise..." (p. 497). The researchers hypothesized that this strategy would not work

under all conditions because accessibility experiences influence the conclusions made from bringing these alternatives to mind. They created a hypothetical KIA paradigm (using event stimuli) in which participants either had to generate many thoughts regarding how the event could have turned out differently from the feedback they received (*10-thought* condition) or few thoughts regarding alternatives (*2-thought* condition). Participants subsequently were required to judge (on a scale from 0 - 100%) the probability that the event could have had a different outcome. Results from the 10-thought versus 2-thought conditions demonstrated what Sanna, N. Schwarz, and Stocker (2002) referred to as a "backfire effect;" that is, participants in the 10-thought condition judged the probability of a different outcome occurring for the event as less likely than participants in the 2-thought condition.[10]

Sanna, N. Schwarz, and Small (2002) found a similar backfire effect when they manipulated the number of thoughts across the type of thoughts (i.e., thinking about the known outcome vs. a different outcome). In their hypothetical KIA design, participants were divided into easy (2-thoughts condition) or difficult (10-thoughts condition) thought conditions, and they were required to produce these thoughts for either the known outcome of the event or for an alternative outcome of the event. The researchers claimed that, if accessibility experiences qualify the accessible content, then the difficulty of

[10] Sanna, N. Schwarz, and Stocker (2002) are not the only researchers to use this type of manipulation, as several studies have explored the effects of extended efforts to retrieve memories of childhood on adults' assessments of the completeness of their autobiographical memories. Research by Belli, Winkielman, Read, Schwarz, and Lynn (1998) demonstrated effects of experimental manipulations on individuals' assessments of the completeness of their memories of childhood in general: Individuals asked to recollect numerous childhood experiences during a brief experiment tended to rate their memories of childhood as being less complete than individuals asked to recollect only a few childhood experiences. Presumably the former participants were surprised at the difficulty of recollecting the required number of childhood experiences, which led them to revise downward their estimates of the completeness of their memories of childhood.

producing 10 thoughts (as compared to the 2-thought condition) for an alternative outcome should increase the hindsight bias (e.g., "It couldn't have happened any other way!"), whereas the difficulty of producing 10 thoughts for the known outcome should reduce the KIA effect (e.g., "It's hard to think of why it actually did turn out this way!"). The results supported the researchers' hypotheses: The participants in the 10-thoughts condition rated the known outcome as more likely than participants in the 2-thoughts condition when the thoughts pertained to a different outcome, whereas participants in the 10-thoughts condition who listed reasons for the known outcome subsequently rated that outcome as less likely than participants in the 2-thought condition. Additionally, the researchers found that this pattern of results was not specific to listing thoughts per se, as they found the same effects when they manipulated difficulty through contraction of the corrugator (brow) muscle (i.e., tensing of the brow is thought to produce the subjective experience of effort). That is, participants who contracted their corrugator muscle while thinking of the known outcome subsequently rated the known outcome as less likely than participants who did not contract this muscle, whereas the participants who tensed their corrugator muscle while thinking of the alternative outcome later rated the known outcome as more likely than participants who had not contracted their brow muscle.

Taken together, the results presented thus far in the biased reconstruction subsection indicate that, at least in terms of moderating the size of the objective effect, subjective experience plays an important role in hindsight bias. However, the results all come from hypothetical KIA paradigms and it is important to consider whether similar patterns can be demonstrated for memory designs. Sanna and N. Schwarz (2003) included a memory paradigm in their investigation of debiasing the KIA effect by using

the manipulation of alternatives in the presence/absence of discrediting information regarding the production of these alternatives. Prior to the 2000 US Presidential elections, participants were asked to predict the percentage of people who would vote for Bush, Gore, or another candidate (the total of all three had to add to 100%). Following the determination of Bush's win, participants were split into three conditions and asked to re-rate the options (i.e., Bush, Gore, or another candidate) as they would have prior to learning the election outcome. Participants in the *prediction-hindsight* group simply re-rated the options, whereas the participants in the *12-thoughts* condition were instructed to list 12 thoughts regarding how the election could have turned out in a different way prior to their re-rating of the options. Similar to participants in the 12-thoughts group, the participants in the *12-thoughts/attribution* condition were instructed to list 12 details relating to how the election could have turned out differently prior to their re-rating of the options, but in an attempt to discredit the difficulty of producing these 12 thoughts the researchers included the following information/task:

> Thank you for listing your thoughts. We realize this was an extremely difficult task that only people with good knowledge of politics may be able to complete. As background information, may we therefore ask you how knowledgeable you are about politics? (p. 291).

Sanna and N. Schwarz (2003) hypothesized that they would find the typical hindsight bias (prediction-hindsight condition), but that this bias would be larger for participants who listed 12 alternatives (12-thoughts condition). Further, they predicted that the hindsight bias would be smaller for participants who had the difficulty of producing 12 thoughts discredited (12-thoughts/attribution condition) because this

manipulation in essence creates an "easy thoughts" condition (e.g., Sanna, N. Schwarz, & Stock, 2002). As expected the results demonstrated a typical hindsight bias, and although listing 12 thoughts regarding how the event may have turned out differently did not increase the magnitude of the effect (as was predicted), having the difficulty of generating the 12 thoughts discredited did lead to an elimination of the KIA effect.

Similar to Sanna, N. Schwarz, and colleagues' approach to the KIA effect (Sanna & N. Schwarz, 2003; Sanna, N, Schwarz, & Small, 2002; Sanna, N. Schwarz, & Stocker, 2002), S. Schwarz and Stahlberg (2003) proposed a biased reconstruction view of hindsight bias. More specifically, they argued against the idea of automatic assimilation of feedback information by claiming that participants in a KIA paradigm use the feedback that they are given as a basis for recreating their pre-outcome response to an item. S. Schwarz and Stahlberg also tackled the issue of memory versus hypothetical designs by claiming that their approach could account for data from both types of paradigms. That is, they proposed that the magnitude of the bias is greater in hypothetical paradigms because every item given feedback in such a design is "eligible" to produce the KIA effect. However, because pre-feedback judgments are made in a memory design, not all items given feedback will be vulnerable to the effect; participants in a memory design can either remember their original responses or reconstruct them, and therefore only the items for which participants are unable to recall their original judgments are eligible to show the hindsight bias. Further, S. Schwarz and Stahlberg contended that the feedback information given to participants does not interfere with correct recall within a memory design. Specifically, they expect hit rates to be equivalent between the feedback and control conditions of a memory paradigm because hindsight bias is only produced on

trials for which the original judgment cannot be remembered – the KIA effect is produced because forgetting the original judgments leaves items that received feedback open to being biased by the correct information during reconstruction of the original response.

To test their assertions regarding the two different types of designs, S. Schwarz and Stahlberg (2003) created a within-subjects KIA paradigm that incorporated both a memory and hypothetical design. In the first part of the experiment participants answered a set of questions; feedback was given for two-thirds of the items and participants were instructed to provide their pre-feedback judgments (memory design). Following this re-rating task, the participants were given another set of questions in which two-thirds of the items were accompanied by feedback, and their task was to answer the questions as they would have without benefit of the feedback (hypothetical design). As expected, the results showed a larger KIA effect for the hypothetical than the memory design.[11] Further, S. Schwarz and Stahlberg also found no difference in the number of hits between the feedback and control conditions of the memory design.

One of S. Schwarz and Stahlberg's (2003) main criticisms of the automatic assimilation approach to hindsight bias is that it does not factor in the use of meta-cognitive processes. In their biased reconstruction view of the KIA effect, "hindsight distortions will only occur when people forget their original prediction and – while

[11] Interestingly, although the size of the hindsight bias was larger in the hypothetical design, the effect size was larger in the memory design. However, it is difficult to interpret this result because S. Schwarz and Stahlberg (2003) did not analyze their hypothetical design data in a typical manner. Specifically, the items that were used as hypothetical and memory stimuli were counterbalanced across participants, and the researchers analyzed the hypothetical design data by yoking the hypothetical data of one participant with the (pre-feedback) memory data of a different participant (e.g., if participant 1 received questions 1-18 in the hypothetical portion of the experiment, then s/he would be yoked with the pre-feedback responses of a participant who answered questions 1-18 in the memory portion of the experiment). As the researchers themselves noted, this manner of analyzing the data makes it difficult to interpret the differences in effect sizes because the effect size for the memory design is based on within-subjects data, whereas for the hypothetical design it is based on between-subjects data.

reconstructing it – have reason to believe that their initial estimate must have been close to what is now the known outcome" (p. 398). To investigate the use of meta-cognitive processes, S. Schwarz and Stahlberg manipulated participants' beliefs in five different ways regarding how close they had been in the foresight judgment to the feedback information they subsequently received: 1) *prediction quality: good* participants were told that their foresight judgments were very close to the feedback information, 2) *prediction quality: too high* participants were informed that their foresight judgments were too high in comparison to feedback, 3) *prediction quality: too low* participants were told that their foresight judgments were too low in comparison to the feedback information, 4) *prediction quality: poor* participants were told that their foresight judgments were quite poor because their judgments were often much too high or low, and 5) *no information* participants were not given any information regarding the quality of their foresight judgments.

Overall, S. Schwarz and Stahlberg (2003) predicted that a higher hindsight bias would be found for participants who had been led to believe that their foresight judgments were relatively good (i.e., close to the feedback information), whereas a smaller effect should be found when participants believed that their original judgments were quite poor in comparison to the correct answers. In general, the researchers found that the hindsight bias and effect size were larger in the prediction quality: good condition, in comparison to the pooled poor conditions (i.e., prediction quality: too low, too high, and poor). However, the magnitude of the effect was only marginally lower in the prediction quality: poor condition, in comparison to the (pooled) prediction quality: too high/low conditions. S. Schwarz and Stahlberg concluded that, although the results

were in line with the biased reconstruction view of the hindsight bias (i.e., a larger bias found when participants believed their original judgments were close to the feedback information), the data also suggest that more than a recreation of original judgments based on personal beliefs is occurring for the KIA items (e.g., the data cannot "totally exclude" a memory impairment approach). More specifically, if the KIA effect was due solely to recreating foresight judgments based on personal beliefs about prior performance then there should have been a stronger distinction between the "poor" and "quite poor" conditions regarding the level of the hindsight bias (therefore, other factors, such as memory impairment, play some role in the effect).

The foremost advantage of the biased reconstruction approach over the automatic assimilation hypothesis is that it recognizes the importance of subjective experience and meta-cognition in the creation (and/or moderation) of the hindsight bias. It is unfortunate that, thus far, the researchers who have proposed a biased reconstruction view of the KIA effect have not addressed the issue of whether there is *accompanying* subjective phenomenology to the effect (and what the factors may be that influence this phenomenology). From the results of the five experiments presented in this paper, it would be difficult to argue that the impact of subjective factors (e.g., accessibility experiences) is similar for both the objective and subjective measures of the effect because this proposal would imply that changes in performance would be similar on the two measures (i.e., perhaps not the same magnitude, but at least in the same direction). A descriptive examination of the measures of effect size for the KIA effect across the five experiments seems to indicate no support for the idea that the size of the hindsight bias necessarily matches the type of subjective experience; for example, the effect size for the

KIA items switching to the correct answer on Test 2 is comparable in Experiment 2 (trivia items: $d = 1.06$) and Experiment 5 (Wordies: $d = 1.13$), yet evidence of subjective experience was found only in Experiment 5.

Although there is nothing inherent in the biased reconstruction approach that would mandate equivalent changes across both objective and subjective performance for the KIA effect (and therefore it is conceivable to think that the approach perhaps could be expanded to include an explanation regarding the presence/absence and type of accompanying subjective experience), the approach, as it currently stands, suffers from some of the same drawbacks as the automatic assimilation hypothesis. That is, even though a biased reconstruction view emphasizes the reconstructive nature of the hindsight judgment task, it still appears to treat memory as a thing; recollection of the original judgment either does or does not occur, and when it does not occur reconstruction of the original judgment is necessary (S. Schwarz & Stahlberg, 2003). Further, no mention is made as to what specific factors may influence recollection of original judgments, and whether these same factors may influence the reconstruction of pre-feedback responses. Therefore, even though the biased reconstruction approach has the potential to surpass previous theories, in terms of handling the different hindsight bias patterns of results (i.e., in comparison to the automatic assimilation approach), rather than proposing changes to this existing theory that would enhance its application to the area, a different approach is presented in the next subsection. This new approach to explaining the bias is not specific to the KIA effect, but its main advantage is that it is a general theoretical framework that can be applied to all aspects of memory and cognition (and, therefore is more parsimonious than the theories that have been proposed thus far for the effect).

A comprehensive memory approach. The attributional approach of Jacoby and colleagues (e.g., Jacoby & Kelley, 1987; Jacoby, Kelley, & Dywan, 1989; Jacoby & Whitehouse, 1989) mentioned in the introduction of this paper is one (general) example of a comprehensive approach to memory and its various processes. In broad strokes, the attribution approach advocates for a constructivist approach to memory; for example, a feeling of remembering occurs when information available in the present is attributed to some source from the past (Kelley & Jacoby, 1993). A major factor in this attributional process is the fluency of processing, and more specifically, it is fluent processing of an item/episode that produces a feeling of familiarity. For instance, if an item on a recognition test is fluently processed, then some inference must be made about the source of that fluency: If the fluency of processing for that item is ascribed to the past (e.g., exposure to the item on a study list) then this attribution gives rise to a feeling of familiarity and the item will be judged as "old" (Jacoby & Whitehouse, 1989; Jacoby et al. 1989).

As its name suggests, the attributional approach is simply that – an approach (as compared to a detailed theory). However, this approach is noteworthy because it strongly emphasizes the reconstructive nature of memory, as well as the multiple factors that can influence the reconstruction of past events (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Jacoby & Kelley, 1987). One of the important advantages of the attributional approach is its emphasis on the idea that the use of the same source of information across tasks will not necessarily result in the same observed outcome for each task (e.g., relying on familiarity can lead to both accurate and inaccurate performance, depending on the nature of the task).

An attributional approach can be seen as focussing on how available information is used to make subjective judgments, depending on the task at hand. For instance, Begg, Robertson, Gruppuso, Anas, and Needham (1996) proposed that the KIA effect may be due, at least in part, to basing judgments on familiarity: Facts learned during an experiment feel familiar, and this familiarity is misattributed to prior knowledge. Begg et al. used three different tests following a study phase to examine what they called the illusory-knowledge effect. Participants in the *total-recall* test were simply asked to answer as many question as they could (number of questions answered with and without study), whereas in the *knew-it* test participants were instructed to answer only the questions they believed they could have answered prior to the study phase. In the *learned-it* test participants were told to respond to a question only if they had learned the answer during the study phase (and had not known the answer prior to the experiment). A measure of the KIA effect was obtained by comparing the learned-it and knew-it tests to the total-recall test. In general, a hindsight bias was found for participants in the knew-it test, but not the learned-it test. Begg et al. claimed that the lack of bias in the learned-it test, along with the KIA effect in the knew-it test, indicated that the effects of feedback depend on where attention is focussed. That is, participants are capable of correctly identifying items that they have just learned when their task is to deal with this new information in the present. It may be that the familiarity produced from feedback results in a hindsight bias when attention is focussed on the past, but that this same familiarity leads to correctly identifying new (learned) items when attention is focussed on the present: What you attribute the familiarity to depends on the goal you are trying to accomplish (Jacoby & Kelley, 1987).

As mentioned at the beginning of this subsection, the attributional approach is not a theory per se (i.e., it is more a collection of general principles that loosely have been applied to memory in general), and therefore a common criticism of the approach is that, in its current form, it is unfalsifiable; its generality enables it to explain many different types of results, but often only after-the-fact. However, there is a relatively recent comprehensive approach to memory that is similar in nature to some of the main tenets of the attributional approach, but which contains a more detailed framework. Specifically, Whittlesea's (2002, 2003, 2004) unified theory of memory – the Selective Construction And Preservation of Experience (SCAPE) – is based on the assumption that there is only one memory system. Specifically, the same processes that are responsible for recollecting an event from the past are also responsible for performing a skilled cognitive task. According to the SCAPE account, "all mental events can be understood as an interaction between memory and the environment" (Whittlesea & Leboe, 2000, p. 102), and this interaction takes place in what has been named the stimulus complex (i.e., purpose of task, physical properties of environment, quality of information, etc.).

Relatedly, in keeping with the unitary memory system principle of SCAPE, Whittlesea (Leboe & Whittlesea, 2002; Whittlesea & Leboe, 2000) claimed that remembering is necessarily an inferential process because "there is no mental content that occurs in a true act of remembering that cannot occur in an act of imagining" (Whittlesea & Leboe, p. 103). For example, the type of perceptual and semantic detail that can be found in a recollection of a perceived event also may be present in a recollection of an imagined event. Therefore, an important implication of the SCAPE account is that both illusions of remembering and veridical recollection are created in the same manner;

whether an individual correctly infers an event to be a memory or imagination depends on what is supported by the interaction of current stimuli, environmental cues, and the intention of the individual (cf. SMF, e.g., Johnson et al., 1993). This treatment of veridical and illusory memory is an important feature for the hindsight bias because, in certain situations, the bias involves two different types of illusory memories; an inaccurate reconstruction of pre-feedback knowledge and a false belief of remembering for that inaccurate reconstruction.

Within the SCAPE account, the construction of mental experiences and responses has two main components; production and evaluation (Whittlesea, 2003, 2004; Whittlesea & Leboe, 2000). The production function includes the creation of motor and mental responses, and this construction is based on the stimulus complex. Specifically, Whittlesea (2003) argued that the production function is controlled by the transfer-appropriate processing (TAP) principle (e.g., Franks, Bilbrey, Lien, & McNamara, 2000), in that, the interaction of current and past processing plays a large part in determining performance. For example, the TAP hypothesis argues that the closer the match between the processes used at encoding and retrieval, the better performance will be at retrieval (see Franks et al., 2000; Morris, Bransford, & Franks, 1977; Rajaram, Srinivas, & Roediger, 1998 for a more detailed discussion of the TAP hypothesis). The evaluation component monitors production by assessing performance "to infer the nature of the previous experience or knowledge of a stimulus [and this] inference is based on the coherence of current processing and on the context and task in which the stimulus is processed" (Whittlesea & Leboe, 2000, p. 103). In short, the production function results in performance, whereas the evaluation function results in the subjective experience of

that performance.

Contrary to many dual-process models, Whittlesea (2002, 2003; Leboe & Whittlesea, 2002) argued that recollection and familiarity are the result of the same processes, and that the differences found between feelings of recollection and familiarity stem from differences in the available information that each is based upon.[12] Whittlesea noted that familiarity and recollection are often defined as being the result of automatic (unconscious) and controlled (conscious) processes, respectively, but he argued that neither of these subjective experiences had a defining feature; for example, both recollection and familiarity can occur with or without awareness of the processes. Leboe and Whittlesea (2000) attempted to illustrate the automatic component of recollection by having participants study target items that were paired with either semantically related associates or rhyming associates; at test, half of the targets were presented with their studied associates, and the other half of the target items were paired with the associates that they had not been studied with (e.g., a target studied with a rhyming associate might have been paired with its unstudied semantic associate). The test also included new pairs of targets with semantic or rhyming associates, and participants were asked to make recognition judgments under one of three conditions: 1) inclusion condition (say "old" to any target from the study phase), 2) exclude-meaning condition (say "old" only to targets that had been studied with a rhyming associate), and 3) exclude-rhyme condition (say "old" only to targets that had been studied with a semantic associate).

---

[12] The SCAPE account makes no strong distinctions along the lines of the "traditional" dichotomies, such as episodic/semantic, declarative/procedural, remembering/knowing, and so forth. A detailed discussion of the SCAPE account for each dichotomy is beyond the range of the present paper. However, overall, Whittlesea (2002) has argued that each dichotomy has fuzzy boundaries, and similar to Gruppuso et al.'s (1997) functionalist approach, he claimed that "processing that is labelled clear recall on one occasion may instead be labelled familiarity on another, when one has the opportunity to experience another way of remembering" (p. 343).

Leboe and Whittlesea (2000) found that, overall, participants in the exclude conditions did fairly well at both accepting and rejecting the instructed target words; however, the manipulation of test context had a significant effect on the discrimination judgments. More specifically, participants in both of the exclude conditions were much more likely to accept an appropriate old target item if that target was in the same context as it had been during the study phase (i.e., paired with the same rhyming or semantic associate with which it had been studied). Additionally, participants in both exclusion tasks were more likely to correctly reject an old inappropriate target item if that target word appeared in the same context at test as at study. Leboe and Whittlesea contended that these results showed that the ability to recall an event is not "unilaterally" under the influence of an intention to remember. That is, participants in both of the exclusion tasks were intentionally attempting to recollect the studied contexts of the target items (the necessary information for correctly completing the recognition judgments), but this intention was influenced by the test context. The researchers were not trying to deny the importance of intention for the act of recall, but rather they emphasized that intention is not necessarily the defining feature or cause of recall.

In general, Whittlesea (2002) claimed that the feeling of remembering an experience typically occurs through two main "routes." The first mechanism of remembering involves the perception of discrepancy – a concept that pre-dates the formalized SCAPE account and is referred to as the discrepancy-attribution hypothesis (e.g., Whittlesea, 1993, 2002; Whittlesea & Williams, 2001a). Whittlesea argued that individuals are constantly constructing and evaluating responses (mental and motoric) to the stimuli surrounding them, and that this evaluation function produces one of three

perceptions; coherence (all elements fit together), incongruity (some elements conflict with each other in an "identifiable way"), or discrepancy (similar to incongruity, but the source of the conflict is not readily identifiable). According to the discrepancy-attribution hypothesis, the feeling of familiarity is a consequence of the perception of discrepancy. Specifically, individuals can detect differences between how they expect to perform on a stimulus and how they actually perform, and if their actual performance is more fluent than they expected (referred to as "surprising fluency") they may attribute this fluency to some source in the past. However, the attribution of the perception of discrepancy to either a source in the past or the present depends on a multitude of factors, such as prior knowledge of the stimuli and the present conditions/context. Therefore, the feeling of familiarity is not an automatic result of discrepancy, but rather it is the result of an inferential process that is triggered by the detection of discrepancy.

To gather evidence for the discrepancy-attribution hypothesis, Whittlesea and Williams (1998; Experiment 3) had participants study three types of stimuli: 1) natural words (e.g., *table*), 2) orthographically regular nonwords (e.g., *hension*), and 3) orthographically irregular nonwords (e.g., *lictpub*). At test, participants were required to say each item aloud and judge whether that item had been studied or was new. Results from the new items showed that there were more false alarms for the regular nonwords than for the natural words or the irregular nonwords (hereafter referred to as the hension effect). Whittlesea and Williams claimed that the regular nonwords produced higher false alarms because the fairly fluent pronunciation of these items led participants to expect that they could do more with them, such as produce a meaning. But, when participants' expectations of the regular nonwords were not met (e.g., a meaning could not be

produced), their evaluation of their processing led to a perception of discrepancy. Additionally, the participants were unable to determine the true source of their surprising fluency – that is, that the fluency was due to the orthographic regularity of the regular nonwords – and consequently they incorrectly attributed the fluency of these items to having seen them during the study phase.[13]

The second route to remembering occurs through a "special" perception of coherence (i.e., integrality). To start with some background, in the process of testing the perception of discrepancy, Whittlesea and Williams (2001b) developed a sentence-stem paradigm in which they had participants study a list of words (e.g., *boat*); at test, participants were given sentence stems that were followed by either a new or old probe word (Experiment 1). Whittlesea and Williams manipulated whether the sentence stem was predictive (high-constraint) of the probe word (e.g., *The stormy seas tossed the...boat*) or whether it was merely consistent (low-constraint) with the probe (e.g., *She saved her money and bought a...boat*). They also manipulated the timing of the probe word so that half of the trials had a pause between the presentation of the sentence stem and the probe word, and the other half of the trials contained no pause. The researchers found an interaction between the predictiveness and pause manipulations: Probes presented with a predictive stem were more likely to be judged old than probes presented with consistent stems, but only if there had been a pause between the sentence stem and the probe. Whittlesea and Williams maintained that the consistent stems simply prepared

[13] In a series of follow-up experiments, Whittlesea and Williams (2000) demonstrated that the hension effect was not due to the inclusion of either the natural words or the irregular nonwords (Experiments 3a-e), and that the effect could be reduced if participants were given a task to complete with the regular nonwords (e.g., rhyming judgment, Experiment 4). Whittlesea and Williams concluded that these results showed that what is important for producing the effect is the properties of regular nonwords themselves, as well as the context in which they are processed (e.g., simple recognition vs. rhyming task and recognition).

participants to incorporate a terminal word (probe), leading the participants to experience these types of events as "merely coherent." In contrast, the predictive stems created an expectation for the outcome of the sentence, but because the predictive context only restricted the potential outcomes (i.e., because more than one word could fit the stem) "the expectation is framed in terms of the meaning of the stem, not in anticipation of the specific meaning of the probe" (p. 16).

The sentence-stem paradigm that Whittlesea and Williams (2001b) used to investigate the perception of discrepancy demonstrated that the probe words paired with consistent stems felt "merely coherent" for participants, and subsequently did not lead to a feeling of familiarity. However, by modifying the sentence-stem paradigm so that the sentence stems were presented with the target items in the study phase (referred to as the sentence-stems-in-training paradigm), Whittlesea (2002) found that a special sense of coherence – what he refers to as a perception of integrality – could lead to feelings of remembering. More specifically, he argued that presenting the stems with the target words in the study phase led participants to create "themes," and that these themes produced definite expectations about the terminal words of the stems in the test phase. Further, Whittlesea claimed that the level of sentence constraint impacted the definite expectations formed at study; low-constraint items led participants to form definite *specific* expectations, whereas the high-constraint sentences let to definite *general* expectations. That is, studying low-constraint sentences led to a specific definite expectation because the theme of the sentence is formed by the terminal word (i.e., a variety of terminal words are consistent with the sentence, therefore the theme of the sentence is not clear until the presentation of the final word). Conversely, the theme of

high-constraint sentences is present in the sentence itself, and therefore the definite expectation created at study is general rather than specific because the expectation is not tied to the presentation of that final word (i.e., as long as it is consistent with the sentence, the terminal word has little "impact" on the theme).

Whittlesea (2002) argued that, at test in the sentence-stems-in-training paradigm, the presentation of the same terminal word for a low-constraint studied stem validated participants' definite specific expectations, and this validation led to a perception of coherence (which does not give rise to claims of recognition, Whittlesea, 2004). However, the validation of the definite general expectations for the high-constraint sentences led participants to experience a perception of integrality (e.g., that the parts "belong" together); participants may use this perception of integrality to infer that the sentence and terminal word had been experienced in the study phase, which results in a conscious state of remembering (in this context, reflected by a higher false alarm rate for high-constraint sentences).

One of the core advantages of applying the SCAPE framework to the KIA effect is that a major feature of the framework revolves around the idea that subjective phenomenology is not merely a "by-product" of the production of past experiences; the evaluation function "can produce subjective experiences that are predictable or entirely unpredictable from objective properties of performance, such as fluency or similarity to prior experiences" (Whittlesea, 2004, p. 892). This uncoupling of objective and subjective experience is the first necessary ingredient for explaining the differential rates of the subjective phenomenology component of the effect found across the five experiments reported in this paper. That is, as mentioned in a previous section, the

objective measure of hindsight bias does not appear to be directly related to whether there is also an accompanying subjective experience, and therefore it is important to propose an explanation of the effect that does not necessarily bind these two different measures of the KIA effect together (although this is not meant to imply that the two measures may not be influenced in a similar manner under some conditions).[14]

The application of the SCAPE framework to hindsight bias becomes a little more complicated in regards to the explanation for why subjective phenomenology is found under only certain conditions; is it a perception of discrepancy or a perception of integrality that leads participants to sometimes claim that they remember having given the feedback information during the first test? In both the trivia KIA paradigm (Experiments 1 and 2; no evidence of subjective phenomenology) and the Wordies paradigm (Experiments 4 and 5; higher claims of conscious remembering for feedback KIA items) the stimuli were always presented with the two possible responses at Test 1 and Test 2 (correct answer and plausible foil). Therefore, unlike the sentence-stem and sentence-stems-in-training paradigms, there are no novel items presented/integrated on the final test (i.e., no equivalent items to the novel or re-paired terminal words). However, as discussed in the previous section, a key difference between the Wordies and trivia paradigms is the interaction between the feedback tasks and the nature of the

[14] A SCAPE explanation of the objective measures of the KIA effect is not explicitly included in this subsection because the focus of the present work is on the subjective nature of the bias (i.e., the goal of the research was not to measure differences in the objective measures of hindsight bias across paradigms). However, a SCAPE explanation of the objective measure of hindsight bias would involve principles that have already been discussed, such as source monitoring and TAP. Because the SCAPE framework is not specific to the KIA effect (as were automatic assimilation and biased reconstruction), and it separates objective and subjective performance, an account of the subjective phenomenology can be proposed separately from an account of the objective measures. However, this separation is not meant to imply that the different types of information that factor into producing the objective measure also will not play a part in producing the subjective phenomenology (e.g., it is likely that, under certain conditions, the information that feeds into creating an increased hindsight bias may also result in a heightened subjective experience).

stimuli. Therefore, it is important to look at how these factors may play into producing perceptions of integrality and/or discrepancy.

Focusing specifically on KIA trials in a trivia KIA experiment (e.g., items that switched to the correct answer on Test 2), it could be argued that there is in fact a large overlap between the feedback and control items in the sort of information available for making the objective and subjective judgments. For example, because the trivia items were difficult (and because participants had to guess a lot on Test 1) it is not surprising that participants had a difficult time trying to reconstruct their Test 1 response on the final test, and this difficulty is present for both feedback and control items. The feedback items, though, have an additional "source" of information over control items (i.e., the feedback phase), which has the potential to play a role in both the objective and subjective judgments. This feedback phase leads to a higher rate of moving to the correct answer on the second test for the feedback items – due to principles such as source monitoring, (e.g., confusing the correct answers shown in feedback with responding in Test 1) – but the nature of the feedback tasks/stimuli combination does not add information that would support conclusions of "remembering" or "knowing." For example, participants were aware that they were exposed to feedback between the tests, and therefore had at least one basis for explaining why the correct response may have come to mind at Test 2 (which would eliminate any perception of discrepancy, but not necessarily eliminate a bias toward giving the correct information in Test 2). Because the feedback phase in the trivia KIA paradigms made no attempt to provide participants with an explanation for *why* a particular response was the correct answer to a question, the feedback items (like the control items) were likely still experienced as "difficult" on the

second test (i.e., having been exposed to the correct answers did not provide participants with "evidence" or reasons for why they may have chosen the correct answer on Test 1).

Unlike in the trivia KIA paradigm, the information that is unique to the feedback items (feedback phase) in a Wordies paradigm is likely to impact the evaluation process in a way that does sometimes lead to the conclusion that the correct solutions have been remembered from the first test. Specifically, the understanding-of-the-solution feedback task (Experiments 4 and 5) supplied participants with the opportunity both to learn the correct answers to the word puzzles and the steps (reasons) that lead to those correct answers. In Test 2, as in the trivia paradigm, participants are aware that they were given feedback for some of the items (which, as stated previously, they may use as a basis for explaining why the correct solutions came to mind at Test 2); however participants may fail to appreciate the impact that learning *how* to get to the correct solutions (training) has on their performance. That is, at Test 2 the participants experience a perception of integrality for the feedback items: The correct solutions to the KIA feedback items in Test 2 seem like they "belong" with the puzzles (e.g., Whittlesea, 2002) because the participants understand why those responses are the correct solutions (e.g., "I know I was shown that correct answer by the experimenter, but I remember how I solved the puzzle, and I'm sure I did that in the first test").

Obviously the construction of subjective phenomenology (or lack of subjective phenomenology) in a hindsight bias experiment is not as straight-forward as described in the preceding paragraphs; for example, learning the steps for why a solution to a word puzzle is correct during the feedback phase will not always lead to a feeling of remembering for that item in the final test. As mentioned earlier in this subsection, the

SCAPE approach integrates principles from a variety of areas (Whittlesea, 2004), such as TAP, fluency attribution (e.g., Jacoby & Whitehouse, 1989), and – central to the present discussion – source monitoring. Therefore, even though a participant in a Wordies paradigm may experience a perception of integrality for a feedback KIA item, s/he will only conclude that s/he is remembering having given the correct solution to that item if the perception of integrality is attributed to some source in the past and, more specifically, to having chosen that correct response in Test 1.

Although the application of the SCAPE framework appears to be a promising method of interpreting the combination of objective and subjective measures of the KIA effect, more research is needed before strong conclusions can be drawn. For example, I have argued that participants in the Wordies paradigm sometimes claim to remember in hindsight having given the correct information in foresight (for KIA trials) because they experience a perception of integrality. Nevertheless, future research may show this claim to be mistaken (i.e., perhaps it's a sense of discrepancy that is being misattributed to the past), or at least not how the feeling of remembering occurs across all situations and/or manipulations (in some situations it may be the perception of integrality that leads to the subjective experience component, whereas under other conditions it's the perception of discrepancy causing the phenomenology of the bias). Additionally, the SCAPE framework itself is relatively new, and Whittlesea (2004) noted that "undoubtedly, some of its assumptions are wrong and others oversimplified, and yet others are simply missing" (p. 906).[15] However, compared to the other approaches to the KIA effect that

[15] Whittlesea (2004) also noted that the two main criticisms of the SCAPE account are: 1) it is not as "well-specified" as other theories of memory, and 2) it has been accused of being a post-hoc explanation of behaviour. I believe that the first criticism will begin to be resolved with future research because, as of yet,

have been discussed in this section, it could be argued that the benefits of using a SCAPE account of memory and cognition to explain the subjective and objective measures of the hindsight bias outweigh any current drawbacks of the account.

A Related Phenomenon: The forgot-it-all-along effect

Although there are many (metamemory) effects that are related to the hindsight bias, such as the consistency of attitudes/beliefs (e.g., Ross, 1989), unintentional plagiarism (e.g., Landau & Marsh, 1997), feelings-of-knowing (e.g., Koriat & Levy-Sadot, 1999), and false-memory-for-false-memory (e.g., Marsh & Hicks, 2001), in the interest of brevity only one of the relevant phenomena will be discussed in this section; the forgot-it-all-along effect.

Schooler and colleagues (Schooler, 1999, 2001; Schooler Ambadar, & Bendiksen, 1997; Schooler, Bendiksen, & Ambadar, 1997) described two interesting case studies in which women reported full-blown recovered-memory experiences. What makes these two cases particularly remarkable is that in each a close confidant (e.g., former husband) of the woman involved reported that the woman had talked about the abuse during the period of supposed amnesia. Schooler and his coauthors speculated that during the recovered-memory experience the women remembered the abuse in a different way than they had previously (e.g., more completely, more episodically, or with a qualitatively different interpretation), such that the experience of remembering was qualitatively different from their previous recollections of the abuse, and that this in turn gave rise to what they termed a *forgot-it-all-along* (FIA) effect (named in reference to the KIA

---

few researchers have attempted to apply a SCAPE framework to their data (i.e., as more and more people apply the SCAPE framework to their results, the more refined and developed it will become). As to the second criticism, I agree with Whittlesea (2004) that many approaches to memory and cognition suffer from this issue. Additionally, I think that, as we learn more about how to apply the framework, it may become possible to increase SCAPEs predictive capabilities.

effect). That is, having recalled an event in manner X may cause one to forget having previously recollected it in manner Y.

Arnold and Lindsay (2002) developed a laboratory analogue that captures some aspects of Schooler's FIA mechanism (1999, 2001; see also Joslyn, Loftus, McNoughton, & Powers, 2001; Padilla-Walker & Poole, 2002, for related experimental work on the FIA effect). The prediction was that if remembering sexual abuse in qualitatively different ways on two occasions can lead a person to fail to remember the prior instance of remembering, then an analogous effect may occur if individuals are led to recall innocuous laboratory materials in qualitatively different ways on two occasions. In Arnold and Lindsay's Experiment 1, participants studied a list of homophones, each accompanied by a biasing context word (e.g., hand–*palm*, gun–*fire*). Participants were then tested on a subset of the study list, with some target items being cued with the studied-context word (e.g., hand: p***m) and others cued with an other-context word (e.g., flames: f**e). Subsequently, participants were tested on all of the studied items, this time with the studied-context cues given as recall prompts, and after each word was recalled they judged whether or not they had recalled that word on the first test. As predicted, participants were more likely to forget their prior recall of the words cued with other-context words on Test 1 (i.e., words with a change in context between tests) than of words cued with studied-context words on Test 1 (i.e., words with the same context on both tests). These result were taken as evidence that remembering a past event in a different way can lead one to fail to remember a prior instance of recalling that event.

In three additional studies, Arnold and Lindsay (2002) eliminated various alternative explanations for their findings. For example, the results of Experiment 1 might

be attributed to weaker Test 1 remembering of items in the other-context condition, rather than to the qualitative mismatch between Test 1 and Test 2 recall of those items. In Experiment 2, the studied-context versus other-context cues were manipulated in both recall tests; a FIA effect was found for items cued on Test 1 with studied-context cues and on Test 2 with other-context cues (as well as for items cued on Test 1 with other-context cues and on Test 2 with studied-context cues). Experiment 3 revealed that the FIA effect is not restricted simply to manipulating the dictionary definition of the to-be-recalled words (e.g., from hand-*palm* to tree-*palm*), but that it can also be obtained with more subtle shifts in meaning from Test 1 to 2. For example, participants who on Test 1 recalled "palm" in response to the cue "The fortune teller traced the lifeline on the p*** of his hand" and recalled the same studied item on Test 2 in response to the cue "He used his p*** to swat the fly" were less likely to remember their Test 1 recall of that item than were participants who were given the same cues on both tests. Finally, Experiment 4 showed that the effect was not merely due to participants judging whether they had seen the Test 2 retrieval cues on Test 1 (as opposed to judging whether they had recalled the target word). To this end, multiple short study lists and free recall procedures were used for Test 1, followed by cued recall in Test 2. Even though there were no retrieval cues in Test 1, participants more often remembered their Test 1 recollection of target words if they were cued in Test 2 with cues that matched the studied context of the targets than if they were tested with cues that did not match the studied context (i.e., cues that matched the non-studied meaning of the target words).

The most relevant FIA study, regarding the present topic of the accompanying subjective phenomenology of an effect, was conducted as a follow-up experiment

(Arnold & Lindsay, in press), and the question of interest focused on the issue of confidence; that is, when participants incorrectly judged that they had not recalled an item on Test 1, were they confident that they had *not* recalled that item, or were they simply unsure that they had recalled it? After making each "Yes/No" judgment of prior remembering decision (i.e., "Did you recall that target in Test 1, yes or no?"), participants were shown a screen with a 6-point confidence scale; (1) very low, (2) quite low, (3) somewhat low, (4) somewhat high, (5) quite high, and (6) very high. The participants were instructed to choose the option that best described their confidence for the judgment decision that they had just completed, and the experimenter stressed that there was no right or wrong answer to this confidence rating task (i.e., that participants should use the full range of the scale to select the option that best reflected their level of confidence for each individual trial). The results demonstrated that the FIA effect is not simply due to guessing: When participants incorrectly claimed that they had not previously recalled a context-change item (vs. an item that was tested with the same context on both tests), they were often quite confident that they had not previously remembered that item. Indeed, more than half of the time (61.67%) in the changed-context condition participants rated their incorrect "No" judgments in the high-range of the confidence scale (compared to 41.1% in the same-context condition). These confidence data demonstrated that participants were not simply choosing the "No" option for the judgment of prior recall because of low confidence (e.g., "I guess I'll say 'No,' but only because I have to give either a 'Yes' or 'No' answer"). Rather, the confidence judgment data suggest that participants in the FIA experiments are often quite confident in their erroneous belief that they had not previously recalled an item.

Although the FIA effect was named after the KIA effect, it would be premature to assume that these two effects are merely opposite ends of the same continuum, but it is reasonable to compare the two phenomena. For instance, both the KIA and the FIA effects revolve around retrospective knowledge: Participants are instructed to focus on their past so as to perform a task in the present. More specifically, both effects arise when participants are unable to assess/reconstruct their prior performance (i.e., level of pre-feedback knowledge for KIA, and prior recall for FIA) to the extent that is necessary to accurately complete their present task. Further, it is the addition of information (feedback for KIA) or change in information (context-change for FIA) that hinders performance; participants are fully aware that they must ignore the additional/changed information but it appears that they are unable to do so.

Similar to one of the key aspects of the production function of memory in the SCAPE framework (Whittlesea, 2002, 2004), Arnold and Lindsay (2002, in press) proposed that the FIA effect can be understood in terms of the notion of TAP (e.g., Morris et al., 1977). As mentioned earlier, the idea is that current thoughts about an item will be a poor cue for a prior instance of recalling that item if the current way of thinking about the item differs from how it was thought of during the prior recall. Although the nature of the tasks are different, a similar argument could be made for the KIA effect. Specifically, for the feedback items that participants answered incorrectly on Test 1, although the surface features of Test 1 and Test 2 match (i.e., the same stimuli are presented with the same possible alternatives), the feedback phase will have "cued" the participants to think about these feedback items in a different way than they had thought about them on the first test. Further, although there are many factors that influence

whether participants will claim to have given that feedback information in foresight – for example, whether they attribute the coming-to-mind of the correct information as solely due to the feedback phase, or whether they can reconstruct their thought processes from the first test that support the conclusion that they actually had chosen the incorrect response – only the feedback items will be influenced in a systematic manner by having been thought about in a different way subsequent to the first test. Therefore, a TAP explanation of the objective KIA data does not need to argue that only the feedback items will change on Test 2 (i.e., control items may also be changing between tests), but rather that only the feedback items will change in a systematic way (toward the feedback information).

Measuring the subjective experience component of the FIA effect obviously cannot be accomplished in the same manner as the R-JK-G judgments used with the KIA effect in the present experiments: It is not reasonable to ask participants in a FIA experiment if they "remember" that they did not recall an item on the first test. However, the results from the confidence rating FIA paradigm do indicate that the effect is not due to simple guessing, as participants are often quite confident about their incorrect "no" judgments of prior recall. What is not clear, though, is whether this pattern of confidence data would be found across different types of FIA paradigms or if, like the data for the subjective measure of hindsight bias, this pattern of confidence judgments would only be found under specific circumstances. Specifically, the confidence judgment was implemented in a "context sentence-as-cue" FIA paradigm; participants studied the target words with a context sentence and subsequently they were tested with either the same sentences on both tests, or a different context sentence was used on Test 1 and the studied

sentence was used on Test 2. However, the FIA effect also has been demonstrated with a "context word-as-cue" design (Arnold & Lindsay, 2002) and it is possible that this type of paradigm would not result in the same pattern of confidence judgments (e.g., it may parallel the lack of subjective phenomenology found in the trivia KIA paradigm).

Although the hindsight bias is a well-documented phenomenon, investigation of the FIA effect has just begun and further research is warranted before strong conclusions regarding the link between the two effects can be substantiated. Nevertheless, it would be a mistake to ignore the similarities between the two effects because the research conducted in one arena could lend understanding (and potential avenues of exploration) to the other, especially regarding the separability and nature of the subjective and objective components for each effect.

Summary and Conclusions

The goal of the present research was not to discredit the KIA effect. Indeed, the effect has been replicated numerous times and across a variety of situations (e.g., Christensen-Szalanski, & Willham, 1991; Fischhoff, 1975; Hoffrage & Pohl, 2003). The purpose of the five experiments presented in this paper was to emphasize the distinction between the objective and subjective measures that comprise the hindsight bias, and to provide evidence against the idea that an objective measure of the effect is automatically accompanied by a *feeling* of having remembered/known the information prior to receiving the feedback information. As the results of the present experiments demonstrated, the KIA effect sometimes has an accompanying component of subjective phenomenology, but the results also showed that it cannot be assumed that the measure used to characterize the effect (i.e., a numerical move toward, or switch to, the feedback

information on the final test) specifies anything about the subjective phenomenology of the bias.

Separating the objective and subjective components of the hindsight bias is important for a variety of reasons. One problem of assuming the co-occurrence of subjective experience is that this (mistaken) assumption begins to compound as more and more research is conducted. For example, if researchers assume that their participants really believe that they had known the feedback information in foresight, then any manipulation that the researchers subsequently implement to modulate that KIA effect will be assumed also to modulate the subjective experience of the effect (i.e., either increasing or decreasing the accompanying sense of previously having known the correct information). However, a more important problem with the coupling of the objective and subjective components is that this approach to the KIA effect leads to a mediocre understanding of the phenomenon: Investigating the situations under which individuals do (or do not) have an accompanying belief of possessing the feedback knowledge in foresight would tell us more about the effect itself, as well as inspiring additional lines of research (e.g., Whittlesea, 2004). Relatedly, investigating the separate objective and subjective components of the hindsight bias would contribute to our general understanding of memory processes, in that it would require both general (e.g., SCAPE) and more specific (e.g., biased reconstruction) theoretical frameworks of memory to explain the differing patterns of results found across the measures of objective and subjective performance.

References

Arnold, M. M., & Lindsay, D. S. (in press). Remembrance of Remembrance Past. *Memory.*

Arnold, M. M., & Lindsay, D. S. (2002). Remembering remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 521-529.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*, 610-632.

Begg, I. M., Robertson, R. K., Gruppuso, V., Anas, A., & Needham, D. R. (1996). The illusory-knowledge effect. *Journal of Memory and Language, 35*, 410-433.

Belli, R. F., Winkielman, P., Read, J. D., Schwarz, N., & Lynn, S. J. (1998). Recalling more childhood events leads to judgments of poorer memory: Implications for the recovered/false memory debate. *Psychonomic Bulletin & Review, 5*, 318-323.

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language, 28*, 563-580.

Bonds-Raacke, J. M., Fryer, L. S., Nicks, S. D., & Durr, R. T. (2001). Hindsight bias demonstrated in the prediction of a sporting event. *Journal of Social Psychology, 141*, 349-352.

Brain Teasers! (n.d.). Retrieved September 8, 2003, from http://www.billsgames.com/brain-teasers/

Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and human decision processes, 48*, 147-168.

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

Dehn, D. M., & Erdfelder, E. (1998). What kind of bias is hindsight bias? *Psychological Research, 61*, 135-146.

Dodson, C. S., & Johnson, M. K. (1996). Some problems with the process-dissociation approach to memory. *Journal of Experimental Psychology: General, 125*, 181-194.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 523-533.

Dunn, J. C. (2004). Remember-Know: A matter of confidence. *Psychological Review, 111*, 524-542.

Eldridge, L. L., Sarfatti, S., & Knowlton, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review, 9*, 139-145.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288-299.

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

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

Franks, J. L., Bilbrey, C. W., Lien, K. G., McNamara, T. P. (2000). Transfer-appropriate processing (TAP) and repetition priming. *Memory & Cognition, 28*, 1140-1151.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*, 309-313.

Gardiner, J. M., & Conway, M. A. (1999). Levels of awareness and varieties of experience. In B. H. Challis & B. M. Velichkovsky (Eds.), *Stratification in cognition and consciousness* (pp. 237-254). Amsterdam, Netherlands: John Benjamins Publishing Company.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory and Cognition, 18*, 23-30.

Gardiner, J. M., Kaminska, Z., Dixon, M., & Java, R. I. (1996). Repetition of previously novel melodies sometimes increases both remember and know responses in recognition memory. *Psychonomic Bulletin & Review, 3*, 366-371.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory, 10*, 83-98.

Goethals, G. R., & Reckman, R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology, Vol. 9*, 491-501.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 259-278.

Halford, K. W., & Griffith, U. (2002). "How has the week been for you two?" Relationship satisfaction and hindsight memory biases in couples' reports of relationship events. *Cognitive Therapy & Research, 26*, 759-773.

Hardt, O., & Pohl, R. F. (2003). Hindsight bias as a function of anchor distance and anchor plausibility. *Memory, 11*, 379-394.

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

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & et al. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory and Cognition, 16*, 533-538.

Henkel, L. A., Franklin, N., & Johnson, M. K. (2000). Cross-modal source monitoring confusions between perceived and imagined events. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 321-335.

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 566-581.

Hoffrage, U., & Pohl, R. F. (2003). Research on hindsight bias: A rich past, a productive present, and a challenging future. *Memory, 11*, 329-335.

Holmes, J. B., Waters, H. S., & Rajaram, S. (1998). The phenomenology of false memories: Episodic content and confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1026-1040.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language, 30*, 513-541.

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review, 5*, 705-709.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Psychology Bulletin, 13*, 314-336.

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

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118*, 126-135.

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13-47). Mahwah, NJ: Erlbaum.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.

Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 482-526). New York: Guilford Press.

Joslyn, S., Loftus, E. F., McNoughton, A., & Powers, J. (2001). Memory for memory. *Memory & Cognition, 29*, 789-797.

Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologia, 98*, 127-140.

Kelley, C. M., & Jacoby, L. L. (1993). The construction of subjective experience: Memory attributions. In M. Davies & G. W. Humphreys (Eds.), *Consciousness: Psychological and philosophical essays. Readings in mind and language* (Vol. 2, pp. 74-89). Oxford, UK: TJ Press Ltd.

Kelley, C. M., & Jacoby, L. L. (2000). Recollection and familiarity: Process-dissociation. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 215-228). London: Oxford University Press.

Koriat, A., & Levy Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483-502). New York: The Guilford Press.

Landau, J. D., & Marsh, R. L. (1997). Monitoring source in an unconscious plagiarism paradigm. *Psychonomic Bulletin & Review, 4*, 265-270.

Leboe, J. P., & Whittlesea, B. W. A. (2002). The inferential basis of familiarity and recall: Evidence for a common underlying process. *Journal of Memory & Language, 46*, 804-829.

Lieberman, J. D., & Arndt, J. (2000). Understanding the limits of limiting instructions: Social psychological explanations for the failures of instructions to disregard pretrial publicity and other inadmissible evidence. *Psychology, Public Policy, and Law, 6*, 677-711.

Lindsay, D. S., & Johnson, M. K. (1991). Recognition memory and source monitoring. *Bulletin of the Psychonomic Society, 29*, 203-205.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52*, 297-318.

Loftus, E. F. (1979). The malleability of human memory. *American Scientist, 67*, 312-320.

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.

Marsh, R. L., & Hicks, J. L. (2001). Output monitoring tests reveal false memories of memories that never existed. *Memory, 9*, 39-51.

Mazursky, D., & Ofir, C. (1990). "I could never have expected it to happen": The reversal of the hindsight bias. *Organizational Behavior and Human Decision Processes, 46*, 20-33.

Morris, C. D., Bransford, J. D., and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179-195). London: Oxford University Press.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior, 19*, 338-368.

Padilla-Walker, L. M., & Poole, D. A. (2002). Memory for previous recall: A comparison of free and cued recall. *Applied Cognitive Psychology, 16*, 515-524.

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

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89-102.

Rajaram, S., Hamilton, M., & Bolton, A. (2002). Distinguishing states of awareness from confidence during retrieval: Evidence from amnesia. *Cognitive, Affective, & Behavioral Neuroscience, 2*, 227-235.

Rajaram, S., Srinivas, K., & Roediger, H. L. III (1998). A transfer-appropriate processing account of context effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 993-1004.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341-357.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review, 111*, 588-616.

Sanna, L. J., & Schwarz, N. (2003). Debiasing the hindsight bias: The role of accessibility experiences and (mis)attributions. *Journal of Experimental Social Psychology, 39*, 287-295.

Sanna, L. J., Schwarz, N., & Small, E. M. (2002). Accessibility experiences and the hindsight bias: I knew it all along versus it could never have happened. *Memory & Cognition, 30*, 1288-1296.

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

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments and Computers, 20*, 206-217.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime User Guide*. Pittsburgh: Psychology Software Tools Inc.

Schooler, J. W. (1999). Seeking the core: The issues and evidence surrounding recovered accounts of sexual trauma. In L. M. Williams (Ed.), *Trauma and memory* (pp. 203-216). Thousand Oaks, CA, USA: Sage Publications.

Schooler, J. W. (2001). Discovering memories in light of meta-awareness. *The Journal of Aggression, Maltreatment and Trauma, 4*, 105-136.

Schooler, J. W., Ambadar, Z., & Bendiksen, M. (1997). A cognitive corroborative case study approach for investigating discovered memories of sexual abuse. In J. D. Read & D. S. Lindsay (Eds.), *Recollections of trauma: Scientific evidence and clinical practice* (pp. 379-387). New York: Plenum Press.

Schooler, J. W., Bendiksen, M., & Ambadar, Z. (1997). Taking the middle line: Can we accommodate both fabricated and recovered memories of sexual abuse? In M. A. Conway (Ed.), *Recovered memories and false memories* (pp. 251-292). New York: Oxford University Press.

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

Sharpe, D., & Adair, J. G. (1993). Reversibility of the hindsight bias: Manipulation of experimental demands. *Organizational Behavior and Human Decision Processes, 56*, 233-245.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174-215.

Stalhberg, D., & Maass, A. (1998). Hindsight bias: Impaired memory or biased reconstruction? In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 8, pp. 105-132). Chichester, England UK: John Wiley & Sons, Inc.

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

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*, 1235-1253.

Whittlesea, B. W. A. (2002). Two routes to remembering (and another to remembering not). *Journal of Experimental Psychology: General, 131*, 325-348.

Whittlesea, B. W. A. (2003). On the construction of behavior and subjective experience: The production and evaluation of performance. In J. S. Bowers & C. J. Marsolek (Eds.), *Rethinking implicit memory* (pp. 239-260). London: Oxford University Press.

Whittlesea, B. W. A. (2004). The perception of integrality: Remembering through the validation of expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 891-908.

Whittlesea, B. W. A., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. *Journal of Experimental Psychology: General, 129*, 84-106.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica, 98*, 141-165.

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 547-565.

Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 3-13.

Whittlesea, B. W. A., & Williams, L. D. (2001b). The discrepancy-attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. Journal of Experimental Psychology: Learning, Memory, & Cognition, 27, 14-33.

Williams, A. D. (1992). Bias and debiasing techniques in forensic psychology. *American Journal of Forensic Psychology, 10*, 19-26.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*, 616-641.

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

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1341-1354.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130*, 361-379.

Appendix A

Table A1

The mean proportion of response judgment R-JK-G designations (Experiments 1 - 5) for the natural log transformed data for feedback and control items

Condition	Feedback			Control		
	R	JK	G	R	JK	G
Exp't 1						
Different Toward[a]	-.59	-1.35	-1.89	-.81	-1.08	-1.53
Switch I - C[b]	-1.27	-1.06	-.70	-1.06	-.88	-.54
Exp't 2						
Switch I - C	-1.20	-1.30	-.63	-1.13	-1.28	-.35
Exp't 3						
Different Toward	-.81	-1.19	-1.89	-.64	-1.47	-1.50
Switch I - C	-1.08	-1.25	-.97	-1.21	-.93	-.50
Exp't 4						
Switch I - C	-.93	-1.08	-1.09	-1.76	-1.02	-.40
Exp't 5						
Switch I - C	-1.17	-1.10	-.81	-1.57	-1.15	-.40

[a]Different Toward = items given a number that moves toward the correct answer on Test 2, but does not switch sides of the number scale.

[b]Switch I - C = items that switch from the incorrect response on Test 1 to the correct response on Test 2.

Table A2

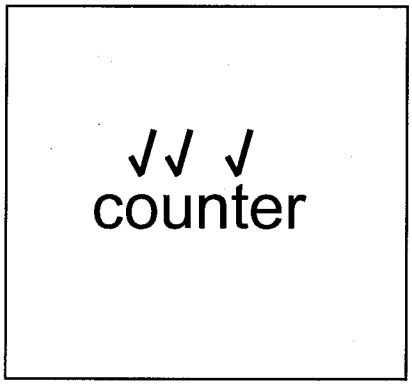
The mean proportion of number judgment R-JK-G designations (Experiments 1and 3) for the natural log transformed data for feedback and control items

Condition	Feedback			Control		
	R	JK	G	R	JK	G
Exp't 1						
Different Toward[a]	-1.75	-1.69	-.40	-1.91	-1.43	-.43
Switch I – C[b]	-1.88	-1.76	-.03	-1.30	-1.30	.00
Exp't 3						
Different Toward	-2.11	-1.25	-.49	-1.96	-1.58	-.32
Switch I - C	-1.90	-1.72	-.18	-1.50	-1.40	-.11

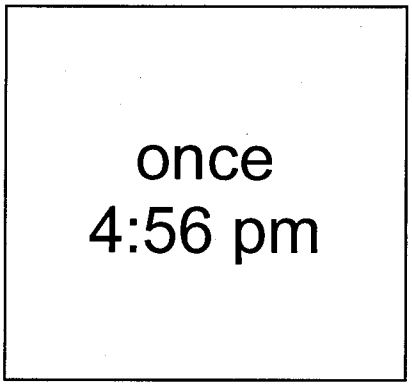
[a]Different Toward = items given a number that moves toward the correct answer on Test 2, but does not switch sides of the number scale.

[b]Switch I - C = items that switch from the incorrect response on Test 1 to the correct response on Test 2.

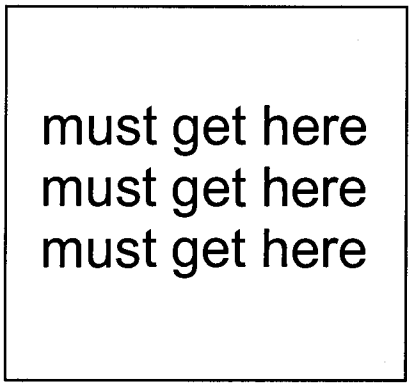
Appendix B



solution: check-out counter



solution: once upon a time



solution: the three musketeers