Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 2, 187–194

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.2.187

Illusions of Competence in Monitoring One's Knowledge During Study

Asher Koriat
University of Haifa

Robert A. Bjork
University of California, Los Angeles

The monitoring of one's own knowledge during study suffers from an inherent discrepancy between study and test situations: Judgments of learning (JOLs) are made in the presence of information that is absent but solicited during testing. The failure to discount the effects of that information when making JOLs can instill a sense of competence during learning that proves unwarranted during testing. Using a paired-associates task, the authors examined aspects of the cue–target relationships that seemed likely contributors to such illusions of competence. These aspects have the potential to create differential strengths of a priori and a posteriori associations, that is, the probability with which a cue, when presented alone, elicits the corresponding target versus the perceived association between the cue and the target when both are present. The authors argue that the former has the greater influence on later recall, whereas the latter has the greater influence on JOLs.

Previous work on judgments of learning (JOLs) indicates that such judgments are both moderately accurate in predicting future memory performance and generally sensitive to manipulations that affect actual learning and memory performance. However, several dissociations have been observed between predicted performance (JOLs) and actual performance (e.g., Benjamin, Bjork, & Schwartz, 1998; Carroll, Nelson, & Kirwan, 1997; Simon & Bjork, 2001), and these dissociations provide important clues for the mechanisms underlying JOLs. In this article, we focus specifically on the conditions that give rise to an overestimation of one's future memory performance—that is, to illusions of competence. Students, for example, exhibit such illusions when they hold unduly high expectations about their future test performance (see Dunning, Johnson, Ehrlinger, & Kruger, 2003; Metcalfe, 1998). In the present research, we set out to explore one general mechanism that might contribute to such illusions.

Examination of the literature reveals that JOLs elicited following study are generally well calibrated and do not exhibit the marked overconfidence bias that is typically observed, for example, in retrospective confidence judgments (see McClelland & Bolger, 1994). Nevertheless, in several studies using paired-associate learning, researchers have found that JOLs were inflated (on the first study block) compared with recall performance (Dunlosky & Nelson, 1994; Koriat, Sheffer, & Ma'ayan, 2002; Mazzoni & Nelson, 1995; Schneider, Visé, Lockl, & Nelson, 2000). This overconfidence bias has largely been confined to item-by-item JOLs, that is, to JOLs elicited following the study of each item. When participants provide aggregate judgments for the list as a whole, overconfidence is reduced, and there is sometimes underconfidence (Koriat et al., 2002; Mazzoni & Nelson, 1995).

Certain aspects of our prior findings suggest, however, that some paired-associate items are susceptible to exaggerated JOLs and that an examination of interitem differences might be instructive about the conditions that foster inflated predictions. A similar examination of interitem differences has proved useful in the study of the feeling of knowing (e.g., Koriat, 1995; Koriat & Lieblich, 1977).

Factors Contributing to Illusions of Competence

The mechanism we think contributes to inflated JOLs for certain items stems from a fundamental difference between the conditions of learning and the conditions of testing. On a typical memory test, people are presented with a question and are asked to produce the answer. In contrast, in the corresponding learning condition, both the question and the answer appear in conjunction, meaning that the assessment of one's future memory performance occurs in the presence of the answer. This conjunction can create a perspective bias: To predict accurately, a learner needs to adopt the perspective of the examinee, but doing so requires detaching oneself from the perspective of the learner, discounting what one now knows. The difficulty of achieving such a change of perspective is a potential source of overconfidence.

Indeed, several researchers have used the notion of the "curse of knowledge" to describe the tendency to be biased by one's own knowledge when judging the perspective of a more ignorant other (Birch & Bloom, 2003; Camerer, Lowenstein, & Weber, 1989; Keysar & Henly, 2002). As an example of the incapacitating effect of one's knowledge, Newton (1990) asked participants (*tappers*) to tap out the rhythm of a familiar song to *listeners* and to predict the likelihood that the listeners would successfully identify the song. Although tappers' predictions averaged 50%, the actual success rate of listeners averaged less than 3%. This result, as well as other results reviewed by Pronin, Puccio, and Ross (2002), demonstrates the difficulty that people have in discounting their privileged experience. Keysar and his associates (e.g., Keysar & Henly, 2002)

Asher Koriat, Department of Psychology, University of Haifa, Haifa, Israel; Robert A. Bjork, Department of Psychology, University of California, Los Angeles, California.

This research was supported by Grant No. 97-34 from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. We are grateful to Yaffa Lev and Erez Ofek for programming the experiments, to Hadas Gutman, Michal Vind, and Sarah Roper for conducting the experiments, and to Limor Sheffer for the statistical analyses.

Correspondence concerning this article should be addressed to Asher Koriat, Department of Psychology, University of Haifa, Haifa 31905, Israel. E-mail: akoriat@research.haifa.ac.il

also reported evidence indicating that speakers overestimate the effectiveness of their communication, expecting their addressees to understand their intentions more than is warranted.

We argue that learners, in a similar manner, are susceptible to a perspective bias: They may fail to discount what they know during study in predicting what they will know at test. Indeed, Kelley and Jacoby (1996) observed that participants, after solving a given anagram, were quite successful in predicting the difficulty that that anagram would pose for other participants, whereas such predictions were much less accurate when made in the presence of the solution to the anagram. In like manner, we argue that learners should find it difficult to escape the influence of a presented answer—but that the adverse effects of this difficulty should be more pronounced for some items than for others, as we describe.

Relevant Prior Research

A similar idea regarding the discrepancy between the context of learning and the context of testing has been advanced by T. O. Nelson and Dunlosky (1991) and Dunlosky and Nelson (1994; see also T. O. Nelson, Narens, & Dunlosky, 2004), who found that JOLs made at a delay following study are far more accurate in predicting eventual recall than are JOLs made immediately after study. However, this delayed-JOL effect occurs only when JOLs are cued by the stimulus term of a paired associate and not when they are cued by an intact stimulus-response pair (Dunlosky & Nelson, 1992). Nelson and Dunlosky proposed that the condition in which JOLs are delayed and cued by the stimulus alone approximates the eventual criterion test, which requires access to information in long-term memory, whereas JOLs made in response to intact cue–target pairs (or immediately after study) are made in the actual presence of the target (or its presence in short-term memory), situations that do not approximate the conditions at the time of the final test.

In studies of the delayed-JOL effect, the primary focus has been on resolution or relative accuracy, that is, on the extent to which participants' JOLs discriminate between items that are eventually recalled and those that are not. The focus of this article, in contrast, is on bias in absolute accuracy (cf. calibration) operationalized as the extent to which people accurately predict the magnitude of eventual memory performance. However, Nelson and Dunlosky have also examined calibration and found that delayed JOLs prompted by the cue alone produced better absolute accuracy than did immediate JOLs or delayed JOLs prompted by intact cue–target pairs (Dunlosky & Nelson, 1992, 1994; T. O. Nelson & Dunlosky, 1991). More important for our purposes, Dunlosky and Nelson (1997) found that delayed JOLs were consistently higher when prompted by the cue–target pair than when they were prompted by the cue alone, and they proposed that this effect might be a type of hindsight effect (Fischhoff, 1975): When both the cue and target are presented together, they evoke an "I knew it all along" feeling.

Goals of the Present Research

Our point of departure in trying to understand the processes that underlie inflated JOLs is to focus on interitem differences in the nature and strength of the association between the cue and target. Previous studies have established that the degree of semantic relatedness between the members of the pair is one of the most potent determinants of JOLs (Connor, Dunlosky, & Hertzog, 1997; Dunlosky & Matvey, 2001; Koriat, 1997). In fact, several findings suggest that semantic relatedness is overweighted as a cue for JOLs relative to extrinsic factors such as the circumstances of learning or encoding strategies (Carroll et al., 1997; Koriat, 1997; Shaw & Craik, 1989). In general, however, semantic-associative relatedness has been found to affect JOLs and recall to roughly the same extent (e.g., Dunlosky & Matvey, 2001; Hirshman & Bjork, 1988; Koriat, 1997).

In an effort to unravel the source of inflated JOLs, we draw a subtle but important distinction between two aspects of relatedness. Following Koriat (1981), we distinguish between a priori and a posteriori relatedness. A priori relatedness refers to the likelihood that the cue word in a paired associate will bring to mind the target word rather than any of the other potential responses. This type of relatedness is crucial for the test situation, when the learner is presented with the cue and asked to recall the corresponding target. A posteriori relatedness, in contrast, refers to the perceived relationship between the cue and the target when both are present, as is the case at the time of study, when JOLs are typically solicited. A priori relatedness is best measured by word-association norms, that is, by the probability that the cue word will elicit the target word as its first associate. In contrast, a posteriori relatedness is best measured by subjective judgments of the degree of relatedness between the cue and the target when both are present. The crucial difference is that the existence of other potential associates of the prime is critical in determining the degree of a priori relatedness but is entirely immaterial for a posteriori relatedness.

To the extent that JOLs are affected by the presence of the target (or answer) during study, they should be inflated when the strength of a posteriori relatedness is inordinately high relative to that of a priori relatedness. That is, we expect an illusion of competence when the presence of the target highlights aspects of the cue that are less likely to come forward in the presence of the cue alone.

In the experiments reported below, we manipulated attributes of paired associates that we assume affect the relative strength of a posteriori compared with a priori relatedness. To illustrate, and drawing on Koriat's (1981) materials, suppose that people are asked to judge the relative probabilities with which the second term in each of the following pairs is given as a response to the first term in word-association norms: *lamp–light*, *find–seek*, *sell–buy*, and *beautiful–nice*. The probabilities for the four pairs, respectively, are .706, .025, .564, and .028 (Palermo & Jenkins, 1964). People who have actually been asked to guess these probabilities greatly underestimated the differences among the pairs. When cue and target are presented together (e.g., *find–seek*), in our view, people tend to focus on a posteriori relatedness, ignoring the role of other likely (and competing) responses to the cue (e.g., *lose*). Therefore, pairs that have only a weak a priori association (e.g., *find–seek)* will sometimes be perceived as quite strongly related, producing inflated JOLs. In fact, learners perceive some relationship between words even if the normative association between those words is zero (see Fischler, 1977).

Experiment 1: The Effects of A Priori Associative Relatedness

In Experiment 1, we examined how JOLs and recall increase as a function of the degree of associative relatedness between the cue and the target. As just noted, we assume that even a weak a priori association tends to be perceived as a moderate association when both members of a pair are present, and that participants tend to perceive a relationship even between words that are unrelated according to word-association norms. Hence we expected JOLs to increase less sharply with associative strength than should actual recall. The pattern would be such that both unrelated pairs and weakly associated pairs should engender excessive JOLs compared with actual recall.

Many previous studies of the monitoring of learning during study have demonstrated marked effects of word relatedness on both JOLs and recall. However, in none of the previous articles have the results been discussed in terms of the distinction between a priori and a posteriori relatedness. As noted earlier, it is critical to distinguish between two ways in which relatedness is defined. In a number of previous studies, word pairs were classified as "related" and "unrelated" on the basis of subjective ratings of relatedness such as memorability judgments (Koriat, 1997; Experiment 1), ease-of-learning judgments (Dunlosky & Matvey, 2001), or ratings of "how easy or difficult it would be to come up with something in common between the two members of the pair" (Rabinowitz, Ackerman, Craik, & Hinchley, 1982, p. 690).

Understandably, such ratings are prompted by the entire cue–target pairs, and therefore are likely to capture the same type of a posteriori relatedness that is assumed to underlie JOLs. Hence, when relatedness is operationalized in this manner, there is no reason, in our view, to expect its effects on JOLs to be weaker than those observed for actual recall. Indeed, a reanalysis of the experiments referred to as Studies 4–8 and 10–11 in Koriat et al.'s article (2002), in which pairs were classified as easy or difficult on the basis of subjective ratings of memorability, revealed no significant interaction between difficulty and measure in any of these seven studies. As we discuss later (Experiment 3), pairs such as *citizen–tax* or *nurse–wife*, which have zero a priori association, would no doubt receive high ease-of-learning or memorability ratings.

On the other hand, studies in which relatedness was defined in terms of word association norms have mostly yielded results that are consistent with our hypothesis. In an experiment reported by Koriat (1997, Experiment 2), for example, the list of paired associates included 35 related pairs with a priori association of .05 or more according to word association norms and 35 unrelated pairs with zero a priori association. A reanalysis of the data indicated that recall for the related and unrelated pairs averaged 67.3 and 19.2, respectively, on the first presentation of the list, whereas the corresponding mean JOLs were 64.8 and 28.9. A Measure (JOL vs. recall) × Associative Relatedness analysis of variance (ANOVA) yielded $F(1, 23) = 9.79$, $MSE = 91.78$, $p < .005$, for the interaction. JOLs were significantly higher than recall for the unrelated pairs, $t(23) = 2.59$, $p < .02$, but not for the related pairs, $t(23) = 0.60$, *ns*, in which the observed difference was in the opposite direction. A similar interaction is apparent in results obtained by Connor et al. (1997, Experiment 3), but only for three of the four data sets reported. The results of a more recent study by Hertzog, Kidder, Powell-Moman, and Dunlosky (2002) do not exhibit the expected interaction.

We designed Experiment 1, then, to test whether associative strength has a weaker effect on JOLs than on actual recall, as we expected. We used a list of paired associates that included unrelated, weakly associated, and strongly associated pairs.

Method

Participants. Twenty-four English-speaking undergraduates enrolled in the overseas program of the University of Haifa, Haifa, Israel, were paid NIS 25 (~U.S.$5) for participating in the experiment.

Materials[1]. We constructed a list of 60 word pairs so as to include 20 unrelated, 20 weakly associated, and 20 strongly associated pairs. We defined associative strength as the probability of occurrence of the second word of a pair (target) as an associate of the first word (cue) among college students. We took the weak-association and strong-association pairs from Koriat (1981, Experiment 1). Their average probabilities of association were .065 (range = .025–.118) and .564 (range = .408–.706), respectively, and the unrelated pairs had zero associative strength.

Apparatus. The experiment was conducted on a Silicon Graphics workstation. The stimuli were displayed on the computer screen. JOLs and recalled responses were both spoken orally by participants and then entered by the experimenter on a keyboard.

Procedure. Participants were instructed that they would have to study 60 paired associates and would have to indicate their JOLs about each pair as soon as it disappeared from the screen. They were told that in the test phase, they would see each stimulus word in turn and would be asked to recall the corresponding response word.

During the study phase, the stimulus and response words were presented at the center of the screen side by side for 4 s. Participants were instructed to study each pair so that later they would be able to recall the second word in each pair when the first was presented. They were urged to use the entire 4 s for studying. The pair was replaced after 500 ms by the statement *Probability to Recall*. Participants reported their estimate orally on a 0%–100% scale. During the test phase, which began about 1 min after the end of the study phase, the 60 stimulus words were presented one after the other for up to 8 s each. Participants had to say the response word aloud within the 8 s allotted. The experimenter scored the response, and 1 s afterward, a beep was sounded and the next stimulus word was presented. Order of presentation of the items was randomly determined for each participant for each of the two phases of the experiment.

Results

Mean predicted recall (JOL) and actual recall percentages are presented in Figure 1 as a function of associative strength. Predicted recall ($M = 61.09$) matched actual recall ($M = 58.07$) closely, and both yielded strong effects of associative strength. A similar pattern of results was reported by Dunlosky and Matvey (2001) and Koriat (1997). However, as predicted, associative strength had a weaker effect on predicted recall than it had on actual recall. A two-way ANOVA, Measure (JOL vs. recall) × Associative Strength (unrelated, weak association, and strong association) yielded $F(1, 23) = 1.22$, $MSE = 268.33$, *ns*, for measure; $F(2, 46) = 269.9$, $MSE = 147.03$, $p < .0001$, for associative strength; and $F(2, 46) = 18.21$, $MSE = 106.05$, $p < .0001$, for the interaction.

[1] All materials used in this study are available from Asher Koriat on request.

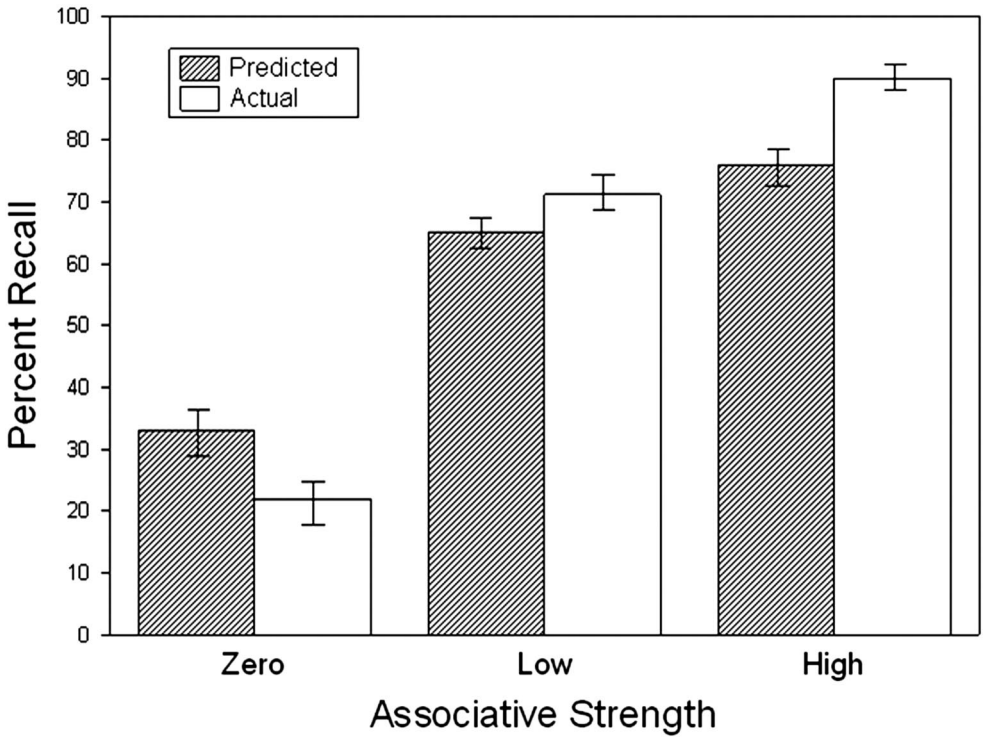

Figure 1. Mean predicted recall (judgments of learning) and actual recall as a function of associative strength (Experiment 1A). Error bars represent ± 1 *SEM*.

As can be seen in Figure 1, the interactive pattern is evident both in comparing the unrelated pairs with the weakly associated pairs as well as in comparing the two levels of association for the related pairs. Thus, a two-way ANOVA using only the unrelated and weakly related pairs yielded $F < 1$, for measure; $F(1, 23) = 344.98$, $MSE = 115.04$, $p < .0001$, for associative strength; and $F(1, 23) = 12.00$, $MSE = 145.21$, $p < .005$, for the interaction. A similar ANOVA using only the weak- and strong-association pairs yielded $F(1, 23) = 14.71$, $MSE = 162.74$, $p < .001$, for measure; $F(1, 23) = 67.56$, $MSE = 78.68$, $p < .0001$, for associative strength; and $F(1, 23) = 7.01$, $MSE = 51.52$, $p < .05$, for the interaction. JOLs were significantly lower than recall for the strong-association pairs, $t(23) = 6.60$, $p < .0001$, but the reverse was true for the zero-association pairs, $t(23) = 2.32$, $p < .05$. We expected the weakly associated pairs to also yield inflated JOLs, but this expectation was not borne out. The JOL-recall comparison for these pairs yielded $t(23) = 1.67$, *ns*.

Discussion

The finding that a priori associative strength exerts a weaker effect on JOLs than it does on recall is consistent with the idea that even a weak associative link, as measured by free association norms, is perceived as a moderately strong link when both members of a pair are presented together. However, the absolute values of predicted and actual recall in Experiment 1 are not entirely consistent with our predictions, because only the unrelated pairs, not the weakly associated pairs, yielded exaggerated JOLs.

Experiment 2: Forward and Backward Associations

In Experiment 2, we investigated the effects of another attribute: associative direction. Consider the pair *cats–kittens*, for example. Whereas the probability of *kittens* eliciting *cats* is .72, according to word association norms, the probability that *cats* elicits *kittens* is only .02. Assuming that backward associations are less beneficial for recall than forward associations (D. L. Nelson, McKinney, Gee, & Janczura, 1998; D. L. Nelson & Zhang, 2000), we expected inflated JOLs when the pairs were arranged in a backward direction in comparison with when they were arranged in a forward direction. That is, the presence of the response (*kittens*) along with the stimulus (*cats*) is likely to emphasize those attributes of the stimulus that are shared with the response and thus enhance the processing fluency of the entire pair (see Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin et al., 1998).

Method

Participants. Twenty English-speaking undergraduates enrolled in the overseas program of the University of Haifa were paid NIS 25 (~ U.S.$5) for participating in the experiment.

Materials. We used a list of 24 word pairs with asymmetric association. We divided the pairs into two equal sets that were matched in terms of the strength of the forward and backward associations (according to Palermo & Jenkins, 1964). The means of the associative strength in the forward and backward directions were .397 and .020, respectively, for Set A, and .396 and .021, respectively, for Set B. We assigned one set to the forward condition (i.e., with the strongest association being from the cue word to the target word) and the other to the backward condition, and we

counterbalanced the assignment across participants. In addition, we selected 24 unrelated pairs (zero associative strength) from the same norms.

Apparatus. The experiment was conducted on an IBM-compatible personal computer. The stimuli were displayed on the computer screen, and each participant's JOL and recall responses were entered by the experimenter on the keyboard.

Procedure. The procedure was the same as that of Experiment 1, except that each pair was presented for 5 s during the study phase. Participants were informed that the list would contain 48 paired associates.

Results

Mean predicted recall (JOL) and actual recall are plotted in Figure 2 for the forward and backward pairs. These means support our prediction: The direction of association had a weaker effect on JOLs than it did on memory performance. A two-way ANOVA, Measure (JOL vs. recall) × Direction (forward vs. backward) on these means yielded $F(1, 19) = 6.58$, $MSE = 165.12$, $p < .05$, for measure; $F(1, 19) = 11.58$, $MSE = 188.37$, $p < .005$, for direction; and $F(1, 19) = 18.73$, $MSE = 68.39$, $p < .0005$, for the interaction. The forward pairs yielded perfect absolute accuracy (cf. calibration), with mean JOLs (78.1) being practically identical to mean recall (78.8), $t(19) = 0.20$. The backward pairs, in contrast, produced inflated JOLs, with mean JOLs (75.7) being substantially higher than mean recall (60.3), $t(19) = 4.09$, $p < .001$. In fact, direction of association had a numerically small and nonsignificant effect on JOLs, $t(19) = 1.16$, *ns*, whereas it had a sizable effect on recall, $t(19) = 4.00$, $p < .001$.

The unrelated pairs also produced a certain degree of overconfidence: JOLs and recall for these pairs averaged 37.32 and 24.28, respectively, $t(19) = 3.48$, $p < .005$. This pattern is consistent with the results of Experiment 1, suggesting that participants tend to perceive some degree of association between pairs that have zero a priori association according to word association norms.

Discussion

In discussing the results of Experiment 2, it might be useful to borrow a distinction between two theories of semantic priming that have led to extensive research and controversy (see Hutchison, 2003). According to the associative view, priming in a lexical decision task depends strictly on the cue-to-target associative strength. This strength should determine the extent to which activation spreads from the cue to the target. In contrast, according to the featural view, priming is based purely on the shared overlap in features between a prime and a target.

In terms of such a distinction, one might propose that whereas recall success depends more heavily on the directional, cue-to-target associative link, JOLs rely more heavily on the global overlap between the cue and the target. Hence direction of association should be less critical for JOLs than it is for recall, as was found to be the case. Indeed, the existence of backward priming in lexical decision (Koriat, 1981) has been taken as evidence for the featural view of semantic priming (see Table 3 in Hutchison, 2003).

Experiment 3: A Priori Versus A Posteriori Associations

Experiments 1 and 2 indicated that associative strength and associative direction exert weaker effects on JOLs than they do on

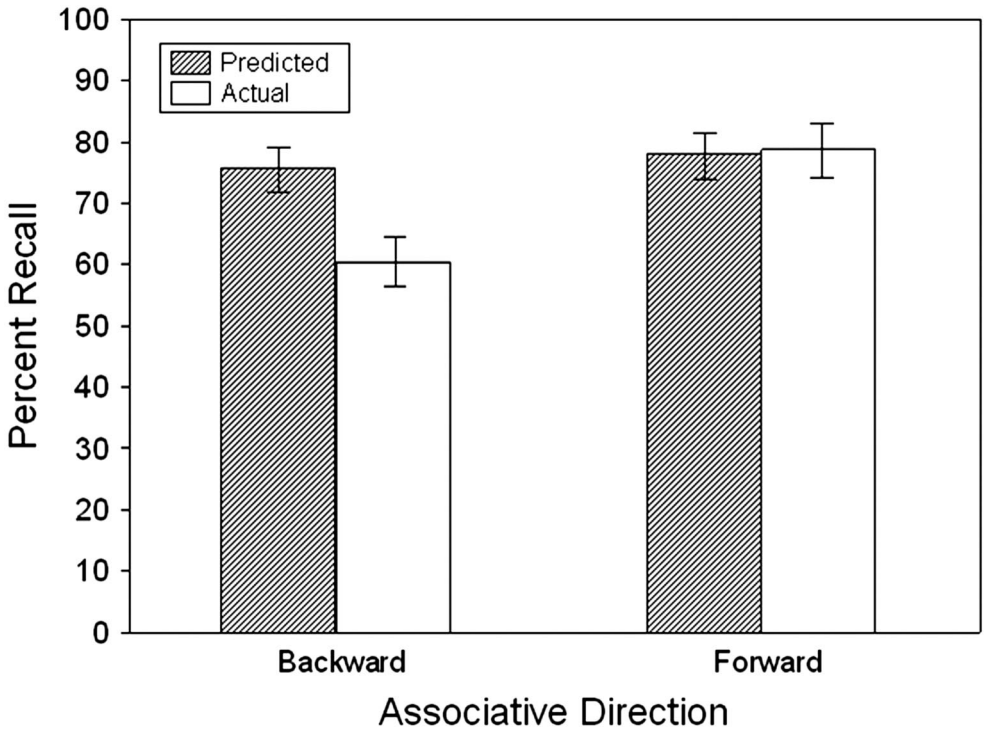

Figure 2. Mean predicted recall (judgments of learning) and actual recall for the forward-associated and backward-associated pairs (Experiment 2). Error bars represent ± 1 *SEM*.

recall. In Experiment 3, we extended the investigation to pairs for which the word–word association is purely a posteriori, that is, pairs in which neither word is an a priori associate of the other but that appear related when presented together. To illustrate, a pair such as *nurse–wife* has been found to yield semantic facilitation in lexical decisions (Fischler, 1977) despite the fact that the associative connection between the two words is zero. Assuming that the relationship between the members of a purely a posteriori pair is perceived only when both words appear together, we would expect these pairs to yield inflated JOLs. This would not be the case for a priori pairs in which the cue and target words have a direct, preexisting association.

Method

Participants. Sixteen Hebrew-speaking undergraduates at the University of Haifa were paid NIS 20 (~U.S.$4) for participating in the experiment.

Materials. We compiled a list of 72 Hebrew word pairs, consisting of 24 pairs with a high a priori association, 24 purely a posteriori pairs, and 24 unrelated pairs,. We took the 24 high-association pairs from Hebrew word association norms (Breznitz & Ben-Dov, 1991). We chose them so that the target word was a common response to the cue word. The average probability of association across the 24 pairs was .21 (SD = .10, range = .11–.53). Associative strength was determined by the probability of occurrence of the target as a response to the cue word (when only one response was solicited).

The 24 a posteriori pairs were selected by two judges to be semantically or associatively related, but their a priori association according to the norms was zero. Examples (translated from Hebrew) are: *bed–night*, *clean–soap*, and *laugh–humor*.

Finally, the 24 unrelated pairs were chosen so that they had zero association and were also judged by the two judges as having low association. Consistent with our hypothesis, the latter criterion was the more difficult to meet: Only two pairs from the norms met that criterion; the remainder of the unrelated pairs had to be chosen on intuitive grounds.

Apparatus and procedure. The apparatus and procedure were the same as those of Experiment 1 except that each word pair appeared on the screen for 3.5 s during the study phase, and 6 s were allowed for responding during the test phase.

Results

Mean predicted recall and actual recall are plotted in Figure 3 for the a priori, a posteriori, and unrelated pairs. A two-way ANOVA, Measure (JOL vs. recall) × Pair Type, on these means yielded $F(1, 15) = 13.69$, $MSE = 267.85$, $p < .005$, for measure; $F(2, 30) = 156.01$, $MSE = 163.92$, $p < .0001$, for pair type; and $F(2, 30) = 4.42$, $MSE = 105.01$, $p < .05$, for the interaction. A similar ANOVA including only the a priori and a posteriori pairs also yielded a significant interaction, $F(1, 15) = 10.20$, $MSE = 88.45$, $p < .01$. As apparent from the figure, mean JOLs closely matched mean recall for the a priori pairs, $t(15) = 0.9$, *ns*, but the a posteriori pairs yielded inflated JOLs, $t(15) = 3.79$, $p < .005$, as did the unrelated pairs, $t(15) = 3.77$, $p < .005$.

In sum, JOLs were well calibrated for the a priori pairs, whereas the purely a posteriori pairs produced a marked illusion of knowing. The unrelated pairs also yielded an illusion of knowing, consistent with the results of Experiments 1 and 2.

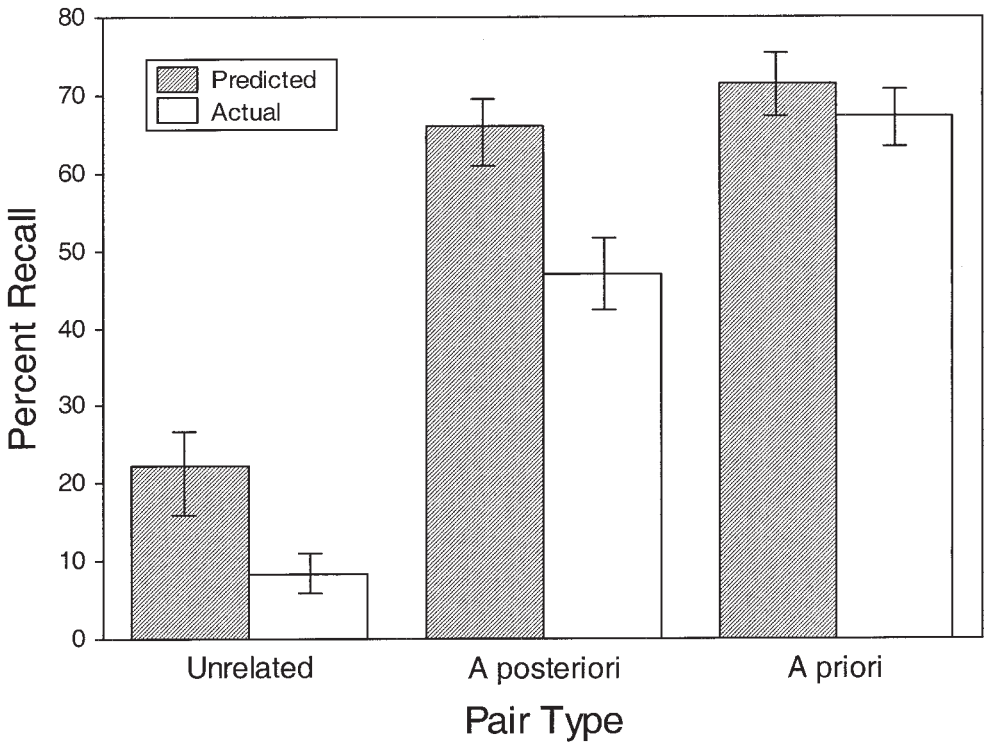

Figure 3. Mean predicted recall (judgments of learning) and actual recall for the unrelated, a posteriori, and a priori pairs (Experiment 3). Error bars represent ± 1 *SEM*.

Discussion

The results of Experiment 3 provide further support for the idea that perceived relatedness between the stimulus and response terms affects JOLs beyond whatever effects relatedness might have on actual recall. Both a priori and purely a posteriori pairs have a high degree of perceived association when both members of the pair are present. The difference between them is that in purely a posteriori pairs, the cue word, when presented alone, evokes other stronger associates that compete with the target word, not only preventing the target word from appearing in word association norms but also impairing the likelihood of its being produced on tests of cued recall.

General Discussion

In this study, we manipulated characteristics of to-be-learned materials—paired associates, in this case. The results from these manipulations suggest that when the to-be-learned materials trigger associations during study that are weak or absent during subsequent test, participants are prone to illusions of competence when predicting their own future recall.

It is important to stress that the overconfidence we observed is not simply a standard feature of JOLs. By and large, JOLs do not exhibit an overconfidence bias and, in fact, for many of the items used in this study, JOLs were very well calibrated. Thus, it is not the presence of the answer per se that produces overconfident JOLs but, rather, the presence of an answer that elicits a posteriori associations between cue and target that are inordinately strong relative to the a priori association between those words.

Present Processing and Future Performance

The selective occurrence of the overconfidence bias for such pairs reinforces our broader claim that the bias ensues from the tendency to overgeneralize from present processing to future processing. As Bjork (1999) noted, individuals are prone to interpret current performance as evidence of learning even when current performance is propped up by local conditions (such as massed or predictable practice) that will not be present on a later test of learning (e.g., Simon & Bjork, 2001).

In support of the idea that JOLs monitor aspects of current processing, Koriat, Bjork, Sheffer, and Bar (2004) have recently found that when participants expected a test either immediately after study, a day after study, or a week after study, their JOLs were completely indifferent to the expected retention interval, although actual recall exhibited a steep and typical forgetting function (see Carroll et al., 1997, for similar indifference of JOLs to retention interval). This pattern resulted in markedly inflated JOLs for a week's delay, in which participants predicted over 50% recall, whereas actual recall was less than 20%. In our view, the overconfidence demonstrated in the present study derives also from the tendency to rely on current processing, which in this case takes the form of giving undue weight to associations that are activated by the presence of the to-be-memorized target during encoding.

A unique feature of the overconfidence bias investigated in this study is that it is inherent in the learning process itself. Learning requires exposing learners to new information that they are expected to recall or use in the future. To the extent that they have to monitor the degree of mastery of each item during encoding (and, perhaps, to allocate learning resources accordingly), they should be prone to unwarranted high JOLs when the a posteriori associations activated by the item are inordinately strong.

Foresight Bias and Hindsight Bias

We have come to think of the overconfidence observed in the present study as resulting from a foresight bias that is related to, but that differs from, the extensively researched hindsight bias. The hindsight bias refers to the tendency to distort the memory of a previously made judgment after acquiring the correct answer (Fischhoff, 1975; for reviews, see Christensen-Szalanski & Willham, 1991; Hawkins & Hastie, 1990). We see the foresight bias as a kind of mirror image: Unlike the hindsight bias, which occurs when the recall of one's past answer is made in the presence of the correct answer, the foresight bias occurs when predictions about one's success in recalling the correct answer are made in the presence of that answer. Because both biases seem to reflect the failure to escape the influence of the correct answer, it is tempting to speculate that a similar mechanism underlies both biases. There are, however, important differences between the two biases. Chief among those, apart from the fact that one involves postdictions (hindsight) and the other involves predictions (foresight), is that the hindsight bias constitutes a memory distortion, whereas the foresight bias constitutes a metacognitive bias. Although there has been research documenting the influence of metacognitive judgments on the magnitude of the hindsight bias (Werth, Strack, & Förster, 2001), the foresight bias, as we have defined it, calls for a theoretical analysis within the framework of metacognition rather than within memory per se.

References

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.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55–68.

Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science, 14,* 283–286.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII—Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

Breznitz, S., & Ben-Dov, G. (1991). *Norms for word associations in Hebrew*. Unpublished manuscript, University of Haifa, Haifa, Israel.

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

Carroll, M., Nelson, T. O., & Kirwan, A. (1997). Tradeoff of semantic relatedness and degree of overlearning: Differential effects on metamemory and on long-term retention. *Acta Psychologica, 95,* 239–253.

Christensen-Szalanski, J. J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes, 48,* 147–168.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12,* 50–71.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1180–1191.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20,* 374–380.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33,* 545–565.

Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy? *Journal of Memory and Language, 36,* 34–49.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12,* 83–87.

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

Fischler, I. (1977). Associative facilitation without expectancy in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 18–26.

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

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17,* 209–225.

Hirshman, E. L., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 484–494.

Hutchison, K. (2003). Is semantic priming due to association strength or featural overlap? A "micro-analytic" review. *Psychonomic Bulletin & Review, 12,* 82–87.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35,* 157–175.

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science, 13,* 207–212.

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory & Cognition, 9,* 587–598.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology*: *General, 124,* 311–333.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349–370.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133,* 643–656.

Koriat, A., & Lieblich, I. (1977). A study of memory pointers. *Acta Psychologica, 41,* 151–164.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131,* 147–162.

Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1263–1274.

McClelland, A. G. R., & Bolger, F. (1994). The calibration of subjective probabilities: Theories and models 1980–94. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 453–482). Chichester, England: Wiley.

Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality and Social Psychology Review, 2,* 100–110.

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review, 105,* 299–324.

Nelson, D. L., & Zhang, N. (2000). The ties that bind what is known to the recall of what is new. *Psychonomic Bulletin & Review, 7,* 604–617.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science, 2,* 267–270.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods, 9,* 53–69.

Newton, L. (1990). *Overconfidence in the communication of intent: Heard and unheard melodies.* Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Palermo, D. S., & Jenkins, J. J. (1964). *Word association norms: Grade school through college.* Minneapolis, MN: University of Minnesota Press.

Pronin, E., Puccio, P., & Ross, L. (2002). Understanding misunderstanding: Social psychological perspectives. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 636–665). Cambridge, England: Cambridge University Press.

Rabinowitz, J. C., Ackerman, B. P., Craik, F. I. M., & Hinchley, J. L. (1982). Aging and metamemory: The roles of relatedness and imagery. *Journal of Gerontology, 37,* 688–695.

Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning (JOL) task. *Cognitive Development, 15,* 115–134.

Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging, 4,* 133–135.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 907–912.

Werth, L., Strack, F., & Förster, J. (2001). Certainty and uncertainty: The two faces of the hindsight bias. *Organizational Behavior and Human Decision Processes, 87,* 323–341.

Received December 10, 2001
Revision received September 14, 2004
Accepted September 15, 2004 ■