Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test

ASHER KORIAT
University of Haifa, Haifa, Israel

and

ROBERT A. BJORK
University of California, Los Angeles, California

Monitoring one's knowledge during study is susceptible to a *foresight bias* (Koriat & Bjork, 2005). Judgments of learning (JOLs) are inflated whenever information that is present at study and absent, but solicited, at test, such as the targets in cue–target paired associates, highlights aspects of cues that are less apparent when those cues are presented alone. The present findings demonstrate that foresight bias can be alleviated by study–test experience (Experiment 1), particularly test experience (Experiments 2 and 3), and by delaying JOLs after study (Experiment 4) and that both foresight bias and its alleviation have behavioral consequences, as measured by study time allocation (Experiment 5). Collectively, the findings suggest that overconfidence and misallocation of study time arise from a mismatch that is inherent to education—that the answer is present at study and absent at test—and that alleviating the problem requires creating conditions at study that sensitize learners to retrieval conditions at test.

Recently, we (Koriat & Bjork, 2005) described an illusion of competence that arises from what we termed a *foresight bias*. Such a bias, in our view, derives from an inherent discrepancy between the standard conditions of learning and the standard conditions of testing, a difference that seems innocuous but is the source of distorted judgments and overconfidence. On a typical memory test, people are presented with a question (e.g., "What is the capital of Belgium?") and are asked to produce the answer. In contrast, in the corresponding learning condition, both the question and the answer generally appear in conjunction (e.g., "The capital of Belgium is Brussels"), meaning that the assessment of one's future memory performance occurs in the presence of the answer. This difference has the potential of creating a *perspective bias*. That is, the learner, who needs to adopt the perspective of the examinee, may find it difficult to detach himself or herself from the perspective of the learner, because doing so requires discounting what he/she now knows. The failure to discount potential answers during learning results in a foresight bias—that is, unduly high predictions of one's future performance.

Several discussions in the literature have used the phrase *curse of knowledge* to capture the difficulty in discounting one's privileged knowledge and experience in judging what a more ignorant other knows or should know (Birch & Bloom, 2003; Camerer, Loewenstein, & Weber, 1989; Keysar & Henly, 2002; see Pronin, Puccio, & Ross, 2002, for a review). Participants, for example, tend to overestimate the effectiveness of their communications, assuming, more than is warranted, that the recipients will understand the intentions behind the message (Keysar & Henly, 2002; Newton, 1990). We argue that a similar curse of knowledge might underlie the unduly high expectations that students often hold about their future test performance (see Dunning, Johnson, Ehrlinger, & Kruger, 2003; Metcalfe, 1998).

How might such prediction inflation be remedied? In the next sections, we will describe our approach to that question; we will review some relevant findings, and we will lay out the theoretical background that motivates our investigation.

The Foresight Bias and Its Determinants

In our experiments, we used a simple paired-associates task to investigate the foresight bias hypothesis. Participants studied a list of paired associates and made judgments of learning (JOLs) after the study of each pair. Pre-

vious studies have established that a critical determinant of both JOLs and recall in this task is the degree of preexperimental association between the two words (see, e.g., Carroll, Nelson, & Kirwan, 1997; Connor, Dunlosky, & Hertzog, 1997; Dunlosky & Matvey, 2001; Koriat, 1997). As far as the foresight bias is concerned, however, a distinction must be drawn between two types of associative relations: a priori associations and a posteriori associations. A priori association refers to the probability with which the cue word, when presented alone, brings to mind the target word, as indexed by word association norms. A posteriori association, in contrast, refers to the perceived association between the cue and the target when *both* are present. Our argument is that a priori associations are the more critical for performance during cued recall when the cue (or question) appears alone but that a posteriori associations may have the greater influence on participants' predictions of their own future performance when both the cue (question) and the target (answer) are presented in conjunction. When the strength of a posteriori associations is high, relative to that of the corresponding a priori associations, as is often the case, participants will be susceptible to overconfidence—that is, to a foresight bias.

The critical difference between the two types of association is that potential associates of the cue *other than the one studied* are crucial in determining the degree of a priori relatedness but are immaterial for a posteriori relatedness (Koriat, 1981). Thus, the perceived a posteriori association between two words may sometimes be quite high when the strength of the a priori association is relatively low (even zero for certain *purely a posteriori* word pairs, such as *nurse–wife* or *citizen–tax*; see Koriat & Bjork, 2005). This discrepancy occurs if the target word, when presented together with the cue word, highlights aspects of the cue word that are not apparent when that word appears alone.

Consider, for instance, backward associations. There exist many asymmetrically associated word pairs for which the backward association, from the target to the cue, is strong, whereas the association from the cue to the target is relatively weak. According to D. L. Nelson, McEvoy, and Schreiber's (1998) norms, for example, the likelihood of *cheddar* eliciting *cheese* in the word association task is .92, whereas that of *cheese* eliciting *cheddar* is only .05. Backward-associated pairs (e.g., *cheese–cheddar*) are, therefore, expected to induce an illusion of competence, because the association from the target to the cue inflates the a posteriori relatedness, relative to the a priori relatedness. Indeed, using word pairs with asymmetrical associations, Koriat and Bjork (2005, Experiment 2) obtained results consistent with the foresight bias hypothesis: When these pairs were presented for study in a forward direction, so that the dominant association was from the cue to the target, mean JOLs were practically identical to mean recall (78.1% and 78.7%, respectively); when the words appeared in a backward direction, however, the respective means were 75.7% and 60.3%. In fact, direction of association had a nonsignificant effect on the participants' predictions of their later recall, even though it had a sizable effect on their actual recall.

Additional evidence consistent with the foresight bias hypothesis has come from the observation that pairs with a very weak a priori cue–target association tend also to produce inflated JOLs. Presumably, when both members of a pair are presented together during study, even a mild a priori association between them tends to be perceived as a moderate association. Thus, cue–target associative strength has been found to have a weaker impact on predicted than on actual recall (Koriat & Bjork, 2005, Experiment 1). Furthermore, pairs that were specially constructed to have *zero* a priori association but high a posteriori association (*purely a posteriori* pairs) yielded a marked illusion of competence (Experiment 3). Altogether, these results support the hypothesis that an illusion of competence is likely when a posteriori associations are strong, relative to a priori associations.

Assuming that learners are susceptible to illusions of competence that ensue from the foresight bias, how might such illusions be remedied? This question is important because illusions of competence may have serious consequences. A student who falls prey to a foresight bias in preparing for an exam, for example, is likely to stop studying prematurely, expecting to do better on the exam than he/she ends up doing. In certain job settings, such as air traffic control, nuclear plant operation, and police/military operations, where on-the-job mistakes can be truly disastrous, illusions of competence can have far more serious consequences. How, then, can such illusions of competence be overcome? To answer this question, we must first examine the factors that affect JOL accuracy.

The Determinants of JOL Accuracy

A commonly held assumption among metacognition researchers is that JOLs should be accurate to the extent that processing fluency at the time of making JOLs incorporates the same demands as the later criterion—retrieval (see Benjamin, Bjork, & Schwartz, 1998; Groninger, 1979; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Koriat & Ma'ayan, 2005). Conditions that produce an illusion of competence tend to involve different demands at study and test. Such was also the case for the conditions that we found to yield a foresight bias: They all involved associations that were activated at study but were absent at test. Such associations enhance the processing fluency of the pair as a whole, thus misleading the JOLs made at study.

To remedy the illusion of competence associated with the foresight bias, subjective experience must be educated by providing learners with mnemonic cues that are more pertinent to the testing situation, as opposed to those that are more specific to the study situation. A good example of how metacognitions may be remedied is provided by Benjamin's (2003) study (see also Guttentag & Carroll, 1998). As is well known (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989), participants falsely predict that they will recognize high-frequency words better than low-frequency words but postdict (correctly) the reverse pattern. However, in Benjamin's study, when participants were required to make postdictions after having had a

study–test opportunity (i.e., to judge whether they would have recognized rejected items if they had been presented in the study list), their predictions of future performance shifted toward superior performance on the low-frequency words.

In the present research, we explored two debiasing procedures, both of which we assumed would enhance the learner's sensitivity to mnemonic cues pertaining to retrieval fluency. The first involved providing the learner with test experience, similar to what was done in Benjamin's (2003) study. Several experiments have suggested that repeated testing improves JOL accuracy (e.g., Begg et al., 1989; King, Zechmeister, & Shaughnessy, 1980; Koriat, 1997; Leonesio & Nelson, 1990; Lovelace, 1984). Thus, in Experiment 1, we examined whether the presentation of a list for several study–test blocks can help reduce or eliminate the illusion of competence associated with the foresight bias. In Experiments 2 and 3, we compared the contributions of study and test opportunities to the alleviation of this illusion. If the foresight bias derives from the inherent discrepancy between the study and the test situations, test experience should be more effective in alleviating this bias than is study experience.

The second debiasing procedure we investigated (Experiment 4) rests on the delayed-JOL effect reported by T. O. Nelson and Dunlosky (1991; see also Dunlosky & Nelson, 1992, 1994). They found that JOLs (when prompted by the cue alone) were considerably more accurate when they were delayed until shortly after study than when they were made immediately after study. Arguably, the solicitation of JOLs some time after study, in the presence of the cue alone, makes the judgment situation more similar to the recall (test) situation (Weaver & Kelemen, 2003) and provides learners with a diagnostic cue—retrieval fluency—for discriminating between items that are more or less likely to be recalled. Indeed, Koriat and Ma'ayan (2005) reported evidence suggesting that whereas immediate JOLs are based on the feedback from encoding operations, delayed JOLs tend to be influenced primarily by the ease with which to-be-remembered items are accessed, and ease of access is a much more valid cue for subsequent recall when JOLs are delayed than when they are made immediately after study (see also T. O. Nelson, Narens, & Dunlosky, 2004). On the basis of these suggestive results, we hypothesized that delaying JOLs should help learners overcome the foresight bias that characterizes immediate JOLs.

Experiment 5, finally, extended investigation to the control function of metacognition (see T. O. Nelson & Narens, 1990) by attempting to show that the foresight bias has behavioral consequences for the strategic regulation of study time. Assuming that JOLs drive the allocation of study time during self-paced learning (see Nelson & Leonesio, 1988), the foresight bias should also be reflected in study time allocation. Assuming, further, that practice helps alleviate the foresight bias, the effects of such practice should also be reflected in changes of study time allocation.

EXPERIMENT 1

Experiment 1 tested the hypothesis that repeated study–test practice reduces the foresight bias. The materials were the same as those in Koriat and Bjork's (2005) Experiment 2—that is, pairs of English words with an asymmetric strength of forward and backward associations. Backward pairs, as has been indicated earlier, have been found to precipitate an illusion of competence. The question is whether the feedback inherent in a study–test cycle serves to inform participants' metacognitive assessment of the relative recallability of forward and backward pairs. We expected that there would be an illusion of competence for the backward pairs on the first study–test cycle but that this illusion should dissipate after the learners have had further opportunities to study and recall the list of items. In addition to reducing the overconfidence bias for backward pairs, increased practice should also improve JOL resolution in general—that is, accuracy in monitoring the relative recallability of different items.

Method

Participants. Twenty-four English-speaking undergraduates enrolled in the University of Haifa overseas program were paid for participating in the experiment.

Materials. The list was the same as that used in Experiment 2 of Koriat and Bjork (2005). Briefly, it included 24 word pairs with asymmetric associations, divided into two equal sets that were matched in terms of the strength of the forward and backward associations (the means for associative strength in the forward and the backward directions were .397 and .020, respectively, for one set and .396 and .021, respectively, for the second). One set was assigned to the forward condition, with the strongest association being from the cue word to the target word, and the other was assigned to the backward condition, with the assignment being counterbalanced across participants. In addition, the list included 24 unrelated pairs, each with zero associative strength.

Apparatus. The experiment was conducted on an IBM-compatible personal computer. In this experiment, as well as in Experiments 2–4, the stimuli were displayed on the computer screen, and JOLs and recalled responses were spoken orally by the participants and then entered by the experimenter on a keyboard.

Procedure. The experiment included four study–test cycles. The participants were instructed that they would have to study 48 paired associates and, for each pair, to assess the chances that they would be able to recall the target word in response to the cue word in a subsequent test, which would take place immediately after presentation of the whole list.

During the study phase, the two words appeared side by side for 5 sec and were replaced by the phrase "Probability to Recall." The participants then provided JOLs on a 0%–100% scale, expressing the probability of future recall, and the next pair was presented as soon as the experimenter recorded the response on the keyboard.

During the test phase, the 48 cue words were presented one after the other for 8 sec each. The participants had to say the response word aloud within the 8 sec allotted. The full study–test cycle was repeated three more times. The order of presentation of the pairs was randomly determined for each participant for each study and test phase.

Results and Discussion

The effects of direction of association on JOLs and recall during the first study–test cycle. We first will

report the results for the first study–test cycle, to verify that we replicated the foresight bias. The means for predicted and actual recall during the first study–test cycle are shown in Figure 1.

The results support the hypothesized foresight bias for the backward pairs. Although a strong effect of associative direction was found for actual recall, there was little or no corresponding effect on the participants' recall predictions. A two-way measure (JOL or recall) × associative direction (forward or backward) ANOVA yielded $F(1,23) = 3.27$, $MS_e = 390.08$, n.s., for measure; $F(1,23) = 41.12$, $MS_e = 77.87$, $p < .0001$, for direction; and $F(1,23) = 24.76$, $MS_e = 59.90$, $p < .0001$, for the interaction. For the forward pairs, the participants' predictions closely matched their actual recall [$t(23) = 0.17$, n.s.], whereas for the backward pairs, there was an illusion of competence, with predicted recall (74%) markedly inflated, in comparison with actual recall (58%) [$t(23) = 2.96$, $p < .01$].

In principle, the difference in the recall of forward and backward pairs on the first test could reflect, in part or in whole, guessing—because the response terms in the forward and backward pairs were the first associates of the corresponding stimulus terms 40% and 2% of the time, respectively. The participants were not forced to respond, however, and the details of their responding suggest that they withheld responding when no response came to mind. There were 42% omission responses for the unrelated items, for example, which suggests that the participants avoided guessing. The corresponding numbers for backward and forward pairs were 22% and 12%, respectively ($p < .001$). Assuming that omission responses reflect confidence (or lack thereof) in the responses that come to mind (Koriat & Goldsmith, 1996), it appears that forward and backward pairs induce similar metacognitive feelings during learning but give rise to different feelings during remembering.

The effects of practice on JOLs and recall. Did practice help in alleviating the foresight bias for the back-
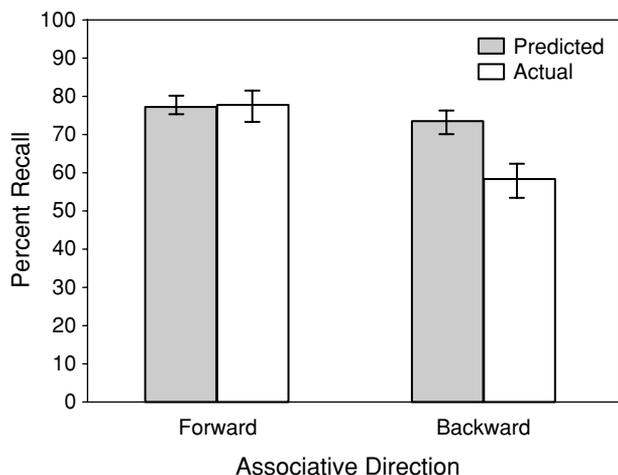

Figure 1. Experiment 1: Mean predicted recall (judgment of learning) and actual recall for the forward and backward pairs in Presentation 1. Error bars represent ±1 *SEM*.

ward pairs? Figure 2 (top panel) presents mean JOLs and recall for the backward and forward pairs as a function of presentation. A three-way ANOVA on these means yielded $F(1,23) = 6.72$, $MS_e = 330.16$, $p < .05$, for measure; $F(1,23) = 37.89$, $MS_e = 96.63$, $p < .0001$, for direction; and $F(3,69) = 81.19$, $MS_e = 131.67$, $p < .0001$, for presentation. The following interactions were also significant: measure × presentation, $F(3,69) = 16.77$, $MS_e = 120.88$, $p < .0001$; direction × presentation, $F(3,69) = 10.94$, $MS_e = 35.64$, $p < .0001$; direction × measure, $F(1,23) = 12.64$, $MS_e = 41.00$, $p < .01$; and the triple interaction $F(3,69) = 9.41$, $MS_e = 35.02$, $p < .0001$.

The measure × presentation interaction apparently derives from the underconfidence-with-practice (UWP) effect (Koriat, Sheffer, & Ma'ayan, 2002): Practice in studying a list of items increases underconfidence. This pattern is clearly seen in the results for the unrelated pairs (Figure 2, bottom panel). These pairs yielded a slight degree of overconfidence on the first presentation [$t(23) = 2.19$, $p < .05$], possibly because the learners could perceive some relationship between two words that, according to association norms, were unrelated (see Koriat & Bjork, 2005). The participants' initial overconfidence, however, gave way to a marked underconfidence from the second study–test cycle on: Whereas in the first presentation, JOLs and recall averaged 38.25 and 27.70, respectively, the corresponding means for the fourth presentation were 78.73 and 93.06, respectively. A presentation × measure ANOVA yielded $F(3,69) = 251.37$, $MS_e = 101.60$, $p < .0001$, for presentation; $F(1,23) = 9.58$, $MS_e = 490.76$, $p < .01$, for measure; and $F(3,69) = 28.95$, $MS_e = 78.22$, $p < .0001$, for the interaction.

The remaining interactions, all involving associative direction, seem to reflect the observation that the illusion of competence associated with the backward pairs decreased with practice to the extent that it disappeared completely by the fourth presentation. For that presentation, a two-way ANOVA yielded effects that were still significant for measure [$F(1,23) = 11.02$, $MS_e = 45.03$, $p < .01$] and direction [$F(1,23) = 8.57$, $MS_e = 17.41$, $p < .01$], but not for the interaction ($F < 1$). Because of the possibility of a ceiling effect on the fourth presentation (8 participants demonstrated perfect recall), the ANOVA was repeated for the third presentation, and here, too, the interaction was not significant ($F < 1$). As can be seen in Figure 2 (top panel), the triple interaction was mostly due to the results for the first two presentations. For these presentations, the relatively large improvement in the recall of backward pairs was not paralleled by a corresponding increase in JOLs. In sum, the results support the hypothesis that providing participants with further study–test opportunities helps reduce the foresight bias.

Changes in monitoring resolution with practice. Another way to examine the effects of practice on the extent of the foresight bias is to focus on resolution—that is, on the cross-item correspondence between JOLs and recall, as indexed by a within-participants gamma correlation (see T. O. Nelson, 1984). To compare resolution for the backward and the forward pairs, it was necessary
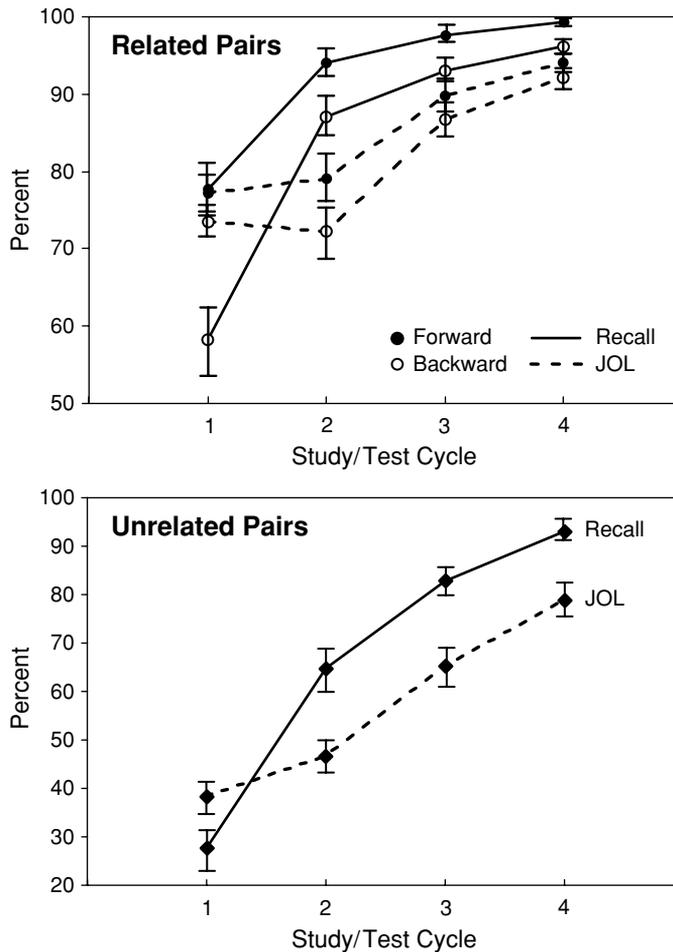
Figure 2. Experiment 1: Mean judgment of learning (JOL) and recall as a function of presentation. The top panel presents the results for the forward and backward pairs, whereas the bottom panel presents the results for the unrelated pairs. Error bars represent ±1 *SEM*.

to include the unrelated pairs in each analysis, in order to allow for greater variability in both recall and JOLs. Thus, gamma correlations were calculated twice for each presentation—first, by including only the forward and unrelated pairs, and second, by including only the backward and unrelated pairs. Because, as was noted above, several participants achieved perfect recall on the fourth presentation (rendering gamma noncomputable), the analyses were confined to the first three presentations (3 participants for whom one or more correlations were still indeterminate were excluded from the analyses).

The means of the correlations for the remaining participants are plotted in Figure 3. A presentation × associative direction ANOVA on the gamma values yielded $F(2,40) = 13.12$, $MS_e = 0.073$, $p < .0001$, for presentation; $F(1,20) = 18.60$, $MS_e = 0.011$, $p < .001$, for direction; and $F(2,40) = 2.18$, $MS_e = 0.0097$, $p < .15$, for the interaction.

These results disclose several trends. First, resolution improved with practice, consistent with previously re-

ported results (e.g., Begg et al., 1989; King et al., 1980; Koriat, 1997; Leonesio & Nelson, 1990; Lovelace, 1984). Second, resolution was generally poorer when the backward pairs were included in the analysis than when the forward pairs were included. Finally, there was a trend suggesting that the improvement in resolution as a result of practice was larger for the backward than for the forward pairs, which supports the idea that practice can remedy the poor resolution produced by the inflated JOLs associated with backward pairs.

EXPERIMENT 2

Experiment 2 was designed to evaluate the separate contributions of study and test experience toward reducing illusions of competence. According to the foresight bias hypothesis, this illusion derives from the inherent discrepancy between the learning and the testing situations. Hence, repeated tests should be particularly useful in providing learners with the kind of experience that is rela-
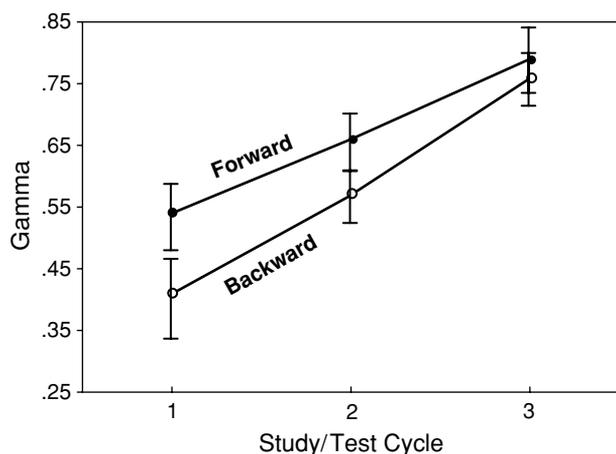
Figure 3. Experiment 1: Mean gamma correlations for the forward (+ unrelated) pairs and for the backward (+ unrelated) pairs as a function of presentation for Presentations 1–3. Error bars represent ±1 *SEM*.

tively protected from the misleading effects of a posteriori associations. Repeated study trials, in contrast, are less likely to do so, unless the participants engage in covert retrieval attempts. Indeed, previous studies suggest that repeated testing enhances JOL accuracy more than does repeated study (King et al., 1980). There is evidence, however, that prior study opportunities, even without explicit testing, also enhance JOL accuracy (Lovelace, 1984).

Experiment 2 used a between-participants design to compare a study condition and a test condition. The comparison of interest is shown below, where S denotes a study cycle, T a test cycle, and SJ a study cycle during which JOLs are solicited:

Study group: SJ – S – S – S – SJ – T

Test group: SJ – T – S – T – SJ – T.

Thus, the two groups differed only in the second and fourth cycles, during which the pairs were either presented intact for study (study group) or tested using cued recall (test group). A comparison of JOLs elicited in the first block with those elicited in the fifth block permits evaluation of the relative contribution of study and test experience to the alleviation of the foresight bias.

In order to generalize our conclusions to other item attributes that may also cause a foresight bias, the discrepancy between a priori and a posteriori associations was manipulated in Experiment 2 by using *purely a posteriori* pairs (Koriat & Bjork, 2005, Experiment 3). The study list included pairs with a moderate a priori cue–target association, plus two types of pairs with zero association according to the norms: purely a posteriori pairs, which appear related when shown together despite their having no a priori cue–target relationship, and completely unrelated pairs.

Experiment 2 was designed to test three hypotheses: (1) that the purely a posteriori pairs will exhibit exaggerated JOLs on the first study–test cycle, relative to the other pairs;

(2) that the foresight bias for the a posteriori pairs should be reduced by test practice; and (3) that study practice should be less effective than test practice in reducing that bias.

Method

Participants. Thirty-two Hebrew-speaking University of Haifa undergraduates participated in the experiment; 18 took part for course credit, and 14 were paid for their participation.

Materials and Apparatus. The list (in Hebrew) was the same as that used in Koriat and Bjork (2005, Experiment 3). It consisted of three classes of 24 pairs each: pairs with a moderate a priori association, purely a posteriori pairs, and unrelated pairs. The a priori pairs were selected from Hebrew word association norms so that the target word was a common response to the cue word (mean associative strength was .21). The 24 a posteriori pairs were selected so that they were judged by two judges to be semantically or associatively related but their a priori association, according to the Hebrew norms, was actually zero (e.g., *bed–night*, *clean–soap*, and *laugh–humor*). The 24 unrelated pairs had either zero association, according to the norms (no special effort was made to screen out pairs according to possible a posteriori associations), or in the absence of sufficient norms, consisted of words that were randomly paired but judged by the experimenter to be unrelated. The experiment was conducted on a Silicon Graphics personal computer.

Procedure. The procedure was the same as that in Experiment 1, except for the following. An SJ procedure was used in Blocks 1 and 5, in which each pair appeared on the screen for 2.5 sec and the participants reported JOLs at the end of each study trial. An S procedure was used in Block 3, in which each intact pair was presented for study for 2.5 sec, and a T procedure was used in Block 6, in which only the cue word was presented for a maximum of 6 sec and the participants had to say the response word within the 6 sec allotted. The procedure for Blocks 2 and 4 differed between the two groups. In the study group, an S procedure was used in both blocks (as in Block 3), whereas in the test group, a T procedure was used.

The participants were given specific instructions before each block, informing them about the type of task expected (S, SJ, or T). The order of presentation of the items was randomly determined for each participant for each block.

Results

Performance of the test group. Figure 4 presents the results for the test group, which can be used to address the first two hypotheses outlined earlier. With regard to the first hypothesis, which concerns the first block, JOLs were well calibrated for the a priori pairs but were inflated for both the purely a posteriori pairs and the unrelated pairs, consistent with previous results (Koriat & Bjork, 2005). Thus, a measure (JOL or recall) × pair type (a priori, a posteriori, or unrelated) ANOVA yielded $F(1,15) = 13.65$, $MS_e = 323.79$, $p < .005$, for measure; $F(2,30) = 186.42$, $MS_e = 137.70$, $p < .0001$, for pair type; and $F(2,30) = 7.81$, $MS_e = 89.78$, $p < .005$, for the interaction. A similar ANOVA including only the a priori and a posteriori pairs also yielded a significant interaction [$F(1,15) = 14.54$, $MS_e = 93.45$, $p < .005$]. For the a priori pairs, mean JOLs roughly matched mean recall [averaging 73.40 and 68.09, respectively; $t(15) = 1.23$, n.s.]. The a posteriori pairs, in contrast, yielded a marked illusion of competence: JOLs and recall averaged 67.16 and 43.42, respectively [$t(15) = 4.33$, $p < .001$]. The unrelated pairs also produced inflated JOLs [with JOL and recall averaging 21.66 and 9.99, respectively; $t(15) = 3.08$, $p < .01$].
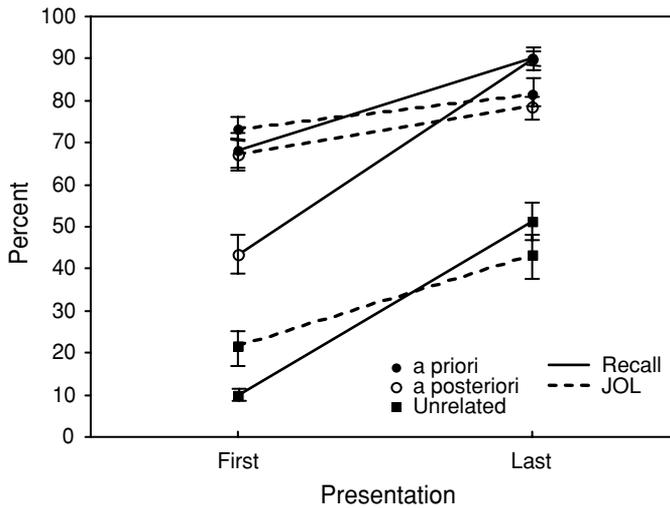
Figure 4. Experiment 2: Mean judgment of learning (JOL) and recall for the first two blocks and last two blocks of the test group. The results are plotted separately for the a priori, a posteriori, and unrelated pairs. Error bars represent ±1 *SEM*.

The second hypothesis, that test experience should reduce the illusion of competence for the a posteriori pairs, was tested by comparing the results for the first and the last study–test cycles (both involving an SJ–T block) of the test group. Focusing only on the a priori and a posteriori pairs (see Figure 4), a three-way measure (JOL or recall) × pair type (a priori or a posteriori) × block (first or last) ANOVA yielded $F < 1$ for measure; $F(1,15) = 22.95$, $MS_e = 101.55$, $p < .001$, for pair type; and $F(1,15) = 175.83$, $MS_e = 87.83$, $p < .0001$, for block.

Several interactions, however, were also significant, including the block × measure interaction [$F(1,15) = 33.12$, $MS_e = 144.40$, $p < .0001$], a result that is consistent with the UWP effect (see Experiment 1 and Koriat et al., 2002). More important, the triple interaction was also significant [$F(1,15) = 15.43$, $MS_e = 58.15$, $p < .005$]. Whereas in the first block the a posteriori pairs yielded inflated JOLs, no such effect was obtained for the last block, where both the a priori and the a posteriori pairs yielded a slight underconfidence of a similar magnitude. For this block, a measure (JOL or recall) × pair type (a priori or a posteriori) ANOVA yielded $F < 1$ for pair type; $F(1,15) = 13.49$, $MS_e = 116.96$, $p < .005$, for

measure; and $F < 1$ for the interaction. Thus, test experience helps remedy the foresight bias.

Comparing the effects of study and test experience. We will consider finally the last hypothesis, that study experience should be less effective than test experience in debiasing JOLs for the purely a posteriori pairs. This hypothesis could not be evaluated directly because, in the study group, recall was tested only in the last cycle, but some clues can be gained by comparing the study and the test groups on the last study–test cycle. Table 1 presents the pertinent means. As is apparent in Table 1, even though recall performance for the a posteriori pairs in the study condition was inferior to performance for the a priori pairs, JOLs were practically identical for the two types of pairs. The test condition, in contrast, did not produce such a foresight bias. Indeed, a two-way measure × pair type ANOVA yielded a significant interaction for the study group [$F(1,15) = 5.61$, $MS_e = 38.33$, $p < .05$], but not for the test group ($F < 1$). Although this pattern of results is consistent with the hypothesis that test opportunities are more effective than study opportunities in debiasing JOLs, it should be noted that the interactive pattern observed for the study group derived from differences in recall, rather than from differences in JOLs.

Table 1
Means and Standard Errors of the Means for JOL and Recall for the Last Study–Test Cycle for the Study and Test Groups, Presented Separately for the A Priori, A Posteriori, and Unrelated Pairs (Experiment 2)

| | Study Group | | | | Test Group | | | |
| | JOL | | Recall | | JOL | | Recall | |
Pair Type	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
A priori	79.1	4.6	93.5	1.4	81.5	3.3	90.1	1.7
A posteriori	77.8	4.3	84.9	3.2	78.5	2.8	89.8	2.7
Unrelated	54.3	5.1	53.2	5.6	43.4	5.0	51.2	4.6

Changes in monitoring resolution as a result of study and test experience. For the reason mentioned above, we also could not compare directly the effects of study and test experience on resolution. We should note, however, that for the test group, gamma correlations calculated across all the pairs improved from the first SJ–T block to the last SJ–T block, averaging .60 and .80, respectively [$t(15) = 4.11, p < .001$]. For the study group, the respective mean correlation for the last SJ–T block was only .59, significantly lower than what was obtained in that block for the test condition [$t(30) = 3.08, p < .005$] and at the same level as the corresponding correlation from the first SJ–T block in the test condition. This pattern suggests that monitoring resolution was enhanced by test experience, but not by study experience.

Discussion

The results of Experiment 2 yielded support for the three hypotheses that were tested. First, JOLs were markedly inflated for the purely a posteriori pairs in the first block of the test group. The inflation amounted to 24%, in comparison with only 5% for the a priori pairs. Second, consistent with the results of Experiment 1, the difference in the foresight bias between the a posteriori and the a priori pairs disappeared by the final study–test cycle, suggesting that study–test experience helps remedy the foresight bias. Finally, comparison between the study and the test conditions provides support for the idea that test experience alleviates the foresight bias more effectively than does study experience. In fact, monitoring accuracy in the last study–test cycle of the study group was no better than what was achieved by the test group on the first study–test cycle, a result that disagrees with Lovelace's (1984) finding (see also Connor et al., 1997) that study opportunities without explicit testing can enhance JOL accuracy.

EXPERIMENT 3

Experiment 3 was an attempt to replicate the results of Experiment 2 with a somewhat different procedure. First, to achieve some generalization, Experiment 3 used asymmetrically associated pairs, as in Experiment 1, but this time the pairs were in Hebrew. Second, because both JOLs and recall were quite high in the last study–test cycle of Experiment 2 (which may have masked possible effects of study experience), the intervening blocks between the initial (SJ) and final (SJ–T) cycles were reduced in Experiment 3 from three to only one. Accordingly, the design of Experiment 3 was as follows:

Study group: SJ – S – SJ – T

Test group: SJ – T – SJ – T.

Method

Participants. The participants were 48 Hebrew-speaking University of Haifa undergraduates; 13 participated for course credit, and 35 were paid.

Materials. The list included 54 Hebrew pairs with asymmetric associations. Of these, 27 pairs were presented in a forward direction, and 27 were presented in a backward direction. For the forward set, mean associative strength in the forward and backward directions was .616 and .013, respectively, whereas for the backward-presented set, these means were .631 and .029, respectively. In addition, 27 unrelated pairs, selected from the Breznitz and Ben-Dov (1991) norms to have zero associative strength, were included.

The compilation of the asymmetrical pairs was based on a preliminary norming study in which Hebrew-speaking participants were asked to write down the first word that came to mind in response to each of a series of stimulus words. The stimulus words had been selected so as to increase the likelihood of identifying asymmetrically associated pairs. Two samples of participants ($n = 63$ and $n = 60$, respectively), each participant receiving either 50 or 60 stimulus words, took part in the norming study.

Procedure and Apparatus. The procedure was the same as that in Experiment 2, except that Cycles 3 and 4 were eliminated. The experiment was conducted on an IBM-compatible personal computer.

Results

Table 2 presents mean JOLs and recall for the study and test groups for the forward, backward, and unrelated pairs for the first and final study–test cycles.

Performance on the first study–test cycle for the test group. The results for the first study–test cycle of the test group confirm the occurrence of the foresight bias for the Hebrew asymmetrical pairs. Whereas JOLs underestimated recall (by 13.0%) for the forward pairs, JOLs overestimated recall performance (by 6.3%) for the backward pairs. A measure (JOL or recall) $\times$ associative direction (forward or backward) ANOVA yielded a significant interaction [$F(1,23) = 46.29, MS_e = 48.30, p < .0001$]. Thus, the effects of associative direction were weaker for JOLs than for recall.

Comparing the effects of study and test experience on JOLs. For the study group, the effects of associative direction amounted to 4.4% on JOLs elicited during the first study cycle and remained about the same (4.3%) following study experience. For the test group, in contrast, this effect amounted to 6.3% on JOLs elicited during the first study–test cycle and increased to 10.1% following test experience. A three-way condition (study or test) $\times$ block (first or third) $\times$ associative direction (forward or backward) ANOVA of JOLs yielded a near-significant triple interaction [$F(1,46) = 3.96, MS_e = 11.89, p < .053$]. Separate two-way block (first or third) $\times$ associative direction (forward or backward) ANOVAs yielded a significant interaction for the test group [$F(1,23) = 5.86, MS_e = 15.21, p < .05$], but not for the study group ($F < 1$). Thus, test experience, but not study experience, increased JOLs' sensitivity to associative direction.

The foresight bias on the last study–test cycle. Focusing on the last study–test cycle, we examined whether test experience reduced the foresight bias more than did study experience. The effects of associative direction were similar for JOLs and recall in the test group, amounting to 10.1% and 11.8%, respectively, whereas in the study group, they were weaker for JOLs (4.3%) than for recall (8.7%). Two-way associative direction (forward or backward) $\times$ measure (JOL or recall) ANOVAs yielded a significant interaction for the study group [$F(1,23) = 4.65, MS_e = 25.28, p < .05$], but not for the test group ($F < 1$).

Table 2
Means and Standard Errors of the Means for JOL and Recall for
the First and Last Study–Test Cycles for the Study and Test Groups,
Presented Separately for the Forward, Backward, and Unrelated Pairs
(Experiment 3)

| | First Cycle | | | | Last Cycle | | | |
| | JOL | | Recall | | JOL | | Recall | |
Pair Type	M	SEM	M	SEM	M	SEM	M	SEM
				Study Group				
Forward	75.8	2.9			79.4	2.4	94.5	1.1
Backward	71.5	3.1			75.1	2.4	85.9	2.5
Unrelated	27.0	3.0			38.6	3.6	51.1	3.9
				Test Group				
Forward	75.1	3.1	88.2	2.8	84.7	2.9	97.5	0.6
Backward	68.8	3.1	62.5	4.5	74.6	3.2	85.7	2.8
Unrelated	28.3	3.8	16.8	2.2	36.7	4.2	44.8	4.8

Changes in monitoring resolution as a result of study and test experience. For the test group, the JOL–recall gamma correlations averaged .64 for the first SJ–T cycle and .78 for the last cycle [$t(23) = 3.03, p < .01$]. For the study group, in contrast, the correlation for the last cycle was only .62, significantly lower than the respective correlation for the test group [$t(46) = 3.92, p < .001$] and similar to the corresponding correlation for the first SJ–T cycle in the test group.

Discussion

The results of Experiment 3 replicated the findings from Experiment 2, using different materials and a modified procedure. The foresight bias that was observed for the first study–test cycle of the test group decreased on the second cycle, and there was a suggestion that test experience was more effective than study experience in alleviating the inflated JOLs associated with the backward-associated pairs.

Did study experience also help alleviate the foresight bias? The results of Experiment 3, like those of Experiment 2, failed to support this possibility. This failure is consistent with the idea that the foresight bias derives from the inherent discrepancy between the study and the test situations and that the key to reducing this bias lies in providing learners with test opportunities that help them overcome the contaminating effects of a posteriori associations that are activated by the presence of the target.

EXPERIMENT 4

In Experiment 4, we examined a second procedure that has the potential of reducing foresight bias: soliciting JOLs a few trials after study, rather than immediately after study. T. O. Nelson and Dunlosky (1991; see also Dunlosky & Nelson, 1992, 1994) found that delayed JOLs were considerably more accurate in predicting subsequent recall than were immediate JOLs. The opportunity for covert recall induced by delayed JOLs, when prompted by the stimulus alone, apparently provides participants with a more valid cue for discriminating between items that are and are not likely to be recalled on the subsequent test.

Indeed, a recent study (Koriat & Ma'ayan, 2005) suggests that (1) as the solicitation of JOLs is increasingly delayed, a shift occurs in the basis of JOLs from reliance on encoding fluency (the ease with which an item is committed to memory) toward greater reliance on retrieval fluency (the ease with which the target comes to mind in response to the cue) and (2) retrieval fluency after some delay is a much better predictor of subsequent recall than is encoding fluency.

In addition, as has been argued by Spellman and Bjork (1992; see also Kimball & Metcalfe, 2003), the covert retrieval of the associated response on JOL trials prompted by the stimulus alone is not only more diagnostic when such trials are delayed (see Koriat & Ma'ayan, 2005), but also more consequential. It is more consequential because the act of covert retrieval is itself a potent learning event, making a given response more recallable on the final test than it would have been otherwise, but only if the covert retrieval process is difficult enough to exercise the retrieval processes that will be required at the time of the final cued recall test. We expected, therefore, that delaying JOLs prompted by the stimulus alone might reduce the *curse of knowledge*, helping learners mitigate the effects of inflated a posteriori associations that stem from the presence of the target during study.

In Experiment 4, the participants studied forward-related, backward-related, and unrelated word pairs, and for half of the items in each class, they provided JOLs immediately after studying each pair, whereas for the other half, they made JOLs after some delay.

Method

Participants. Forty Hebrew-speaking undergraduates at the University of Haifa (32 of them female and 8 male) were paid for participating in the experiment.

Materials and Apparatus. A new list of 60 word pairs with unidirectional association was compiled from Hebrew word association norms for college students collected by Rubinsten, Anaki, Henik, Drori, and Faran (2005).[1] These pairs were divided into two equal sets that were matched in terms of the strength of the forward and backward associations. The means of associative strength in the forward and the backward directions were .40 and .04, respectively, for Set A and .40 and .03, respectively, for Set B. One set was assigned

to the forward direction, and the other was assigned to the backward direction, with the assignment being counterbalanced across participants. In addition, 30 unrelated pairs (zero associative strength) were included. The experiment was conducted on a PC.

Procedure. The 90 pairs were ordered randomly for each participant, with the restriction that each set of 30 successive pairs included 10 forward, 10 backward, and 10 unrelated pairs. Of these, 5 pairs of each type were assigned to the immediate-JOL condition, and the remaining pairs were assigned to the delayed-JOL condition.

The instructions for the study phase were similar to those used in the previous experiments, but the participants were told about the difference between immediate and delayed JOLs. Each study trial began with a cross at the center of the screen, accompanied by a beep. The cross (for 500 msec) was replaced by a presentation of the cue–target pair for 2.5 sec. For the immediate-JOL condition, after the disappearance of the pair, only the cue word was shown, together with a JOL prompt that appeared at the bottom of the screen: "Probability to recall (0%–100%)?" The delayed-JOL pairs, in contrast, were followed simply by the next trial. For these pairs, the JOL prompt appeared only after all 30 pairs in a block had been presented: The cue word was shown together with the JOL prompt. The order of JOL elicitation for the latter pairs was such that the stimulus word for the first 5 pairs studied (in a block of 30) appeared first in random order, then those of the next 5 pairs, and so on.

In the test phase, the 90 cue words appeared in a random order, and the participants had to say aloud the response word within 6 sec. This test phase was added only after the first 20 participants had been run, so that the test phase data are available only for 20 participants.

Results

We first will examine the foresight bias for the immediate-JOL condition, confining the analysis to the 20 participants for whom recall data were available. Immediate JOLs for the forward and backward pairs averaged 84.0 and 80.8, respectively, whereas the respective means for recall were 77.6 and 48.5. A measure (JOL or recall) $\times$ pair type (forward or backward) ANOVA yielded $F(1,19) = 29.88$, $MS_e = 251.07$, $p < .0001$, for measure; $F(1,19) = 42.61$, $MS_e = 122.45$, $p < .0001$, for pair type; and $F(1,19) = 30.81$, $MS_e = 108.73$, $p < .0001$, for the interaction. Thus, JOLs were relatively well calibrated for the forward pairs [$t(19) = 1.98$, n.s.], but were markedly inflated for the backward pairs [$t(19) = 6.41$, $p < .0001$].

A similar analysis was carried out for delayed JOLs. Delayed JOLs averaged 79.9 and 68.7 for the forward and backward pairs, respectively, whereas the respective means for recall were 75.6 and 56.7. A measure (delayed JOL or recall) $\times$ pair type (forward or backward) ANOVA yielded $F(1,19) = 5.98$, $MS_e = 222.97$, $p < .05$, for measure; $F(1,19) = 22.89$, $MS_e = 199.06$, $p < .0001$, for pair type; and $F(1,19) = 6.58$, $MS_e = 45.61$, $p < .05$, for the interaction. Thus, delayed JOLs also evidenced a weaker effect of associative direction on JOLs than they did on recall.

With respect to the *consequences* of delayed- versus immediate-JOL trials, it is of interest to note that for the backward pairs, final cued recall was higher in the delayed-JOL condition than in the immediate-JOL condition (56.7 vs. 48.5). The target response was clearly more recallable on immediate- than on delayed-JOL trials, meaning that the participants gave themselves more "presentations" of the cue–target pairing in the immediate case, but the

more difficult covert recall in the delayed case, when it succeeded, apparently had much greater (positive) consequences on the test of final cued recall—that is, was a more potent learning event.

Although delaying JOLs did not eliminate the foresight bias altogether, it did reduce it markedly: Whereas immediate JOLs yielded an overconfidence bias amounting to 6.4% and 32.3% for the forward and backward pairs, respectively, the respective values for delayed JOLs were 4.3% and 12.0%. A three-way condition (immediate or delayed) $\times$ measure (JOL or recall) $\times$ associative direction (forward or backward) ANOVA yielded a significant triple interaction [$F(1,19) = 9.92$, $MS_e = 82.88$, $p < .01$].

The effects of delaying JOLs can be seen in Figure 5, which presents the results for JOLs across all 40 participants. Delayed JOLs were overall lower than immediate JOLs (see also Koriat & Shitzer-Reichert, 2002), but the effects of associative direction were, indeed, stronger for delayed JOLs than for immediate JOLs, mimicking the effect typically found for recall. A condition (immediate or delayed) $\times$ pair type (forward or backward) ANOVA yielded $F(1,39) = 8.68$, $MS_e = 154.25$, $p < .01$, for condition; $F(1,39) = 25.75$, $MS_e = 64.91$, $p < .0001$, for pair type; and $F(1,39) = 4.26$, $MS_e = 56.79$, $p < .05$, for the interaction.

Previous work on the delayed-JOL effect has focused primarily on resolution, rather than on calibration, and has documented that delaying JOLs improves discrimination between recalled and unrecalled items. The results of Experiment 4 fit this pattern. Two JOL–recall gamma correlations were calculated for each participant, one including
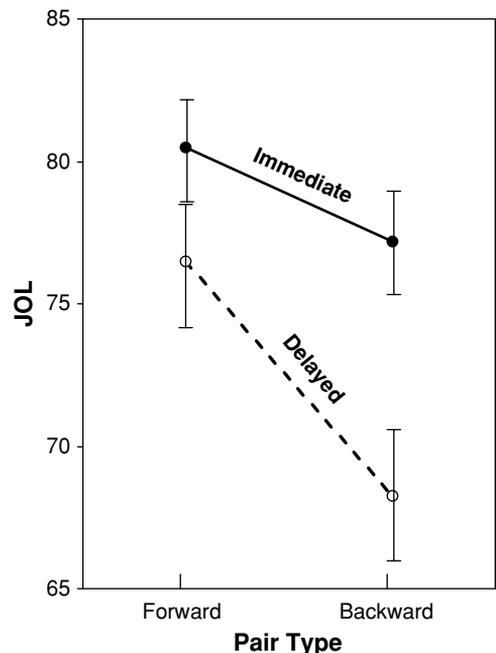

Figure 5. Experiment 4: Mean immediate and delayed judgments of learning (JOLs) for the forward and backward pairs. Error bars represent ±1 *SEM*.

only immediate-JOL items and the other including only delayed-JOL items (45 items in each case). These correlations averaged .64 and .86, respectively, across participants [$t(19) = 4.23$, $p < .001$].

Discussion

Consistent with our hypothesis, delaying JOLs alleviated foresight bias, although not entirely. Presumably, in comparison with immediate JOLs, delayed JOLs prompted by the cue alone provide participants with cues that are more pertinent to the test situation—especially, perhaps, a lesser ability to recall (covertly) the response members of the backward pairs than those of the forward pairs—thus reducing the contaminating effects of a posteriori associations.

EXPERIMENT 5

Finally, Experiment 5 extended the investigation to the control functions of metacognition. It examined whether the foresight bias and its alleviation by increased practice are reflected in the strategic allocation of study time to different items in the list.

A common assumption among students of metacognition is that monitoring informs and drives control operations (Koriat & Goldsmith, 1996; T. O. Nelson & Narens, 1990). Hence, the foresight bias may be expected to have detrimental effects on the strategic regulation of study time under self-paced learning conditions, causing learners to allocate insufficient study time to items with inflated a posteriori associations. Thus, from that perspective, Experiment 5 represents an attempt to demonstrate that the foresight bias and its alleviation as a result of practice (Experiment 1) are not mere epiphenomena but actually influence the allocation of study time between items.

The design of the experiment was similar to that in Experiment 1, except that study time, not JOL, was the measure of interest. We expected the effects of associative direction to mimic those found for JOLs in Experiment 1: On the first presentation of the list, the learners should spend roughly the same amount of time studying the backward and the forward pairs, owing to foresight bias, but with increased practice, they should allocate relatively more study time to the backward pairs.

Method

Participants. The participants were 24 English-speaking undergraduates enrolled in the University of Haifa overseas program. They were paid for their participation.

Materials. The same list of stimuli as that in Experiment 1 was used, with the exception that 16 filler pairs were added in order to reduce the likelihood of a ceiling effect attributable to the self-paced study procedure. Eight of the filler pairs were unrelated, and 8 were related.

Apparatus and Procedure. The apparatus was the same as that in Experiment 2. The stimuli were displayed on the computer screen, and recall scoring was entered by the experimenter.

The procedure was identical to that in Experiment 1, with two exceptions. First, a self-paced procedure was used: The participants were told that they should study each item for as long as they needed and that, when they were through studying, they should press the left key of the mouse, at which time the next pair would appear on the screen. The participants were instructed to invest exactly as much time as they needed for studying each word pair, not more and not less. They were also instructed that their success in performing the task would depend on their success in recalling as many words as possible during testing while keeping the total time invested in studying the entire list as short as possible. They were informed that the list included 64 paired associates and that the test involved cued recall.

Second, no JOLs were solicited. Thus, during the study phase, each word pair remained on the screen until the left key of the mouse was pressed, and 1 sec thereafter, the next pair was shown. The procedure for the test phase was the same as that in Experiment 1.

Results and Discussion

Effects of associative direction on study time and recall during the first cycle. Assuming that study time was affected by the participants' JOLs, study time should be relatively insensitive to the direction of association on the first study trial, whereas recall testing should evidence better memory for the forward than for the backward pairs. Indeed, this is precisely the pattern found. As can be seen (Figure 6, top panel), mean study time was practically identical for the forward and the backward pairs on the first presentation, and both means were substantially lower than the mean time spent studying the unrelated pairs. Memory performance, in contrast, as shown in the bottom panel, yielded a different pattern, with performance on the backward pairs markedly inferior to performance on the forward pairs. In fact, it is striking to note that study times for the backward pairs were similar to the study times for the forward pairs, whereas the recall levels for the backward pairs were similar to the recall levels for the unrelated pairs.

Several analyses confirmed these impressions. An overall two-way pair type (forward, backward, or unrelated) × measure (study time or recall) ANOVA yielded a significant interaction [$F(2,46) = 57.43$, $MS_e = 71.78$, $p < .0001$]. A t test confirmed that less time was spent studying the backward pairs than studying the unrelated pairs [$t(23) = 6.08$, $p < .0001$] but that there was no difference between the amount of time spent studying the backward and the forward pairs [$t(23) = 0.30$, n.s.]. As far as recall is concerned, however, the backward pairs yielded memory performance inferior to that for the forward pairs [$t(23) = 6.50$, $p < .0001$]. Recall was still slightly but significantly better for the backward pairs than for the unrelated pairs [$t(23) = 3.14$, $p < .01$].

The effects of practice on study time and recall. Given that the backward pairs received unduly short study times on the first presentation, we can now examine whether study–test practice helped the participants adjust the allocation of study time to reflect the relatively low recall of the backward pairs.

As can be seen in the top panel of Figure 6, study time decreased with additional presentations for all pair types. After the first presentation, however, relatively more time was allocated to the study of backward pairs than to the study of the forward pairs. Considering only the data from the first two presentations, a presentation × pair type (forward or backward) ANOVA yielded $F(1,23) = 2.91$, $MS_e = 11.15$, $p < .11$, for presentation; $F(1,23) =$
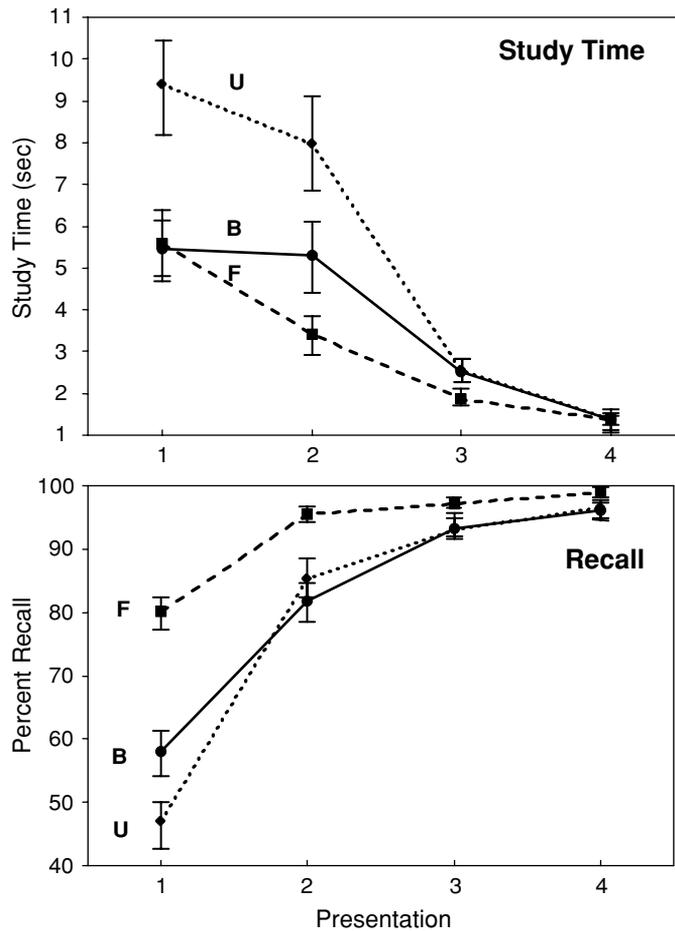
Figure 6. Experiment 5: Mean study time (top panel) and recall (bottom panel) for the forward (F), backward (B), and unrelated (U) pairs as a function of presentation. Error bars represent ±1 *SEM*.

11.71, $MS_e = 1.63$, $p < .01$, for pair type; and $F(1,23) = 8.71$, $MS_e = 2.81$, $p < .01$, for the interaction. A similar ANOVA comparing backward and unrelated pairs yielded $F < 1$ for presentation; $F(1,23) = 32.49$, $MS_e = 8.01$, $p < .0001$, for pair type; and $F(1,23) = 2.64$, $MS_e = 3.77$, $p < .12$, for the interaction. These results suggest a shift in study time allocation from the first to the second presentation in the direction of increased emphasis on the backward pairs, relative to the other pair types.

The recall data indicate that the backward pairs not only continued to exhibit inferior recall, relative to the forward pairs across presentations, but also were no better recalled than were unrelated pairs. Considering only Presentations 2–4, a presentation × pair type ANOVA comparing recall for backward and forward pairs yielded $F(1,23) = 15.81$, $MS_e = 104.57$, $p < .001$, for pair type. On the other hand, a similar ANOVA comparing backward and unrelated pairs yielded $F < 1$ for pair type. Thus, the allocation of relatively more study time to the backward pairs did not compensate for their low recall. This finding is consistent, perhaps, with the labor-in-vain effect (Mazzoni & Cornoldi, 1993; Nelson & Leonesio, 1988), which refers

to the observation that large increases in self-paced study time can yield little or no increase in the subsequent likelihood of recall.

GENERAL DISCUSSION

Illusions of competence are quite common in everyday life. Several studies (e.g., Dunning et al., 2003; Metcalfe, 1998) have established that people have little insight into their limitations and incompetence and that they tend to overestimate their expertise and talent. Such illusions can result in frustrations and disappointment, as when a student receives a grade on an exam that is much lower than he or she expected. They can also have serious consequences in real-world job contexts, where mistakes can cause great harm to one's self and others. An important challenge, then, is to find ways to eliminate or reduce such illusions.

In the present research, we focused on one mechanism that is a potential contributor to inflated predictions of one's own future memory performance. This mechanism, which we have labeled *foresight bias*, derives from an

inherent discrepancy between study and test situations: When people assess their future recall during study, they do so in the presence of information that not only will be absent during testing, but also they will be required to produce. The failure to discount that information during study may lead to a foresight bias—inflated predictions of one's future memory. The present results demonstrate some of the ways in which such illusions of competence can be reduced.

Conditions That Produce a Foresight Bias

It is important to emphasize that for many of the items used in these and related experiments, JOLs were relatively well calibrated. JOLs are not routinely inflated by the presence of an answer but, rather, are inflated only when the presence of the answer (target) brings to the fore aspects of the cue that are less likely to come forward when the cue appears alone. That is, the tendency to overestimate one's own future performance does not occur across the board but is confined to a circumscribed set of items. It is not the presence of the answer per se that produces inflated JOLs; rather, they are produced when the a posteriori associations activated by the intact cue–target pair are inordinately strong, relative to the a priori associations from the cue to the target—creating, perhaps, an unwarranted sense that the answer is "natural" or "obvious."

The Foresight Bias and Its Amelioration

The present results also support the contention that the foresight bias derives specifically from the discrepancy between the study and the test situations. What the debiasing procedures explored in this study had in common is that they provided the participants, when making recall predictions, with cues that were pertinent to the testing situation. It is of interest to note, however, that the procedure used in Experiments 2 and 3, unlike the delayed-JOL procedure in Experiment 4, required that JOLs in the final blocks be solicited following a study trial (i.e., in the presence of the to-be-remembered target). Why, then, did study experience in these blocks not wipe out the effects of previous testing? This question, of course, can also be raised with regard to previous results demonstrating a beneficial effect of testing opportunities on monitoring (e.g., Benjamin, 2003; Guttentag & Carroll, 1998; King et al., 1980). Although it *does* appear that test experience carries over to a study situation in which both the cue and the target appear together, it may also be the case that the methodological constraint of having to elicit JOLs in the typical study situation (in response to the cue–target pair) results in an underestimation of the potential effects of test experience. Such an underestimation may explain why the observed differences between the study and the test groups in Experiments 2 and 3 were not very strong.

This comment brings forward a fundamental dilemma. As was indicated earlier, the foresight bias is inherent in the learning process, meaning that its debiasing must necessarily be limited. That is, it is not reasonable to expect a learner to both memorize a new piece of information and, at the same time, discount completely the activations evoked by that information. The tasks of memorizing new information and of monitoring its future recallability seem, in that sense, to pose conflicting demands on the learner.

The Control Function of Metacognition

Finally, the potential consequences of the foresight bias on the strategic allocation of a learner's study time, as demonstrated by the results of Experiment 5, merit comment. Such results are consistent with the view of metacognition researchers (e.g., Koriat, 2000; T. O. Nelson, 1996) that the importance of metacognitive monitoring is that it exerts measurable effects on behavior. Monitoring errors and inaccuracies may, therefore, have serious consequences. Experiment 5, for example, showed that foresight bias, indeed, had its expected consequences on the amount of time allocated to different items: During the first presentation of forward and backward pairs, about the same amount of study time was allocated to each, despite the fact that backward pairs consistently yield inferior recall, relative to forward pairs.

It is also instructive that the recall of backward pairs in Experiment 5 was about the same as the recall of unrelated pairs but more study time was allocated to the unrelated pairs. The allocation of more study time to the unrelated pairs suggests that the participants *did* adopt a strategy of allocating more time to more difficult pairs, perhaps to compensate for their difficulty (Nelson & Leonesio, 1988; Son & Metcalfe, 2000). Thus, had the participants been able to discount the effects of a posteriori associations, they would have been expected to compensate also for the difficulty of the backward pairs by investing more study time in these pairs than they did in the forward pairs. It is encouraging, though, that with repeated study–test practice, relatively more study time was allocated to the backward pairs, suggesting that debiasing procedures actually *can* remedy ineffective strategic regulation of learning resources, at least to some degree.

CONCLUDING COMMENT

The present findings add to the growing body of research on the nature and consequences of metacognitive monitoring. The picture that emerges is that the reading that learners take of the degree to which they have or have not acquired skills and knowledge has multiple consequences and is, arguably, as important as the actual level of skill or knowledge that they have attained. Broadly, these findings suggest that an important key to optimizing both individual learning and formal instruction is metacognitive sophistication. Both the monitoring and control functions of metacognition are central to how effectively information is studied and learned, but as is illustrated by the present results and multiple other results in the literature, one's subjective experience can be interpreted in ways that are misguided and the source of illusions of comprehension.

REFERENCES

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

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, **31**, 297-305.

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.

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

Camerer, C., Loewenstein, 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.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology & 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, & 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 & Language*, **33**, 545-565.

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.

Groninger, L. D. (1979). Predicting recall: The "feeling-that-I-will-know" phenomenon. *American Journal of Psychology*, **92**, 45-58.

Guttentag, R., & Carroll, D. (1998). Memorability judgments for high- and low-frequency words. *Memory & Cognition*, **26**, 951-958.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 22-34.

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

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition*, **31**, 918-929.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, **93**, 329-343.

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

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. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness & Cognition*, **9**, 149-171.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 187-194.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, **103**, 490-517.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory & Language*, **52**, 478-492.

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.

Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy: Insights from the processes underlying judgments of learning in children. In P. Chambres, M. Izaute, & P.-J. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 1-17). New York: Kluwer.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 464-470.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 756-766.

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General*, **122**, 47-60.

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

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme and word fragment norms*. Available at w3.usf.edu/FreeAssociation/Intro.html.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, **51**, 102-116.

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., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 676-686.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation*: *Advances in research and theory* (Vol. 26, pp. 125-173). San Diego: Academic Press.

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.

Pronin, E., Puccio, C., & Ross, L. (2002). Understanding misunderstanding: Social psychological perspectives. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 636-665). Cambridge: Cambridge University Press.

Rubinsten, O., Anaki, D., Henik, A., Drori, S., & Faran, Y. (2005). Norms for free associations in the Hebrew language. In A. Henik, O. Rubinsten, & D. Anaki (Eds.), *Word norms for the Hebrew language* [in Hebrew] (pp. 17-34). Beer-Sheva: Ben-Gurion University of the Negev.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 204-221.

Spellman, B. A., & Bjork, R. A. (1992). Technical commentary: When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, **3**, 315-316.

Weaver, C. A., III, & Kelemen, W. L. (2003). Processing similarity does not improve metamemory: Evidence against transfer-appropriate monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1058-1065.

NOTE

1. These norms became available in the course of this study.