Journal of Experimental Psychology: General
1997, Vol. 126, No. 4, 349–370

0096-3445/97/$3.00

Monitoring One's Own Knowledge During Study: A Cue-Utilization Approach to Judgments of Learning

Asher Koriat
University of Haifa

How do people monitor their knowledge during acquisition? A cue-utilization approach to judgments of learning (JOLs) is outlined, distinguishing 3 types of cues for JOLs: intrinsic, extrinsic, and mnemonic. In 4 experiments using paired-associates learning, item difficulty (intrinsic) exerted similar effects of JOLs and recall. In contrast, the extrinsic factors of list repetition, item repetition within a list, and stimulus duration affected JOLs less strongly than recall, supporting the proposition that extrinsic factors are discounted in making JOLs. Although practice impaired calibration, increasing underconfidence, it did improve resolution (i.e., the recall–JOL correlation). This improvement was seen to reflect a shift in the basis of JOLs with practice, from reliance on intrinsic factors, towards greater reliance on mnemonic-based heuristics.

There has been a great deal of interest in recent years in the metacognitive processes that supervise and control various aspects of information processing and behavior (see Metcalfe, 1996; Metcalfe & Shimamura, 1994; Nelson, 1996; Reder, 1996; Schwartz, 1994). This interest spans across different areas of psychology, including memory research (e.g., Johnson, Kounios, & Reeder, 1994; Koriat, 1993; Koriat & Goldsmith, 1996; Metcalfe, 1993; Nelson & Narens, 1990), developmental psychology (Butterfield, Nelson, & Peck, 1988; Schneider, 1985), aging research (Bäckman & Lipinska, 1993; Burke, MacKay, Worthley, & Wade, 1991; Koriat, Ben-Zur, & Sheffer, 1988), neuropsychology (Funnell, Metcalfe, & Tsapkini, 1996; Janowsky, Shimamura, & Squire, 1989; Schacter, 1991; Shimamura & Squire, 1986; Sohlberg, Mateer, & Stuss, 1993), social psychology (Ross, 1997), decision making (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Keren, 1991), and forensic psychology (Luus & Wells, 1994; Sporer, Penrod, Read, & Cutler, 1995). A great deal of research has been carried out on several types of metacognitive judgments that mediate performance during learning and remembering. These include ease of learning judgments elicited before study (e.g., Leonesio & Nelson, 1990), judgments of learning made during study (e.g., Mazzoni & Cornoldi, 1993), feeling of knowing judgments made before or during the attempt to retrieve information from memory (e.g., Koriat, 1995; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992) and subjective confidence in the correctness of the retrieved information (Juslin, 1994; Kelley & Lindsay, 1993; Koriat & Goldsmith, 1996). Among the issues addressed are how these metacognitive judgments are formed, how valid or accurate they are in mirroring actual knowledge, how they govern strategy choice, and what are their eventual effects on performance. One possible reason for the upsurge of interest in the area of metacognition is, perhaps, that research in this area may have some bearing on the long-standing issue of the relationship between consciousness and behavior (see Barnes, Nelson, Dunlosky, Mazzoni, & Narens, in press; Koriat, in press; Nelson, 1996).

This article is specifically concerned with judgments of learning (JOLs), that is, judgments made by participants at the end of a learning trial regarding the likelihood of remembering the acquired information on a subsequent memory test. In a typical experiment on JOLs, participants are instructed to memorize a list of paired associates as preparation to recall the target word (*response term*) when presented with the cue word (*stimulus term*). Following each study trial, they are asked to make predictions regarding the likelihood of recalling the response term during test. In the test phase the participants are presented with each of the cue words in turn, and their recall of the corresponding target words is tested. In several experiments using this procedure, participants have been found to be moderately accurate in their predictions (e.g., Arbuckle & Cuddy, 1969; Dunlosky & Nelson, 1994; Lovelace, 1984; Mazzoni & Nelson, 1995; Zechmeister & Shaughnessy, 1980). Furthermore, the results suggest that participants allocate study time in accordance with their JOLs, spending more time studying those items that are judged to be less likely to be remembered (Mazzoni & Cornoldi, 1993; Mazzoni, Cornoldi, & Marchitelli, 1990; Nelson & Leonesio, 1988).

Two questions about JOLs suggest themselves: (a) What is the basis of JOLs? (b) What is the reason for the accuracy of JOLs in predicting actual memory performance? With regard to the first question, one simple hypothesis is the *direct-access hypothesis* (see King, Zechmeister, & Shaughnessy, 1980): People assess the future recallability of an item by reading the strength of the memory trace that is formed

The experiments were conducted at the Institute of Information Processing and Decision Making, University of Haifa. I am grateful to Ravit Levy for her help in all stages of the research program, to Limor Sheffer for the analyses of the data, and to Galit Aharonovich and Rachel Shitzer for conducting the experiments.

Correspondence concerning this article should be addressed to Asher Koriat, Department of Psychology, University of Haifa, Haifa 31905, Israel. Electronic mail may be sent via Internet to akoriat@psy.haifa.ac.il.

following study (see Arbuckle & Cuddy, 1969; Dunlosky & Nelson, 1994; Mazzoni & Nelson, 1995; Nelson & Narens, 1990). This hypothesis assumes that participants can monitor directly trace strength and can also assess on-line the moment-to-moment increase in trace strength that occurs as more time is spent studying an item. One implication of this view is that the effects of various factors on JOLs (e.g., number of study trials, encoding strategies, etc.) are mediated by the effects of these factors on degree of learning (see Figure 1a in Mazzoni & Nelson, 1995).

The direct-access hypothesis parallels a similar hypothesis that has been proposed with regard to feeling-of-knowing (FOK) judgments elicited following recall failure (see discussions in Koriat, 1993, 1994; Nelson & Narens, 1990; Schwartz, 1994). This hypothesis affords a straightforward answer to the second question about JOL: What is the reason for the accuracy of JOLs in predicting actual memory performance? According to this hypothesis, the accuracy of JOLs follows from the fact that both subjective and objective indexes of knowing are directly affected by the strength of the memory trace, and therefore it is only natural to expect a strong correspondence between them. Occasional findings of a weak correlation between predicted and actual recall performance, however, need not be necessarily fatal for the direct-access view, because recall can be assumed to depend on other factors besides the strength of the memory trace at the time of making JOLs (see Nelson & Dunlosky, 1991).

The direct-access view (or *target-strength view;* see Schwartz, 1994) and its implications can be illustrated by the following statement from Cohen, Sandler, and Keglevich (1991):

> Different words in a list receive encodings of different strengths, which result in item representations in memory of different strengths. These strength differences are detected by the subject and transformed into recall probability ratings. Because the probability of retrieval is related to the strength of item representations, these ratings have predictive power for subsequent free recall. (p. 524)

An alternative account of the basis of JOLs and their accuracy is provided by the cue-utilization view. This view, which is more congenial with some of the recent approaches in metacognition, assumes that JOLs, like other metacognitive judgments such as FOKs and subjective confidence, are inferential in nature: JOLs are based on the implicit application of rules or heuristics in order to achieve a reasonable assessment of the probability that the information in question will be recalled or recognized at some later time (see, e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989). Thus, in making judgments of learning or judgments of knowing, participants do not monitor directly the strength of the memory trace of the item in question, but use a variety of cues that are generally predictive of subsequent memory performance. These cues may include the person's belief about his or her general memory efficacy (see Hertzog, Dixon, & Hultsch, 1990), the perceived memorial consequences of various characteristics of the study situation such as number of study trials and encoding strategies (Begg et al., 1989; Zechmeister & Shaughnessy, 1980), the type of memory test expected (Mazzoni & Cornoldi, 1993), the previous task-specific experience (Hertzog et al., 1990; King et al., 1980; Mazzoni & Cornoldi, 1993; Schneider, 1986), the perceived relative difficulty of the study items in question (Arbuckle & Cuddy, 1969), and so on. Because JOLs are based on inferences and heuristics, their accuracy is not guaranteed but depends on the empirical correlation between the cues used and the criterion memory test. In general, JOLs are accurate as long as the cues used at the time of making the judgments are consistent with the factors that affect subsequent performance on the criterion memory test. This may be true by and large for many of the cues. However, in some situations the cues for JOLs may lack any predictive validity, in which case little correspondence would be expected between JOLs and memory performance. In other cases still, systematic discrepancies may be found between predicted and actual recall, either because the inferences drawn are based on an incorrect theory or belief or because of some of the biases inherent in the application of particular heuristics (see e.g., Heath et al., 1994).

The present work was conducted within the cue-utilization approach to JOL. It is based on the distinction between three general classes of cues for JOLs: intrinsic, extrinsic, and mnemonic. The first class includes cues involving characteristics of the study items that are perceived to disclose the items' a priori ease or difficulty of learning. Several inherent attributes of the studied material can serve as predictors of the preexperimental difficulty of learning or remembering different items. In the case of paired associates, for example, degree of associative relatedness between the members of a pair can serve as a powerful predictor of memory performance (e.g., Rabinowitz, Ackerman, Craik, & Hinchley, 1982). Similarly, in the case of single words, imagery value is a relatively effective diagnostic of memorability (Begg et al., 1989; Groninger, 1979).

Indeed, normative judgments of ease or difficulty of learning are predictive of the relative recallability of different items (e.g., Arbuckle & Cuddy, 1969; Begg et al., 1989; Leonesio & Nelson, 1990; Nelson & Leonesio, 1988; Underwood, 1966). For example, Nelson and Leonesio (1988) found a mean within-person correlation of about .50 between ease-of-learning judgments and recall for both a list of trigrams and a list of paired associates. Thus, a priori estimates of ease of learning can serve as effective predictors of recallability. In fact, JOLs of one participant have been found to be moderately predictive of the memory performance of another participant, possibly because of commonly shared difficulty characteristics of items (Lovelace, 1984; Lovelace & Marsh, 1985; Rabinowitz et al., 1982; Underwood, 1966).

The second class consists of extrinsic factors that pertain either (a) to the conditions of learning, or (b) to the encoding operations applied by the learner. The former category includes number of times an item has been studied (Lovelace, 1984; Zechmeister & Shaughnessy, 1980), presentation time (Mazzoni et al., 1990), massed versus distributed repetition of items (Dunlosky & Nelson, 1994; Zechmeister & Shaughnessy, 1980), and so on. The latter category includes encoding operations performed by the learner, such as level of processing (Rabinowitz et al., 1982; Shaw & Craik, 1989), interactive imagery (Begg et al., 1989; Begg, Vinski, Frankovich, & Holgate, 1991; Dunlosky & Nelson, 1994), and so on.

Both intrinsic and extrinsic factors can affect JOLs directly, through the explicit application of a particular rule or theory. For example, a person may deduce that a three-letter trigram is more likely to be recalled when it forms a word than when it is nonsense (e.g., Underwood, 1966) or that an item seen twice is more likely to be remembered than an item seen once (e.g., Zechmeister & Shaughnessy, 1980). However, these cues may also influence JOLs indirectly, through their effects on the third class of cues, to be considered now.

The third class comprises internal, mnemonic indicators that may signal for the participant the extent to which an item has been learned and will be recalled in the future. Recent discussions in the literature have emphasized the importance of internal, subjective cues as a source of information. It has been argued that the phenomenal experiences that accompany information processing serve as input for judgmental processes of many kinds (see e.g., Schwarz & Clore, 1996; Strack, 1992). In the case of metacognitive judgments such as FOK, JOL, or retrospective confidence, several mnemonic cues have been considered, including the accessibility of pertinent information (Dunlosky & Nelson, 1992; Koriat, 1993; Morris, 1990), the ease with which information comes to mind (Kelley & Lindsay, 1993; Koriat, 1993; Mazzoni & Nelson, 1995), cue familiarity (Metcalfe, Schwartz, & Joaquim, 1993; Reder, 1987; Reder & Ritter, 1992; Reder & Schunn, 1996), the ease of processing of a presented item (Benjamin & Bjork, 1996; Begg et al., 1989), the memory for its ease of acquisition (Lovelace, 1984), and the memory for the outcome of previous recall attempts (see Gardiner, Passmore, Herriot, & Klee, 1977; King et al., 1980; Mazzoni & Cornoldi, 1993). Each of these internal cues can support a heuristic for predicting future recall.

An important advantage of internal, mnemonic cues as predictors of memory performance is that they are generally sensitive to both intrinsic and extrinsic factors that affect degree of learning. As far as intrinsic factors are concerned, Begg et al. (1989), for example, obtained results suggesting that the effects of several intrinsic stimulus attributes (e.g., concreteness–abstractness, frequency of occurrence) on JOLs are mediated by the effects of these attributes on ease of processing. With regard to the effects of extrinsic factors, the extensive work of Jacoby and his associates suggests that fluency of processing and experienced familiarity are affected by previous exposure (Jacoby & Dallas, 1981; Jacoby & Kelley, 1987; Jacoby, Kelley, & Dywan, 1989; Kelley & Lindsay, 1993). For example, advance exposure to the answers of general information questions was found to reduce recall latencies of these answers, and this reduction may mediate the stronger confidence that accompanies their recall (Kelley & Lindsay, 1993; Morris, 1990; Nelson & Narens, 1990). Perhaps, then, the validity of metacognitive judgments in predicting the accuracy of recall and recognition performance is mediated in part by reliance on mnemonic cues such as accessibility and ease of processing (see Koriat, 1993; Nelson & Narens, 1990).

In sum, most of the cues that can serve as the basis for item-by-item JOLs can be grouped into three classes: intrinsic, extrinsic, and mnemonic. As sketched in Figure 1, both intrinsic and extrinsic cues may affect JOLs directly, but they may also exert their effect indirectly through their influence on any of the internal, mnemonic cues. More important, the direct and mediated effects are assumed to entail qualitatively different processes: The direct effects of intrinsic and extrinsic cues are likely to involve an analytic inference that applies the person's a priori theory about the memorial consequences of a variety of factors. The effects of internal, mnemonic cues on JOLs, in contrast, are assumed to generally entail a nonanalytic, implicit inference that uses a global heuristic rather than a logical, conscious deduction (see Jacoby & Brooks, 1984; Kelley & Jacoby, 1996a; Koriat, 1994).

What are the implications of the cue-utilization approach to JOL? Unlike the direct-access view of JOL, the cue-utilization approach as was sketched previously allows for a greater variability in the determination of JOLs and their accuracy. Thus, one implication is that the relative weight of different cues in determining JOLs may differ from one condition to another and may also change with practice studying the same list of items. For example, there is evidence that extrinsic factors that affect degree of learning also affect metacognitive judgments when manipulated within persons but not when manipulated between persons (Begg et al., 1989; Carroll & Nelson, 1993). Presumably, a within-person manipulation increases awareness of the potential contribution of these factors and motivates participants to use the extrinsic cues discriminately. Also, it has been proposed that the basis of JOLs may change during learning. For example, Hertzog et al. (1990), who had participants study three word lists, observed that first-list predictions of overall recall correlated with perceived memory self-efficacy, but as the participants studied more lists, this correlation weakened and was replaced by a stronger correlation with recall performance on the preced-

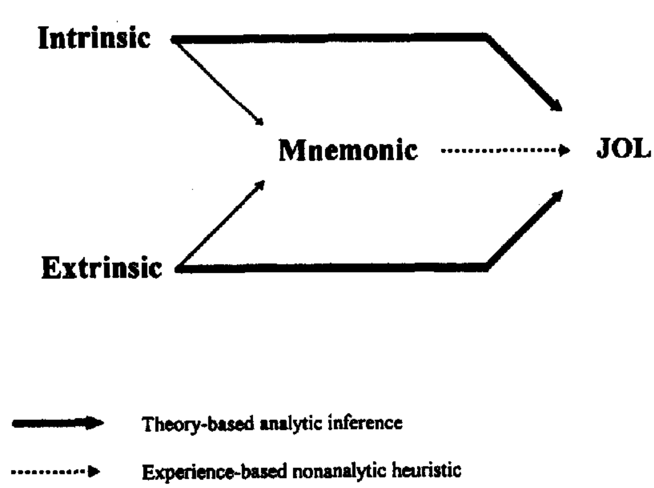

Figure 1. A schematic model of the effects of intrinsic, extrinsic, and mnemonic cues on judgments of learning (JOLs).

ing list. Other researchers (Begg et al., 1989; King et al., 1980; Lovelace, 1984) obtained results suggesting that during initial study, JOLs reflect primarily differences in the preexperimental, intrinsic characteristics of the items, whereas with subsequent presentations, JOLs reflect the ease with which the items retrieve their prior encodings. In terms of the conceptual framework advanced in this article, these results suggest increased reliance on mnemonic–experiential cues with increased practice.

Some of the changes in the basis of JOLs may be expected to affect JOL's accuracy. Thus, the increased reliance on mnemonic cues with practice may be expected to improve JOL accuracy because such cues reflect the effects of past experience and can serve as a good basis for memory predictions. This has indeed been found to be the case (Begg et al., 1989; King et al., 1980; Lovelace, 1984; Mazzoni et al., 1990). JOL accuracy has also been found to increase with overlearning (Leonesio and Nelson, 1990) and with repeated testing (King et al., 1980; Lovelace, 1984; Lovelace & Marsh, 1985; Mazzoni & Cornoldi, 1993; Mazzoni et al., 1990; Spellman & Bjork, 1992). The findings of Nelson and Dunlosky (1991; see also Dunlosky & Nelson, 1992) regarding the delayed JOL effect also suggest that both the basis of JOLs and their accuracy differ depending on the stage at which JOLs are measured.

The present study examined two hypotheses regarding the relative contribution to JOLs of the three types of cues described previously: intrinsic, extrinsic, and mnemonic. The first hypothesis concerns the relative weight of intrinsic versus extrinsic factors. It is proposed that JOLs are comparative in nature: They tend to focus on the relative recallability of different items within a list and are less sensitive to factors that affect overall performance (see Begg et al., 1989; Shaw & Craik, 1989). Therefore, in general, the effects of extrinsic factors should be discounted in predictions of recall relative to those of intrinsic factors. Several observations are consistent with this general prediction: Whereas positive correlations have been consistently observed between JOLs and various indexes of item difficulty, extrinsic factors appear to be discounted in judgments of future recall, and, in fact, sometimes fail to have any effect at all. For example, Cutting (1975) tested recall of a list of words learned under a shallow or a deep orienting task. Orienting tasked affected actual memory performance more than it affected recall predictions. Shaughnessy (1981) found elaborative rehearsal to yield better recall than maintenance rehearsal, but there was no difference between the two conditions in immediate predictions of recall. In the study by Rabinowitz et al. (1982), recall performance was better under interactive imagery instructions than under learning instructions, but memory predictions were nearly equal for the two conditions. Dunlosky and Nelson (1994) also found that JOLs elicited immediately after study were not affected by either imagery versus rote rehearsal instructions or by distributed versus massed repetitions. Delayed JOLs, however, were sensitive to these manipulations, as is discussed later.

Thus, as far as the typical, immediate JOLs are concerned, these seem to be relatively insensitive to the extrinsic factors. At the same time, however, immediate JOLs have been found to be predictive of actual recall in each of the conditions investigated presumably because of their reliance on the intrinsic properties of the items studied. In the Rabinowitz et al. (1982) study, for example, although JOLs were not affected by encoding conditions, they were very sensitive to the degree of relatedness between the two members of the pairs (an intrinsic factor). Similarly, Shaw and Craik (1989) found that although level of processing exerted a dramatic effect on recall, it had only a mild effect on JOLs. However, participants were nevertheless quite successful in predicting the relative recallability of different words. They proposed that people "are largely unaware of memory effects associated with different mental processes but are somewhat sensitive to the effects associated with different materials" (p. 134).

Of more direct relevance to the proposed hypothesis are the results reported recently by Carroll, Nelson, and Kirwan (1997). Participants studied a list of related and unrelated paired associates. The related pairs were studied to a criterion of two correct recalls, whereas the unrelated pairs were studied to a criterion of eight correct recalls. This manipulation resulted in better recall performance for the (overlearned) unrelated pairs than for the related pairs. JOLs, on the other hand, displayed the opposite effect: They were higher for the related pairs, suggesting that the intrinsic factor of semantic relatedness is overweighted relative to that of the extrinsic factor of degree of learning.

In sum, although there are some exceptions to the rule (e.g., Begg et al., 1991), the results on the whole are consistent with the hypothesis that JOLs elicited immediately after study are sensitive to intrinsic cues pertaining to the relative difficulty of different items in a list and are less sensitive to extrinsic cues pertaining to the circumstances of learning or to encoding strategies.

This hypothesis was tested in Experiments 1–3 with regard to the effects of one of the best studied extrinsic factors affecting degree of learning: number of repetitions. The intrinsic factor investigated was the relative a priori difficulty of different paired associates, as manipulated by the degree of associative relatedness between the members of a pair. Number of repetitions offers an interesting prediction with regard to JOLs: If the effects of repeated presentations of an item are discounted in the computation of JOLs, then JOLs should be expected to lag behind recall performance with increased practice studying the same items. Thus, the correspondence between predicted and actual recall may, in fact, deteriorate with practice, with overlearning resulting in underconfidence in one's performance. This prediction is somewhat at odds with the general belief that practice should improve overall performance. It also runs counter to the general tendency for overconfidence reported in many calibration studies (e.g., Allwood & Granhag, 1996; Lichtenstein, Fischhoff, & Phillips, 1982).

Whereas the first hypothesis concerns the calibration of JOLs, the second concerns monitoring resolution. In this article, resolution or *discrimination accuracy* (Yaniv, Yates, & Smith, 1991) refers to the accuracy in monitoring the relative recallability of different items. It is proposed that repeated practice with a list of items enhances sensitivity to interitem differences (Begg et al., 1989; Leonesio & Nelson, 1990) and should therefore improve the predictive validity of JOLs as indexed by the correlation between JOL and recall across items. Consistent with this proposition, Leonesio and Nelson found the accuracy of JOLs in predicting recall performance after 4 weeks to be higher for items learned to a criterion of four correct trials than for items learned to a criterion of one correct trial. Other researchers reported increased JOL accuracy (i.e., improved resolution, with repeated study, repeated test trials, or both; Begg et al., 1989; King et al., 1980; Lovelace, 1984; Mazzoni et al., 1990).

It is proposed that the improved resolution with repeated study of a list of items derives specifically from a change in the basis of JOLs, from a greater reliance on intrinsic cues toward a greater reliance on internal-mnemonic cues. This change reflects a general shift that occurs with a practice from a theory-based analytic inference toward reliance on an *experience-based nonanalytic heuristic* (to borrow the terminology of Kelley & Jacoby, 1996a; see also Strack, 1992). During the initial study of a list of items, the JOLs associated with different items reflect primarily the direct assessment of the preexperimental difficulty of the items on the basis of some preconception about the memorial consequences of different stimulus attributes (e.g., word frequency, concreteness–abstractness, etc.; see Begg et al., 1989). With increased practice learning these items, participants become increasingly more sensitive to internal cues associated with different items such as the ease of processing of the items or the relative accessibility of the target in question. Assuming that these cues are generally relevant to recall performance, this shift should result in improved accuracy of JOLs, accompanied by a reduced contribution of intrinsic properties to JOLs.

In sum, practice learning a set of items is expected to yield divergent effects on calibration and resolution. On the one hand, it should impair calibration by increasing the discrepancy between mean overall JOL and mean overall recall. This impairment, in the form of increased underconfidence, is assumed to result from the tendency to discount extrinsic factors. On the other hand, it should improve the discrimination between items that are likely to be recalled and those that are not. This improvement is expected to stem from the increased reliance on internal, mnemonic cues in making JOLs.

All of the experiments to be reported used lists of Hebrew paired associates for which some preexperimental estimate of learning difficulty was available. In the study phase, participants studied each pair for a predetermined study time and, following the completion of the study period, judged the probability that they would be able to recall the target term when presented with the cue term. The test phase involved cued recall in which all cues were presented one after the other, and participants were asked to recall the target corresponding to each of the cues. Degree of learning was manipulated in Experiments 1 and 2 by repeating the list for several study–test blocks, whereas in Experiment 3 it was manipulated by repeating some of the items within the same list. In Experiment 4 degree of learning was manipulated by presenting different pairs for different study durations.

Experiment 1

Experiment 1 included two conditions. In the experimental (Same) condition, participants were presented with the same list of paired associates for two study–test presentations. In the control (Different) condition, participants received two different lists in the two presentations. This design allows examination of the changes that occur in recall and JOL as a result of learning the same list twice and as a result of general transfer of learning. Both of these may be considered extrinsic factors. The intrinsic factor was the difficulty of the items as judged by another group of participants. To allow a wide range of difficulty ratings, the word pairs used varied greatly in the semantic relatedness between the members of the pairs.

Method

Participants. Thirty-two Hebrew-speaking undergraduates (27 women and 5 men) participated in the experiment, 30 for course credit and 2 for payment.

Stimuli. The stimulus lists were constructed on the basis of the results of a preliminary experiment designed to obtain data on the perceived relative difficulty of the items. For that experiment, 100 Hebrew stimulus–response pairs were compiled, which represented a wide range of associative relatedness. Thirty participants were presented with these pairs in a random order. They were instructed to imagine that 100 people had been asked to memorize the pairs so that they could later recall the response word when shown the stimulus word. The participants were asked to estimate, for each pair, how many people would be likely to recall the correct response. Mean memorability score ranged from 22% (for *citizen-fox*) to 99% (for *cow-milk*). The median memorability rating was 45% and was used to define the pairs as high difficulty (below 45%) or low difficulty. The pairs were divided into two lists of 50 pairs each, which were matched on memorability ratings. One list (List A), with an average memorability score of 53% included 25 high-difficulty and 25 low-difficulty pairs, whereas the second list (List B), with an average memorability score of 52% included 24 high-difficulty pairs and 26 low-difficulty ones.

Apparatus. The experiment was controlled by a Digital Microvax Lab II work station (Digital Equipment Corp., Maynard, MA). The stimuli were displayed on a computer screen, and JOLs and recall scoring were entered by the experimenter on a separate monitor.

Procedure. The experiment involved two study–test blocks (presentations). In the study phase of each block, participants were instructed that they would have to study 50 paired associates and would have to indicate their JOLs about each item as soon as it disappeared from the screen. They were told that in the test phase they would be shown each stimulus word in turn and would be asked to recall the response words.

During the study phase, the stimulus and response pairs were presented at the center of the screen side by side for 5 s. The participants were instructed to study each pair so that later, in the second stage, they would be able to recall the second word in each pair when the first is presented. Participants were urged to use the entire 5 s for studying. The pair was then replaced by the following question: "What are the chances that you will recall the second word when presented with the first word?" The participant reported his or her estimate orally on a 0%–100% scale, and the experimenter entered the data on her monitor. The order of presentation of the pairs was randomly determined for each participant.

During the test phase, the 50 stimulus words were presented one after the other for 10 s each in a new random order. The participants had to say the response word within the 10 s allotted. A synonym (e.g., *anxiety* instead of *fear*) or a very closely related word (e.g., *foxes* instead of *fox*) were scored as correct responses.

In the Same condition, the list of paired associates was repeated across the two blocks, whereas in the Different condition, different lists were used in each of the two blocks. There were 16 participants in each condition. In the Same condition, 8 participants received List A and 8 received List B in both presentations, whereas in the Different condition, the order of the two lists was counterbalanced across each group of 8 participants.

Results

I shall first examine the hypothesis regarding the relative contribution of the intrinsic and extrinsic factors to JOLs and recall. Then I examine the hypotheses regarding the effects of presentation on both calibration and resolution.

The contribution of intrinsic and extrinsic factors to JOLs and recall. It was proposed that the effects of extrinsic factors on JOL are discounted relative to those of intrinsic factors. In the present study, the two types of factors were represented by item difficulty and study presentation, respectively. Figure 2 presents mean JOLs and mean actual recall as a function of rated item difficulty (top panel) and presentation (bottom panel). The results are plotted separately for the Same and Different conditions.

In the following analyses the contrast between predicted and actual performance was treated as a repeated factor (labeled *Measure*) in the analyses of variance (ANOVA). This treatment allowed assessment of the possible interaction between this factor and the two manipulated factors. Consider first the results for the intrinsic factor: item difficulty. As can be seen in Figure 2 (top panel), participants were somewhat underconfident in their judgments: Whereas actual recall performance averaged 67.2 across conditions and presentations, JOL averaged only 55.5 (but see next). Both JOLs and recall were strongly affected by item difficulty, but the effect was somewhat stronger for JOLs than for recall: A three-way ANOVA, Condition × Difficulty × Measure yielded $F(1, 30) = 27.29$, $MSE = 322.81$, $p < .0001$, for measure; $F(1, 30) = 220.80$, $MSE = 295.47$, $p < .0001$, for difficulty; and $F(1, 30) = 6.42$, $MSE = 82.23$, $p < .05$, for the Measure × Difficulty interaction. Thus, there was a slight tendency for item difficulty to be overweighted in the determination of JOLs: Whereas easy items had an advantage of 29% over difficult items in recall, they had an advantage of 35% in JOLs.

Turning next to the effects of presentation, the extrinsic factor, the results, as expected, were very different for the Same and Different conditions (Figure 2, bottom panel). The Different condition yielded relatively small effects of presentation that were not consistent across the two measures. For the Same condition, in contrast, both recall and JOL increased with presentation, and as expected, the increase was larger for recall than for JOL.

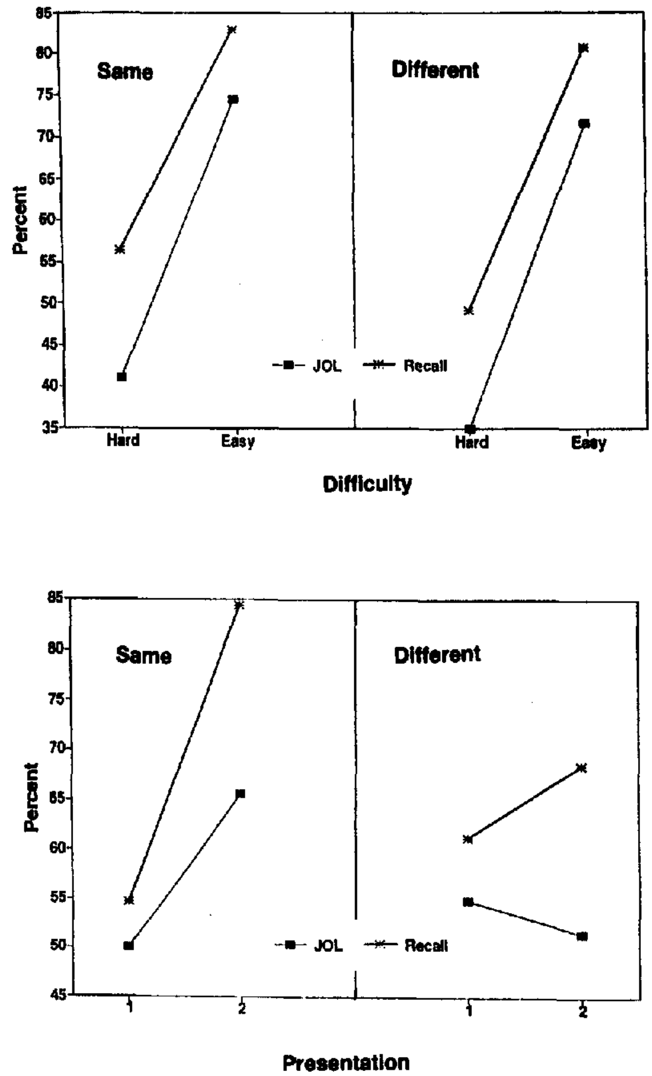

Figure 2. Mean judgment of learning (JOL) and recall as a function of rated item difficulty (top panel) and presentation (bottom panel) for the Same and Different conditions (Experiment 1).

A Condition × Presentation × Measure ANOVA on these data yielded $F(1, 30) = 27.29$, $MSE = 322.81$, $p < .0001$, for measure; $F(1, 30) = 95.99$, $MSE = 100.57$, $p < .0001$, for presentation; $F(1, 30) = 69.35$, $MSE = 100.57$, $p < .0001$, for the Condition × Presentation interaction; and $F(1, 30) = 18.87$, $MSE = 126.60$, $p < .0001$, for the Measure × Presentation interaction. A separate two-way Presentation × Measure ANOVA for the Same condition yielded $F(1, 15) = 11.07$, $p < .005$, for the interaction: Whereas the effect of presentation amounted to 30% for recall, it amounted to only 16% for JOL. Thus, unlike the intrinsic factor of item difficulty, the extrinsic factor of presentation is discounted in the determination of JOLs.

The Different condition yielded a pattern of results that is difficult to interpret. The effects of presentation on actual recall were weaker for this condition than for the Same condition, as indicated by a significant Condition × Presentation interaction, $F(1, 30) = 47.51$, $p < .0001$. However, they were nevertheless significant, $F(1, 15) = 7.95$, $p < .05$. If this effect is due to transfer of learning, then it would seem that such transfer effects are not revealed by JOLs. In fact, a Presentation × Measure ANOVA for the Different condition also yielded $F(1, 15) = 7.82$, $p < .05$, for the interaction.

A question of interest is whether the discounting of the effects of repetition occurs equally for both easy and difficult items. The effects of practice were stronger for harder-to-learn than for easier-to-learn items. Thus, a Difficulty × Presentation ANOVA for recall yielded a significant interaction for the Same condition, $F(1, 15) = 51.97$, $p < .0001$, but not for the Different condition, $F(1, 15) = 1.26$. Focusing only on the Same condition, a three-way ANOVA, Measure × Difficulty × Presentation yielded significant effects for the Difficulty × Presentation interaction, $F(1, 15) = 77.14$, $p < .0001$; for the Measure × Difficulty interaction, $F(1, 15) = 4.94$, $p < .05$; and for the triple interaction, $F(1, 15) = 10.99$, $p < .005$. Separate Measure × Presentation ANOVAs for the difficult and easy items indicated a significant interaction only for the difficult items, $F(1, 15) = 13.81$, $p < .005$, but not for the easy items, $F(1, 15) = 2.05$. Thus, the tendency to discount the effects of repeated study was mostly due to the difficult items.

Calibration. The results presented so far indicate that the effects of repeated practice are stronger for recall than for JOLs. If participants' JOLs are well calibrated on the first presentation of a list, the differential effects of practice on recall and on JOLs should result in increased underconfidence with practice. This trend is captured in Figure 3, which depicts the calibration curves for Presentations 1 and 2 for both the Same and the Different conditions. The data are plotted according to the procedure commonly used in calibration studies (see Lichtenstein et al., 1982): The probability judgments are grouped into 10 levels, 0–10, 11–20, . . . , 91–100. The percentage of correct recall is plotted against the mean JOL, computed across participants; perfect calibration is indicated by the diagonal line. The plots reiterate the previously noted effects: There is an increase in underconfidence with increasing practice, with the effect being somewhat stronger for the Same than for Different conditions. In fact, participants were relatively well calibrated in the first presentation, but became more underconfident in the second presentation. Thus, mean over/underconfidence for each participant, computed as the weighted mean of the differences between the mean JOL and the percentage of correct recall for each category (see Lichtenstein et al., 1982), averaged −.049 and −.184 for Presentations 1 and 2 of the Same condition and −.062 and −.167, respectively, for the Different condition.

Resolution. The results presented so far indicate that increased practice with a list of items impairs the monitoring of future recallability, as indexed by the correspondence between the absolute levels of JOLs and recall. Did practice also impair resolution (i.e., the discrimination between recalled and nonrecalled items)?

A simple measure of resolution is the within-person gamma correlation between JOLs and recall (see Koriat & Goldsmith, 1996; Nelson, 1984). These correlations averaged .58 and .67 for Presentations 1 and 2, respectively, for the Same condition and .59 and .65, respectively, for the Different condition. All correlations were significantly different from zero at the .0001 level. A two-way ANOVA on these correlations yielded $F < 1$, for condition; $F(1, 30) = 4.28$, $p < .05$, for presentation; and $F < 1$, for the

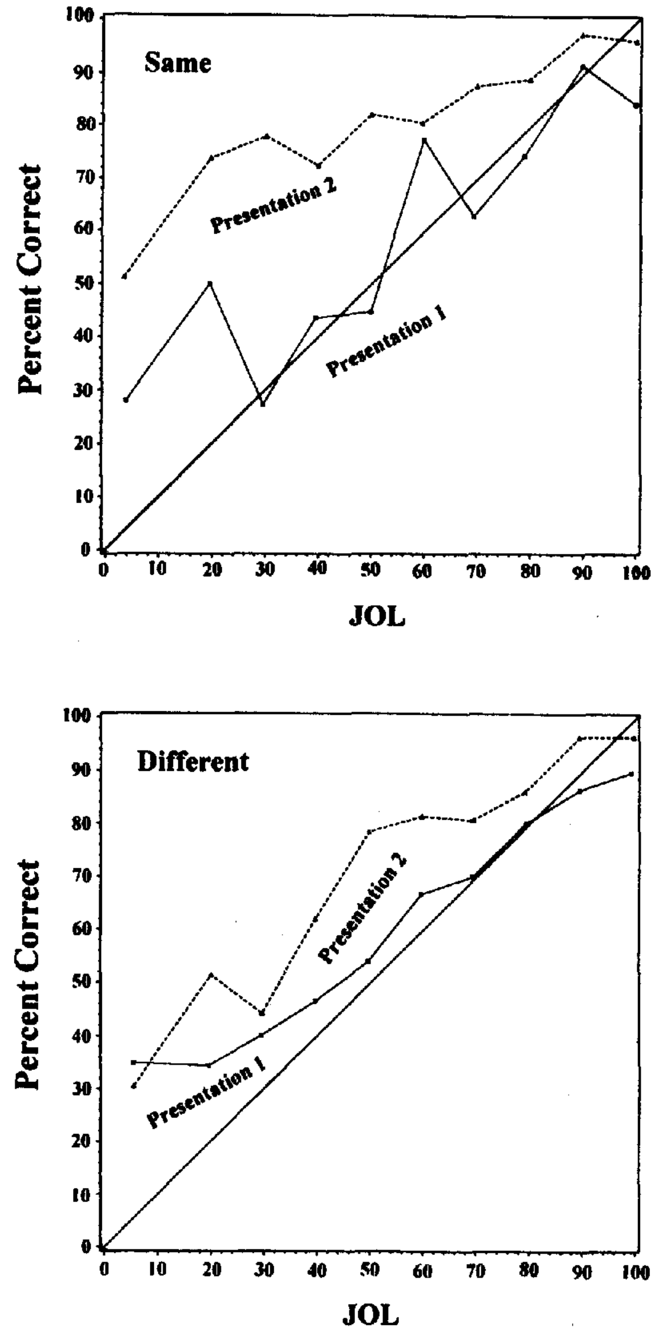

Figure 3. Calibration curves plotting percent correct recall as a function of judgment of learning (JOL) for presentations 1 and 2 for the Same (top panel) and Different conditions (bottom panel). The diagonal lines indicate perfect calibration (Experiment 1).

interaction. Thus, there was a significant improvement in JOL accuracy from Presentation 1 to Presentation 2. Because this improvement occurred for both conditions, it would seem to be due to the general experience with the task.

In sum, the results of Experiment 1 disclose the predicted pattern: Practice seems to impair calibration while improving resolution. That is, it widens the discrepancy between mean overall JOLs and mean overall recall, but at the same time increases the success with which JOLs differentiate between items that are subsequently recalled and those that are not.

What is the source of the improved resolution? The mean gamma correlation between JOLs and item-difficulty scores (with high scores on item difficulty designating higher memorability ratings) averaged .54 and .59 for Presentations 1 and 2, respectively, for the Different condition, and .54 and .46, respectively, for the Same condition. A Presentation × Condition ANOVA on these correlations yielded $F(1, 30) = 12.87$, $p < .01$, for the interaction. Whereas the JOL-difficulty correlation did not change with presentation for the Different condition, $F(1, 15) = 3.49$, *ns*; it decreased significantly from Presentation 1 to Presentation 2 for the Same condition, $F(1, 15) = 10.60$, $p < .01$. Thus, the improved resolution with practice in the Same condition is accompanied by a tendency of participants to rely less heavily on the judged difficulty of the items in assessing the probability of prospective recall.

Discussion

The results of Experiment 1 indicated a differential effect of degree of learning on recall and JOL: The repeated presentation of stimuli had a milder effect on JOLs than it had on recall. In contrast, judged item difficulty, if anything, had a stronger effect on JOL than it had on recall. These results are not consistent with the direct-access view of JOLs according to which JOLs monitor trace strength. If this were so, both recall and JOL would have been expected to be similarly affected by both intrinsic and extrinsic factors.

The results, however, can be accommodated by the cue-utilization view according to which JOLs are based on a number of cues that are generally predictive of recall. It was proposed that JOLs are more sensitive to cues that discriminate between items and therefore are affected more strongly by intrinsic factors than by extrinsic factors. Consistent with this proposition, the results of Experiment 1 indicated that the effects of practice were discounted in making judgments about future recallability. Note that the discounting of the effects of learning was evident for difficult items but not for easy items, perhaps because difficult items were those that benefited most from the repeated presentation of the list. Thus, at least for difficult items there was a tendency for increased underconfidence with increased learning.

The differential effects of list repetition are consistent with previous findings indicating that participants tend to discount the contribution of extrinsic factors when they judge the prospective recall of items (e.g., Cutting, 1975; Rabinowitz et al., 1982; Shaughnessy, 1981; Shaw & Craik, 1989). What is peculiar about the results of Experiment 1, however, is that practice impairs calibration: Whereas participants were relatively well calibrated on the first presentation, they were markedly underconfident on the second presentation. This pattern is counterintuitive because participants must have been in a better position to judge the overall chances of recall after having experienced the entire sequence of study and test. Should further exposures to the list improve calibration by reducing underconfidence, or should they impair calibration further? Should providing participants with feedback about the correctness of their responses mitigate against the increased underconfidence with learning? These questions were addressed in Experiment 2.

An important finding of Experiment 1 is the differential effects that practice had on calibration and resolution (or absolute and relative monitoring, respectively; see Nelson & Dunlosky, 1991). Thus, despite the fact that the correspondence between mean overall JOL and mean overall recall decreased from the first to the second presentation, the cross-item correlation between JOLs and recall increased. It was proposed that the improved resolution is due to a greater reliance on mnemonic–experiential cues than on intrinsic cues. Consistent with this proposition is the observation that JOLs became less dependent on perceived item difficulty from the first to the second presentation. Apparently JOLs made after the first exposure to the items are based primarily on people's preconceptions about the memorial consequences of the intrinsic characteristics of items, but with increased practice JOLs are based more heavily on experience-based heuristics. Experiment 2, which used four presentations of the same list, affords a closer look at the effects of practice on the correlational pattern between JOLs, item difficulty, and recall.

Experiment 2

The procedure of Experiment 2 differed from that of Experiment 1 in three respects. First, the same list (including 70 pairs) was repeated for four study–test presentations. This procedure allows examination of the question whether the underconfidence effect decreases or increases with successive presentations. If increased practice with the same stimuli improves monitoring, then mean JOL should tend to match that of actual recall as participants learn more about their performance. If, on the other hand, participants' predictions continue to discount the effects of learning in their JOLs, then practice should not reduce underconfidence and may even increase it. The use of four repetitions of the list also permits a more detailed examination of the changes in relative monitoring (or resolution) that were observed in Experiment 1. As detailed earlier, it is expected that participants' initial JOLs reflect primarily the preexperimental, perceived difficulty of the items. With increased practice, however, internal cues that are more closely tuned to the specific idiosyncratic experience with the encoding and remembering of the items come to make a stronger contribution to JOLs, thereby improving overall resolution. If such is the case, then the JOL-recall correlation should increase with practice, whereas the JOL-difficulty correlation should decrease systematically with practice.

Second, a between-person manipulation was included involving feedback. Half of the participants received feedback about the correctness of their responses, whereas the remaining participants were not given any feedback, as in Experiment 1. The question is whether feedback can help reduce the expected impairment in calibration with practice studying the same items. It is also important to find out whether feedback also improves resolution (i.e., the JOL-recall correlation).

Finally, whereas in Experiment 1 the measurement of item difficulty was based on subjective ratings, in Experiment 2 item difficulty was defined in terms of word-association norms. Items were defined as difficult or easy, respectively, depending on whether the stimulus elicited the response word with a probability of 0 or with a probability that exceeded .05.

Method

Participants. Twenty-four Hebrew-speaking undergraduates (17 women and 7 men) participated in the experiment for course credit. Twelve were assigned randomly to the feedback condition and 12 were assigned to the no-feedback condition.

Materials. Seventy word pairs were selected on the basis of Hebrew word-association norms. They were chosen to represent two levels of associative relatedness. For 35 pairs (related) the stimulus word elicited the response word as a first associate in 5%–20% of the cases, whereas for the remaining pairs (unrelated) the respective value was 0%.

Procedure. The apparatus was the same as in Experiment 1. The procedure was also the same except for the following: The experiment consisted of four study–test blocks. In the study phases of each block, the stimulus–response pair was displayed for 4 s. In the test phases, the stimulus members were displayed for 8 s each. The participant was expected to recall the corresponding response term within the 8 s allotted, and the experimenter scored the response as correct or wrong and entered the score on her computer terminal.

The procedure was identical for the feedback and no-feedback conditions except that in the feedback condition a 30-ms, high-pitch tone was sounded when the response was incorrect or when 8 s elapsed. The next stimulus was then shown 500 ms thereafter. In the no-feedback condition, no sound feedback was given, and the next stimulus appeared after a 530-ms interval. Feedback participants were informed about the significance of the sound.

Results

The effects of presentation, difficulty, and feedback on JOL and recall. The results of Experiment 2 were analyzed in terms of the effects of four factors: condition (feedback vs. no-feedback), presentation (1–4), associative relatedness (unrelated vs. related), and measure (JOL vs. recall). All factors except the first were within-person factors. A four-way ANOVA for a mixed design yielded $F < 1$ for condition; $F(3, 66) = 206.33$, $MSE = 100.24$, $p < .0001$, for presentation; $F(1, 22) = 149.54$, $MSE = 701.82$, $p < .0001$, for associative relatedness; and $F(1, 22) = 8.57$, $MSE = 606.67$, $p < .01$, for measure. Four interactions were also significant: Presentation × Measure, $F(3, 66) = 21.23$, $MSE = 64.22$, $p < .0001$; Condition × Measure, $F(1, 22) = 4.95$, $MSE = 606.67$, $p < .05$; Presentation × Associative Relatedness, $F(3, 66) = 26.32$, $MSE = 58.53$, $p < .0001$; and Presentation × Measure × Associative Relatedness, $F(3, 66) = 6.19$, $MSE = 34.56$, $p < .001$.

I first examined the effects of the feedback manipulation on JOLs and recall. As in Experiment 1, participants were underconfident overall, with JOL and recall averaging 60% and 67%, respectively. However, the underconfidence effect was obtained mostly in the no-feedback condition, for which the respective figures were 54.4, and 67.4, whereas in the feedback condition, these figures were 64.7, and 66.5, respectively. Thus, interestingly, feedback affected JOLs without affecting actual recall performance, thereby reducing underconfidence.

Turning next to the effects of presentation, Figure 4 shows mean JOL and mean actual recall as a function of presentation for the feedback and non-feedback conditions. It can be seen that although feedback helped to boost JOLs across all presentations, the effects of presentation were weaker for JOLs than for recall for both the feedback and the no-feedback conditions. Indeed, the Presentation × Measure interaction was significant for the no-feedback condition, $F(3, 33) = 7.43$, $p < .001$, as well as for the feedback condition, $F(3, 33) = 18.57$, $p < .0001$. Thus, the results of both conditions displayed a trend of increased underconfidence with practice.

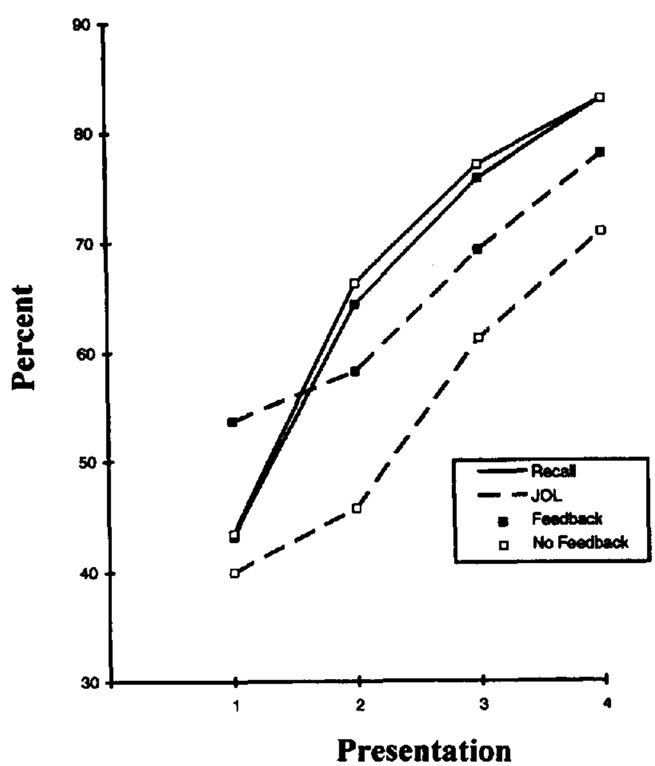

Figure 4. Mean judgment of learning (JOL) and recall as a function of presentation for the Feedback and No-Feedback conditions (Experiment 2).

As far as the effects of associative relatedness are concerned, these were equally strong for both JOLs and recall. The effects of associative relatedness interacted with those of presentation and measure as follows: The increase from Presentation 1 to Presentation 4 was stronger for recall (39.6) than for JOLs (27.5), and this interactive pattern was more pronounced for the unrelated pairs (52.0 and 33.4, respectively) than for the related pairs (27.3 and 21.7, respectively). Note, however, than a Measure × Presentation ANOVA yielded a significant interaction for both the unrelated pairs, $F(3, 69) = 24.08, p < .0001$, and the related pairs, $F(3, 69) = 7.27, p < .0005$.

In sum, then, the results support the predicted increase in underconfidence with practice. This increase was found for both the feedback and the no-feedback conditions and for both the associatively related and associatively unrelated pairs.

Resolution. I now turn to the examination of the cross-item correspondence between JOL and recall. In Experiment 1, the gamma correlation between JOL and recall was found to increase from Presentation 1 to Presentation 2, whereas that between JOL and item difficulty decreased. A similar pattern was clearly observed in Experiment 2, as may be seen in Figure 5. This figure plots mean gamma correlations between JOL and recall, on the one hand, and between JOL and associative relatedness (difficulty), on the other hand, as a function of presentation. The results are plotted separately for the no-feedback condition (left panel) and for the feedback condition (right panel), with associative relatedness treated as a dichotomous variable distinguishing between related and unrelated pairs. Three participants for whom one or more of the correlations was indeterminate were excluded from the analysis.

Let us focus first on monitoring resolution as reflected in the JOL-recall correlations. A Condition (feedback vs. no feedback) × Presentation ANOVA on these correlations yielded $F < 1$ for condition, $F(3, 57) = 26.81, p < .0001$, for presentation, and $F < 1$, for the interaction. Gamma was relatively high even in the first presentation (.66) but it reached .89 in the last presentation. Feedback did not affect monitoring resolution, and the improvement in resolution with practice was equally observed regardless of the feedback condition.

In parallel, the correlations between JOL and associative relatedness, on the other hand, decreased systematically with presentation. A similar two-way ANOVA on these correlations yielded $F(1, 19) = 2.79$, *ns*, for condition; $F(3, 57) = 31.60$, $p < .0001$, for presentation; and $F < 1$, for the interaction. The correlations were somewhat higher for the feedback (.85) than for the no-feedback (.77) condition, but for both conditions JOLs became less dependent on associative relatedness with increased practice: The pertinent gammas dropped from .93 to .73, almost exactly the reverse pattern from that observed for the JOL-recall correlations.

In sum, the correlational pattern depicted in Figure 5 indicates that for both feedback conditions the increased

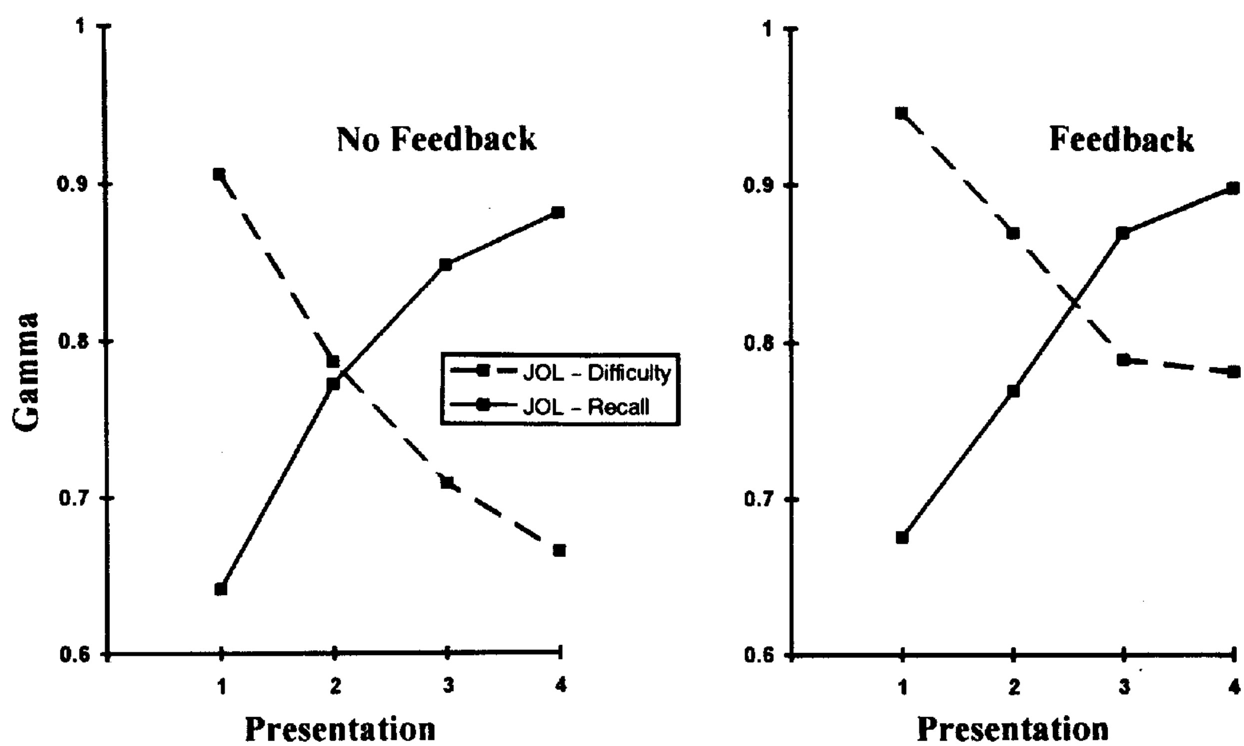

Figure 5. Mean within-subject gamma correlations between judgment of learning (JOL) and item difficulty, on the one hand, and between JOL and recall, on the other hand, plotted as a function of presentation for the no-feedback (left panel) and feedback (right panel) conditions (Experiment 2).

monitoring resolution with practice is accompanied by a decrease in the JOL-difficulty correlation. This pattern is consistent with the idea that the basis for JOLs changes with practice from a theory-based inference that relies predominantly on intrinsic cues towards a greater reliance one experiential–mnemonic cues that can serve as valid cues for JOLs.

Assuming that the improvement in JOL accuracy is due to a greater reliance on mnemonic cues, a question of interest, then, is whether these cues are commonly shared by all participants, as is possibly the case with the intrinsic cues that affect JOLs primarily before study or during initial study. Thus, one possibility is that both types of cues are commonly shared but are not perfectly correlated: Participants initially base their judgments on one set of commonly shared stimulus attributes (e.g., associative relatedness), but memorability actually varies with another set of commonly shared stimulus attributes, and it is these latter attributes that come to play a more dominant role in the computation of JOLs by participants in the later study trials. If so, then one person's actual recall should be as effectively predicted by another person's JOLs (other's JOLs) as by the person's own JOLs (own JOLs). In that case, the improvement in JOL accuracy with practice should be found even when own JOLs are replaced by other's JOLs. Alternatively, the item-specific mnemonic cues that are used during the later study presentations may be idiosyncratic in nature and are based on privileged access to personal aspects of encoding and remembering. In that case, only one's own JOLs evidence increased predictive validity with practice.

To examine this question, a procedure previously used by Lovelace (1984) was applied. Each participant was randomly yoked with another participant from the same feedback condition, and the JOLs reported by each participant on the four study trials were replaced by the corresponding JOLs of the other participant. The JOL-recall gamma correlations were then calculated for each participant and presentation, as before, and their means across participants were .63, .59, .62, and .58 for Presentations 1–4, respectively. These results clearly indicate that unlike own JOLs, the other's JOLs correlations do not increase with practice and remain low throughout. A two-way ANOVA comparing the effects of presentation for own and other's JOL, yielded $F(1, 20) = 50.04$, $p < .0001$, for the own versus other contrast, and $F(3, 60) = 13.84$, $p < .0001$, for the interaction between this factor and presentation. Note that the JOL-recall correlation was similar for own and for other's JOLs in the first presentation, the corresponding means, across the two conditions averaging .66 and .63, $F < 1$. This result suggests that idiosyncratic cues do not contribute to JOLs in the first presentation. In the fourth presentation, in contrast, the correlation between recall and own JOLs was much higher (.89 across the two conditions) than the corresponding correlation with other's JOLs (.58).

If theory-based JOLs rely on commonly shared beliefs about the effects of intrinsic factors, whereas mnemonic cues tend to be idiosyncratic, then participants may be expected to make similar JOLs to the same items during initial study, but with increased practice they should tend to diverge. This seems indeed to be the case: For each presentation, the correlations across items were calculated between the JOLs reported by each participant and those reported by each of the remaining 21 participants. (Two participants were eliminated because they gave JOLs of 100 to all items in one or more presentations.) The 21 correlations were then averaged to obtain one mean for each participant. These means averaged .63, .57, .51, and .43 for Presentations 1, 2, 3, and 4, respectively, $F(3, 63) = 102.54$, $p < .0001$. Thus, intersubject consensus in the assignment of JOLs to different items was initially high and decreased monotonically with successive presentations.

To further explore the changes in the correlational pattern that took place during learning, the following analysis was carried out: For each participant all gamma correlations among 9 variables were calculated across items. The 9 variables were associative relatedness (difficulty), and the JOL and recall values for each of the four presentations. These correlations were calculated after splitting JOLs at each participant's median for each presentation, so that all participating variables were dichotomous.[1] The means of these gamma correlations (see Table 1) were then submitted to a multidimensional scaling analysis using 1 − Gamma as an estimate of distance. The results yielded a two-dimensional solution with a stress of .11. This solution is depicted in Figure 6. The results reveal several trends: First, it can be seen that Presentations 2–4, JOLs on one presentation are more strongly correlated with recall on the previous test (averaging .97 across the three presentations) than with recall on the subsequent test (averaging .92), $F(1, 20) = 18.02$, $p < .001$ (see also Lovelace, 1984).[2] This difference

Table 1
Mean Within-Person Correlations Between JOLs, Recall, and Difficulty

Variable	1	2	3	4	5	6	7	8	9
1. JOL1	—	.792	.761	.722	.748	.710	.716	.735	.938
2. JOL2	.792	—	.901	.884	.960	.871	.855	.900	.846
3. JOL3	.761	.901	—	.919	.925	.973	.934	.931	.764
4. JOL4	.722	.896	.919	—	.851	.946	.979	.968	.746
5. Recall1	.748	.960	.925	.851	—	.905	.914	.911	.815
6. Recall2	.710	.871	.973	.946	.905	—	.959	.915	.754
7. Recall3	.716	.855	.934	.979	.914	.959	—	.867	.768
8. Recall4	.735	.900	.931	.968	.911	.915	.867	—	.745
9. Difficulty	.938	.846	.764	.746	.815	.754	.768	.745	—

Note. The integers attached to judgments of learning (JOLs) and Recalls refer to number of presentation. All JOL variables in this table are dichotomous.

[1] Gamma values are generally lower when the participating variables are multivalued than when they are dichotomous. Because both recall and associative relatedness are dichotomous variables, JOLs were collapsed into two categories also, so that the distances in the multidimensional scaling analysis would not be contaminated by differences in the number of categories of the participating variables.

[2] The corresponding gamma correlations, calculated over the full range of JOLs, were .94 and .84, respectively, $F(1, 20) = 63.80$, $p < .0001$.

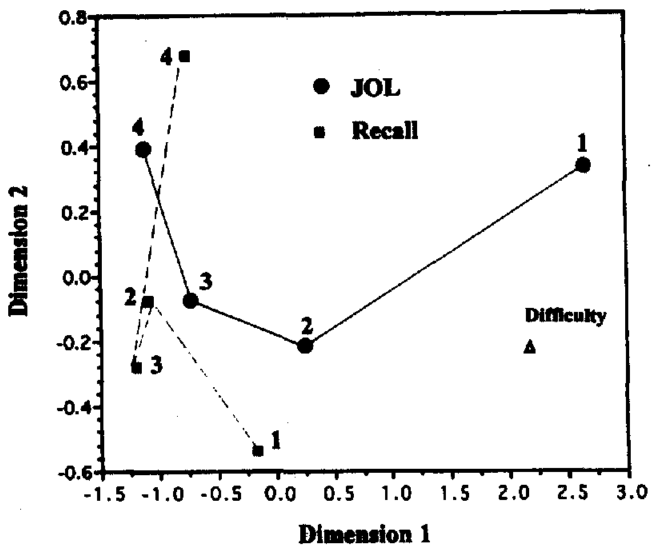

Figure 6. A two-dimensional representation of the mean within-person correlations between judgments of learning (JOLs), recall, and difficulty. The integers designate number of presentation (Experiment 2).

suggests that JOLs do not simply monitor the current strength of the memory trace but are based on cues, among them those pertaining to the outcome of the previous recall opportunity.

Second, Figure 6 suggests the interesting possibility that predicted and actual memory performance represent two orthogonal dimensions, with the horizontal and vertical axes representing changes in JOL and changes in recall, respectively. A similar pattern has been reported by Koriat and Lieblich (1977) with regard to the feeling of knowing: An analysis of memory pointers (word definitions) in terms of the memory states they tend to precipitate yielded two orthogonal dimensions: the likelihood of evoking or suggesting the correct target and the likelihood of engendering a feeling of knowing.

Finally, Figure 6 depicts the trend noted earlier, that with increasing practice JOLs become less heavily dependent on item difficulty and more closely related to actual memory performance (see also King et al., 1980; Lovelace, 1984). This trend is consistent with the idea that with increased practice JOLs come to rely more heavily on mnemonic cues and less on the intrinsic qualities of the stimuli.

Discussion

The results of Experiment 2 replicated most of the trends observed in Experiment 1. First, practice learning the same stimuli had a milder effect on JOLs than it had on recall. Whereas participants were relatively well calibrated on the first presentation, they evidenced underconfidence from the second presentation on. The underconfidence effect did not diminish with further presentations and, in fact, survived even on the fourth presentation of the list. As in Experiment 1, the tendency to discount the effects of learning was more pronounced for difficult than for easy items, but here it was significant for both types of items.

Although learning appeared to impair calibration, it clearly improved resolution: With repeated presentations of the list, participants became increasingly more accurate in distinguishing between items that were likely to be recalled and those that were not. The systematic increase in JOL accuracy, as indexed by the JOL-recall correlation, was accompanied by a systematic reduction in the size of the correlation between JOL and associative relatedness. This pattern suggests that in the course of the experiment, participants became increasingly tuned to item-specific mnemonic cues that disclose the memorability of different pairs.

One mnemonic cue that appears to have contributed to the improved accuracy of JOL is the memory for previous recall attempts. This is indicated by the observation that the best predictor of JOLs in the later part of the experiment is the recall performance on the preceding test. Of course, it is possible that what is operating as a mnemonic cue is not the direct retrieval of the specific recall episode but some less articulate cue such as ease of processing (see Begg et al., 1989) or ease of access (see Morris, 1990) that is highly correlated with recallability on the preceding trial. In either case the yoking procedure suggests that the mnemonic cues concerned are idiosyncratic rather than commonly shared, possibly reflecting privileged personal knowledge about item-specific encoding and decoding operations.

Feedback about the correctness of the responses helped to reduce the overall discrepancy between mean predicted and mean actual recall. However, somewhat surprisingly, the Presentation × Measure interaction was obtained for the feedback condition as well, suggesting that explicit feedback was not sufficient to eliminate the tendency to discount the effects of learning. On the other hand, feedback exerted only a small effect on resolution. Possibly, the major effect of feedback was to increase the participants' awareness of the improvement in their memory performance with presentation; not to increase sensitivity to interitem differences in memorability.

Finally, item difficulty exerted similar effects on JOLs and recall, unlike what was found in Experiment 1. The discrepancy may be due to the difference in the way item difficulty was operationalized in the two experiments (but see Experiments 3 and 4). In any case, the effects of the intrinsic factor of item difficulty were not discounted like those of the extrinsic factor of list repetition.

Experiment 3

The previous experiments tested the hypothesis that list repetition affects recall more than it affects JOLs. Experiment 3 extended investigation of this hypothesis to the repetition of items within a list. It examined the question whether the number of times an item has been studied within a list exerts weaker effects on JOLs than on recall, similar to what has been found for list repetition. It should be noted that Zechmeister and Shaughnessy (1980) have reported that twice-presented items elicit higher JOLs than once-presented items, but their data do not permit a direct comparison between the effects of item repetition on JOL and on recall.

In this experiment, a list of 72 paired associates was presented for study, with an equal number of items appearing once, twice, or three times during list presentation. Participants indicated their JOLs in each trial, and their recall was finally tested. JOLs are expected to evidence reduced sensitivity to number of study trials despite the fact that this variable can help discriminate between different items within the same list.

Method

Participants. Twenty-four Hebrew-speaking undergraduates (20 women and 4 men) participated in the experiment for course credit.

Materials. Seventy-two word pairs were selected on the basis of a preliminary study in which 12 judges were presented with 100 word pairs covering a wide range of degree of relatedness and were asked to estimate how many people, out of 100 people who are asked to memorize the pairs, would be able to recall the response term when presented with the cue term. The 72 pairs selected were divided into three sets of 24 pairs each, matched on memorability ratings. Each set included 12 easy (memorability mean of 50% or more) and 12 difficult items. For each participant, each set was assigned to one level of number of presentations (1, 2, and 3), with the assignment counterbalanced across participants.

Procedure. The apparatus was the same as in Experiment 1. The procedure was also the same except for the following: The experiment included one study block involving 144 trials, followed by a 72-trial test. For the study phase, all pairs were presented in a random order with the restriction that at least four pairs separated between two presentations of the same pair. Each pair was displayed for 5 s, and participants announced their JOLs as soon as the pair disappeared from the screen. (Thus, each repetition of an item was followed by a JOL.) In the test phase, the 72 stimulus members were displayed in a random order for a maximum of 10 s each, and the participant was expected to recall the corresponding response term within that period.

Results

Calibration. It should be noted that recall can only be analyzed as a function of number of presentations, whereas JOLs can be analyzed for each presentation for items presented once, twice, or three times. However, because participants did not know in advance how many times each item was going to be shown, it is meaningful to compare the effects of presentation (first, second, or third) for JOLs with the effects of number of presentations for recall.

Consider first the data for JOLs. Mean JOLs are plotted in Figure 7 as a function of presentation for items that were presented once, twice, or three times. It can be seen that JOLs generally increase with repeated presentations. Collapsing across all items, a one-way ANOVA yielded $F(2, 46) = 42.95$, $MSE = 83.77$, $p < .0001$, for presentation.

Percent recall also increased as a function of number of presentations, $F(2, 46) = 52.48$, $MSE = 147.36$, $p < .0001$, but as can be seen in Figure 8, the effects of presentation are stronger on recall than on JOL: A Measure (recall vs. JOL) × Number of Presentations (3) ANOVA yielded $F(1, 23) = 2.87$, $MSE = 677.31$, $p < .10$, for measure; $F(2, 46) = 81.36$, $MSE = 133.13$, $p < .0001$, for presentation; and $F(3, 46) = 5.10$, $MSE = 98.0$, $p < .01$, for the interaction. Thus, although participants' mean JOLs perfectly matched mean recall for items presented once, subsequent presentations of these items are associated with underconfidence regarding their future recall.

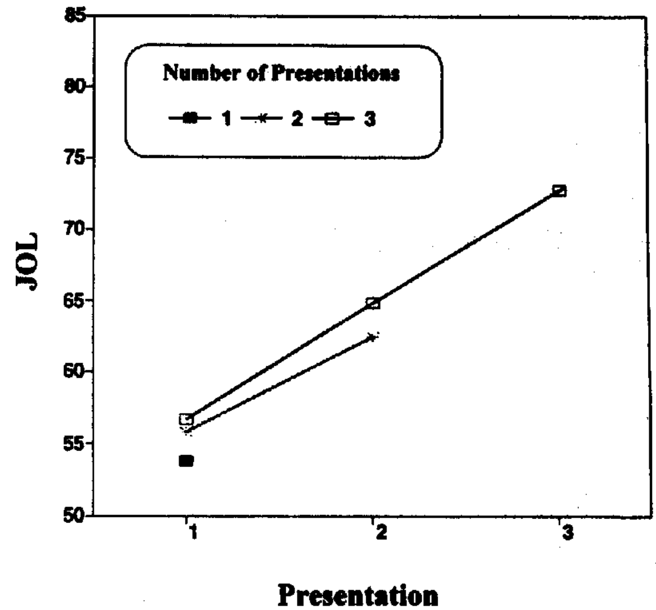

Figure 7. Mean judgment of learning (JOL) as a function of presentation for items presented once, twice, and three times (Experiment 3).

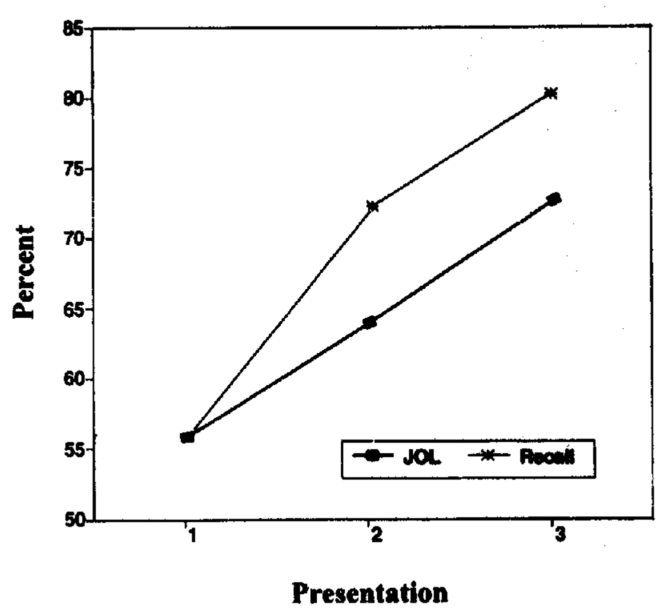

Figure 8. Mean judgment of learning (JOL) and recall as a function of presentation (Experiment 3).

Turning next to the intrinsic factor of item difficulty, JOLs for difficult and easy items averaged 47.3 and 80.6, respectively. The respective means for percent recall were 52.1 and 86.2. A Measure × Difficulty ANOVA yielded $F(2, 46) = 2.87, p < .10$, for measure; $F(1, 23) = 182.97, p < .0001$, for difficulty; and $F < 1$ for the interaction. Thus, it would seem that unlike the extrinsic factor of repetitions, the intrinsic factor of item difficulty exerted equivalent effects on JOLs and recall.

Resolution. To evaluate the accuracy of JOLs in predicting recall, the JOLs reported by each participant on the last presentation of each item were taken as the effective JOLs. Gamma correlations calculated over the entire list of 72 items between recall and last JOL averaged .59, significantly different from zero, $t(23) = 16.08, p < .0001$. The respective correlation between JOL and difficulty was .62, $t(23) = 28.70, p < .0001$, and that between recall and difficulty was .56, $t(23) = 16.19, p < .0001$.

Unlike the finding of Experiments 1 and 2 that list repetition improves resolution, the results of Experiment 3 indicated no improvement as a function of within-list item repetition: The JOL-recall gammas calculated separately over each set of 24 items with the same number of presentations averaged .69, .45, and .61 for items presented once, twice, or three times. The corresponding JOL-difficulty gamma correlations decreased with number of presentations. These correlations, were .67, .59, and .58, $F(2, 46) = 4.10, p < .05$, for Presentations 1, 2, and 3, respectively.

Discussion

The results of Experiment 3 concur with those of the previous experiments in indicating that whereas intrinsic cues exert similar effects on JOLs and recall, the extrinsic cues associated with item repetition affect JOLs less than they affect recall. More important, this interactive pattern was observed despite the fact that item repetition was manipulated within list and thus could have served to help the discrimination between different items. Apparently, although JOLs are more sensitive to between-item discrimination, they tend to reflect more strongly item differences that are associated with their intrinsic characteristics than with those that ensue from extrinsic effects.

The resolution results are difficult to interpret. Although item repetition might be expected to increase sensitivity to item differences, the within-list manipulation of number of presentations may have masked sensitivity to item-specific differences. The observation that the JOL-difficulty correlation decreased from Presentation 1 to Presentation 3 suggests that there was nevertheless a shift from the predominant reliance on intrinsic cues towards increased reliance on mnemonic cues.

Experiment 4

Experiment 4 extended investigation to another extrinsic factor: presentation time. In all of the previous experiments, degree of learning was manipulated by varying number of study trials, with each trial involving the same presentation duration. In Experiment 4, in contrast, degree of learning was manipulated by varying the presentation duration of different items in the list. The experiment was similar to Experiment 3 except that each pair of words appeared for one trial only, but presentation duration, manipulated within the same list, varied between 2, 4, and 8 s. The question is whether the effects of presentation duration, like those of item repetition, are discounted in the recall predictions. Results suggestive of this possibility have been reported by Groninger (1979). The second question is whether JOL accuracy increases with the amount of study time, even when presentation duration is manipulated within list. This has been found to be the case by Groninger (1979) for the recall of a list of words when study time was manipulated between participants.

Method

Participants. Thirty Hebrew-speaking undergraduates (21 women and 9 men) participated in the experiment for course credit.

Materials. The stimuli were the 72 word pairs used in Experiment 3. These were divided into three matched sets of 24 pairs, as in Experiment 3.

Procedure. The apparatus was the same as in Experiment 3. The procedure was also the same except for the following. For the study phase each of the three sets of pairs was assigned to one of three presentation durations, 2, 4, and 8 s, with the assignment counterbalanced across participants. The entire list was then presented in random order so that presentation duration of the stimulus–response pairs varied randomly within the list. Participants were told that different pairs would be presented for different durations, and that they should spend the entire presentation duration to memorize each pair. Participants indicated their JOLs as soon as the pair disappeared from the screen. In the test phase, the stimulus members of all pairs were displayed in a new random order, each appearing for 5 s.

Results

The effects of presentation duration and difficulty on JOLs and recall. A three-way ANOVA, Presentation Duration × Difficulty × Measure (JOL vs. recall) yielded $F(1, 29) = 1.11$, $MSE = 1{,}019.06$, *ns*, for measure; $F(2, 58) = 14.38$, $MSE = 112.84$, $p < .0001$, for duration; and $F(1, 29) = 439.17$, $MSE = 442.74$, $p < .0001$, for difficulty. The Duration × Measure interaction was also significant, $F(2, 58) = 4.64, MSE = 62.95, p < .05$, as was that between difficulty and measure, $F(1, 29) = 6.42, MSE = 300.50, p < .05$. The triple interaction was not significant, $F < 1$.

As can be seen in Figure 9, the Duration × Measure interaction derives from the fact that the effects of presentation duration were stronger for recall than for JOLs, consistent with the hypothesis that the effects of extrinsic factors are discounted in recall predictions. Thus, the overall effect of presentation duration (from 2 to 8 s) amounted to 10% for recall but only to 4% for JOL. The interaction between presentation duration and measure was found for difficult items, $F(2, 58) = 3.59, p < .05$, but not for easy items, $F(2, 58) = 2.16$.

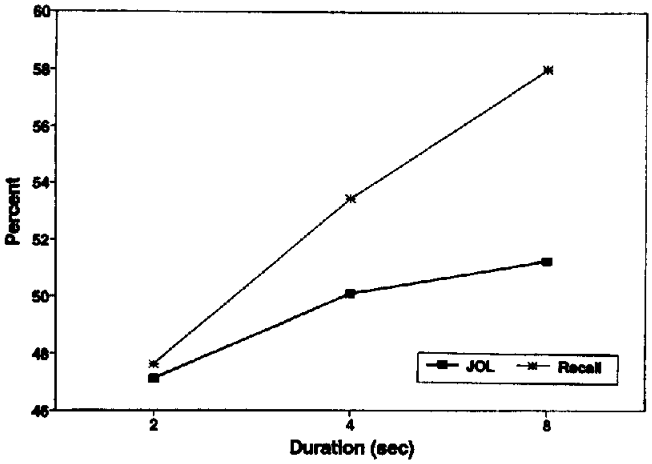

Figure 9. Mean judgment of learning (JOL) and recall as a function of presentation duration (Experiment 4).

The main effects of item difficulty were somewhat inconsistent with those observed in the previous experiments. Here the effects of difficulty were in fact, stronger for recall than for JOL: The difference between difficult and easy items amounted to 51.1 for recall and 41.8 for JOL.

Resolution. JOL-recall gamma correlations calculated over the entire list of 72 items averaged .68, significantly different from zero, $t(29) = 21.19$, $p < .0001$. The respective correlation between JOL and difficulty was .65, $t(29) = 34.48$, $p < .0001$, and that between recall and difficulty was also .65, $t(29) = 24.27$, $p < .0001$.

When the JOL-recall gammas were computed separately over each set of 24 items with the same presentation duration, they averaged .72, .65, and .66 for presentation durations of 2, 4, and 8 s, respectively. A one-way ANOVA on these data yielded $F(2, 58) = 1.56$, *ns*. Thus, there was no sign that the accuracy of JOL is better for items presented for longer study times than for those presented for shorter study times. The respective correlations between JOLs and difficulty were .67, .65, and .64, respectively, $F < 1$.

Discussion

Experiment 4 differed from the previous experiments in that the manipulated variable was presentation duration rather than number of presentations. Note, however, that according to the total time hypothesis (see Cooper & Pantle, 1967), both variables should be expected to have similar effects on memory performance because what matters is the total time spent studying an item, regardless of the number of presentations used. Indeed, the results parallel those obtained with number of presentations: Presentation duration affected JOLs less strongly than it affected recall. Once again, the pattern obtained is consistent with the idea that extrinsic factors are discounted in predictions of recall.

The resolution data yielded little effects of presentation duration. Again it is possible that this is because study time was manipulated within list. Perhaps participants could achieve a finer discrimination between items presented for longer study times than between items presented for shorter study times if study time were manipulated between lists. Indeed Groninger (1979), who manipulated study time between participants found a stronger relationship between predicted recall and actual recall of a list of words when the words were presented at a 6-s rate than when they were presented at a 3-s rate.

General Discussion

The present study disclosed several interactions concerning predicted and actual recall. First, whereas intrinsic factors exerted roughly equivalent effects on JOLs and recall, extrinsic factors had consistently weaker effects on JOLs than they had on recall, resulting in increased underconfidence with degree of learning. Second, although repeated study impaired calibration, it nevertheless consistently improved resolution (i.e., the ability to discriminate between items that are likely to be recalled and those that are not). Finally, the correlational pattern between JOLs, recall, and item difficulty changed systematically with practice in a manner suggesting decreased reliance on intrinsic cues and increased reliance on internal, mnemonic cues. Let me first discuss the general conceptual framework underlying the present study before elaborating on the specific findings.

The Cue-Utilization Approach to JOLs

The interactive pattern of results described previously can be more readily accommodated by the cue-utilization view of JOLs, than by the trace-access view. These two views differ with regard to the two basic questions about JOLs, first, what is the basis for JOLs, and second, what is the reason for their accuracy.

The trace-access view (see Cohen et al., 1991; King et al., 1980; Schwartz, 1994) offers a straightforward answer to both questions. According to this view, JOLs are based on the direct monitoring of the strength of the item's memory trace at the time of making JOLs. Although this assumption seems overly simple, it still pervades many of the discussions in the literature and, in fact, may have served as the catalyzing force for investigating feeling of knowing judgments during learning. The advantage of the trace-access approach is that it explains JOL accuracy: JOLs are accurate in predicting recall because both JOLs and recall are sensitive to memory strength. If this is correct, then any factor that enhances degree of learning should have similar effects on both JOLs and recall (see Mazzoni & Nelson, 1995). Of course, the correlation between JOLs and recall performance is not expected to be perfect, first, because JOLs and recall are tested at different points in time and, second, because recall possibly depends on other factors besides trace strength. Nevertheless, some degree of correspondence between the two variables may be expected both across items and across conditions. Thus, for example, item repetition or study time should have the same effects on recall and JOLs, which they did not.

The cue-utilization view, in contrast, assumes that JOLs are inferential in nature. They are based on a variety of beliefs and cues that are more or less predictive of future memory performance. Thus, in making JOLs, people may rely on theories or beliefs about their own general or specific memory abilities (e.g., "I have a poor memory for names"), about the relative difficulty of different memory tests (e.g., recall is more difficult than recognition), about the memorial consequences of different encoding operations and of different study and test situations, and so on. These beliefs, in conjunction with some pertinent cues about the task and about the difficulty of different items, can help participants make sensible memory predictions even prior to actual study. In addition, people can rely on mnemonic–experiential cues that are more or less diagnostic of future memory performance. These may include familiarity, fluency of processing, the ease with which information comes to mind, and so on.

The cue-utilization approach to JOLs is consistent with some of the current views regarding the basis of other metacognitive judgments such as FOK and subjective confidence. FOK judgements have been said to depend on the application of domain-specific knowledge to make educated inferences about future memory performance (e.g., Nelson, Gerler, & Narens, 1984) and on mnemonic–experiential cues such as the familiarity of the stimulus that serves to probe the solicited target (e.g. Metcalfe et al., 1993; Reder, 1987) or the accessibility of pertinent partial information about the target (Koriat, 1993). Subjective confidence too has been assumed to depend on the consideration and weighting of supporting and contradicting evidence (e.g., Koriat, Lichtenstein, & Fischhoff, 1980) or on such experiential cues as the ease with which an answer is reached (Kelley & Lindsay, 1993; Nelson & Narens, 1990). In a similar manner, JOLs elicited during the study phase may also rest on theory-based considerations or on such mnemonic cues as familiarity or ease of processing rather than on the direct monitoring of the strength of the underlying memory trace.

If JOLs are inferential in nature, then their accuracy is not guaranteed but should vary with the predictive validity of the cues under the particular conditions in question. Furthermore, some systematic discrepancies may be expected between JOLs and recall because of incorrect beliefs or to biases inherent in the heuristics used. Such discrepancies have indeed been reported with regard to FOK judgments elicited following recall failure (e.g., Koriat, 1995; Koriat & Leiblich, 1977; Metcalfe et al., 1993; Reder & Ritter, 1992; Reder & Schunn, 1996; Schwartz & Metcalfe, 1992), as well as with regard to retrospective subjective confidence (e.g., Chandler, 1994; Fischhoff, Slovic, & Lichtenstein, 1977; Kelley & Lindsay, 1993). With regard to JOLs too, the results of the present study join with those of previous findings in demonstrating systematic discrepancies between JOLs and memory performance. These discrepancies have important implications both regarding the possible of JOLs and regarding the factors that contribute to their accuracy and inaccuracy. Let us examine some of these implications.

The Multiple Basis of JOLs

The present study is predicated on the assumption that JOLs are multiply determined (for a similar view regarding FOK judgments, see Leonesio & Nelson, 1990; Nelson et al., 1984): A variety of cues can feed into the computation of JOLs, and the extent to which each of them is used may vary from one situation to another. As noted previously these cues are best revealed when JOLs are found to deviate systematically from actual memory performance.

For example, Begg and his associates (Begg et al., 1989, 1991) documented several such deviations that seem to derive from incorrect beliefs: Participants expected better memory performance following interactive imagery than following separate imagery and following generate instructions than read instructions, but neither of these variables actually affected memory performance. Such incorrect beliefs are often best reflected in JOLs made during initial study and may change following experience with the task (e.g., Begg et al., 1989; Herrmann, Grubs, Sigmundi, & Grueneich, 1986).

Dissociations between JOLs and memory performance may also disclose the operation of particular heuristics. Thus Begg et al. (1989) found that JOLs are accurate with regard to the comparison between concrete and abstract words: Concrete words yielded higher JOLs and better recognition memory than abstract words. In contrast, JOLs were inaccurate with regard to the effect of word frequency because common words yielded higher JOLs than rare words, but the opposite pattern was found for recognition memory. These results were seen to support the idea that JOLs are based on the ease of processing of an item at the time of making the judgment. Concrete words and high frequency words are easier to process and are therefore associated with higher JOLs. The implication, then, is that different factors either enhance or reduce JOL accuracy depending on whether or not they affect JOLs in the same way that they affect memory performance.

Another example of a dissociation that seems to ensue from the use of a particular heuristic was reported by Benjamin, Bjork, and Schwartz (in press). Capitalizing on an earlier finding by Gardiner, Craik, and Bleasdale (1973), they found that initial difficulty in generating an answer to a question, as indexed by recall latency, correlated positively with the probability of recalling that answer in a later free-recall test, but negatively with JOLs about eventual recall. These results suggest that the ease with which answers are accessed is one of the mnemonic cues for JOLs. This cue is a generally good predictor of future memory performance (see Koriat, 1993; Leonesio & Nelson, 1990) but may be invalid in some situations.

Another finding reported in Narens, Jameson, and Lee (1994) also supports the contribution of ease of access to JOLs: Subthreshold presentation of target words in a paired-associates task increased JOLs without affecting final recall. Perhaps the effects of priming are transitory and therefore influence JOLs but not later recall performance.

Needless to say, the dissociations mentioned previously, as well as other dissociations reported in the literature (e.g., Cutting, 1975; Mazzoni & Nelson, 1995; Shaughnessy, 1981; Zechmeister & Shaughnessy, 1980) are more difficult to accommodate by the trace-access view of JOLs than by the cue-utilization approach.

Furthermore, on the whole, the results of the present study also suggest that JOLs are determined by a multiplicity of factors and cannot be explained in terms of any one factor alone. In this article I proposed a tentative classification of the various factors that affect item-by-item JOLs, as well as a distinction between two modes of influence. Thus, putting aside such factors as the person's beliefs about his or her memory efficacy (e.g. Hertzog et al., 1990), it was proposed to distinguish between three classes of factors: intrinsic, extrinsic, and mnemonic. Furthermore, it was argued that intrinsic and extrinsic factors can affect JOLs directly, but they may also affect JOLs indirectly, through the mediation of mnemonic factors. Let me now examine the results from the perspective of these distinctions.

The Contribution of Intrinsic and Extrinsic Factors

The results of the present study support the usefulness of distinguishing between the effects of intrinsic and extrinsic factors. In all four experiments differences in the intrinsic factor of item difficulty had substantial effects on both JOLs and recall. This was true whether the estimate of item difficulty was based on subjective judgments (Experiments 1, 3, and 4) or on word-association norms (Experiment 2). Whereas Experiment 1 yielded somewhat stronger effects of item difficulty on JOLs than on recall, Experiment 4 yielded the opposite pattern, and Experiments 2 and 3 yielded similar effects for both criteria. On the whole, then, the intrinsic factor of item difficulty seems to have equivalent effects on JOLs as it does on recall. Therefore, simply basing JOLs on the perceived relative difficulty of the items is likely to contribute to the accuracy of JOLs.

In contrast to the effects of item difficulty, those of the extrinsic factors were consistently discounted across the four experiments. This was true for the effects of list repetition (Experiments 1 and 2), within-list item repetition (Experiment 3), and study time (Experiment 4). All three manipulations yielded significant effects on JOLs, but these were weaker than those obtained for recall which resulted in increased underconfidence with increasing degree of learning.

The tendency to discount the effects of extrinsic factors in predictions of memory performance is generally consistent with the results of previous studies (e.g., Cutting, 1975; Rabinowitz et al., 1982; Shaughnessy, 1981; Zechmeister & Shaughnessy, 1980). Some of these studies indicate that extrinsic factors such as type of encoding exert stronger effects on JOLs when manipulated within person than when manipulated between persons (see e.g., Begg et al., 1989; 1991), possibly because a within-person manipulation enables participants to use extrinsic cues discriminately. Note, however, that in all of the experiments of the present study, the extrinsic factors were manipulated within person, and in Experiments 3 and 4, they were also manipulated within the same list. Although this latter manipulation could help discrimination between items, its effects on JOL were nevertheless weaker than its effects on recall.

The results of Experiments 1 and 2 are particularly surprising because of the nature of the extrinsic factor manipulated: number of presentations. It might have been expected that practice should improve calibration because participants learn more about the task and about their performance with increasing repetitions. However, the general effect of practice was to impair calibration by increasing underconfidence. More important, in Experiment 2 this trend was equally observed whether or not participants were given feedback about the correctness of their answers. Feedback did help reduce the overall discrepancy between mean JOLs and mean recall, but nevertheless the tendency to discount the effects of repetition was observed even in the feedback condition. Possibly, feedback makes participants aware that their overall performance is actually better than they have predicted, but this increased awareness does not get to be extrapolated to subsequent presentations.

The Contribution of Mnemonic Cues: Changes in the Basis of JOLs and Their Accuracy With Practice

Although practice learning the same items impaired calibration in Experiments 1 and 2, it nevertheless resulted in a systematic improvement in resolution (i.e., in the cross-item correlation between JOLs and recall). This was true in Experiment 1, and also in both the feedback and the no-feedback conditions of Experiment 2. It would seem that the primary benefit of practice is to increase sensitivity to cues that discriminate between items rather than to those that characterize all items alike. More important, the improvement in the predictive validity of JOLs with practice was accompanied by a systematic reduction in the size of the correlation between JOLs and item difficulty.

What is the source of these changes? According to the present formulation, they reflect the increased reliance on internal, mnemonic cues that are diagnostic of the relative recallability of different targets. In the first presentation of the list, JOLs are primarily based on the perceived a priori difficulty of the items (see Figure 5) and should not be very different from the kind of ease-of-learning (EOL) judgments elicited before learning (see Leonesio & Nelson, 1990). Indeed, in all four experiments initial (or only) JOLs were highly correlated with the item-difficulty measure, and these correlations were consistently higher than the respective correlations with recall on the first (or only) test.

With practice learning the same items, JOLs become increasingly more tuned to internal cues that reflect the degree to which the studied items have been mastered. The evidence for the increased reliance on mnemonic cues with practice comes from two observations. The first is the correlational pattern noted previously, in which JOLs become increasingly more accurate and at the same time exhibit lesser dependency on the items' a priori difficulty. The second observation is that the improved accuracy of JOLs appears to be due to the use of idosyncratic, item-specific cues rather than because of cues that are commonly shared by all participants. The idosyncratic cues apparently reflect privileged access to personal aspects of encoding and remembering.

The observation that JOLs on one trial were most strongly associated with performance on the immediately preceding trial suggests the possibility that the increased resolution of JOLs with practice is particularly due to repeated testing. This may explain why resolution did not increase in Experiment 3 with repeated study of the items. Perhaps the opportunity to retrieve the response enhances the likelihood of future retrievals and at the same time provides valid internal cues for JOLs (see Benjamin & Bjork, 1996). If this is indeed the case, then repeated testing may be one way in which the subjective experience underlying JOLs can be educated (see Jacoby, Bjork, & Kelley, 1994). It should be noted, however, that JOL accuracy has also been found to increase with repeated study without repeated testing (e.g., Lovelace, 1984), though part of this improvement may also be due to implicit retrieval of the response during study.

The idea that improved resolution is specifically due to mnemonic cues pertaining to ease of retrieval may also explain the delayed JOL effect reported by Nelson and Dunlosky (1991). They found JOLs to be considerably more accurate when they are delayed until shortly after study than when they are made immediately after study. Possibly, when JOLs are delayed, participants rely more heavily on cues pertaining to the ease with which the target can be retrieved. Two observations are consistent with this interpretation. First, more participants in the delayed than in the immediate condition reported attempting recall the correct response when making the JOL. Second, the delayed JOL effect occurs only when JOL is cued by the stimulus alone not when cued by the entire stimulus–response pair. Possibly, the presentation of the response deprives participants of ease of access as a basis for making JOLs in much the same way that when participants are asked to evaluate anagram difficulty in the presence of the solution, they are deprived of the subjective experience of solving the anagram as a basis for their judgments (Kelley & Jacoby, 1996a; see next). These observations suggest that delayed JOLs are accurate because participants can make use of mnemonic cues concerning the accessibility of the target or partial clues about it. As mentioned earlier, Dunlosky and Nelson (1994) also observed that although immediate JOLs were relatively insensitive to encoding manipulations that affected eventual recall, delayed JOLs did vary with such manipulations. Apparently, then, delayed JOLs are more diagnostic of future recall because they rely more heavily on mnemonic cues.

Recent results by Carroll et al. (1997) are also consistent with this interpretation and suggest that delaying JOLs increases the relative weight of mnemonic as opposed to intrinsic factors. In their study, mentioned earlier, related paired associates that were studied to a criterion of two correct recalls evoked higher JOLs than unrelated pairs studied to a criterion of eight correct recalls, whereas eventual recall, tested after 2 or 6 weeks, evidenced the opposite pattern. However, when JOL ratings were delayed by 1 day, they yielded a similar pattern to that obtained for eventual recall, namely, a greater sensitivity to amount of learning than to semantic relatedness. The greater sensitivity of delayed JOLs to number of presentations possibly reflects a heavier reliance on mnemonic cues that are correlated with number of presentations. It is unlikely that this effect is due to the direct use of the extrinsic factor of number of presentations (e.g., "I should recall this item, because I have seen it so many times") because this would imply better memory for the study history of an item a day after study than a short while after study.

Two Modes of Influence?

An important assumption of the conceptual scheme presented in this article concerns the distinction between two modes of influence on JOLs: the rule-based influence underlying the direct effects of intrinsic and extrinsic factors and the heuristic-based effects underlying the reliance on mnemonic cues. This distinction parallels that advanced by others (Brown & Siegler, 1993; Jacoby & Brooks, 1984; Kelley & Jacoby, 1996a; Strack, 1992). The two types of influence are assumed to involve qualitatively different processes.

Consider, for example, the finding that JOLs are affected by the person's belief about his or her general memory efficacy (see Hertzog et al., 1990). This effect is likely to be based on the deliberate use of one's theory about one's own memory. In a similar manner, one may judge that concrete words ought to be remembered better than abstract words and that words encountered several times should be less likely to be forgotten than those encountered just once. Such beliefs can affect JOLs directly, through the application of a particular theory about the influence of various factors on memory performance.

Intrinsic and extrinsic factors, however can influence JOLs indirectly through their effects on internal, mnemonic cues. Thus, for example, previous encounters with a stimulus make its later processing faster (see Jacoby & Dallas, 1981) and may therefore enhance JOLs through their effects on fluent processing. In general, many of the intrinsic and extrinsic cues that affect actual memory performance also leave their mark on internal–experiential qualities of processing such as perceptual fluency and the ease with which information comes to mind, and these qualities can serve as relatively valid, global cues for predictions of subsequent memory performance (see Kelley & Jacoby, 1996a, 1996b).

How do the two modes of influence on JOLs differ? Generally speaking, the direct effects of intrinsic and extrinsic factors tend to entail an analytic, deliberate, and conscious deduction, such as the inference that "I ought to remember" a particular item because it is very easy. This deduction is based on a domain-specific, a priori theory. In contrast, the effects of mnemonic cues tend to be based on the application of a general purpose nonanalytic heuristic that yields a global "feeling of knowing" or hunch. In discussing the fluency heuristic, Jacoby and his associates (e.g., Jacoby, 1991; Jacoby, Toth, & Yonelinas, 1993) invoked this distinction in contrasting the effects of automatic and intentional uses of memory. They proposed that intentional, recollection-based uses afford greater personal control than familiarity-based processes. Kelley and Jacoby (1996a) applied the distinction between theory-based and experience-based processes to the analysis of how participants evaluate the difficulty of anagrams for others: Difficulty judgments may be based on one's a priori theory about the factors that affect anagram difficulty in general or on the participant's own experience attempting to solve the anagram.

In terms of Kelly and Jacoby's contrast, the effects of practice observed in Experiments 1 and 2 can be seen to reflect a general shift from theory-based to experience-based judgments. Ease-of-learning (EOL) judgments made by participants before study as well as JOLs made by participants after the first presentation of an item are heavily based on participants' a priori beliefs, whereas those elicited, for example, following the fourth study presentation in Experiment 2 are more heavily based on experiential–mnemonic cues. This is not to say that EOL judgments are not influenced by mnemonic cues: In the same way that judgments of anagram difficulty (for others) can be based on the person's own attempt to solve the anagrams, so judgments regarding the difficulty of different items in our experiments can also be based on observing one's own success in attempting to commit the items to memory. In fact, Begg et al. (1989) proposed that JOLs in general are based on ease of processing and reported results indicating that JOLs are affected by the same variables as judgments of ease of processing. These findings, however, do not necessarily indicate that ease of processing, as a mnemonic–experiential cue underlies all JOLs, because ease of processing judgments possibly reflect the contribution of theory-based processes as well.

An alternative interpretation to the changes that occur with practice (see Figure 6) is that JOLs are based on the same mnemonic cue across all presentations but that the determinants of that mnemonic cue change with practice. Although this possibility cannot be ruled out, it is difficult to see how it can accommodate the two observations noted previously, first, that the JOL-difficulty correlation diminishes systematically with practice (Experiments 1 and 2) and, second, that the improvement in JOL accuracy with practice seems to ensue from greater reliance on idiosyncratic cues. Future work, however, can explore further the possibility of a change from theory-based to mnemonic-based effects by taking advantage of some of the expected differences between these two modes of influence. Thus, as noted above, it has been proposed that theory-based judgments allow participants greater control in escaping the influence of factors that are seen to be irrelevant than experience-based judgments (Kelley & Jacoby, 1996a). Also, experience-based judgments have been assumed to be easier to reach than theory-based judgments (Kelley & Jacoby, 1996a; Strack, 1992). It would be instructive to explore the possibility that both of these differences distinguish between JOLs elicited earlier and later during learning.

As far as the accuracy of JOLs is concerned, it should be stressed that both theory-based and heuristic-based JOLs can be accurate to different degrees, but their accuracies and inaccuracies possibly derive from different sources. The accuracy of theory-based JOLs should depend on the adequacy of the theory or the belief in question and on its applicability to the particular circumstances studied. Sometimes feedback may help rectify wrong beliefs, resulting in improved JOL accuracy (Herrmann et al., 1986). The accuracy of heuristic-based JOLs, on the other hand, should depend on the validity of the pertinent internal cues in predicting actual memory performance. This validity generally reflects the extent to which the effective cue in question (e.g., fluency or accessibility) is affected by the same factors and in the same way as the criterion memory performance (see Begg et al., 1989). These differences in the determinants of JOL accuracy can also be exploited in distinguishing between the two modes of influence.

Methodological Implications: Distinguishing Calibration and Resolution

A final note concerns methodology. The differential effects of practice on calibration and resolution highlight the need to distinguish between absolute and relative measures of the correspondence between subjective and objective measures of performance (see, e.g., Keren, 1991; Lichtenstein et al., 1982; see also Brown & Siegler's, 1993, distinction between metric properties and mapping properties). Examinations of this correspondence in the context of decision-making studies place a particular emphasis on calibration (i.e., on the absolute correspondence between assessed probability and the likelihood that the response is correct; see Allwood & Granhag, 1996; Arkes, Christensen, Lai, & Blumer, 1987; Juslin, 1994; Keren, 1991). Thus, the typical finding in the literature is that of overconfidence in the correctness of one's decisions. Students of metamemory, in contrast, have placed a far heavier emphasis on resolution, that is, on the cross-item correlation between metacognitive judgments and actual memory performance (see Nelson, 1984; Schwartz & Metcalfe, 1994). In fact, in the great majority of studies of metamemory, calibration cannot even be evaluated because metacognitive judgments have been solicited on a rating scale or on a dichotomous scale (but see Nelson & Dunlosky, 1991; Koriat, 1994, 1995). Thus, for example, in some of the JOL studies, data on predicted and obtained memory performance were collected on different scales, and, therefore, conclusions regarding the relative insensitivity of JOLs to certain extrinsic factors had to be based on certain assumptions about the underlying scales of measurement (e.g., Shaw & Craik, 1989).

The results of the present study stress the importance of soliciting metacognitive judgments in the form of probabilities and of analyzing the data in terms of calibration and resolution conjointly. Clearly there are conditions for which calibration is critical and others for which resolution is important and calibration is immaterial. Therefore it is important to examine both indexes when drawing conclusions on the general accuracy of metacognitive judgments.

References

Allwood, C. M., & Granhag, P. A. (1996). Considering the knowledge you have: Effects of realism in confidence judgments. *European Journal of Cognitive Psychology, 8,* 235–256.

Arbuckle, T. Y., & Cuddy, L. L. (1969) Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81,* 126–131.

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes, 39,* 133–144.

Bäckman, L., & Lipinska, B. (1993). Monitoring of general knowledge: Evidence for preservation in early Alzheimer's disease. *Neuropsychologia, 31,* 335–345.

Barnes, A. E., Nelson, T. O., Dunlosky, J., Mazzoni, G., & Narens, L. (in press). An integrative system of metamemory components involved in retrieval. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII—Cognitive regulation of performance: Interaction of theory and application.* Cambridge, MA: MIT Press.

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

Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition, 19,* 487–497.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (in press). The mismeasure of memory: When retrieval fluency is misleading. *Journal of Experimental Psychology: General.*

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review, 100,* 511–534.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30,* 542–579.

Butterfield, T. O., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling-of-knowing. *Developmental Psychology, 24,* 654–663.

Carroll, M., & Nelson, T. O. (1993). Effects of overlearning on the feeling of knowing are more detectable in within-subject than in between-subject designs. *American Journal of Psychology, 106,* 227–235.

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.

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & Cognition, 22,* 273–280.

Cohen, R. L., Sandler, S. P., & Keglevich, L. (1991). The failure of memory monitoring in a free recall task. *Canadian Journal of Psychology, 45,* 523–538.

Cooper, E. H., & Pantle, A. J. (1967). The total time hypothesis in verbal learning. *Psychological Bulletin, 68,* 221–234.

Cutting, J. E. (1975). Orienting tasks affect recall performance more than subjective impressions of ability to recall. *Psychological Reports, 36,* 155–158.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOLs) and the delayed-JOL effect. *Memory & Cognition, 20,* 373–380.

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(4),* 545–565.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552–564.

Funnell, M., Metcalfe, J., & Tsapkini, K. (1996). In the mind but not on the tongue: Feeling of knowing in an anomic patient. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 171–194). Hillsdale, NJ: Erlbaum.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1,* 213–216.

Gardiner, J. M., Passmore, C., Herriot, P., & Klee, H. (1977). Memory for remembered events: Effects of response mode and response-produced feedback. *Journal of Verbal Learning and Verbal Behavior, 16,* 45–54.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

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

Heath, L., Tindale, S. R., Edwards, J., Posavac, E. J., Bryant, F. B., Henderson-King, E., Suarez-Balcazar, Y., & Myers, J. (1994). *Applications of heuristics and biases to social issues.* New York: Plenum Press.

Herrmann, D. J., Grubs, L., Sigmundi, R. A., & Grueneich, R. (1986). Awareness of memory ability before and after relevant memory experience. *Human Learning, 5,* 91–107.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging, 5,* 215–227.

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

Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of comprehension, competence, and remembering. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 57–81). Washington, DC: National Academy Press.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception and concept learning. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 1–47). New York: Academic Press.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 3,* 306–340.

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

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

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 2,* 1–16.

Janowsky, J. S., Shimamura, A. P., & Squire, R. L. (1989). Memory and metamemory: Comparisons between frontal lobe lesions and amnesic patients. *Psychology, 17,* 3–11.

Johnson, M. K., Kounios, J., & Reeder, J. (1994). Time course studies of reality monitoring and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1409–1419.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57,* 226–246.

Kelley, C. M., & Jacoby, L. L. (1996a). Adult egocentrism:

Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35,* 157–175.

Kelley, C. M., & Jacoby, L. L. (1996b). Memory attributions: Remembering, knowing, and feeling of knowing. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 287–308). Hillsdale, NJ: Erlbaum.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1–24.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77,* 217–273.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology, 95,* 329–343.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639.

Koriat, A. (1994). Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In J. Metcalfe & P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 115–135). Cambridge, MA: MIT Press.

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. (in press). Illusions of knowing: A window to the link between knowledge and metaknowledge. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions.* London: Sage.

Koriat, A., Ben-Zur, H., & Sheffer, D. (1988). Telling the same story twice: Output monitoring and age. *Journal of Memory and Language, 27,* 23–39.

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

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

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

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

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge, England: Cambridge University Press.

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

Lovelace, E. A., & Marsh, G. R. (1985). Prediction and evaluation of memory performance by young and old adults. *Journal of Gerontology, 40,* 192–197.

Luus, C. A. E., & Wells, G. L. (1994). Eyewitness identification confidence. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 348-361). Cambridge, England: Cambridge University Press.

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.

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition, 18,* 196–204.

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.

Metcalfe, J. (1993). Novelty monitoring, metacognition and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review, 100,* 3–22.

Metcalfe, J. (1996). Metacognitive processes. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition: Memory* (Vol. 10, pp. 381–407). San Diego, CA: Academic Press.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 851–861.

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing.* Cambridge, MA: MIT Press.

Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 223–232.

Narens, L., Jameson, K. A., & Lee, V. A. (1994). Subthreshold priming and memory monitoring. In J. Metcalfe & P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 71–92). Cambridge, MA: MIT Press.

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., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113,* 282–300.

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, and Cognition, 14,* 676–686.

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

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19,* 90–138.

Reder, L. M. (1996). *Implicit memory and metacognition.* Hillsdale, NJ: Erlbaum.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 435–451.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45–78). Hillsdale, NJ: Erlbaum.

Ross, M. (1997). Validating memories. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. Brainerd (Eds.), *Memory for everyday and emotional events* (pp. 49–81). Mahwah, NJ: Erlbaum.

Schacter, D. L. (1991). Unawareness of deficit and unawareness of knowledge in patients with memory disorders. In G. P. Prigatano & D. L. Schacter (Eds.), *Unawareness of deficit after brain injury: Clinical and theoretical issues* (pp. 127–151). New York: Oxford University Press.

Schneider, W. (1985). Development trends in the metamemory-memory behavior relationship: An integrative view. In D. L. Forrest Pressley, G. E. Mackinnon, & T. G. Waller (Eds.), *Metacognition, cognition and human performance* (pp. 57–109). New York: Academic Press.

Schneider, W. (1986). The role of conceptual knowledge and metamemory in the development of organizational processes in memory. *Journal of Experimental Child Psychology, 42,* 218–236.

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feeling of knowing. *Psychonomic Bulletin and Review, 1,* 357–375.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1074–1083.

Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93–113). Cambridge, MA: MIT Press.

Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social Psychology: Handbook of basic principles* (pp. 433–465). New York: Guilford Press.

Shaughnessy, J. J. (1981). Memory monitoring accuracy and modification of rehearsal strategies. *Journal of Verbal Learning and Verbal Behavior, 20,* 216–230.

Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging, 4,* 131–135.

Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 452–460.

Sohlberg, M. M., Mateer, C. A., & Stuss, D. T. (1993). Contemporary approaches to the management of executive control dysfunction. *Journal of Head Trauma Rehabilitation, 8,* 45–58.

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

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence–accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315–327.

Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L. I. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 249–275). Hillsdale, NJ: Erlbaum.

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology, 71,* 673–679.

Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probablistic judgment. *Psychological Bulletin, 110,* 611–617.

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society, 15,* 41–44.

Received April 16, 1996
Revision received July 31, 1996
Accepted December 28, 1996 ■