

The Intricate Relationships Between Monitoring and Control in Metacognition: Lessons for the Cause-and-Effect Relation Between Subjective Experience and Behavior

Asher Koriat, Hilit Ma'ayan, and Ravit Nussinson
University of Haifa

Do we run away because we are frightened, or are we frightened because we run away? The authors address this issue with respect to the relation between metacognitive monitoring and metacognitive control. When self-regulation is goal driven, monitoring affects control processes so that increased processing effort should enhance feelings of competence and feelings of knowing. In contrast, when self-regulation is data driven, such feelings may be based themselves on the feedback from control processes, in which case they should decrease with increasing effort. Evidence for both monitoring-based control and control-based monitoring occurring even in the same situation is presented. The results are discussed with regard to the issue of the cause-and-effect relation between subjective experience and behavior.

Keywords: metacognition, subjective experience, monitoring, control, judgments of learning

A long-standing issue in psychology and philosophy concerns the cause-and-effect relation between phenomenal experience and behavior (Baars, 1988; Bargh, 1997; Bless & Forgas, 2000; Flanagan, 1992; Mandler, 2002; Marcel, 1983a, 1983b; Marcel & Bisach, 1988). Whereas many discussions in cognitive psychology assume that subjective experience can play a causal role in influencing behavior, recent findings lend credence to the idea that subjective experience may be based on the feedback from one's own behavior and thus follow rather than precede behavior. Indeed, in reviewing their own work, Kelley and Jacoby (1998) praised the insight owed to the James–Lange view of emotion, according to which “subjective experience can involve an attribution or unconscious inference about effects on performance and so follow from, rather than be responsible for, objective performance” (pp. 127–128). In this article we address the causal links

between subjective experience and behavior within a restricted domain—that of metacognitive monitoring and metacognitive control. We believe that our analysis and results can provide some insight into the general issue of the relation between subjective experience and behavior.

The Cause-and-Effect Relation Between Subjective Experience and Behavior

Most of the discussions of the status of subjective experience in human behavior have centered on the causal role that consciousness might play in guiding behavior (Schwarz & Clore, 1996). The issue that has been addressed concerns the extent to which phenomenal consciousness affects behavior, in general, and “rational” action, in particular. In Posner and Snyder's (1975) conceptual framework, for example, controlled processes, as opposed to automatic processes, were seen to characterize conscious functioning. Block (1995) associated consciousness with the reflective pursuit of one's goals, arguing that without consciousness one loses the “rational control of action.” In Schacter's (1989) model, the conscious system is assumed to function as the gateway to an executive control system that initiates and regulates voluntary action. Only activations that gain access to consciousness can be used by the executive system and can thus influence voluntary activities (see also Marcel, 1986). Jacoby went even further, using voluntary control as a diagnostic of consciousness (e.g., Jacoby, Lindsay, & Toth, 1992; Jacoby, Ste-Marie, & Toth, 1993): By distinguishing between controlled-conscious processes and automatic-unconscious processes, he emphasized the inhibitory function of awareness in opposing influences that could otherwise prevail in memory and behavior (Jacoby, 1999; Jacoby, Jennings, & Hay, 1996).

Somewhat less effort has been invested in exploring the possible causal role of behavior and performance on subjective experience. However, over the years several formulations have been advanced suggesting that subjective experience may actually follow rather

Asher Koriat, Hilit Ma'ayan, and Ravit Nussinson, Department of Psychology, University of Haifa, Haifa, Israel.

Ravit Nussinson has previously published under the name Ravit Levy-Sadot.

Portions of this work were presented at the British Psychological Society (Cognitive Section) Conference, University of Leeds, September 2004, and at the Sixth Tsukuba International Conference on Memory, March 2005. The research was carried out at the Institute of Information Processing and Decision Making, University of Haifa. It was supported by Israel Science Foundation Grant 928/00 and German Federal Ministry of Education and Research Grant BMBF within the framework of German–Israeli Project Cooperation.

We are grateful to Yaffa Lev and Hana Struminger for programming the experiments, to Rinat Gil for her help in conducting and analyzing the experiments, and to Limor Sheffer for her advice in the statistical analyses. We are indebted to Jeffrey Toth and Bruce Whittlesea for their comments on a draft of this article.

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

than precede behavior. The most cited such formulation is the James–Lange theory regarding the relationship between emotional feelings and emotional behavior. According to this view, an exciting or threatening event elicits certain physiological and behavioral reactions. Subjective emotional experience then occurs as a feedback from these reactions. In James's (1884) words,

Common sense says, we lose our fortune, are sorry and weep; we meet a bear, are frightened and run; we are insulted by a rival, are angry and strike. The hypothesis here to be defended says that this order of sequence is incorrect . . . and that the more rational statement is that we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble, because we are sorry, angry, or fearful, as the case may be. (p. 190)

Some aspects of the James–Lange theory were revived by the work of Schachter and Singer (1962). Two factors were assumed to determine different emotions: the physical changes in a person's body and the interpretations that the person gives to those changes in the light of the stimulus situation. Schachter and Singer showed that activation produced by epinephrine could be experienced either as anger or as happiness, depending on the person's attributions. Thus, as with the James–Lange theory, emotional feelings are assumed to emerge in response to bodily changes. However, an important assumption in the Schachter–Singer view, which is taken up later, is that once an emotional feeling has been produced, that feeling can then cause specific actions (see also Carver & Scheier, 1990). This assumption implies that although subjective experience can emerge in response to the feedback from one's own reactions, it can in turn cause other reactions that are compatible with it.

Several discussions in social psychology also imply that one's feelings, attitudes, and beliefs are based on observing one's own behavior. Bem's (1967) self-perception theory maintains that a person's inner states are based on inferences from observations of one's own overt behavior and its context (Bem, 1965, 1966). Such inferences are functionally similar to those that any outside observer could make about that person. A similar proposal was made by Nisbett and Wilson (1977): People's subjective reports about the reasons for their behavior are based on a post hoc explanation of their behavior in terms of their a priori theories about the possible effects of particular stimuli on particular responses.

Several studies carried out in recent years have explored predictions that follow more directly from William James's proposal. These studies suggest that participants can be induced to experience specific emotional feelings by making them adopt certain behavioral expressions and postures: Participants tend to feel happy when they are induced to smile, angry when they are induced to frown, and more sad when they sit in a slumped posture (e.g., Duclos et al., 1989; Laird & Bresler, 1992; Zajonc, 1985; see Adelman & Zajonc, 1989, and Strack & Deutsch, 2004, for reviews). These effects were obtained even when body postures and facial expressions were manipulated unobtrusively (e.g., Stepper & Strack, 1993).

Another line of research in social psychology, which is more closely linked to the work reported in this article, concerns the metacognitive experiences that accompany information processing and behavior. Underlying that research is the assumption that people's judgments are sometimes based on the retrospective inspection of their own cognitive processes and performance,

particularly the ease or fluency with which information is encoded or retrieved (for reviews, see Benjamin & Bjork, 1996; Schwarz, 2004; Winkielman, Schwarz, Fazendeiro, & Reber, 2003). In a classic study (Schwarz et al., 1991), participants who were asked to recall 12 examples of their own assertive behavior subsequently rated themselves as less assertive than participants who had to recall only 6 such examples. Arguably, the effort needed to retrieve many examples led participants to the inference that they were not very assertive. Similarly, participants who were asked to recall 12 childhood events subsequently rated their childhood memory as poorer than participants who had to recall only 4 events (Winkielman, Schwarz, & Belli, 1998). Stepper and Strack (1993), who had participants recall 6 examples of assertive behavior, observed that those who did so while contracting the corrugator muscle (producing an expression associated with a feeling of effort) subsequently judged themselves as less assertive than those who contracted the zygomaticus muscle (producing an expression associated with a feeling of ease). These and many similar experiments support the idea that people's judgments are influenced by the feedback from their own performance and behavior.

Of still greater affinity to the proposal to be detailed below is the work of Jacoby, Kelley, and Whittlesea. Jacoby, Kelley, and their associates (e.g., Jacoby & Dallas, 1981; Kelley & Jacoby, 1998) provided ample evidence in support of their view that subjective experience is formed as a result of a process in which the fluent processing of a stimulus is attributed (or misattributed) unconsciously to a previous encounter with the stimulus or to its perceptual qualities. For example, the subjective experience of familiarity or visual brightness is based on the interpretation of variations in one's own performance. Whittlesea and his associates (Whittlesea, 1997, 2003) also incorporate the assumption that the specific subjective feelings experienced are based on the interpretation of one's own performance in the light of one's intuitive theory (see General Discussion).

The Relationship Between Metacognitive Monitoring and Metacognitive Control

The experimental work to be reported in this article concerns the relation between monitoring and control in metacognition. To introduce the logic underlying that work, we shall draw an analogy from emotional behavior. As noted earlier, the question raised by William James (1884) is whether we run away because we are frightened or we are frightened because we run away. The first option assumes that the behavioral response to a threatening situation is mediated by the feeling of fear: A conscious or unconscious appraisal of the situation, based on a variety of cues, may give rise to the feeling of fear (Lazarus, 1966), which then leads to escape behavior. The second option is that flight behavior is a direct response to the external situation; it is either automatically triggered by the external circumstances or represents a self-initiated coping response intended to avoid threat. It is the feedback from running away that then causes the subjective feeling of fear.

Because the feeling of fear and the action of running away generally go hand in hand, how can we tell which is the cause and which is the effect? One possible approach is to consider the strength of each of the two variables. Assume that it is indeed the subjective feeling of fear that causes one to run away from the

danger. Then, the faster one runs away, the less fear one should experience after running away. In contrast, if it is the feedback from running away that gives rise to the subjective feeling of fear, then the faster one runs away the *more* fear one should experience. It is this general logic that underlies our investigation of the relationship between monitoring and control processes in metacognition.

The dominant view in current theorizing on metacognition emphasizes the causal link from subjective experience to behavior or, more specifically, from monitoring to control (e.g., Barnes, Nelson, Dunlosky, Mazzone, & Narens, 1999; Koriat & Goldsmith, 1996; Nelson, 1996; Nelson & Narens, 1990; Son & Schwartz, 2002). Metacognitive monitoring refers to the subjective assessment of one's own cognitive processes and knowledge, whereas control refers to the processes that regulate cognitive processes and behavior. In their analysis of the relationship between monitoring and control, Nelson and Narens (1990, 1994) proposed a distinction between an object level and a metalevel. The metalevel is assumed to monitor the processes that take place at the object level and control them accordingly. Thus, for example, during the study of new material, learners are assumed to monitor subjectively the degree of learning and to allocate further learning resources according to the monitoring output.

The idea that metacognitive feelings affect metacognitive control derives from a functional approach to metacognition, which emphasizes the adaptive value of putting subjective monitoring to use in regulating one's own behavior (e.g., Hart, 1965; Koriat & Goldsmith, 1996; Nelson, Dunlosky, Graf, & Narens, 1994). This approach can be illustrated by Hart's analysis of the feeling of knowing (FOK) that is experienced when one attempts to retrieve a solicited item from memory. Hart (1965, 1967) stressed the functional value of FOK as an internal monitor that signals whether the solicited piece of information is stored in memory. According to him, FOK

can serve as an indicator of what is stored in memory when the retrieval of a memory item is temporarily unsuccessful or interrupted. If the indicator signals that an item is not in storage, then the system will not continue to expend useless effort and time at retrieval; instead, input can be sought that will put the item into storage. Or if the indicator signals that an item is in storage, then the system will avoid redundantly inputting information that is already possessed. (Hart, 1965, p. 214)

The functional view of monitoring reflected in this quote derives its impetus from two general observations in metacognition: first, that people are generally accurate in monitoring their knowledge, and second, that controlled processes appear to be tuned to the output of subjective monitoring.

With regard to the first observation, many studies have demonstrated positive correlations across items between subjective and objective indexes of knowing, suggesting that by and large, people can monitor the relative accuracy of their knowledge. This has been found to be the case across a variety of metacognitive judgments: Judgments of learning (JOL) made about different items during study are moderately predictive of the relative future recall or recognition of these items (e.g., Arbuttle & Cuddy, 1969; Dunlosky & Nelson, 1994; Koriat, 1997; Koriat, Sheffer, & Ma'ayan, 2002; Lovelace, 1984; Mazzone & Nelson, 1995; Zechmeister & Shaughnessy, 1980). Similarly, FOK judgments elicited

following a recall failure are predictive of the likelihood of recalling the illusive target at some later time or recognizing it among distractors (Gruneberg & Monks, 1974; Hart, 1965; Koriat, 1993; Schwartz & Metcalfe, 1994). Finally, confidence judgments in an answer are generally diagnostic of its correctness (e.g., Koriat & Goldsmith, 1996; Robinson, Johnson, & Herndon, 1997). Admittedly, dissociations between subjective and objective indexes of knowing have been observed in some circumscribed situations to the extent that metacognitive judgments were undiagnostic or even counterdiagnostic of actual memory performance (Benjamin, Bjork, & Schwartz, 1998; Chandler, 1994; Koriat, 1995; Leippe, 1980; Metcalfe, Schwartz, & Joaquim, 1993; Reder & Ritter, 1992). However, these are the exception rather than the rule.¹

The second observation concerns the control component of metacognition. Several observations suggest that metacognitive judgments play a critical role in the strategic regulation of information processing and behavior, thus highlighting the functional value of their accuracy. For example, as is discussed in detail below, when learners are allowed to control the time spent studying each item in a list, they generally allocate more time to items associated with lower than with higher ease-of-learning (EOL) or JOL ratings (see Son & Metcalfe, 2000). This observation has been taken to indicate that learners monitor degree of learning and use their JOLs as a basis for regulating the allocation of study time to different items (see Dunlosky & Hertzog, 1998; Nelson & Leone-sio, 1988). With regard to FOK judgments, several findings suggest that a positive FOK drives memory search: Participants spend more time searching for an elusive memory target when they feel that the target is available in memory than when they feel that it is not available (e.g., Barnes et al., 1999; Costermans, Lories, & Ansay, 1992; Gruneberg, Monks, & Sykes, 1977; Nelson & Narens, 1990). In addition, Reder (1987; see also Nhuyvanisvong & Reder, 1998) observed that the preliminary FOK associated with a question guides the strategy of question answering. Finally, confidence judgments in the correctness of retrieved information have also been assumed to play a role in guiding memory reports: In reporting about a witnessed past event, people tend to volunteer or withhold a piece of information that comes to mind depending on their subjective confidence in its correctness (Koriat & Goldsmith, 1996).

Taken together these observations suggest the following story, which we shall dub *Story 1*: The fact that metacognitive judgments are generally accurate in predicting memory performance makes them a useful basis for regulating information processing. Such regulation should have an adaptive value in terms of improving the effectiveness of cognitive performance. According to Story 1, then, metacognitive feelings play a mediating role similar to that of fear in the first option mentioned earlier: Once such feelings have been formed on the basis of whatever cues available, they can be

¹ This aspect of metacognitive accuracy, which is labeled *resolution* or *relative accuracy* (see Koriat et al., 2002; Nelson & Dunlosky, 1991), is commonly indexed by a within-participant gamma correlation between metacognitive judgments and actual memory performance (Nelson, 1984). In contrast to resolution, calibration (or absolute accuracy), which refers to the correspondence between mean metacognitive judgments and mean actual performance and reflects the extent to which metacognitive judgments are realistic, tends to be quite poor, generally exhibiting overconfidence (see Metcalfe, 1998).

used to guide action (Koriat, 2000; Koriat & Levy-Sadot, 1999; Nelson & Narens, 1990). In fact, it would seem that the postulated causal link from monitoring to control is responsible in part for the recent upsurge of interest in metacognition. This interest derives from the conviction that metacognitive feelings are not mere epiphenomena but play a causal role in influencing and guiding one's own behavior (Koriat & Goldsmith, 1996; Nelson, 1996).

In this article, however, we explore the possibility that the basic observations mentioned above may also be telling a different story, which we shall dub *Story 2*. In contrast to the commonly assumed "monitoring affects control" hypothesis (Nelson & Leonesio, 1988) underlying *Story 1*, *Story 2* emphasizes the reverse causal link from control to monitoring. It asserts the following: First, the correlation between monitoring and control processes derives from the fact that metacognitive judgments are based on the feedback from the outcome of control operations. This is like the idea that the feeling of fear is based on the feedback from running away. The implication is that monitoring does not precede controlled action but follows it, being retrospective rather than prospective in nature. Second, the accuracy of monitoring judgments in predicting actual memory performance derives precisely from the fact that these judgments are based on the feedback from the outcome of control operations. Thus, it is not because monitoring judgments are accurate that they are used as a basis for strategic control. Rather, it is because metacognitive judgments rely on the feedback from control operations that they are generally accurate.

The assumption underlying *Story 2* can be illustrated by the following quote about the FOK, which may be contrasted with the quote from Hart presented earlier: "It is by attempting to search for the solicited target that one can judge the likelihood that the target resides in memory and is worth continuing to search for" (Koriat, 1995, p. 312). The assumption in this quote is that people do not consult their FOK in order to decide whether to search for a solicited memory target. Rather, they start searching their memory for the target, and when retrieval fails, their FOK is based on the feedback from the retrieval attempt (e.g., the amount and ease of access of partial information; see Koriat, 1993). Therefore, monitoring follows control, and although FOK judgments are prospective in their intention (involving predictions of future performance), they are retrospective in their basis.

To foreshadow, we do not see the two stories depicted above as being mutually exclusive, and in fact, we shall show that evidence consistent with both of them can be found in one and the same situation. However, although we present results in support of each of the two stories, our first aim in this article is to promote *Story 2* by examining certain paradoxical predictions that follow from the postulated causal effects of control on monitoring. Our second aim is to clarify the conditions under which monitoring drives control processes and those in which monitoring is based on the feedback from such processes. Finally, we shall try to clarify the intricate relationships between monitoring and control that ensue when both *Story 1* and *Story 2* are combined.

The conceptual scheme proposed here is assumed to apply to metacognitive judgments in general. However, the experimental work to be reported (Experiments 1–6) will focus on JOLs elicited during learning because these judgments allow better opportunities for the investigation of most of our propositions. The final experiment (Experiment 7) is intended primarily to show how the

pattern of results obtained for JOLs extends to confidence judgments. In what follows, we first introduce our conceptual scheme by focusing on several results obtained in the study of monitoring and control processes during learning.

Monitoring and Control Processes During Learning

Nelson and Dunlosky (1991) stated that "the accuracy of JOLs is critical because if the JOLs are inaccurate, the allocation of subsequent study time will correspondingly be less than optimal" (p. 267). This statement (see also Thiede, Anderson, & Theriault, 2003) implies a causal effect of monitoring on control. A classic demonstration of this effect is the relationship between JOLs and study time in self-paced learning (see Nelson et al., 1994): Learners generally allocate more time to difficult items than to easy items (Le Ny, Denhiere, & Le Taillanter, 1972; Zacks, 1969; for a review, see Son & Metcalfe, 2000). Nelson and Leonesio (1988) proposed that the effects of item difficulty are mediated by a monitoring process in which participants first judge the relative ease of learning or recalling different items and then control study time so as to compensate for differences in a priori item difficulty. Indeed, a positive correlation between study time and various indexes of perceived item difficulty has been consistently observed. Thus, in their comprehensive review of the literature, Son and Metcalfe (2000) found that in 35 out of 46 published experimental conditions, learners exhibited a clear preference for studying the more difficult materials.

The findings that prestudy EOL ratings (a) have some validity in predicting the relative recallability of different items under experimenter-paced conditions (e.g., Underwood, 1966) and (b) are inversely related to study time under self-paced conditions are consistent with *Story 1*. A simple model of the underlying process is that participants monitor the difficulty of different items in advance of learning and use the products of their monitoring as a basis for allocating study time to different items. A more dynamic model—the discrepancy-reduction model—was proposed by Dunlosky and Hertzog (1998; see also Dunlosky & Connor, 1997; Nelson & Narens, 1990; Thiede & Dunlosky, 1999): Learners continuously monitor the online increase in encoding strength that occurs as more time is spent studying an item, and cease study when a desired level of strength has been reached. This level, referred to as "norm of study" (Le Ny et al., 1972), is preset on the basis of motivational factors, such as the stress on accurate learning versus fast learning (Nelson & Leonesio, 1988). Thus, in self-paced learning, study continues until the current state of mastery reaches the norm of study.

The discrepancy-reduction model incorporates the test-operate-test-exit (TOTE) feedback loop postulated by Miller, Galanter, and Pribram (1960) to underlie goal-oriented behavior. More generally, it incorporates the control-theory perspective according to which people self-regulate their actions to minimize discrepancies between current states and desired states (see Carver & Scheier, 1990). The model is analogous to the idea that a person regulates the speed of running away from a danger according to the degree of fear, perhaps ceasing to run away when a sufficiently low level of fear (or sufficiently high sense of security) has been attained. Thus, subjective experi-

ence (either metacognitive or affective) is assumed to drive and control self-regulated action.²

The “monitoring affects control” (MC; Nelson & Leonesio, 1988) hypothesis underlying Story 1, which assumes a monitoring-based regulation of study time, encounters several serious difficulties that can be illustrated by the results of an unpublished study (Koriat, 1983). In Experiment 1 of that study, one group of participants studied a list of paired associates under self-paced conditions, whereas another group studied the same list under fixed-time presentation. Participants in the two groups were yoked so that the presentation duration for each fixed-time participant was the same as the average study time spent by the matched self-paced participant on each item. The paired associates were also rated by a different sample of participants on EOL. The correlation across items between mean EOL ratings and mean recall in the fixed condition was high (.76), indicating better recall of the easier items. Surprisingly, it remained high (.82) even for the self-paced group, suggesting that the control over study time failed to eliminate or reduce the contribution of a priori item difficulty to recall. Furthermore, although study time in the self-paced condition was indeed negatively correlated with EOL (–.82), it was also negatively correlated with recall in the fixed condition, and to about the same extent (–.80). Thus, study time appears to be no more than a symptom of item difficulty: It is as predictive of ultimate recall as is EOL, and the relationship is such that the more time is invested in a particular item, the *less* likely it is to be recalled! (See also Mazzoni & Cornoldi, 1993; Mazzoni, Cornoldi, & Marchitelli, 1990; Nelson & Leonesio, 1988).

This pattern of results appears to question the idea that the correlation of JOL with study time reflects a causal effect of monitoring on the ongoing control of learning. Rather, it would seem to lean more toward the view advanced by Begg, Martin, and Needham (1992), that metacognitive “predictions are a form of introspective witness; even when they accurately indicate the state of the system, they have no value for memory” (p. 207).

Experiment 2 of Koriat (1983) yielded yet another intriguing observation. It examined the possibility that participants invest more study time in the more difficult items because they experience the illusion that they do succeed in compensating for differences in item difficulty. If so, then this illusion should be reflected in JOLs elicited after study. Thus, Experiment 2 was a replication of Experiment 1 except that participants made JOLs at the end of each self-paced trial regarding the likelihood of subsequent recall. The results indicated that the self-paced participants were not misleading themselves: Although they allocated more study time to the difficult items, they continued to believe (correctly) that these items were less likely to be recalled than the easy items. In fact, the correlation, across items, between mean JOLs and mean EOLs was .95 for the self-paced group and .94 for the fixed group. This observation naturally raises the question, why do self-paced participants bother to allocate study time differentially if they *know* that the differential allocation of study time is ineffective in reducing the contribution of a priori ease of learning to recall? In fact, in Experiment 4, which we report later, in which participants were presented with the same list for four study-test blocks, this pattern of results was found even on the fourth block. Thus, learners stick to the strategy of differential study time allocation despite the fact that their judgments would seem to suggest that they are aware of the futility of that strategy.

These intriguing observations have led us to consider the possibility that study time actually serves a dual function: It subserves a control function as well as a monitoring function. In what follows, we examine these two functions in the context of the question about the causal relation between monitoring and control.

The Control Function of Study Time

The control function is consistent with what we called Story 1, and it is this function that has been commonly emphasized in most previous research on self-paced learning. Underlying this research is the view that the allocation of study time is *goal driven*: It is used as a strategic tool for regulating memory performance toward the achievements of desired objectives given specific constraints. This view, in fact, is part of the general conception shared by most students of metacognition, in which the person is seen as an

active agent that has at his/her disposal an arsenal of cognitive strategies and devices that can be flexibly applied in order to reach certain goals. The choice of such strategies as well as their online regulation is based on the subjective monitoring of these processes. (Koriat, 2002, p. 263)

Thus, the regulation of study time and effort is but one of the tools that learners use strategically in the service of optimizing their performance, and the output of advance or online monitoring is one of the determinants of the choice and orchestration of these strategic tools.

Indeed, previous research has documented the adaptive, goal-driven nature of study time allocation: Learners invest more study time when they expect a recall test than when they expect a recognition test (Mazzoni & Cornoldi, 1993) and more time when the instructions stress memory accuracy than when they stress speed of learning (Nelson & Leonesio, 1988; Pelegrina, Bajo, & Justicia, 1999). Also, the amount of time allocated to an item increases with the reward for subsequently recalling that item, and with the expected likelihood that the item will later be tested (Dunlosky & Thiede, 1998). As mentioned earlier, learners generally invest more study time in items that are judged to be difficult to remember. However, they tend to choose the easier items for restudy when they are given an easy goal (e.g., to get only a few items correct; Dunlosky & Thiede, 2004; Thiede & Dunlosky, 1999) or under conditions that impose severe constraints on study time (Metcalf, 2002; Son & Metcalfe, 2000). These observations, then, stress the control function of study time as a strategic tool that is used to regulate learning.

The simple prediction from the postulated control function of study time is that for a given item, end-of-study JOLs should increase as more study time is invested in that item. This prediction assumes a causal link between JOLs and study time: JOL is

² The recent work by Metcalfe (2002) and Metcalfe and Kornell (2003) challenges the discrepancy-reduction model of study time allocation. It suggests that the strategy of allocating more study time to the more difficult items is neither normative nor generally descriptive of learners' behavior. First, allocating additional study time to the more difficult items sometimes yields the least return in terms of recall (see also Nelson & Leonesio, 1988). Second, learners tend to allocate most of their efforts to items of medium difficulty. Metcalfe and her associates, however, do endorse the assumption that learners' regulation of study time is guided by their metacognitive judgments in a goal-directed fashion.

assumed to control study time allocation in the same way that fear (or subjective experience in general) may be assumed to control running away (or behavior in general).

The Monitoring Function of Study Time

Let us turn next to the monitoring function of study time, which accords with Story 2. This function becomes evident when we focus on the basis of JOLs rather than on their function. After studying an item, how do people assess its degree of mastery? Assuming that JOLs are based on inference from a variety of cues (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin & Bjork, 1996; Koriat, 1997), one obvious cue for JOLs in the case of self-paced learning is study time, or more generally, memorizing effort. We propose that in self-paced learning, study time allocation is generally *data driven*: Learners spend as much time as an item “calls for” (the question of how learners know what an item calls for is addressed in the General Discussion). When they have then to assess the future recallability of the item, one obvious cue that affects their JOLs is the amount of effort they had to invest in attempting to commit the item to memory. Thus, study time can be seen to represent a rough index of a powerful mnemonic cue that has been emphasized in many discussions of metacognitive judgments: processing fluency (Begg et al., 1989; Benjamin & Bjork, 1996; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Kelley, 1999; Kelley & Jacoby, 1996; Matvey, Dunlosky, & Guttentag, 2001). We propose that learners make use of study time (or memorizing effort) as a cue under the implicit naive theory that an item that is quickly mastered stands a better chance to be recalled than one that takes longer to master. We shall refer to this heuristic as the *memorizing effort heuristic*. A “control affects monitoring” (CM) hypothesis of study time makes the following predictions: First, after one studies an item, the JOL associated with that item should *decrease* with increasing time spent studying it. This is like the idea that fear is caused by running away and that the faster one runs the *more* frightened (or less safe) one should feel. The second prediction derives from the idea that metacognitive judgments are accurate because of their reliance on the feedback from control operations. For this to be true, the memorizing effort heuristic must have some degree of validity in predicting interitem differences in future recall (i.e., resolution; see Koriat, 1997). Hence, it is hypothesized that the more time is invested in an item, the *less* likely it is to be recalled. Finally, the accuracy of JOLs in predicting subsequent recall should be mediated by JOLs’ reliance on memorizing effort.

How does the CM hypothesis explain the intriguing observation suggesting that participants allocate more study time to the more difficult items despite their awareness that this allocation strategy does not compensate for the a priori difficulty of these items? We propose that, in general, participants’ allocation of study time among different study items does not reflect a premeditated policy to invest more study effort in difficult items with the intention either to compensate for their a priori difficulty or to achieve a predetermined norm of study. Rather, the difficulty of an item is monitored ad hoc: Learners invest in an item what it calls for, and it is by realizing that a particular item requires relatively more time and effort to be committed to memory that they “know” that the item is going to be difficult to recall. That is, it is not that learners deliberately invest greater effort in studying a difficult item; it is by investing greater effort in that item that learners know that the

item is difficult.³ This is similar to the idea that it is by running away from a bear that one “knows” that the situation is frightening.

The assumption underlying the CM model of study time is similar to that underlying the accessibility model of FOK (Koriat, 1993). According to that model, it is by searching for a solicited piece of information that one “knows” whether the information is available in memory and worth continuing to search for. Likewise the CM model of study time implies that study experience provides learners with mnemonic cues regarding the likelihood of future recall, and this is true whether learners are allowed to regulate study time or not.

Inherent in the CM model of study time is the idea advanced by Kahneman (1973) in his theory of attention and effort. Kahneman was intrigued by the observation that when participants are presented with a task of intermediate difficulty they do not try as hard as they do when the task is more difficult. He concluded that the effort invested is determined mainly by the intrinsic demands of the task, and people simply cannot try as hard in a relatively easy task as they do when the task becomes more demanding. We propose that, in a similar manner, the allocation of study time in self-paced learning is data driven, determined by the qualities of the items in a bottom-up fashion (see also Pelegrina, Bajo, & Justicia, 2000). Therefore, the amount of time spontaneously allocated to an item reflects its encoding fluency, and encoding fluency is diagnostic of the item’s future recall (Koriat & Ma’ayan, 2005).

In sum, the CM model, which stresses the monitoring function of study time, implies that monitoring *follows* control: The allocation of study time is data driven, and JOLs are based on study time. Therefore, JOLs are expected to *decrease* with study time. This is in contrast to the MC model, which stresses the control function of study time allocation, and leads to the expectation that JOLs should *increase* with the amount of time invested.

How Monitoring and Control Processes Combine

As noted earlier, the assumption underlying the proposed conceptual framework is that the MC and CM models of study time are not mutually exclusive. Rather, study time tends to play both a control function and a monitoring function in self-paced learning: It plays a control function insofar as it is goal driven but a monitoring function insofar as it is data driven. An important theoretical challenge, then, is to specify the reciprocal links that exist between monitoring and control operations or, more generally, between subjective experience and behavior (see Allport, 1993; Dent, 2003).

We propose two general modes in which the MC and CM models can combine in the course of daily life: a sequential mode and a simultaneous mode. In the *sequential mode*, monitoring and control functions alternate in a cascaded pattern, with control following along in the wake of monitoring and the feedback from the control operation serving then as the input for later monitoring, and so on. This mode, as described in the General Discussion, is illustrated for FOK judgments by the results of Koriat and Levy-Sadot (2001) and Vernon and Usher (2003) and, with regard to

³ The reader may be reminded of the adage of the Roman philosopher Seneca: “It is not because things are difficult that we do not dare; it is because we do not dare that they are difficult.”

JOLs, by the results of Son and Metcalfe (2005). These results suggest that monitoring-based control can give way to control-based monitoring.

In the *simultaneous mode*, which is perhaps of greater interest, the MC and CM models occur within the same situation. Indeed, in many real-life situations the amount of effort invested in a task is a joint function of both data-driven and goal-driven factors. A student preparing for an exam may spend an inordinately long time studying a particular segment of the material partly because that segment contains some intrinsic difficulties (data driven) and partly because it is especially interesting or important (goal driven). Presumably, data-driven processes place constraints on goal-driven processes, so that the student cannot invest too much effort in committing an easy item to memory or too little effort in attempting to commit a difficult item to memory (Kahneman, 1973).

The combination of data-driven and goal-driven effects within the same situation presents an interesting theoretical problem, because data-driven effects should result in JOLs *decreasing* with increasing study time whereas goal-driven effects should result in JOLs *increasing* with increased study time. How then are JOLs computed by the learner under conditions that combine both types of contribution? It would seem that an *attribution* process must be postulated in which learners first attribute variations in study time to their source before making JOLs (see Experiments 5 and 7).

Introduction to the Experiments

The experiments to be reported had five aims. The first was to bring to the fore the monitoring function of study time in order to promote the idea that monitoring may be based on the feedback from control operations. Thus, in Experiments 1 and 2 we focused on the type of control-based monitoring that is assumed to occur during study. Experiment 1 examined the idea that JOLs elicited during self-paced learning do not drive study time allocation, as commonly assumed, but are themselves based on study time or memorizing effort under the heuristic that the more time is invested in the study of an item, the *lower* is the likelihood that it will be recalled. Thus, JOLs were expected to decrease with increasing study time. Experiment 1 also evaluated the validity of the memorizing effort heuristic by showing that indeed recall is *inversely* related to study time. In Experiment 2 we challenged the basic assumption of the control view of study time, according to which learners strategically allocate more study time to the more difficult items to meet a predetermined norm of study. In addition, we attempted to show that even when learners are prevented from regulating their own study time, they nevertheless use perceived memorizing effort as a cue for future recall. Experiment 3 explored a further prediction of the monitoring view of study time: Assuming that JOLs are based on the feedback from memorizing effort, they should exhibit less dependence on study time when they are made a few trials after study than when they are made immediately after study.

The second aim was to investigate more closely the processes mediating the accuracy of control-based monitoring in predicting memory performance. Specifically, Experiment 4 capitalized on the findings that the accuracy of JOLs in predicting future recall improves with repeated practice studying the same list of items (see Koriat, 1997). It tested the hypothesis that this improvement derives from (a) increased reliance with practice on the feedback

from memorizing effort and (b) improved diagnosticity of memorizing effort as a cue for recall. It follows that both the negative study time–JOL correlation and the negative study time–recall correlation should increase with practice studying the same list of items. Hence, metacognitive accuracy is correlated with the extent to which metacognitive judgments rely on the feedback from control operations.

The third aim was to bring in the control function of study time and investigate the simultaneous operation of the MC and CM models. In Experiment 5, different incentives were awarded to the recall of different items within the list. This manipulation was expected to bring out the *positive* correlation between study time and JOLs, which is the signature of goal-driven metacognitive regulation. At the same time, however, a *negative* correlation was expected between study time and JOLs for each level of incentive, consistent with the assumption that the allocation of study time between same-incentive items is data driven.

The fourth aim was to explore a situation in which the allocation of study time between same-incentive items is expected to reveal goal-oriented regulation and thus to yield a *positive* study time–JOL correlation. This was done in Experiment 6, which capitalized on the finding that under time pressure learners spend more time studying the easier items (Metcalfe, 2002; Son & Metcalfe, 2000). We argue that time pressure produces a qualitative change in study time allocation from being data driven to being goal driven because learners must, in fact, operate against the data-driven tendency to invest more study time in the more difficult items. Presumably, under time pressure, learners quickly monitor the difficulty of the item before deciding whether to invest more time studying it or quit. Therefore a *positive* correlation between study time and JOLs was expected across items for each incentive level.

The fifth aim was to demonstrate the generality of our conceptual framework by extending investigation to another type of metacognitive judgment: subjective confidence. Experiment 7 was similar in design to Experiment 5 except that participants were timed as they solved problems and made confidence judgments. Increasing the incentive associated with the solution of a problem is expected to increase the time spent on that item as well as the subjective confidence in the correctness of the solution reached. However, unlike this positive correlation, which is a signature of monitoring-based control, the correlation between solution time and confidence is expected to be negative within each level of incentive, suggesting a control-based monitoring in which the time spent solving a problem serves in retrospect as a cue for subjective confidence.

Although much of the experimental work reported in this article concerns monitoring and control processes during learning, we believe that the proposed theoretical framework may hold true for other forms of metacognitive processes, as is illustrated in the General Discussion. Furthermore, we suggest that this framework may also be extended to the study of the cause-and-effect links between subjective experience and behavior in other domains.

Note that in describing our results we borrow the terminology of Brunswick's lens model (Brunswick, 1956), which was used in analyzing the process underlying the perception of the external world. According to that model, perception is centered on distal events and objects in the outside world. These, however, cannot be perceived directly but must be inferred from a variety of proximal cues that impinge on the senses. Therefore, the analysis of perception requires specifying the validity of proximal cues in pre-

dicting distal variables (cue validity; e.g., the correlation between the size of a retinal image and the size of the corresponding distal object), the extent to which different cues are relied on (cue utilization); and the ensuing correspondence between perception and reality (achievement; e.g., the correspondence between the actual and perceived sizes of an object). To the extent that metacognitive feelings are also based on inference from a variety of cues rather than on direct access to memory traces (Koriat, 1997), Brunswick's conceptual framework can also be applied to monitoring processes (see Koriat & Ma'ayan, 2005). Thus, assuming that study time is one of the proximal cues for JOLs, we use the JOL–study time relation as an index of *cue utilization*, the study time–recall relation as an index of *cue validity*, and the JOL–recall relation as an index of *achievement*.

Experiment 1

Experiment 1 tested the basic hypotheses of the monitoring model of study time: first, that learners use study time (or memorizing effort) as a cue for JOLs, so that end-of-study JOLs are inversely related to study time, and second, that memorizing effort is indeed a valid predictor of long-term recall, so that study time is also inversely correlated with recall. Finally, the accuracy of JOLs in predicting delayed recall was evaluated.

The experiment involved the self-paced study of paired associates, with JOLs solicited at the end of each study trial. To evaluate the ecological validity of self-paced study time in predicting long-term recall, cued recall was tested only 4 months later.

Method

Materials. We constructed a list of 60 Hebrew word pairs representing a wide range of associative strength. Associative strength was defined as the probability of occurrence of the second word of a pair (the target) as the first response to the first word (the cue) among college students. For 30 pairs, it was greater than zero according to Hebrew word-association norms (Breznitz & Ben-Dov, 1991) and ranged from .012 to .635 ($M = .144$). The remaining 30 pairs were selected such that the two members were judged intuitively as unrelated. Effort was made to avoid obvious links between words that belonged to different pairs.

A preliminary memorability rating study was conducted to obtain data on the perceived relative difficulty of the items. The 60 pairs were presented in a random order to 19 Hebrew-speaking college students who were instructed to imagine that 100 people had been required to memorize the pairs so that they could later recall the response word when shown the stimulus word. They were asked to estimate, for each pair, how many of them would be likely to recall the correct response. The means of these ratings were subtracted from 100% to obtain difficulty ratings. Mean difficulty ratings ranged from 19.5% to 95.7%, with an average of 56.6%. The Pearson correlation between associative strength and mean difficulty (assuming zero associative strength for the 30 unrelated pairs) was $-.68$.

Participants. Twenty Hebrew-speaking University of Haifa undergraduates participated in the experiment. None of them had participated in the memorability rating study.

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

Procedure. Participants were instructed that they would have to study 60 paired associates so that later (without specifying when) they would be able to recall the second word in each pair when the first was presented. They were told that they should study each pair as long as they needed and should press the left key of the mouse when they were through studying.

The two words appeared side by side, and following the key press they were replaced by the question “Chances to Recall (0%–100%)?” Participants reported orally their estimate of the likelihood of recalling the target in the later cued-recall test. They were instructed that their success in performing the task would depend on their success in recalling as many words as possible during the test while keeping the total time invested in studying the entire list as short as possible. No immediate memory test was given; participants were dismissed, with the explanation that the experiment concerned only their JOLs and not their actual memory performance. All participants were contacted about 4 months later and were invited to participate in a new experiment. This session actually took place on average 130 days after the study phase (range = 118–144 days). They were reminded of the first session, and their memory was tested: The 60 stimulus words were presented one after the other for up to 8 s each, and participants had to say aloud the response word within the 8 s allotted. The experimenter recorded the response, and 1 s thereafter a beep was sounded and the next stimulus word was presented. The order of presentation of the items was randomly determined for each participant for each of the two phases of the experiment.

Results

Cue utilization: Memorizing effort as a cue for JOLs. To examine the predictions of the CM model, all study times were split at the median for each participant. Study times averaged 4.8 s and 10.7 s, respectively, for below-median and above-median items. JOLs for these classes averaged 67.6 ($SD = 14.2$) and 43.9 ($SD = 10.2$), respectively, $t(19) = 8.15$, $p < .0001$, $\eta_p^2 = .78$. Thus, the more time was allocated to the study of an item, the *lower* were JOLs for that item.

A within-participant gamma correlation was also calculated across items between study time and JOLs for each participant. This correlation, averaged across all participants, was negative and significant: $-.42$, $t(19) = 9.59$, $p < .0001$. This result was quite reliable: The correlation was negative for each of the 20 participants ($p < .0001$, by a binomial test).

It might be argued that this correlation simply reflects the fact that items that are perceived to be difficult induce longer study times and also elicit lower JOLs than items that are perceived to be easy. This argument is difficult to refute because of the inherent link between perceived difficulty and study time. Nevertheless, we calculated the within-person Pearson correlation between study time and JOLs with difficulty ratings partialled out. The residual correlation averaged $-.24$ across participants and was highly significant, $t(19) = 5.19$, $p < .0001$. Thus, the correlation between study time and JOLs is not entirely mediated by judged item difficulty (see further evidence in Experiment 4).

Cue validity: The validity of the memorizing effort heuristic. How valid is the memorizing effort heuristic? Percentage recall for items with below-median and above-median study times averaged 10.9 ($SD = 10.6$) and 3.9 ($SD = 3.5$), respectively, $t(19) = 3.36$, $p < .005$, $\eta_p^2 = .37$. Thus, although percentage recall was quite low overall ($M = 7.4\%$, $SD = 6.4$), it *decreased* with increasing study time. The mean within-participant gamma correlation between study time and recall (with $n = 19$, because 1 participant achieved 0% recall) was low ($-.22$) but near significant, $t(18) = 1.89$, $p < .08$. This correlation was negative for 14 out of the 19 participants ($p < .05$ by a binomial test). Thus, there was a trend suggesting that the more time spent studying an item, the less it was likely to be recalled.

The finding that recall decreased with increasing study time may seem surprising, but it is consistent with results reported by Bahr-

ick and Phelps (1987), which seem also to disclose the predictive validity of data-driven variation. Participants studied the Spanish translations of 50 English words, and following the presentation of all English–Spanish pairs, their cued recall was tested. Items not recalled were then included in the next study trial, and this procedure was repeated for several more trials until a participant recalled all the words. Thus, it was possible to calculate for each participant how many study trials each word received.

Bahrick and Phelps (1987) tested cued recall for the words 8 years later. Their results clearly indicated that recall *decreased* with number of study trials. The magnitude of the effect was very impressive (see their Table 3): about 14% recall for pairs that had been presented once or twice and only about 2% recall for pairs that had been presented 11 times or more. These results stand in sharp contrast to the pattern characteristic of a learning curve. In fact, it is the mirror image of a typical learning curve that is obtained when number of study trials is experimentally manipulated.

Achievement: The accuracy of JOLs. Assuming that participants do rely on memorizing effort in making JOLs and that memorizing effort is diagnostic of delayed recall, we might expect JOLs to exhibit some degree of validity in predicting recall. Indeed, recall increased with JOLs, averaging 0.9 ($SD = 2.1$) and 13.2 ($SD = 11.4$), respectively, for items with below-median and above-median JOLs, $t(19) = 5.00, p < .0001, \eta_p^2 = .57$. For the 19 participants with nonzero recall, the gamma correlation between JOLs and recall (see Nelson, 1984) averaged .52, $t(18) = 7.47, p < .0001$. The gamma correlation was positive for 18 out of the 19 participants ($p < .0001$ by a binomial test).

We have argued that the accuracy of JOLs in predicting delayed recall is mediated by the use of study time as a cue for JOLs. Indeed, the Pearson JOL–recall correlation was .22, and when study time was partialled out, the correlation dropped to .16, $t(18) = 2.82, p < .05, \eta_p^2 = .33$, for the difference.

Discussion

The results of Experiment 1 are in line with what we called Story 2, which assumes that JOLs are based on the feedback from study effort. JOLs decreased with increased study time, supporting the memorizing effort heuristic as a basis for JOLs. In addition, the results supported the validity of that heuristic: Recall was inversely related to study time.

Why is it the case that easily learned words are better remembered? This question should be addressed by theories of memory. However, regardless of the explanation, as far as metamemory is concerned, it would seem that learners do exploit this correlation in making JOLs and presumably improve their predictions by doing so. We propose that learners are not aware of the correlation between memorizing effort and future recall and do not use the memorizing effort heuristic as a deliberate, analytic inference. Rather, this heuristic is applied unconsciously to yield a sheer subjective feeling that can serve as the basis of recall predictions (see Koriat, 2000; Koriat & Levy-Sadot, 1999).

The observation that both JOLs and recall decreased with presentation duration stands in sharp contrast with the observation that JOLs and recall generally increase with experimenter-determined presentation duration (Koriat, 1997; Koriat & Ma'ayan, 2005). This contrast highlights the data-driven character of study time allocation (Koriat & Ma'ayan, 2005). A similar

pattern was found in Bahrick and Phelps's (1987) study, in which the number of study trials required to master the items was essentially determined by the items themselves (or by the item–learner interaction). This data-driven character is what gives study time (or number of study trials) its diagnostic value in predicting item memorability. As we shall see later (in Experiments 5, 6, and 7), when study time is goal driven rather than data driven, the functions relating JOL and recall to study time are more similar to those found for the experimenter-controlled than for the self-controlled allocation of study time.

Experiment 2

Experiment 2 was essentially an improved replication of the unpublished study (Koriat, 1983) mentioned earlier. Its primary aim was to challenge the assumption of the MC model that the allocation of study time is used as a strategic tool to compensate for differences in item difficulty (e.g., Mazzoni et al., 1990; Nelson & Leonesio, 1988). Toward that aim, a fixed-rate condition was included in Experiment 2 in addition to a self-paced condition. Each fixed-rate participant was yoked to one self-paced participant so that the mean study time spent by the self-paced participant on each item was assigned to all items for the yoked fixed-rate participant. If the self-paced allocation of study time is indeed guided by the intention to compensate for item difficulty, then the effects of judged item difficulty on JOLs should be strong for fixed-rate participants but weak or even absent for self-paced participants.

A secondary aim was to examine what happens to metacognitive judgments when participants are denied the option to control study time, as occurs in the fixed-rate condition. According to the CM model, it is by studying an item that a person can appreciate the likelihood of recalling that item in the future. This should be the case whether study time is self-paced or fixed. Thus, a fixed-rate presentation (unless it is too fast) should not prevent learners from using perceived memorizing effort as a cue for recall; it only deprives *researchers* of a useful index for the learner's memorizing effort: self-paced study time. Hence, fixed-rate participants should not necessarily exhibit impaired ability to monitor their future memory performance.

Whereas Experiment 1 involved a 4-month retention interval, in Experiment 2 an immediate recall test was used. Because JOLs presumably reflect the participants' immediate feelings, we deemed it important to also evaluate both the accuracy of JOLs and the validity of study time with immediate recall as a criterion. Thus, in Experiment 2 the cued-recall phase followed the study phase in the same session.

Method

Participants. Forty Hebrew-speaking University of Haifa undergraduates participated in the experiment, 8 for course credit and 32 for pay. Participants were assigned alternately to self-paced and fixed-rate conditions according to their order of arrival so as to form 20 pairs of yoked participants.

Materials, apparatus, and procedure. The stimulus materials and the apparatus were the same as in Experiment 1. The procedure for self-paced participants was also the same, whereas for the fixed-rate participants, presentation time for each item was the average study time allocated by his or her yoked self-paced participant. The cued-recall test was administered after a 1.5-min filler task (counting backward at intervals of 3, starting

from a three-digit number), using the same procedure as in Experiment 1. The orders of presentation of the items for the study and test phases were randomly determined for each pair of yoked participants, so that the same random orders were used for both members of the pair.

Results

The effects of item difficulty. According to the MC model, participants allocate more study time to difficult-to-learn items to achieve a preset degree of mastery (e.g., Nelson & Narens, 1990). We tested this assumption by examining the effects of judged item difficulty on self-paced study time and by comparing the effects of study time on JOLs and recall for the self-paced and the fixed-rate conditions.

Focusing first on the self-paced condition, more study time was indeed allocated to difficult items than to easy items: When all items were divided at the median of the difficulty ratings (48.1) into 30 easy items and 30 difficult items, mean study times for the two classes were 3.8 s ($SD = 1.5$) and 7.0 s ($SD = 4.2$), respectively, $t(19) = 4.43$, $p < .001$, $\eta_p^2 = .51$. This pattern replicates previously reported results (e.g., Dunlosky & Connor, 1997; Mazzoni et al., 1990; Nelson & Leonesio, 1988; Zacks, 1969) and is consistent with the view of study time as a strategic tool that is used to regulate memory performance.⁴

Figure 1 (top panel) presents mean recall for easy and difficult items for the self-paced and fixed-rate conditions.⁵ Although there was a substantial effect of item difficulty on recall, the magnitude of this effect hardly differed between the two conditions. Because the self-paced participants spent more time studying the difficult items and less time studying the easy items than the fixed-rate participants, we would have expected weaker effects of judged difficulty on recall for the former participants. However, a Condition (self-paced vs. fixed) \times Difficulty analysis of variance (ANOVA) on recall performance yielded a significant effect for difficulty, $F(1, 38) = 197.75$, $MSE = 134.75$, $p < .0001$, $\eta_p^2 = .84$, but not for condition, $F(1, 38) = 1.32$, $MSE = 424.13$, $p = .26$, or for the interaction ($F < 1$). Thus, the differential allocation of study time by the self-paced participants was ineffective in eliminating or reducing the effects of item difficulty on recall in comparison with the fixed-rate presentation (see also Metcalfe & Kornell, 2003; Pelegrina et al., 2000). These results are consistent with those of Koriat's (1983) unpublished study mentioned earlier and with the labor-in-vain effect documented by Nelson and Leonesio (1988).

More important for the concern of this article are the results for JOLs. Do self-paced participants experience an illusion of control? Although JOLs also evidenced marked effects of item difficulty (Figure 1, bottom panel), these effects too were similar in magnitude for the two conditions: A Condition \times Difficulty ANOVA on JOLs yielded a significant effect for difficulty, $F(1, 38) = 232.26$, $MSE = 101.16$, $p < .0001$, $\eta_p^2 = .86$. The effect of condition approached significance, $F(1, 38) = 3.55$, $MSE = 430.08$, $p < .07$, $\eta_p^2 = .09$, suggesting that the control over study time enhanced JOLs (see Perlmutter & Monty, 1977), but the interaction was again not significant, $F(1, 38) = 1.47$, $MSE = 101.16$, $p = .23$. These results illustrate the paradox discussed in the introduction: Self-paced participants allocate more study time to the more difficult items despite the fact that their JOLs might appear to suggest that they are aware of the futility of the differential allocation of study time.

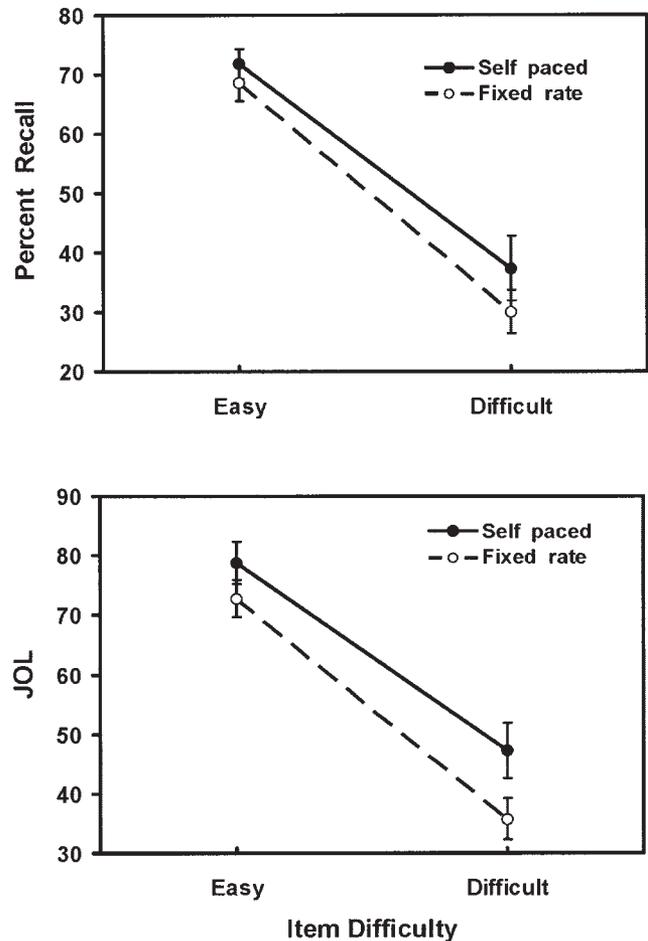


Figure 1. Mean recall (top panel) and judgment of learning (JOL; bottom panel) for easy and difficult items, plotted separately for the self-paced and fixed-rate conditions (Experiment 2). Error bars represent ± 1 standard error of measurement.

⁴ The results of Metcalfe and her associates (Metcalfe, 2002; Metcalfe & Kornell, 2003) indicate that under certain conditions learners tend to invest more of their study time in items of intermediate difficulty. However, in both Experiment 1 and Experiment 2 of this study, self-paced study time increased monotonically with difficulty. Thus, when all items were divided into three classes according to their mean difficulty (see *Method* of Experiment 1), with 20 items in each category, study time for the easy, intermediate, and difficult items averaged 5.4, 8.2, and 9.7 s, respectively, in Experiment 1 and 3.2, 5.8, and 7.1 s, respectively, in Experiment 2. Therefore, for ease of exposition we continue to use the dichotomous division between easy and difficult items.

⁵ Some of the results presented in Figure 1 call for within-subject analyses (e.g., the comparison between easy and difficult items), whereas others call for between-subjects analyses (e.g., the comparison between self-paced and fixed-rate conditions). This is true for most of the figures in this article. Therefore, to avoid confusion, the error bars in all of the figures in this article represent ± 1 SEM around each individual cell mean (which is appropriate for between-subjects analyses). This is true even in those cases where the effects reported in the text are based on within-subject comparisons.

These results raise doubts regarding the presumed goal-driven character of study time allocation but can be readily accommodated by the view that the allocation of study time by self-paced participants is data driven and that JOLs are based on memorizing effort. It is not that learners know that they are less likely to recall a difficult item despite having spent more effort studying it. Rather, it is by investing more effort memorizing an item that they know that it is less likely to be recalled. We shall now examine the evidence for this proposition.

Evidence for the monitoring model of study time. As in Experiment 1, an inverse relationship between end-of-study JOLs and study time was obtained for the self-paced participants: When study times were split at the median for each self-paced participant, JOLs averaged 73.3% ($SD = 17.2$) for below-median study times ($M = 3.1$ s, $SD = 1.2$) and 53.0% ($SD = 18.0$) for above-median study times ($M = 7.6$ s, $SD = 4.5$), $t(19) = 7.79$, $p < .0001$, $\eta_p^2 = .76$.

In parallel, study time was diagnostic of subsequent recall: The recall means for below-median and above-median study times were 62.8% ($SD = 14.9$) and 46.5% ($SD = 21.1$), respectively, $t(19) = 4.16$, $p < .0005$, $\eta_p^2 = .48$. In sum, JOLs as well as recall decreased with increasing study time, consistent with what we called Story 2.

Monitoring processes in the self-paced and fixed-rate conditions. We argued that similar processes underlie metacognitive monitoring under fixed-rate and self-paced conditions. Some support for this proposition is available in Figure 1 (bottom panel), which indicates similar effects of item difficulty on JOLs for the two conditions. Additional support, however, comes from the accuracy of JOLs in predicting recall. The within-person gamma correlation between JOLs and recall was .48 ($p < .0001$) for the self-paced condition and .56 ($p < .0001$) for the fixed-rate condition. The difference between the two correlations was not significant, $t(38) = 0.98$, $p = .34$. This result suggests that a fixed-rate presentation does not impair monitoring accuracy. Presumably participants can still sense the between-item differences in memorizing effort even when presentation rate is fixed.

Discussion

The differential allocation of study time between easy items and difficult items has been taken to suggest a control-theory type model (see Hyland, 1988) according to which study time is regulated to minimize the discrepancy between the current and desired levels of mastery of each item. However, the finding that item difficulty had very similar effects on JOLs in the self-paced and fixed-rate conditions raises doubts about that model. The results also undermine the assumption that in self-paced learning, study continues until perceived degree of learning meets the norm of study (e.g., Le Ny et al., 1972; Nelson & Narens, 1990). If JOLs are assumed to reflect the perceived degree of mastery attained by self-paced participants at the end of a study trial, then the magnitude of JOLs should be about the same for easy and difficult items. But the effects of a priori item difficulty on JOLs were strong and were no weaker for the self-paced participants than for the fixed-rate participants. These results, however, are consistent with the idea that study time is data driven and that JOLs are based on study time rather than vice versa. Indeed, the results of Experiment 2 replicated the finding that JOLs and recall are inversely related to study time. In addition, the comparison of the results for self-paced

and fixed-rate participants lends credence to the proposition that participants rely on memorizing effort as a cue for the feeling of mastery whether or not they are allowed to control the pacing of study.

Experiment 3

In Experiments 1 and 2 we obtained evidence suggesting that learners use memorizing effort as a cue for JOLs, but that evidence was correlational in nature. In contrast, in Experiment 3 we investigated the effects of a manipulation that might be expected to moderate the dependence of JOLs on memorizing effort: soliciting JOLs some time after study rather than immediately after study. Assuming that the memory for the effort invested in mastering an item fades gradually with the passage of time, we would expect delayed JOLs to be less heavily dependent on study time than immediate JOLs.

The procedure was similar to that of previous studies that have contrasted delayed and immediate JOLs (Dunlosky & Nelson, 1994; Nelson & Dunlosky, 1991). Those studies, however, focused on the *delayed-JOL effect*—namely, the observation that JOLs exhibit greater relative accuracy (resolution) in predicting future memory performance when elicited after some delay than when elicited immediately after study. In Experiment 3, in contrast, we focused on the basis of JOLs as it may be disclosed by the study time–JOL correlation.

Method

Materials, apparatus, and procedure. The experiment was conducted on a personal computer, and the materials were the same as in Experiment 1. The procedure was also the same except for the following: The 60 pairs were ordered randomly for each participant with the constraint that each set of 20 successive pairs included 10 easy pairs and 10 difficult pairs. Of these, 5 easy pairs and 5 difficult pairs were assigned to the immediate-JOL condition, and the remaining 10 items were assigned to the delayed-JOL condition. For the immediate-JOL pairs, the study phase was as in Experiment 1, except that the cue word was shown again immediately after the participant had pressed the left mouse button to indicate end of study. Only then was the participant required to indicate JOL. For the delayed-JOL pairs, the cue word appeared after the 20 pairs in a block had been studied. The order of JOL elicitation for these pairs was such that the cue words for the first 5 items studied (in a block of 20) appeared first, in random order, then those of the next 5 items, and so on. The recall phase was exactly as in Experiment 1 except that 6 s were allowed for responding.

Participants. Thirty-four Hebrew-speaking undergraduates from the University of Haifa were paid for participating in the experiment.

Results

For the immediate condition, JOLs and recall averaged 65.8% ($SD = 11.7$) and 43.7% ($SD = 14.7$), respectively. The respective means for the delayed condition were 72.6% ($SD = 13.1$) and 57.7% ($SD = 13.6$). Thus, JOLs were inflated in comparison to recall, $t(33) = 11.84$, $p < .0001$, $\eta_p^2 = .81$, and delayed-JOL items yielded both higher JOLs, $t(33) = 3.09$, $p < .005$, $\eta_p^2 = .22$, and better recall than immediate-JOL items, $t(33) = 7.04$, $p < .0001$, $\eta_p^2 = .60$.

Memorizing effort as a cue for JOLs. Mean JOLs for below-median (short) and above-median (long) study time were calculated for each participant for the immediate and delayed conditions, and their means appear in the left panel of Figure 2. JOLs

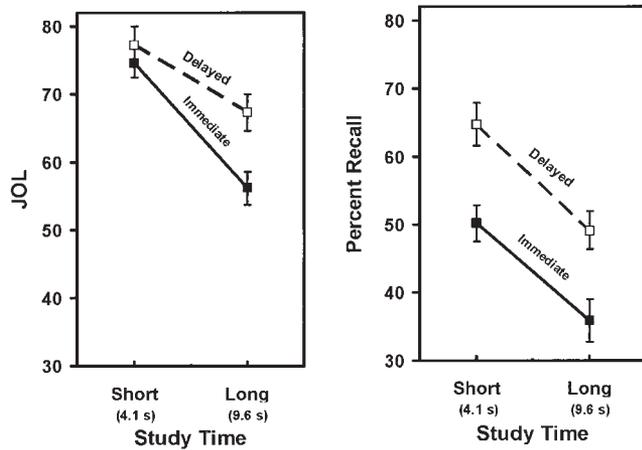


Figure 2. Mean judgment of learning (JOL; left panel) and recall (right panel) for below-median (short) and above-median (long) study times, plotted separately for the immediate and delayed conditions (Experiment 3). Error bars represent ± 1 standard error of measurement.

decreased with increasing study time, consistent with the assumed monitoring function of study time. This decrease, however, was more moderate for the delayed condition than for the immediate condition. A Study Time (short vs. long) \times Condition (immediate vs. delayed) ANOVA yielded significant effects for study time, $F(1, 33) = 84.99$, $MSE = 80.65$, $p < .0001$, $\eta_p^2 = .72$, and for condition, $F(1, 33) = 9.54$, $MSE = 170.34$, $p < .005$, $\eta_p^2 = .22$. The interaction, however, was also significant, $F(1, 33) = 5.40$, $MSE = 113.53$, $p < .05$, $\eta_p^2 = .14$. The effects of study time on JOLs amounted to 18.4% in the immediate condition and to 10.0% in the delayed condition, but both were significant, $t(33) = 9.12$, $p < .0001$, $\eta_p^2 = .72$, and $t(33) = 3.68$, $p < .001$, $\eta_p^2 = .29$, respectively.

The reduced dependence of JOLs on study time for the delayed-JOL condition was also reflected in the within-individual gamma correlations between these two variables across items. These correlations averaged $-.42$ for the immediate condition and only $-.22$ for the delayed condition. Although both correlations were significantly different from zero, $t(33) = 12.28$, $p < .0001$, and $t(33) = 5.51$, $p < .0001$, respectively, the difference between them was significant, $t(33) = 4.58$, $p < .0001$, $\eta_p^2 = .39$.

The validity of the memorizing effort heuristic. As in the previous experiments, study time was a good predictor of recall: Percentage recall decreased with increasing study time, as can be seen in the right panel of Figure 2. Recall was overall better for the delayed-JOL condition ($M = 56.9\%$, $SD = 18.7$) than for the immediate-JOL condition ($M = 43.0\%$, $SD = 18.2$), $F(1, 33) = 43.34$, $MSE = 151.36$, $p < .0001$, $\eta_p^2 = .57$, but the interaction between study time and condition was not significant ($F < 1$). Thus, delaying JOLs reduced reliance on study time as a cue for JOLs but did not affect the validity of study time in predicting recall.

The accuracy of JOLs in predicting recall. The gamma correlations between JOLs and recall averaged $.44$ for the immediate condition and $.79$ for the delayed condition, $t(33) = 7.23$, $p < .0001$, $\eta_p^2 = .61$, consistent with the delayed-JOL effect (e.g., Nelson & Dunlosky, 1991). Thus, delaying JOLs reduced reliance

on study time as a cue for JOLs, and at the same time improved the accuracy of JOLs in predicting recall.

Discussion

The results of Experiment 3 replicated the findings of Experiment 1: JOLs and recall decreased with study time, consistent with the CM model. In addition, the results suggest that the reliance on study time (or memorizing effort) as a cue for JOLs was weaker when JOLs were delayed than when they were immediate. This was so despite the fact that the effects of study time on recall were similar in both cases. These results suggest that the subjective experience gained from the effort invested in studying an item fades away with the passage of time, so that learners are less likely to make use of it when JOLs are delayed than when JOLs are immediate.

The higher accuracy of delayed JOLs, despite their reduced dependence on study time, supports a distinction between two cues for JOLs: encoding fluency and retrieval fluency (see Benjamin & Bjork, 1996). Both of these cues involve control-based monitoring. Koriat and Ma'ayan (2005) reported evidence suggesting that whereas immediate JOLs are based on the feedback from encoding operations (as indexed, e.g., by study time), delayed JOLs tend also to be influenced by the feedback from retrieval attempts, that is, from the ease with which to-be-remembered items are retrieved. Results reported by Nelson, Narens, and Dunlosky (2004) also suggest that delayed JOLs are more accurate than immediate JOLs because they rely on covert recall, which is likely to tap the kind of long-term memory retrieval on which the criterion test itself is based. Note that underlying this explanation is the assumption that monitoring follows control: "People do not monitor the underlying object-level memory system per se, but instead monitor the output from this system" (Nelson et al., 1998, p. 163). Thus, it would seem that people have two different means by which they can appreciate the likelihood of recalling an item in the future: attempting to study the item and attempting to retrieve it (see also Son & Metcalfe, 2005). Both of these imply control-based monitoring. Arguably, however, the feedback from retrieval effort is more diagnostic of future recall than the feedback from memorizing effort.

Experiment 4

Whereas Experiments 1–3 tested the predictions that follow from the CM model with regard to the basis of JOLs, Experiment 4 focused on the predictions regarding the accuracy of JOLs. According to Story 2, the accuracy of metacognitive judgments derives largely from their reliance on the feedback from the control operations involved in learning and remembering. In Experiment 1 we reported results suggesting that the accuracy of JOLs is mediated by their dependence on study time (or memorizing effort). Experiment 4 proceeded to further examine the accuracy of JOLs, focusing on the observation that with repeated study–test cycles of a list of items, the accuracy of JOLs in predicting recall improves (King, Zechmeister, & Shaughnessy, 1980; Koriat et al., 2002; Lovelace, 1984). Koriat (1997) proposed that this improvement occurs because with repeated practice participants rely increasingly on internal, mnemonic cues in making recall predictions. The evidence for this proposal, however, was indirect because no operational measure of mnemonic cues was available in that study.

In contrast, if indeed memorizing effort serves as a mnemonic cue for JOLs, then Koriat's proposal can be tested by showing that the contribution of study time to JOLs increases with practice studying a list of items. Findings supporting this prediction would accord with the assumption of Story 2 that the accuracy of metacognitive judgments is contingent on the degree to which such judgments rely on the feedback from control operations.

In Experiment 4, the paired associates were presented for four self-paced study–test blocks. This allowed us to trace the changes that occurred over the four presentations in (a) the dependence of JOLs on study time (cue utilization), (b) the validity of study time as a predictor of recall (cue validity), and (c) the accuracy of JOLs in predicting recall (achievement). We examined the hypothesis that the improvement in JOL accuracy with practice is mediated by (a) increased reliance on study time as a cue for JOLs and (b) improved diagnosticity of study time as a predictor of recall. It is important to note that the CM model implies that the correlation between study time and JOLs as well as that between study time and recall should become increasingly more *negative* with practice studying the same items. Both of these trends should contribute to the increased accuracy of JOLs with practice.

Koriat (1997) presented evidence suggesting that the mnemonic cues underlying JOLs are idiosyncratic in nature. If memorizing effort is indeed idiosyncratic, reflecting the learner–item interaction, then we should expect the study time–JOL correlation to increase with practice only for a self-paced participant, not for another participant who receives the exact same experimenter-allocated study times as the self-paced participant. To examine this proposition, we included in Experiment 4 an other-paced condition: Each participant in this condition was yoked to one self-paced participant, receiving precisely the same item-by-item study times as those allocated by the self-paced participant. If the study time–JOL correlation is found to increase with practice only for self-paced participants, this would suggest that this increase indeed reflects increased reliance on memorizing effort as an idiosyncratic cue.

Method

Participants. Forty Hebrew-speaking University of Haifa undergraduates participated in the experiment for course credit. They were assigned to the self-paced and other-paced conditions according to their order of arrival so as to form 20 pairs of yoked participants.

Materials, apparatus, and procedure. The list of stimuli and the apparatus were the same as in Experiment 1. The procedure was also the same except for the following: First, the experiment included four study (plus JOL)–test blocks. Second, an other-paced condition was added; each other-paced participant was yoked to one self-paced participant, receiving exactly the same study times to each item in each presentation as that allocated by the yoked self-paced participant on that presentation. Third, at the end of each study block, participants were asked to provide an aggregate estimate of the number of items that they would be able to recall. The results for these estimates are not reported here. There were a few additional minor changes: During the study phase there was a 500-ms interval between the presentation of an item and the JOL probe, and 8 s were allowed for responding during the test phase.

Results

We briefly report several descriptive data on the effects of presentation and item difficulty on JOLs and recall before turning to the main aims of Experiment 4.

The effects of presentation on JOLs and recall. Recall increased with presentation (see Figure 3) and was consistently better for the self-paced participants (78.2%, averaged across the four presentations) than for the other-paced participants (69.8%), suggesting that the option to control the allocation of study time, in itself, enhanced memory performance (see also Mazzoni & Cornoldi, 1993). Mean JOLs across the four presentations were also higher for the self-paced (72.5%) than for the other-paced (64.1%) condition (see also Experiment 2). The results presented in Figure 3 disclose the underconfidence-with-practice effect reported by Koriat et al. (2002): A Measure (recall vs. JOL) \times Presentation ANOVA yielded $F(3, 117) = 29.62$, $MSE = 50.07$, $p < .0001$, $\eta_p^2 = .43$, for the interaction, and when condition was included in the analysis, the triple interaction was not significant ($F < 1$).

The effects of item difficulty. Self-paced participants allocated more study time to the difficult items ($M = 5.1$ s, $SD = 1.8$) than to the easy items ($M = 3.2$ s, $SD = 1.4$), $t(19) = 10.37$, $p < .0001$, $\eta_p^2 = .85$, and did so in each of the four presentations. In parallel, in each presentation recall was better for the easy items than for the difficult items, averaging 88.4% ($SD = 6.2$) and 68.0% ($SD = 18.1$), respectively, across presentations, $t(19) = 6.89$, $p < .0001$, $\eta_p^2 = .71$. Also, as in Experiment 1, difficult items were associated with lower JOLs than easy items despite the fact that they received more study time. This pattern was evident even on the fourth presentation, where study time averaged 1.7 s ($SD = 0.6$) and 2.8 s ($SD = 1.1$) for the easy items and difficult items, respectively, $t(19) = 6.85$, $p < .0001$, $\eta_p^2 = .71$, whereas JOLs averaged 90.9% ($SD = 9.9$) and 78.2% ($SD = 17.6$), respectively, $t(19) = 5.16$, $p < .0001$, $\eta_p^2 = .58$. Thus, even on the fourth presentation, participants allocated more study time to the difficult items although their JOLs might have suggested that they were aware that the differential allocation of study time was ineffective in compensating for the between-item differences in difficulty. These results, however, are consistent with the CM model. Let us now

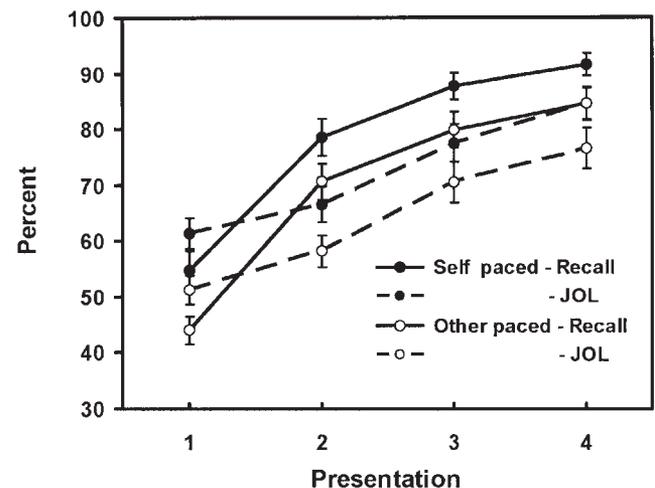


Figure 3. Mean judgment of learning (JOL) and recall as a function of presentation, plotted separately for the self-paced and other-paced conditions (Experiment 4). Error bars represent ± 1 standard error of measurement.

turn to the primary aims of Experiment 4, focusing on the changes that occurred with practice.

Cue utilization: The relationship between study time and JOLs. Figure 4 (top panel) presents mean gamma correlations between study time and JOLs as a function of presentation for the self-paced and other-paced conditions. All self-paced correlations were negative, consistent with the CM model, and all were significant at the .0001 level. In addition, however, two trends were evident. First, the negative correlations were significantly higher for the self-paced than for the other-paced condition, averaging $-.60$ and $-.36$, respectively. A Condition \times Presentation ANOVA yielded $F(1, 38) = 36.38$, $MSE = 0.061$, $p < .0001$, $\eta_p^2 = .49$, for condition. The difference was near significant even on the first presentation, $t(38) = 1.99$, $p < .06$, $\eta_p^2 = .09$.

Second, the interaction between condition and presentation was highly significant, $F(3, 114) = 8.31$, $MSE = 0.015$, $p < .0001$, $\eta_p^2 = .18$: For the self-paced participants, the correlation increased monotonically with presentation, $F(3, 57) = 8.07$, $MSE = 0.019$, $p < .0001$, $\eta_p^2 = .30$, whereas no such systematic increase was evident for the other-paced participants. Thus, the paradox noted in the introduction, of JOLs decreasing with increasing study time in self-paced learning, became, in fact, more pronounced with prac-

tice. The results, however, support the contention that the contribution of mnemonic cues to JOLs increases with repeated presentations of the same list (Koriat, 1997).

In parallel, the correlation between a priori difficulty and JOLs decreased gradually with practice (Figure 4, bottom panel), suggesting that the increased reliance on mnemonic cues with practice was paralleled by decreased reliance on intrinsic cues, such as judged item difficulty (see Koriat, 1997). This pattern was equally observed for the self-paced and other-paced conditions. A Condition (self-paced vs. other-paced) \times Presentation ANOVA on the difficulty-JOL correlation yielded $F(3, 114) = 65.75$, $MSE = 0.009$, $p < .0001$, $\eta_p^2 = .63$, for presentation; $F(1, 38) = 1.04$, $p = .32$, for condition; and $F(3, 114) = 1.58$, $p = .20$, for the interaction.

We interpret this pattern of results to suggest that both self-paced and other-paced participants rely on memorizing effort in assessing degree of mastery but that only for the former participants is study time a relatively reliable measure of their memorizing effort. Furthermore, the mnemonic cues that are responsible for the changes in cue utilization that occur with practice are idiosyncratic in nature rather than being commonly shared (see Koriat, 1997; Nelson, Leonesio, Landwehr, & Narens, 1986).

Cue validity: The predictive accuracy of the memorizing effort heuristic. Figure 5 depicts the mean gamma correlation between study time and recall for each condition as a function of presentation. These means were based only on 14 pairs of participants because 1 participant from each of the remaining 6 pairs yielded perfect recall on the last presentation. For each presentation, the correlation of study time with recall was negative and increased steadily with practice, so that by the fourth presentation it was $-.73$ for self-paced participants. A Condition \times Presentation ANOVA on these correlations yielded $F(1, 26) = 19.84$, $MSE = 0.037$, $p < .0001$, $\eta_p^2 = .43$, for condition; $F(3, 78) = 16.22$, $MSE = 0.026$, $p < .0001$, $\eta_p^2 = .38$, for presentation; and $F(3, 78) = 3.91$, $MSE = 0.026$, $p < .05$, $\eta_p^2 = .27$, for the interaction. Separate one-way ANOVAs for the effects of presentation yielded $F(3, 39) = 18.39$, $MSE = 0.024$, $p < .0001$, $\eta_p^2 = .59$, for the self-paced condition, and $F(3, 39) = 2.61$, $MSE = 0.027$, $p < .07$, $\eta_p^2 = .17$, for the other-paced condition.

Altogether, the results suggest that two changes occur with practice. First, memorizing effort becomes an increasingly valid predictor of recall, and this improvement appears to be due to idiosyncratic aspects of memorizing effort. Second, memorizing effort exerts increasingly stronger effects on JOLs. Possibly both of these changes underlie the improved accuracy of JOLs with practice, which are examined next.

Achievement: The accuracy of JOLs in predicting recall. The JOL-recall gamma correlation is plotted in Figure 6 as a function of presentation for each of the two conditions (based on the 14 matched pairs of participants for whom these correlations were computable). Indeed, the predictive validity of JOLs increased with practice for the self-paced participants, but the other-paced participants also demonstrated a very similar pattern. A Condition \times Presentation ANOVA yielded $F(1, 26) = 1.39$, $MSE = 0.11$, $p = .25$, for condition; $F(3, 78) = 8.04$, $MSE = 0.05$, $p < .0001$, $\eta_p^2 = .24$, for presentation; and $F < 1$ for the interaction. Thus, practice improved monitoring resolution, consistent with previous results (King et al., 1980; Koriat, 1997; Koriat et al., 2002; Mazzoni et al., 1990).

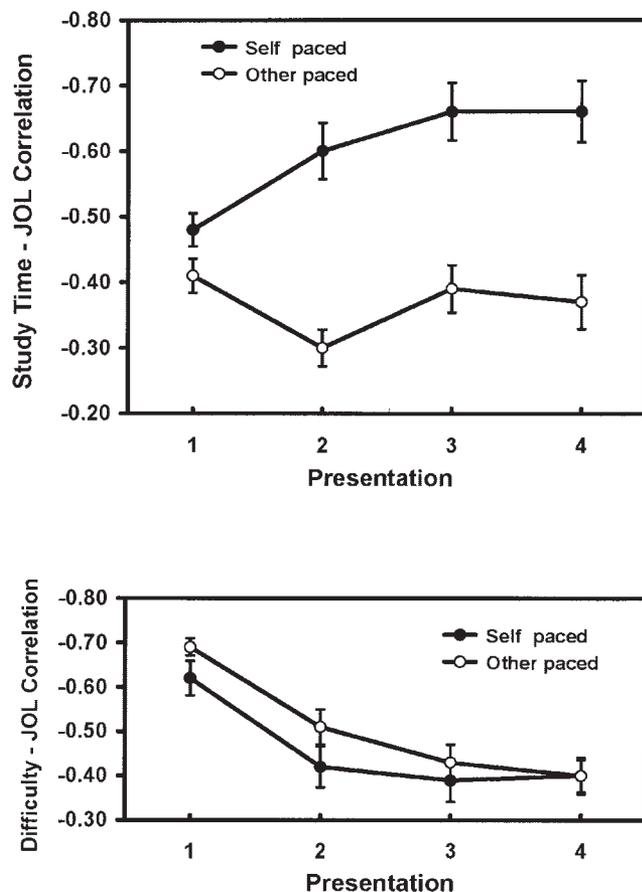


Figure 4. Mean within-participant gamma correlations between study time and judgment of learning (JOL; top panel) and between item difficulty and JOL (bottom panel) as a function of presentation, plotted separately for the self-paced and other-paced conditions (Experiment 4). Error bars represent ± 1 standard error of measurement.

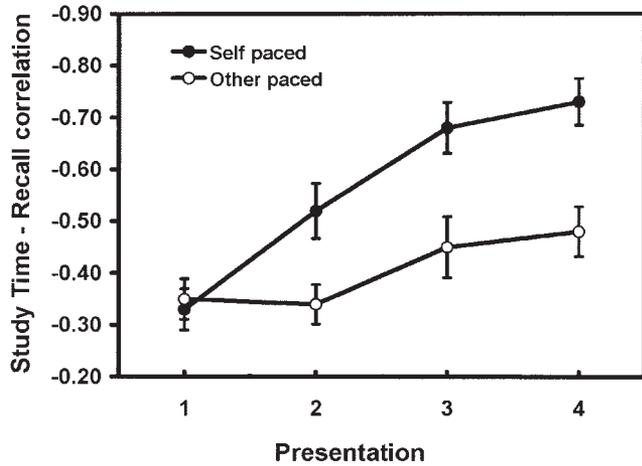


Figure 5. Mean within-participant gamma correlations between study time and recall as a function of presentation, plotted separately for the self-paced and other-paced conditions (Experiment 4). Error bars represent ± 1 standard error of measurement.

The similarity of the results for the self- and other-paced conditions reinforces the conclusion that the effective mnemonic cue in the case of self-paced participants is not study time per se but rather memorizing effort, which can also be used by the other-paced participants. Furthermore, the comparison of the results presented in Figure 6 with those presented in Figures 4 and 5 reinforces the conclusion that the effective cue for JOLs is idiosyncratic in nature and is best disclosed by the self-allocated study time to each item.

Note that the correlation between JOLs and recall for the self-paced condition (.51) was about the same as that between item difficulty and recall (.48) in Presentation 1. Following practice, however, the predictive validity of JOLs surpassed that of judged

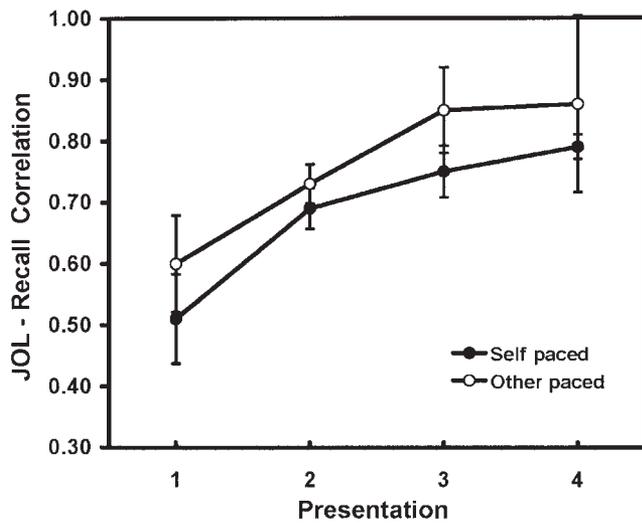


Figure 6. Mean within-participant gamma correlations between judgment of learning (JOL) and recall as a function of presentation, plotted separately for the self-paced and other-paced conditions (Experiment 4). Error bars represent ± 1 standard error of measurement.

difficulty: The respective correlations for Presentation 2 were .69 and .39, respectively, $t(13) = 3.44, p < .005, \eta_p^2 = .48$, and the difference was significant for Presentations 3 and 4 as well.

The nature of the mnemonic cues underlying practice effects. In previous studies participants have been found to assign higher JOLs to items that they had recalled on a previous occasion than to those that they had not (e.g., King et al., 1980; Koriat, 1997; Lovelace, 1984; Mazzoni & Cornoldi, 1993). Is it possible, then, that the practice effects observed in this experiment are solely due to a deliberate reliance on the memory of the outcome of the previous recall opportunity? It is difficult to distinguish experimentally between this type of explicit inference and the type of inference that uses memorizing effort as a cue, because the factors that affect recall in one trial also affect memorizing effort in the next trial. However, we can evaluate the possibility that the negative correlation between study time and JOLs is not due solely to the discrimination between items that were recalled on a previous test and those that were not. To do so, we examined the relationship between study time and JOLs in Presentation 2 for items that had been recalled in Presentation 1 and for those that had not. For each participant, all study times for previously recalled and for previously not recalled items were split at the median. Mean JOLs for below-median and above-median study times are presented in Figure 7 for previously recalled and previously not recalled items. Study times were clearly shorter for items that had been recalled in the previous test ($M = 3.2$ s, $SD = 1.24$) than for those that had not ($M = 7.3$ s, $SD = 3.01$), $t(19) = 7.88, p < .0001, \eta_p^2 = .77$. Also, JOLs were considerably higher for recalled than for unrecalled items, averaging 81.1% ($SD = 11.86$) and 50.4% ($SD = 15.59$), respectively, $t(19) = 10.21, p < .0001, \eta_p^2 = .85$. However, for both types of items JOLs decreased with increasing study time. A Study Time (below vs. above median) \times Previous Recall (recalled vs. not recalled) ANOVA yielded $F(1, 19) = 97.26, MSE = 9.87, p < .0001, \eta_p^2 = .84$, for study time; $F(1, 19) = 100.18, MSE = 195.27, p < .0001, \eta_p^2 = .84$, for previous recall;

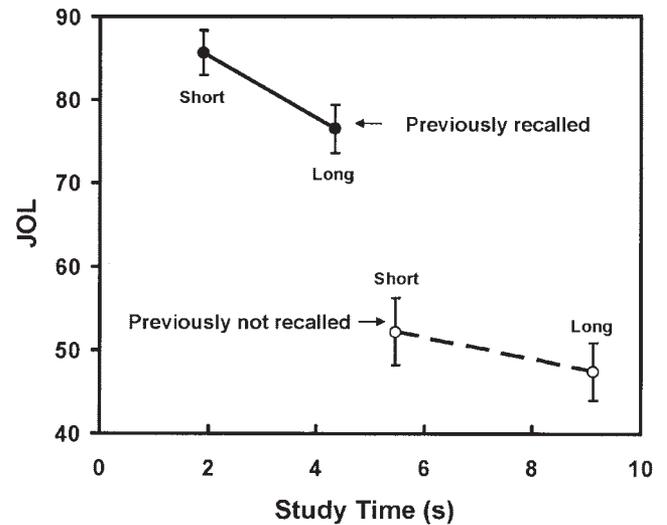


Figure 7. Mean judgment of learning (JOL) for below-median (short) and above-median (long) study time in Presentation 2, plotted separately for items recalled and not recalled in Presentation 1 (Experiment 4). Error bars represent ± 1 standard error of measurement.

and $F(1, 19) = 4.19$, $MSE = 22.66$, $p = .06$, for the interaction. JOLs were higher for below- than for above-median study times for both previously recalled and previously not recalled items, $t(19) = 8.80$, $p < .001$, $\eta_p^2 = .80$, and $t(19) = 3.21$, $p < .01$, $\eta_p^2 = .35$, respectively.

Figure 8 shows that the increase in the negative study time–JOL correlation with practice is preserved even when recall success on the previous test is controlled. In this figure the study time–JOL correlation is plotted as a function of presentation for items recalled and not recalled on the previous presentation (Presentation 4 was not included because of the low number of observations in the “previously not recalled” category). A Presentation (second vs. third) \times Previous Recall (recalled vs. not recalled) ANOVA yielded $F(1, 19) = 3.20$, $MSE = 0.059$, $p < .09$, $\eta_p^2 = .15$, for presentation; $F(1, 19) = 12.83$, $MSE = 0.094$, $p < .005$, $\eta_p^2 = .40$, for previous recall; and $F < 1$ for the interaction.

Discussion

Experiment 4 examined the second general assumption of Story 2, that the accuracy of metacognitive judgments stems in part from the reliance of these judgments on the feedback from control operations. We capitalized on the finding that repeated study–test practice improves the predictive accuracy of JOLs (e.g., Koriat et al., 2002). Assuming that this improvement stems from increased reliance on mnemonic cues pertaining to the processing of the items (Koriat, 1997), we hypothesized that the study time–JOL correlation would increase with practice. That is indeed what was found. In addition, the validity of study time in predicting recall also improved with practice. Thus, the seemingly paradoxical pattern that both JOLs and recall decrease with increased study time was found to intensify with repeated practice.

A comparison of the results for the self-paced participants and other-paced participants indicates that on the one hand, other-paced participants also exhibited improved monitoring with practice, suggesting that they too could benefit from increased reliance

on memorizing effort. On the other hand, however, the correlations with study time suggest that the effective cues used by other-paced participants are not captured by the study time invested by the self-paced participants. This latter result is consistent with Koriat’s (1997) finding suggesting that the mnemonic cues underlying the improvement in monitoring that occurs with practice are idiosyncratic in nature. It would seem, then, that the effective cues underlying JOLs and their improved accuracy in Experiment 4 lie in the idiosyncratic experience that learners gain in attempting to study the items.

We have previously (Experiment 1) considered the possibility that the negative correlations between study time and JOL simply reflect the fact that items that are perceived to be difficult induce longer study times and also elicit lower JOLs than items that are perceived to be easy. Admittedly, this possibility is difficult to rule out because of the inherent link between perceived difficulty and study time. However, the systematic changes that occurred with practice for self-paced participants would seem to argue against it. In fact, whereas the study time–JOL correlation increased with practice, the difficulty–JOL correlation decreased significantly with practice, $F(3, 57) = 29.09$, $MSE = 0.009$, $p < .0001$, $\eta_p^2 = .60$. Thus, it would seem that the study time–JOL correlation is not mediated entirely by a priori judgments of item difficulty.

A final note is in order: Metcalfe’s theory that study time is allocated to the region of proximal learning (Metcalfe, 2002; Metcalfe & Kornell, 2003) predicts that with repeated practice studying a list of items, the region in which study time is selectively allocated should shift toward the more difficult items because more and more of the easy items enter into the learned state. We should note that our data did not indicate such a shift: Study time increased monotonically with difficulty for each presentation (see Footnote 4), and there was no systematic change with practice in the proportion of study time appropriated to the easy, intermediate, and difficult items. However, it is still possible that some of the changes that occur with practice in cue utilization and cue validity are related to the shift in the region of proximal learning that was postulated by Metcalfe.

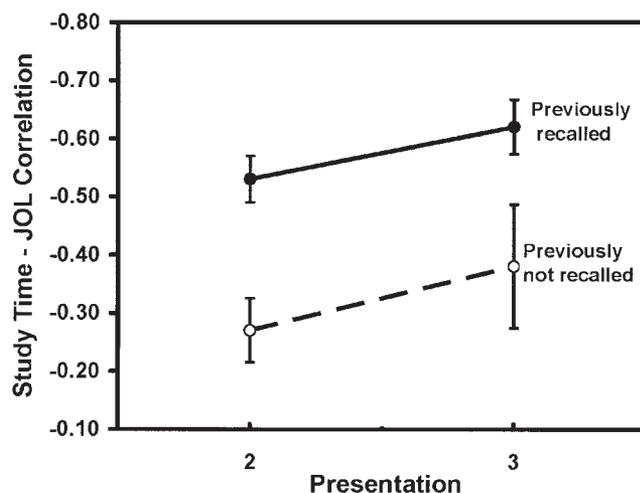


Figure 8. Mean within-participant gamma correlations between study time and judgment of learning (JOL) for Presentation 2, plotted separately for items recalled and not recalled in Presentation 1, and for Presentation 3, plotted separately for items recalled and not recalled in Presentation 2 (Experiment 4). Error bars represent ± 1 standard error of measurement.

Experiment 5

The experiments reported so far have focused on data-driven variation in study time. The aim of Experiment 5 was to bring in the control function of study time, whose signature is a positive correlation between study time and JOLs. Such a correlation is expected when study time is goal driven, regulated by the learner in accordance with specific goals that are extrinsic to the studied items. For example, a student may place a premium on a particular exam, strategically investing more effort in studying for that exam than he or she would otherwise. In that case, the added effort would be expected to instill a *stronger* sense of competence.

To bring to the fore the positive relationship assumed to characterize goal-driven control of study time, we used a differential-incentive condition in Experiment 5 (see Castel, Benjamin, Craik, & Watkins, 2002; Dunlosky & Thiede, 1998): Half of the items were awarded a 1-point bonus for their recall, and the remaining items received a 3-point bonus. The bonus associated with each item was indicated just before the presentation of that item for self-paced study. It was expected that high-incentive items would receive more study time as well as higher JOLs than low-incentive

items, thus resulting in a positive correlation between study time and JOLs across the two sets of items.

It is our proposal, however, that the introduction of goal-driven variation does not preclude the operation of data-driven effects on study time. Therefore, for all items awarded the same incentive, an *inverse* relationship should be found between study time and JOLs. Thus, Experiment 5 is expected to disclose the simultaneous operation of goal-driven and data-driven regulation within the same experiment, revealing both the control and monitoring functions of study time. In addition to the differential-incentive condition, Experiment 5 included a constant-incentive condition in which all items were awarded a bonus of 2 points.

Method

Participants. Thirty-two Hebrew-speaking University of Haifa undergraduates participated in the experiment for course credit. Sixteen participants were assigned randomly to each of the two conditions.

Materials. The list of stimuli was the same as that used in Experiment 1. For the differential condition, the list was divided into two sets of 30 items each, matched on difficulty ratings. One set was awarded an incentive of 1 point, and the other set was awarded an incentive of 3 points; this assignment was counterbalanced across participants. For the constant condition, all items were awarded an incentive of 2 points.

Apparatus and procedure. Participants were told that the experiment concerns the ability of people to allocate learning resources to different topics according to their importance. In the differential condition, participants were instructed to assume that they were studying for an exam in which some of the topics or items were more important to remember than others. The constant-incentive participants were told to assume that they were studying for an exam that was of intermediate importance and that other participants were studying for exams that were either more important or less important than theirs. All participants were told that the importance of each item would be indicated by an incentive value, which refers to the number of points that one earns for correct memory performance on that item. Participants were given self-paced instructions and were told to try to study the list so as to earn as many points as they could but to spend as little time as possible in studying the entire list.

A practice task was used (in both conditions) to familiarize the participants with the requirement to adjust their study effort to the designated incentive.⁶ Four brief stories (three-line paragraphs each) were presented, each on a separate page, with an incentive value indicated at the top of the page—2 for two stories and 1 and 3 for the remaining stories. Participants were instructed to study each paragraph so that they could later answer questions about it. They were told to take into account the importance of each story, as indicated by the incentive value associated with it, but to invest as little time as possible in studying all the stories. Four forced-choice questions were then presented, one about each of the paragraphs.

For the experiment proper, the apparatus was the same as in Experiment 1. The procedure for the study phase was also the same except for the following: The experiment consisted of three study–test cycles. The incentive value awarded to each item remained the same for each participant across the three presentations. On each study trial, the number (1, 2, or 3) designating the incentive value was presented together with a short beep, and 2 s thereafter the study pair was added on the screen. Both the number and the study pair remained on the screen until the participant pressed the left mouse button. The procedure for the test phases was also as in Experiment 1 except that the test phase followed immediately after the study phase.

Results

The effects of incentive on study time, JOLs, and recall. The amount of study time allocated by the differential-incentive group

to 1-point items and 3-point items and by the constant-incentive group (2 points) is plotted in Figure 9 (top panel) as a function of presentation. The respective means for JOLs are plotted in the bottom panel. Differential-incentive participants invested more time in the 3-point items than in the 1-point items, 4.2 s ($SD = 1.7$) and 3.5 s ($SD = 1.5$), respectively, $t(15) = 6.34$, $p < .0001$, $\eta_p^2 = .73$, consistent with the assumed control function of study time. The effect of incentive decreased with presentation but was significant for each of the presentations: $t(15) = 4.28$, $p < .001$, $\eta_p^2 = .55$, for Presentation 1; $t(15) = 3.34$, $p < .005$, $\eta_p^2 = .43$, for Presentation 2; and $t(15) = 3.06$, $p < .01$, $\eta_p^2 = .38$, for Presentation 3. For the constant-incentive group, study time per item averaged 3.1 s across presentations, somewhat less than invested by the differential group.

As expected, incentive level also exerted a significant effect on JOLs. An Incentive \times Presentation ANOVA for the differential condition yielded $F(1, 15) = 9.91$, $MSE = 36.21$, $p < .01$, $\eta_p^2 = .40$, for incentive; $F(2, 30) = 22.14$, $MSE = 84.57$, $p < .0001$, $\eta_p^2 = .60$, for presentation; and $F < 1$ for the interaction. JOLs increased from 63.0 ($SD = 13.6$) for 1-point items to 66.9 ($SD = 11.9$) for 3-point items. A similar ANOVA on recall, in contrast, yielded $F < 1$ for incentive; $F(2, 30) = 227.55$, $MSE = 50.30$, $p < .0001$, $\eta_p^2 = .94$, for presentation; and $F(2, 30) = 1.91$, $MSE = 33.59$, $p = .17$, for the interaction. Mean recall for 1-point and 3-point items averaged 68.8% ($SD = 20.8$) and 68.7% ($SD = 20.7$), respectively. These results are somewhat surprising in the light of those of Castel et al. (2002), who found participants' recall performance to exhibit excellent sensitivity to the incentives attached to recalling different words from 12-word lists.

In sum, increasing incentive from 1 point to 3 points increased study time (from 3.5 to 4.2 s per item across presentations), and at the same time enhanced JOLs (from 63.0% to 66.9%). Thus, increased study time was *positively* associated with increased JOLs, unlike the negative correlation documented in the previous experiments.

The effects of incentive for easy and difficult items. Kahneman (1973) proposed that a person “simply cannot try as hard in a relatively easy task as he does when the task becomes more demanding” (p. 14). Consistent with this proposal, the effects of incentive were less pronounced for the easier pairs than for the more difficult pairs, when these two classes of pairs were defined as in Experiment 1. An Incentive \times Difficulty \times Presentation ANOVA on the results (depicted in Figure 10) yielded significant effects for incentive, $F(1, 15) = 33.90$, $MSE = 0.54$, $p < .0001$, $\eta_p^2 = .69$; for difficulty, $F(1, 15) = 55.12$, $MSE = 1.75$, $p < .0001$, $\eta_p^2 = .79$; and for presentation, $F(2, 30) = 8.40$, $MSE = 8.54$, $p < .005$, $\eta_p^2 = .36$, but the Incentive \times Difficulty interaction was also significant, $F(1, 15) = 15.07$, $MSE = 0.34$, $p < .005$, $\eta_p^2 = .50$, indicating that the incentive manipulation had a more limited effect on the study of the easier pairs than on that of the harder pairs.

The effects of data-driven regulation of study time. We turn next to the effects of data-driven variation in study time, which

⁶ The inclusion of the practice task and the specific instructions used in the experiment were intended to maximize the likelihood of obtaining an effect of incentive on study time. Previous attempts to manipulate incentives in the context of JOL studies have met with limited success in affecting study time (Dunlosky & Thiede, 1998; Le Ny et al., 1972).

follow from the CM model. This can be seen when we focus on each incentive level separately. For all items associated with the same incentive level (1 or 3), study times were split at the median for each participant, and average JOLs for below-median and above-median items were calculated. The mean JOLs across all participants are plotted in Figure 11 for items receiving below- and above-median study times. A similar analysis was carried out for the constant-incentive group across all items, and the results of this analysis are also plotted in Figure 11.

Now the relationship between study time and JOL is *negative*, consistent with the CM model. JOLs were significantly higher for below-median than for above-median study times for both the 1-point and the 3-point items in the differential-incentive group, $t(15) = 5.86, p < .0001, \eta_p^2 = .70$, and $t(15) = 6.59, p < .0001, \eta_p^2 = .74$, respectively, and also for the constant-incentive condition, $t(15) = 5.65, p < .0001, \eta_p^2 = .68$. The slope of the function relating JOLs to study time was -8.23 and -6.95 for the 1-point and 3-point items, respectively, in the differential condition group and -7.22 for the constant group.

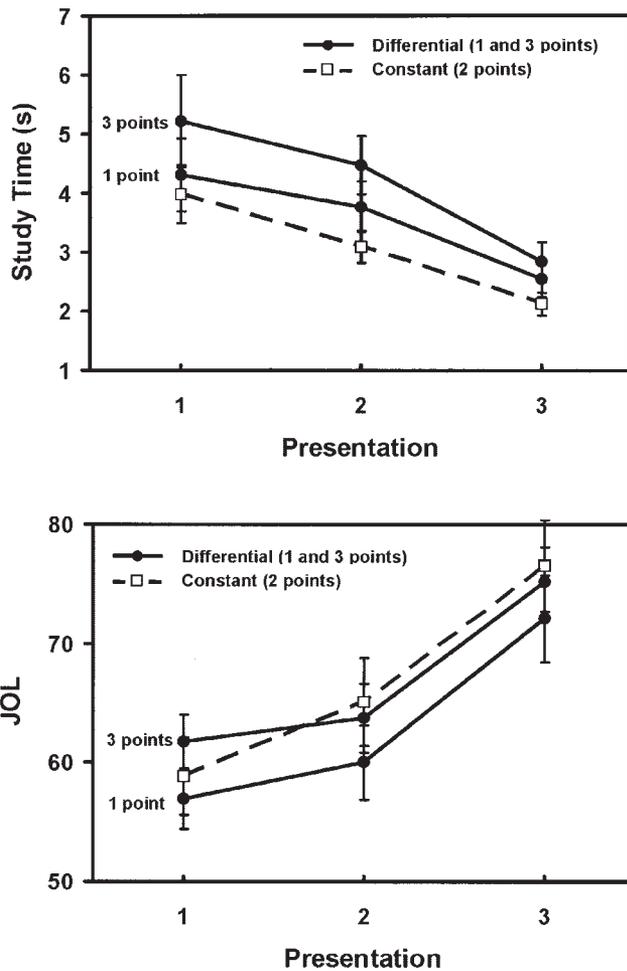


Figure 9. Mean study time (top panel) and judgment of learning (JOL; bottom panel) as a function of presentation, plotted separately for the 1-point and 3-point incentives of the differential-incentive condition and for the constant-incentive (2-point) condition (Experiment 5). Error bars represent ± 1 standard error of measurement.

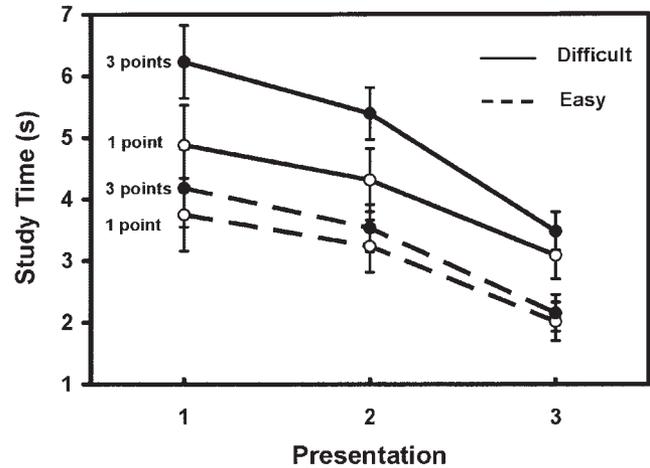


Figure 10. Mean study time as a function of presentation for difficult and easy items in the differential-incentive condition, plotted separately for each incentive level (Experiment 5). Error bars represent ± 1 standard error of measurement.

To allow comparison of these results with those associated with the control function of study time, we have included in Figure 11 a plot of the function relating mean JOLs for incentive levels 1 and 3 in the differential condition (56.9% and 61.7%, respectively) to mean study time allocated in these two levels (4.3 s and 5.2 s, respectively) in Presentation 1.⁷ The relationship that reflects the control function of study time is positive. The slope of the function relating JOLs to study time was 5.26.

The combination of goal-driven and data-driven variation in study time. The combination of goal-driven variation and data-driven variation in study time in the same situation should present a problem for the differential-incentive participants, because increased study time is expected either to increase or to reduce JOLs depending on its source. The results depicted in Figure 11, then, suggest the operation of an *attribution* process in which variations in study time are first attributed to data-driven effects or to goal-driven effects before the implications for JOLs are drawn.

To gain some insight into the nature of this attribution, we examined the possible interaction between the two ways in which study time can affect JOLs. Assuming that participants regulate study time in keeping with the specified incentives, would that regulation spoil the diagnostic value of study time as an index of intrinsic item difficulty? If so, we should expect a weaker dependence of JOLs on study time in the differential-incentive than in the constant-incentive condition. However, the slope of the function relating JOLs to study time was -7.22 for the constant condition, as noted earlier, whereas that for the differential condition (calculated across the two incentives) was -7.49 , suggesting that the utilization of data-driven variation in study time as a basis of JOLs was not impaired by the inclusion of goal-driven, top-down variation.

⁷ The JOL and study time means used in plotting the effects of incentive in this figure were calculated by averaging the respective below-median and above-median means. Therefore, they differ somewhat from the means that entered into the analyses of the effects of incentive reported earlier.

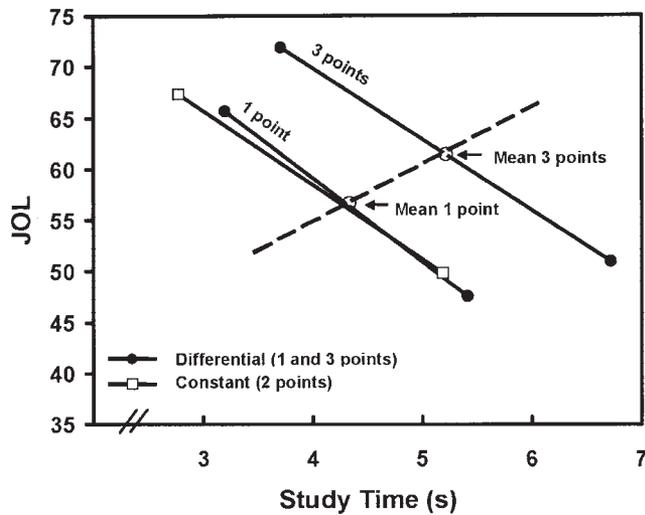


Figure 11. Mean judgment of learning (JOL) for below-median and above-median study time for each incentive level. Plotted also (broken line) is mean JOL as a function of mean study time for each incentive level of the differential-incentive condition (labeled *Mean 1 point* and *Mean 3 points*; Experiment 5).

Further evidence for this conclusion comes from a comparison of the study time–JOL correlation for the differential-incentive condition (calculated across the two incentives) with that for the constant-incentive condition. Figure 12 indicates that for both conditions, this correlation was negative and increased with presentation, replicating the results for the self-paced condition of Experiment 4. There was no indication, however, that the dependence of JOLs on study time was any weaker for the differential-incentive condition: A two-way ANOVA yielded $F(2, 60) = 20.92$, $MSE = 0.21$, $p < .0001$, $\eta_p^2 = .41$, for presentation, and $F < 1$ for both condition and the interaction. Only in Presentation 3 was there a trend in this direction, but this trend was not significant, $t(30) = 0.98$, $p = .34$.

Neither was there any evidence that the predictive validity of study time or JOLs was impaired by the inclusion of differential incentives. Thus, a two-way ANOVA indicated that, as in Experiment 4, the JOL–recall correlation across 31 participants (1 participant attained 100% recall in the last presentation) increased with practice from .51 in Presentation 1 to .83 in Presentation 3. A Presentation \times Condition ANOVA yielded $F(2, 58) = 41.11$, $MSE = 0.02$, $p < .0001$, $\eta_p^2 = .59$, for presentation. There was no effect for condition or the interaction ($F < 1$ for both). Similarly, the study time–recall correlation was negative, as in the self-paced conditions of the previous experiments, and increased from $-.26$ in Presentation 1 to $-.61$ in Presentation 3 across all participants. A similar ANOVA as above yielded $F(2, 58) = 18.74$, $MSE = 0.05$, $p < .0001$, $\eta_p^2 = .39$, for presentation, but again $F < 1$ for both condition and the interaction.

Discussion

The results for the differential condition brought to the fore the positive correlation between study time and JOLs, which is the signature of the control function of study time. Presumably, the differential bonus associated with different items results in the

allocation of more study time to the 3-point items than to the 1-point items and correspondingly in higher JOLs for the former than for the latter items. There are two processes that can bring about such positive correlation. First, increased study time enhances fluent processing, and enhanced fluency can serve as an internal, mnemonic cue to support higher JOLs (see Begg et al., 1989; Benjamin & Bjork, 1996). Second, study time can be used as an extrinsic cue under the belief that an item is more likely to be remembered when it is studied for a longer duration than when it is studied for a shorter duration. Koriat and Ma'ayan (2005) recently provided evidence in support of the former account when study time was experimentally manipulated: Increased presentation duration was found to enhance retrieval fluency, and this increase was sufficient to account for the concomitant increase in JOLs.

In addition, however, the results disclosed the operation of the monitoring function of study time, whose signature is an inverse relationship between study time and JOLs. This latter relationship reflects the retrospective use of study time as a diagnostic cue for JOLs.

How do differential-incentive participants distinguish between the variation in study time that is due to goal-driven effects and that due to data-driven effects? Consider, for example, a learner who invests an inordinately strong effort in studying a certain item, in part because the item turns out to be difficult and in part because its recall is associated with a high incentive. The results suggest that he or she can tease apart the component of study time that is driven by the item from that which is due to the self-control over study time, and assign a negative weight to the former and a positive weight to the latter in computing JOLs. How does one do that? We proposed that an attribution process must be postulated to mediate between study effort and JOLs, so that variations in study time are attributed (in some proportion) to data-driven and to goal-driven differences. The comparison of the results for the differential and constant conditions did not throw light on the nature of the underlying process except to suggest that neither the utilization of data-driven variation in study time as a basis for JOLs

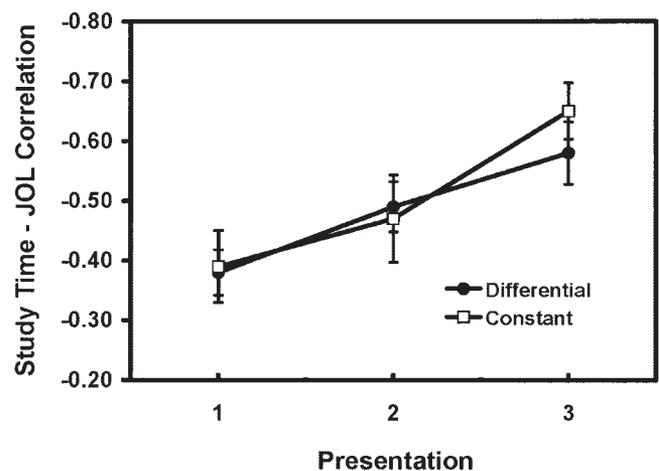


Figure 12. Mean within-participant gamma correlations between study time and judgment of learning (JOL) as a function of presentation, plotted separately for the differential-incentive and constant-incentive conditions (Experiment 5). Error bars represent ± 1 standard error of measurement.

nor the accuracy of JOLs in predicting recall is compromised by the introduction of goal-driven variation (but see Experiment 6).

Experiment 6

Although Experiment 5 highlighted the control function of study time, it did not disclose in full the type of monitoring-based control analogous to that in which we run away because we are frightened. This type of causal relation has been assumed by the discrepancy-reduction approach to underlie the allocation of more study time to the more difficult items.

What are the conditions, then, that produce the kind of monitoring-based control that accords with Story 1? We argue that one such condition is precisely that in which learners are led to invest more time in the *easier* items because in that case their choice most likely reflects a strategically controlled policy, similar to that underlying the choice to invest more time in items associated with higher incentives (Experiment 5). Clearly, the allocation of less study time to the more difficult items cannot be ascribed to bottom-up, data-driven effects; rather it is more likely to stem from top-down processes that, in fact, operate against data-driven processes that invite greater investment in the more difficult items. Therefore, we should expect a *positive* correlation between study time and JOLs across items.

One condition in which learners have been found to allocate more study resources to the easier items is when severe time pressure is imposed (Metcalf, 2002; Son & Metcalf, 2000). When only a limited amount of time is available for study, it might not be wise to concentrate on the difficult items. What is the process underlying the allocation of study time in that case? In order that more study time will be allocated to the easier items, a preliminary assessment of ease of learning must be used as a basis for the decision to continue studying the item or quit. Such assessment may rely on a priori beliefs or on the feedback from the initial attempt to study the item, implying that the decision to continue studying an item involves monitoring-based control rather than control-based monitoring.

This idea was tested in Experiment 6, which was similar in design to Experiment 5 with two exceptions: First, time pressure was imposed. Second, the materials used demanded and also allowed more study time to be allocated, so that learners could, in principle, reach a quick preliminary assessment that the item is too difficult and that given the limited time available it would not be expedient to continue studying it.⁸ We expected Experiment 6 to yield a positive relation between study time and JOLs even within each incentive level.

Learners in Experiment 6 were instructed in advance that there was little chance that they would be able to memorize all the items during the time allotted, and a running counter was displayed during study indicating both the time left and the number of items that were still to be presented. As in Experiment 5, the incentive for recalling each item was manipulated in the differential-incentive condition, whereas in the constant-incentive condition it remained the same for all items. We expected that, as in Experiment 5, both study time and JOLs would increase with increasing incentive, resulting in a positive correlation between them, but unlike in Experiment 5, we expected that within each incentive condition, JOLs would also increase with the amount of study time allotted to an item.

Method

Participants. A total of 48 Hebrew-speaking University of Haifa undergraduates participated in the experiment, 12 for course credit and 36 for pay. Twenty-four participants were assigned randomly to each condition.

Materials. The study list included 22 sets, each consisting of six Hebrew words. Half of the sets (*easy*) were composed of words that belonged to a common semantic domain (e.g., *newspaper, note, letter, library, poem, translation*), whereas the other sets (*difficult*) consisted of unrelated words (e.g., *road, joke, computer, cup, box, glue*). Effort was made to avoid obvious links between words that belonged to different sets. For each set, a test item consisting of five words was constructed by removing one of the words in that set. For the differential condition, half of the items in each difficulty category were assigned to the 1-point condition, and half to the 5-point condition, with the assignment being counterbalanced across participants.

Apparatus and procedure. The apparatus was the same as in Experiment 5. The procedure was similar except for the following: Participants were instructed to study each word set so that when presented with five words from that set they would be able to recall the missing sixth word. They were told that the importance of each set would be indicated by an incentive value—1 or 5 for the differential condition and 3 for the constant condition. Participants were informed that some of the sets would be easier whereas others would be more difficult to study. To create time pressure, we led them to believe that there were altogether 40 sets in the study list, but because they would have only 15 min for study, it is unlikely that they would be able to see all the sets. Because their task was to gain as many points as possible, they were told to try not to spend too much time on each item so that they would have a chance to reach the items that appeared later in the list. In actuality, however, the study phase ended when participants finished studying the 22 sets.

To maintain a severe time pressure throughout the study phase, a running counter was displayed for 5 s following the 4th, the 9th, the 14th, and the 19th sets. The counter consisted of two circles whose areas were gradually covered from one presentation to the next. Participants were told that one circle represented the overall amount of time spent as a proportion of the total amount of time available and the other represented the proportion of studied sets (out of 40). In actuality, the area covered in the time circle amounted to 4/22, 9/22, 14/22, and 19/22, respectively, for the four presentations of the counter, and in the second circle the covered area amounted to 4/40, 9/40, 14/40, and 19/40, respectively.

On each study trial, the incentive value (1, 3, or 5) appeared on the screen for 1 s. A beep was then sounded, after which the set was presented. Both the number and the set remained on the screen until the participant pressed the left mouse key to indicate end of study, and 500 ms thereafter a JOL prompt appeared. Participants were told that only the time used for study proper (from the presentation of the set until the participant pressed the mouse button) would be subtracted from the “allotted” time. After studying the 22 sets, participants were notified by the computer that the allotted time had ended, and then the following question appeared on the screen: “You have studied 22 sets. For how many of them do you think you will be able to recall the missing word?” The data from this aggregate estimate are not reported here. During the test phase, each trial was initiated by a beep, followed by the presentation of the test item on the screen. The test item disappeared when participants said the missing word or when 20 s had elapsed. A paper-and-pencil practice involving two items

⁸ In a pilot study conducted in preparation for Experiment 6, we used the same materials as in Experiment 5 but with the addition of time pressure. The results suggested that participants continued to invest more study time in the difficult items. It was our impression that when the study of an item requires only a few seconds, learners find it more expedient to simply let the study process run to completion rather than interrupt it after an initial assessment that the item is difficult to commit to memory.

preceded the experiment proper. Participants were also given instructions about the counters (circles), with the help of illustrative examples.

Results

The allocation of study time. The study time allocated to the easy and difficult items averaged 13.5 s ($SD = 8.4$) and 9.5 s ($SD = 8.0$), respectively, for the constant condition and 12.9 s ($SD = 5.4$) and 11.6 s ($SD = 5.3$), respectively, for the differential condition. A Difficulty \times Condition ANOVA yielded $F(1, 46) = 8.09$, $MSE = 20.70$, $p < .01$, $\eta_p^2 = .15$, for difficulty and $F < 1$ for condition. The effects of difficulty were somewhat stronger for the constant condition than for the differential condition, but the interaction was not significant, $F(1, 46) = 1.95$, $p = .17$.

The effects of incentive on study time for the differential-incentive group. An Incentive \times Difficulty ANOVA on study time for the differential condition yielded a nonsignificant interaction ($F < 1$). Overall, participants invested more time in the 5-point items ($M = 15.2$ s, $SD = 7.5$) than in the 1-point items ($M = 9.3$ s, $SD = 4.3$), $t(23) = 3.92$, $p < .001$, $\eta_p^2 = .40$, replicating the results of Experiment 5. The effects of incentive were significant for both the easy items, $t(23) = 3.81$, $p < .001$, $\eta_p^2 = .39$, and the difficult items, $t(23) = 3.70$, $p < .005$, $\eta_p^2 = .37$. For the easy items, study time for the high-incentive and low-incentive items averaged 15.4 s ($SD = 7.7$) and 9.9 s ($SD = 4.5$), respectively. The respective values for the difficult items were 15.0 s ($SD = 8.3$) and 8.8 s ($SD = 5.1$).

The relationship between study time and JOLs. We analyzed the results in the same way as in Experiment 5 (see Figure 11). Mean JOLs for slow and fast responses are plotted in Figure 13 for the constant-incentive condition (3 points) and for the high-incentive (5 points) and low-incentive (1 point) items of the differential-incentive condition.

Unlike what was found in Experiment 5, the relationship between study time and JOLs was positive even within each incentive condition, consistent with the control function of study time. For the constant-incentive condition, JOLs were higher for items

with above-median study times ($M = 49.9\%$, $SD = 14.6$) than for items with below-median study times ($M = 26.6\%$, $SD = 21.6$), $t(23) = 5.34$, $p < .0001$, $\eta_p^2 = .55$. A positive relationship was also obtained across the two incentive levels of the differential-incentive condition: JOLs for items with below-median and above-median study times averaged 30.3% ($SD = 23.0$) and 46.0% ($SD = 17.0$), respectively. A Study Time \times Incentive ANOVA yielded significant effects for study time, $F(1, 23) = 10.53$, $MSE = 273.59$, $p < .005$, $\eta_p^2 = .31$, and for incentive, $F(1, 23) = 59.00$, $MSE = 122.77$, $p < .0001$, $\eta_p^2 = .72$, but not for the interaction, $F(1, 23) = 1.41$, $MSE = 70.32$, $p = .25$. JOLs increased with increasing study time for both the high-incentive items, $t(23) = 2.36$, $p < .05$, $\eta_p^2 = .20$, and the low-incentive items, $t(23) = 3.42$, $p < .005$, $\eta_p^2 = .34$. The slope of the function relating JOLs to study time was 2.69 for the constant-incentive condition, and 2.00 and 1.13, respectively, for the 1-point and 5-point items in the differential-incentive condition.

To allow comparison of these results with those associated with the effects of incentive, we have included in Figure 13 a plot of the function relating mean JOLs for the 1-point and 5-point incentives in the differential condition (30.2%, $SD = 19.4$, and 46.3%, $SD = 17.3$, respectively) to mean study time allocated to items at each of these two levels of incentive (9.3 s, $SD = 4.3$, and 15.2 s, $SD = 7.5$, respectively) (see Footnote 7). The slope of the function relating JOLs to study time was 2.72.

The relationship between study time and recall. In the previous experiments, study time was inversely correlated with recall. In Experiment 6, in contrast, participants were expected to invest more time in the judged-easy items (see Metcalfe & Kornell, 2003). Therefore, we expected recall to correlate positively with study time.

Indeed, for the constant-incentive condition, recall was better for items with above-median study time ($M = 31.1\%$, $SD = 26.5$) than for items with below-median study time ($M = 14.0\%$, $SD = 17.8$), $t(23) = 3.27$, $p < .005$, $\eta_p^2 = .32$. The respective means for the differential-incentive condition, calculated across both incentives, were 26.4% ($SD = 22.9$) and 10.3% ($SD = 9.9$), respectively, $t(23) = 3.38$, $p < .005$, $\eta_p^2 = .33$. A Study Time (below median vs. above median) \times Incentive ANOVA yielded significant effects for study time, $F(1, 23) = 7.52$, $MSE = 259.77$, $p < .05$, $\eta_p^2 = .25$, and for incentive, $F(1, 23) = 15.35$, $MSE = 441.03$, $p < .001$, $\eta_p^2 = .40$, but not for the interaction, $F(1, 23) = 1.15$, $p = .30$. Thus, recall increased with incentive, but unlike in Experiment 5, it also increased with increasing study time.

Clearly, the relationship between study time and recall cannot be interpreted solely in a causal sense because participants invested more study time in the easier items in the first place. But when difficulty level was partialled out, the mean correlation between study time and recall for the constant condition dropped from .33 to .22, which was still significant ($p < .005$, with $n = 22$). Thus, perhaps part of the study time–recall correlation is nevertheless due to the control function of study time. Note that the respective correlations between study time and JOLs were .42 and .34 ($p < .001$ and $p < .0001$, respectively, both with $n = 24$).

Data-driven and goal-driven effects. A comparison of the results of Experiment 5 (see Figure 11) and Experiment 6 (see Figure 13) suggests that the introduction of time pressure produced a qualitative change in the allocation of study time from being data driven to being goal driven. To bring to the fore this change, we focused on the constant incentive condition and have plotted in Figure 14 mean JOLs and recall for below-median (short) and

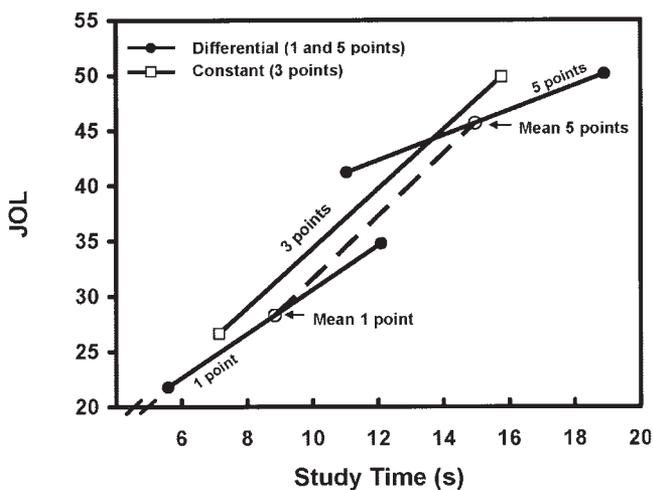


Figure 13. Mean judgment of learning (JOL) for below-median and above-median study time for each incentive level. Plotted also (broken line) is mean JOL as a function of mean study time for each incentive level of the differential-incentive condition (labeled *Mean 1 point* and *Mean 5 points*; Experiment 6).

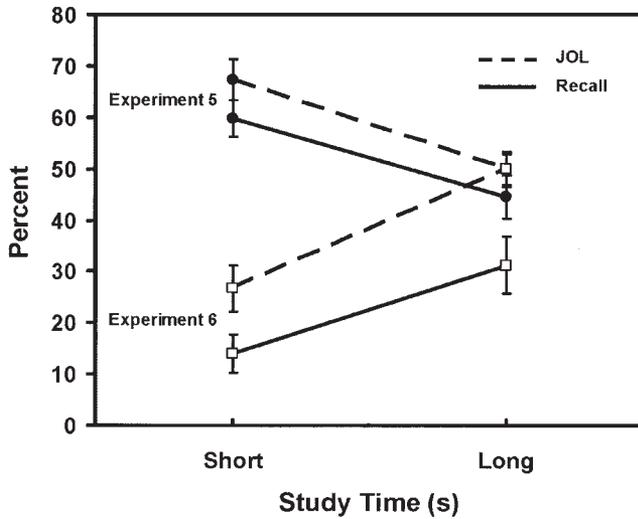


Figure 14. Mean judgment of learning (JOL) and recall for below-median (short) and above-median (long) study times for the constant-incentive condition in Experiments 5 and 6. Error bars represent ± 1 standard error of measurement.

above-median (long) study times in both Experiment 5 (Presentation 1) and Experiment 6. A Study Time (short vs. long) \times Experiment (5 vs. 6) ANOVA yielded a highly significant interaction for both JOLs, $F(1, 38) = 47.48, MSE = 167.52, p < .0001, \eta_p^2 = .56$, and recall, $F(1, 38) = 19.08, MSE = 264.74, p < .0001, \eta_p^2 = .33$.

This interactive pattern is also reflected in the within-person correlations between study time, on the one hand, and JOLs and recall, on the other hand, when the raw study times were used. The pertinent results are presented in Table 1. It can be seen that all the study time–JOL correlations were negative for Experiment 5 and differed significantly from the respective correlations for Experiment 6, which were all positive. The same pattern was observed for the study time–recall correlations.

Discussion

The results of Experiment 6 are consistent with those of Experiment 5 as far as the effects of incentive level are concerned. When incentive level was held constant, however, the correlations be-

tween study time, JOLs, and recall were diametrically opposed to those obtained in Experiment 5. First, participants invested more study time in the easier items rather than in the more difficult items (see Metcalfe, 2002; Son & Metcalfe, 2000), suggesting that learners actually regulated study time in the opposite direction from what data-driven forces may have led them to do. Second, study time correlated positively rather than negatively with JOL and recall performance. Presumably, the initial processing of an item provides diagnostic information about the ease with which an item can be committed to memory, and hence more time was invested in the items that were judged as easier to learn. In addition, increased investment of effort might have contributed to the positive correlation between study time and recall. Altogether, these results are consistent with the MC model, which is expected to hold when study time allocation is goal driven.

Thiede and Dunlosky (1999) proposed a superordinate, strategy-selection stage, in which learners decide whether it is worthwhile to focus on the easier or on the more difficult items. However, the results of Experiments 5 and 6, taken together, suggest that only the allocation of more study time to the *easier* items reflects a premeditated strategy. The allocation of greater effort to the more difficult items, in contrast, is what occurs when study time allocation is left to the mercy of the items. This contrast brings to the fore an important but subtle difference between Experiments 5 and 6 in the presumed role of item difficulty: Whereas in Experiment 5, item difficulty may be said to *dictate* time allocation, in Experiment 6 judged item difficulty may be said to *inform* strategic time allocation.

As in Experiment 5, we might ask, how do learners in the differential condition manage to control and monitor their performance in accordance with the two different types of strategic considerations—incentive level and item difficulty? A comparison of the results for the constant and differential conditions suggests that participants may find it difficult to meet both criteria simultaneously. Thus, as indicated earlier, the constant condition yielded a somewhat stronger effect of item difficulty on study time (an overall effect of 3.9 s) than the differential condition (1.4 s), a stronger effect of study time on JOLs (amounting to 23.3%, compared with 16.3% for the differential condition), and a directionally stronger effect of study time on recall (amounting to 17.1%, compared with 16.1% for the differential condition). This pattern suggests that the allocation of more study time to the easier items is applied more effectively when all items in a list receive the same incentive than when they receive different incentives.

Table 1
Mean Study Time–JOL and Study Time–Recall Gamma Correlations for Each Condition of Experiments 5 and 6, and *t* Tests Comparing These Correlations

Condition	Study time–JOL correlation			Study time–recall correlation		
	Experiment		<i>t</i> test	Experiment		<i>t</i> test
	5	6		5	6	
Constant	-.39 (<i>n</i> = 16)	.38 (<i>n</i> = 24)	<i>t</i> (38) = 7.00, <i>p</i> < .0001	-.23 (<i>n</i> = 16)	.46 (<i>n</i> = 22)	<i>t</i> (36) = 5.69, <i>p</i> < .0001
Differential						
Low incentive	-.40 (<i>n</i> = 16)	.42 (<i>n</i> = 24)	<i>t</i> (38) = 6.72, <i>p</i> < .0001	-.20 (<i>n</i> = 16)	.32 (<i>n</i> = 14)	<i>t</i> (28) = 3.28, <i>p</i> < .005
High incentive	-.50 (<i>n</i> = 16)	.11 (<i>n</i> = 24)	<i>t</i> (38) = 5.24, <i>p</i> < .0001	-.36 (<i>n</i> = 16)	.15 (<i>n</i> = 20)	<i>t</i> (34) = 3.71, <i>p</i> < .001

Note. Participants with perfect recall were excluded; *n* designates number of effective participants for each correlation. JOL = judgment of learning.

This conclusion differs from that suggested by the results of Experiment 5, where the function relating JOLs to study time had a similar (negative) slope for the differential and constant conditions (-7.49 and -7.22 , respectively), suggesting that the contribution of data-driven variation in study time to JOLs was unaffected by the additional variation in incentive. Further research is needed to determine whether this difference reflects a qualitative difference between a situation in which both types of variation induce goal-driven regulation (Experiment 6) and a situation in which one type of variation induces goal-driven regulation whereas the other induces data-driven regulation (Experiment 5).

Experiment 7

The aim of Experiment 7 was to generalize the conceptual framework proposed in this article to another type of metacognitive judgment: subjective confidence. It has been proposed that once an answer to a question has been retrieved or selected, the confidence in that answer is based on the feedback from the process leading up to the solution. Among the cues that have been claimed to contribute to subjective confidence is the time and effort it takes to reach the answer or the decision. As with the memorizing effort heuristic, the assumption is that the greater the effort and the longer the deliberation needed to reach an answer, the lower the confidence in that answer will be. Indeed, several studies have documented an inverse relation between confidence and decision time (e.g., Barnes et al., 1999; Costermans et al., 1992; Kelley & Lindsay, 1993; Nelson & Narens, 1990; Robinson et al., 1997). Although this correlation does not allow specifying which is the cause and which is the effect, Kelley and Lindsay (1993) showed that when response latency is enhanced through priming, confidence judgments also increase accordingly. These results are important in supporting the view of confidence judgments as reflecting control-based monitoring: Possibly, the time to arrive at an answer or a solution is affected by a variety of factors that are inherent in the question or the problem, and once an answer has been reached, the amount of effort and time expended can serve as a cue for the feeling of certainty, much like the way in which study time is assumed to affect JOLs when study time is data driven.

Like study time, however, decision time may also be goal driven: A person is more likely to dwell longer on a question when the motivation for reaching a correct answer is particularly strong. For example, in a forced-choice quiz, a student would be expected to spend more time on questions that score higher. In such cases, variation in decision time should yield a *positive* relationship between deliberation time and confidence in the answer.

The general design of Experiment 7 was similar to that of Experiment 5: Participants attempted to solve several psychometric problems, some of which were associated with a higher incentive (5 points) than others (1 point). After choosing a solution from among distractors, they indicated their confidence in that solution. The variation in incentive was expected to yield a positive correlation between decision time and confidence, consistent with the control function of decision time. Within each incentive level, however, confidence was expected to decrease with increasing decision time, consistent with the presumed monitoring function of decision time.

Method

Participants. Forty-six undergraduate students from the University of Haifa were paid for participating in the experiment.

Materials. A figural matrices task and a figural series task (borrowed from Sheffer, 2003) were used, each including 34 items. These were similar to the items in Raven's Progressive Matrices (Raven, Court, & Raven, 1979) and were modified from out-of-date psychometric entrance tests to Israeli universities. In the figural matrices task, each item consisted of a 3×3 array of symbols, ordered according to a certain principle, with the bottom right-hand symbol missing. Participants had to choose, out of four symbols, the one that logically completed the array. In the figural series task, each item consisted of a series of four or five symbols ordered from left to right according to a certain principle. Participants' task was again to choose the correct symbol that completed the series out of four symbols that appeared beneath.

Four of the items from each task were used for practice. The remaining 30 items, which were used in the experiment proper, were divided into two equal sets that were matched in terms of difficulty (based on available norms). One set was awarded an incentive of 1 point, and the other was awarded an incentive of 5 points for a correct solution, with the assignment of incentive to each set counterbalanced across participants.

Apparatus and procedure. The experiment was conducted on a personal computer. The stimuli and all instructions appeared on the screen. The order of the two tasks was counterbalanced across participants. Participants were told that the experiment evaluated a special version of the University Psychometric Entrance Test and that they should strive to attain the highest score that they could.

Participants were promised a financial reward according to their performance. They were told that normally, a 1-point bonus would be awarded for each correct answer but that occasionally "gift items" would appear for which the bonus would be 5 points for a correct answer. The number of gift items would be determined randomly by the computer (and so the participant should feel lucky to receive a gift item rather than a 1-point item). It was indicated that the level of difficulty of the 1-point and 5-point items was the same. Participants were told that the time allotted to each task was limited and that therefore their success depended on their ability to divide their restricted time wisely among the different items.

Each task began with the four practice items. On each trial, the bonus associated with the question appeared in the upper part of the screen: either "5 points!!!" (in red type) next to an icon of a gift or "1 point" (in green type). After a 1-s interval, the question and the four alternatives were added on the screen and remained until the participant, using the computer's mouse, marked one of the alternatives and confirmed his or her choice. Decision time was measured from the presentation of the problem until the confirmation press. Immediately thereafter, a confidence scale (25%–100%) was added beneath the alternatives, and participants marked their confidence by sliding a pointer on a scale using the mouse. The instructions specified that participants should assess the likelihood that the answer chosen was correct and that 25% indicated chance performance. At the end of each task, participants were given feedback about the number of points they had won.

Results

The effects of incentive on decision time, confidence, and accuracy. There was a large variability in decision time between participants and between items. Mean decision time per participant ranged from 12.2 s to 89.6 s for the figural matrices task and from 14.7 s to 104.2 s for the figural series task. Mean decision time per item ranged from 18.2 s to 60.9 s for the figural matrices task and from 16.3 s to 55.7 s for the figural series task.

Figure 15 (top panel) presents decision time as a function of incentive level (1 vs. 5 points) for each of the two tasks. For both tasks, decision time increased with incentive. An Incentive \times Task

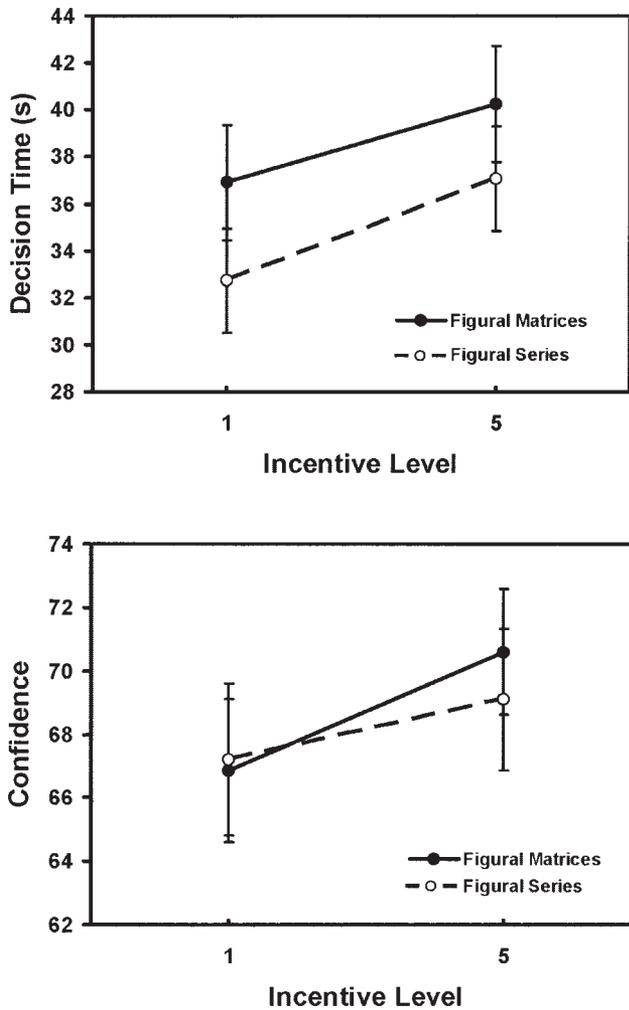


Figure 15. Mean decision time (top panel) and subjective confidence (bottom panel) as a function of incentive level (1 point vs. 5 points) plotted separately for the figural matrices and figural series tasks (Experiment 7). Error bars represent ± 1 standard error of measurement.

ANOVA yielded a significant effect for incentive, $F(1, 45) = 21.05$, $MSE = 32.25$, $p < .0001$, $\eta_p^2 = .32$. Neither the effect of task nor the interaction was significant, $F(1, 45) = 3.12$, $MSE = 198.79$, $p = .08$, and $F < 1$, respectively. The increase in decision time with incentive was significant for both the figural matrices task, $t(45) = 3.23$, $p < .005$, $\eta_p^2 = .19$, and the figural series task, $t(45) = 3.72$, $p < .001$, $\eta_p^2 = .23$.

Confidence also increased with incentive (Figure 15, bottom panel). An Incentive \times Task ANOVA yielded $F(1, 45) = 14.14$, $MSE = 25.76$, $p < .001$, $\eta_p^2 = .24$, for incentive. The effects of task and the interaction were not significant, $F < 1$ and $F(1, 45) = 1.45$, $MSE = 26.80$, $p = .24$, respectively. The increase in confidence with incentive was significant for the figural matrices task, $t(45) = 3.75$, $p < .001$, $\eta_p^2 = .24$, and near significant for the figural series task, $t(45) = 1.66$, $p < .11$, $\eta_p^2 = .06$. Thus, unlike the negative correlation that is typically observed between decision time and confidence, the effects of incentive produced a positive correlation such that increasing incentive from 1 to 5 points increased decision time from 34.8 s ($SD = 13.5$) to 38.7 s ($SD =$

14.1) on average and in parallel increased confidence judgments from 67.0% ($SD = 14.4$) to 69.9% ($SD = 13.1$).

Somewhat surprisingly, increased incentive did not improve actual performance significantly. We should note that performance was much better for the figural matrices task (59.3% correct solutions) than for the figural series task (24.5%, approximately at chance level). Across both tasks, the percentage of correct responses increased slightly from 40.7% for 1-point items to 43.1% for 5-point items. However, an Incentive \times Task ANOVA on the percentage of correct responses yielded a significant effect only for task, $F(1, 45) = 186.38$, $MSE = 298.59$, $p < .0001$, $\eta_p^2 = .81$. Neither the effect of incentive nor the interaction was significant, $F(1, 45) = 1.37$, $MSE = 203.67$, $p = .25$, and $F < 1$, respectively.

It should be noted that participants were markedly overconfident and particularly so for the figural series task: 68.2% confidence compared with 24.5% accuracy. The respective means for the figural matrices task were 68.7% and 59.3%, respectively.

The monitoring function of decision time. As in Experiments 5 and 6, we examined the relationship between confidence and decision time for each of the two incentive levels. First, for all items associated with the same incentive, decision times were split at the median for each participant, and average confidence judgments for below-median items and above-median items were calculated. This was done separately for each task. Then the mean confidence judgments of below-median and above-median items were averaged across both tasks and across all participants; the averages are displayed in Figure 16 for each incentive level.

It can be seen that for each incentive level, confidence judgments decreased as a function of increasing decision time, consistent with the postulated monitoring function of decision time (e.g., Kelley & Lindsay, 1993). A three-way ANOVA, Incentive \times Decision Time (below median vs. above median) \times Task indicated that confidence increased significantly with incentive, $F(1, 45) = 10.79$, $MSE = 58.36$, $p < .01$, $\eta_p^2 = .19$, but decreased significantly with decision time, $F(1, 45) = 96.76$, $MSE = 167.38$, $p <$

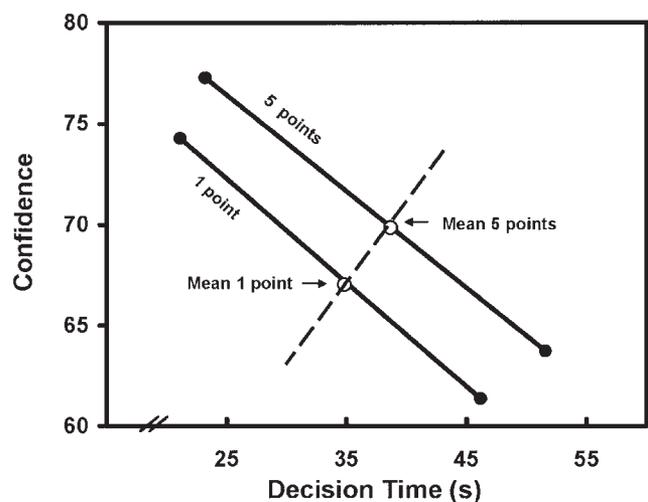


Figure 16. Mean subjective confidence for below-median and above-median study time for each incentive level. Plotted also (broken line) is mean confidence as a function of mean study time for each incentive level of the differential-incentive condition (labeled Mean 1 point and Mean 5 points; Experiment 7).

.0001, $\eta_p^2 = .68$ (with confidence decreasing overall from 75.8% [$SD = 18.5$] for the below-median items to 62.5% [$SD = 14.5$] for the above-median items). Confidence was significantly higher for below-median than for above-median decision times for both the 1-point items, $t(45) = 7.53, p < .0001, \eta_p^2 = .56$, and the 5-point items, $t(45) = 10.22, p < .0001, \eta_p^2 = .70$.

The three-way ANOVA also yielded a significant Task \times Decision Time interaction, $F(1, 45) = 33.25, MSE = 53.63, p < .0001, \eta_p^2 = .42$, indicating a stronger effect of decision time on confidence for the figural matrices task than for the figural series task. The Incentive \times Decision Time interaction, however, was not significant ($F < 1$). For the 1-point items, confidence decreased on average from 73.1% ($SD = 18.5$) to 61.4% ($SD = 12.9$) with increasing decision time, $t(45) = 6.14, p < .0001, \eta_p^2 = .46$. The respective values for the 5-point items were 76.6% ($SD = 15.9$) and 63.4% ($SD = 12.3$), $t(45) = 8.66, p < .0001, \eta_p^2 = .62$.

The slope of the function relating confidence to decision time was -0.51 and -0.48 for the 1-point and 5-point items, respectively. To allow comparison of these results with those associated with the control function of decision time, we have included in Figure 16 a plot of the function relating mean confidence for incentive levels 1 and 5 (67.0% and 69.9%, respectively) to the respective means of decision time (34.8 s and 38.7 s, respectively) (see Footnote 7). The relation reflecting the control function of decision time was positive, with a slope of 0.73.

Note that although the figural series task yielded very low performance, confidence judgments for this task were nevertheless correlated (negatively) with decision time: The within-participant gamma correlation between decision time and confidence averaged $-.20$ for 1-point items, $-.29$ for 5-point items, and $-.22$ across all items, all significant at the .0001 level. The respective correlations for the figural matrices task averaged $-.36, -.45$, and $-.38$, all significant at the .0001 level. Thus, it would seem that participants are influenced by decision latency in making retrospective confidence judgments.

Monitoring accuracy: The correlation between confidence and performance. The results just presented accord with the first assumption of Story 2, that metacognitive judgments are based on the feedback from the outcome of control operations. We now examine the second assumption—that the accuracy of these judgments derives from their reliance on such feedback.

Only the figural matrices task yielded moderate and significant within-person confidence–accuracy gamma correlations: .56 for the 1-point items, .55 for the 5-point items, and .56 across all items, all significant at the .0001 level. The respective correlations for the figural series task were very low and not significant: .10, .06, and .04, respectively. As noted earlier, performance on the figural series task was at chance level.

The results also suggested that the accuracy of confidence judgments derived in part from reliance on decision time. Indeed, for the figural matrices task, gamma correlations between decision time and accuracy averaged $-.40$ for the 1-point items, $-.23$ for the 5-point items, and $-.31$ across all items, all significant at the .0001 level. The respective correlations for the figural series task were low and nonsignificant: .08, $-.04$, and .05, respectively. When the confidence–accuracy correlation was calculated with decision time partialled out, the mean correlation for the figural matrices task dropped from .44 to .35, $t(45) = 3.16, p < .0001, \eta_p^2 = .46$, for the difference. Similar results were found when the analysis was conducted for each incentive level separately. Thus,

decision time explains part, but not all, of the confidence–accuracy correlation.

Discussion

Although Experiment 7 involved retrospective confidence judgments, the results were in agreement with those of Experiment 5, which involved recall predictions. First, increased incentive affected decision time and confidence in the same way: It increased decision time and at the same time enhanced confidence level, thus producing a positive correlation between decision time and confidence. In contrast, for each incentive level, confidence judgments decreased with decision time, as would be expected for control-based monitoring (see Kelley & Lindsay, 1993).

The similarity between the results presented in Figure 16 and Figure 11 is impressive given that the former concerns recall predictions whereas the latter concerns retrospective confidence, and given that study time was in the order of 2–7 s in Experiment 5, whereas decision time was in the order of 20–55 s in Experiment 7. Thus, it is noteworthy that the same conceptual framework can be applied to both types of metacognitive judgments. Note also that like study time in the previous experiments, decision time also proved to have some degree of validity: For the figural matrices task, answers that were reached more quickly had a better chance to be correct than answers that took longer to reach, so that reliance on decision time as a cue was likely to enhance monitoring accuracy (see Robinson et al., 1997). However, participants seemed to rely on decision latency as a cue for confidence even in the case of the figural series task, where decision latency had little diagnostic validity.

General Discussion

In this article, we examined the relationship between monitoring and control in metacognition with an eye to the general philosophical–psychological issue of the role that subjective experience and consciousness might play in behavior. Needless to say, the results presented here bear only indirectly on that issue. However, they suggest one line of research that might be pursued to scratch the surface of this intricate, long-standing issue.

In what follows, we first review our conceptual framework and examine how the reported findings bear on it. We then focus on previous work on metacognition and examine theories and findings that support the view that monitoring informs and drives strategic control, and those that assume that monitoring processes are based themselves on the feedback from control processes. Our review of the monitoring-based-control position is brief, but we dwell somewhat longer on theories and findings that imply control-based monitoring, discussing both the bases and accuracy of metacognitive feelings. We then examine research that bears on how monitoring-based control and control-based monitoring may combine in the course of information processing and behavior. Finally, we mention several open issues that deserve attention in future research.

Review of Our Conceptual Framework and the Pertinent Evidence

Our investigation of the relationship between monitoring and control was based on an analogy from emotional behavior. The

question that has been posed over the years is whether we run away because we are frightened or we are frightened because we run away. This analogy provided the logic for our investigation. If we focus on the *intensity* of fear and the speed of running away, then the correlation between them should be telling about the cause-and-effect relation between fear and running away: If it is the subjective feeling of fear that drives running away, then the faster one runs away the less frightened (or more safe) one should feel. In contrast, if fear is based on the feedback from running away, then the faster one runs away the more frightened (or less safe) one should feel.

Applying this logic to metacognition, we considered the relationship between the amount of effort invested in a task and the ensuing metacognitive feeling. The dominant view in metacognition research emphasizes the goal-oriented function of control operations. The discrepancy-reduction model, for example, which incorporates the TOTE model proposed by Miller et al. (1960) to describe goal-oriented behavior, predicts that JOLs following study should increase with study time. In contrast, if monitoring is based on the feedback from control operations, then JOLs following study should decrease with increased study time.

The CM model. Our first aim in the present article was to promote the idea that monitoring can be based on the feedback from control operations and thus follows rather than precedes control. We proposed that this occurs when control processes are data driven, tuned to the qualities of the task. Under such conditions, the feedback from control operations is likely to provide clues regarding the task in question. This is what happens, for example, when we attempt to judge the weight of an object by lifting it, because the feedback from the effort invested is telling about the intrinsic properties of that object. Such is also what typically happens in self-paced learning, when the amount of time and effort spent studying an item is left to the mercy of that item. Similarly, the FOK associated with the attempt to retrieve an item from memory is also likely to be based on the feedback from the retrieval attempt, because that feedback generally reflects on the specific task at hand. Also, the effort and time spent attempting to reach a decision or an answer typically convey information about the amount of doubt experienced (Adams & Adams, 1961). Thus, to the extent that control processes are data driven, we should expect metacognitive feelings to be based on the feedback from control processes.

What is the evidence in support of the CM model? First, with regard to JOLs, a negative correlation between study time and JOLs, which is the signature of control-based monitoring, was consistently observed. This was true in Experiments 1 and 2, in both the immediate and delayed JOLs of Experiment 3, in each of the four presentations of Experiment 4, and in each of the incentive conditions of Experiment 5. A similar, negative correlation between decision time and confidence was observed in Experiment 7. Thus, a negative correlation between control effort and metacognitive feelings was observed in all self-paced conditions of this study, consistent with the idea that metacognitive judgments are retrospective in nature, based on the feedback from control operations.

Second, with regard to monitoring accuracy, the prediction from the CM model is that to the extent that metacognitive predictions are based on mnemonic cues that reside in the feedback from control operations, their accuracy should be a function of (a) cue

utilization and (b) cue validity. We have discussed cue utilization. Let us consider next cue validity.

In Experiment 1, study time was negatively correlated with recall 4 months later, supporting the validity of the memorizing effort heuristic. The same inverse relationship was observed in Experiments 2 and 3. It was also found for each of the four presentations in Experiment 4 and for each incentive level in Experiment 5. Experiment 7 also yielded a negative correlation between decision time and accuracy for the figural matrices task. These results testify for the viability of the implicit naive theory underlying cue utilization: When control is data driven, control effort is inversely predictive of correct performance.

Given the observations regarding cue utilization and cue validity, it might be expected that metacognitive predictions would be accurate by and large. Indeed, in all of the experiments involving learning (Experiments 1–6), JOLs correlated positively with recall. In addition, in Experiment 7, confidence judgments were diagnostic of the correctness of the solution for the figural matrices task.

The results of Experiment 4 are particularly instructive regarding the intimate link between cue validity, cue utilization, and achievement: Both cue validity and cue utilization increased systematically with repeated study–test cycles, suggesting that with increased practice, both the reliance on study time as a cue for JOLs and the validity of study time in predicting recall increased. Both of these changes seem to contribute to the improved accuracy of JOLs with repeated study–test cycles.

The MC model. Let us turn next to the MC model, which seems to accord better with everyday intuitions. This model assumes that subjective experience informs the initiation and self-regulation of control operations that may in turn change subjective experience. Thus, when we feel that we do not understand a letter that we have just read, we read it again. When we feel that we have not mastered the to-be-remembered material, we spend more time studying it until we feel more confident.

The research reported in this article did not address in full the predictions of the MC model. However, much of the previous research that focused on the presumed causal antecedents of control operations has yielded results suggesting that metacognitive feelings drive and inform control operations (see Son & Schwartz, 2002, for a review). In the present study, in contrast, we focused on the consequences of control operations, because the consequences that are predicted by the MC model can be readily contrasted with those that follow from the CM model. The prediction from the MC model is that metacognitive feelings should increase rather than decrease with the invested effort. We proposed that this should occur when the strategic regulation of control processes is goal driven.

Indeed, in Experiment 5, both study time and JOLs were affected in the same way by incentive so that increased study time correlated with *increased* JOLs. This pattern was replicated in Experiment 6, which involved severe time pressure. An analogous pattern was observed in Experiment 7, in which decision time as well as confidence increased with the incentive associated with correct solution. Thus, when effort was strategically regulated in accordance with the person's goals, a positive correlation was obtained between effort and the ensuing metacognitive feelings.

Further support for the MC model comes from the results of Experiment 6, in which JOLs also increased with study time within each incentive level. This result is indeed what would be expected

when study time is used as a strategic tool toward the achievement of particular goals. But whereas previous discussions have found the evidence for the MC model in the greater allocation of study time to the more difficult items (Dunlosky & Hertzog, 1998; Mazzoni & Cornoldi, 1993; Nelson & Leonesio, 1988), here that model seems to be best revealed under conditions that induced learners to allot more study time to the easier items. Presumably, under severe time pressure, learners had to mobilize effort to counteract the data-driven demand to invest more time in the more difficult items. The hypothesized process in this case is that a fast preliminary monitoring drives greater investment in the easier items, which in turn contributes to further enhancing the JOLs associated with these items.

The combined operation of the CM and MC models. We return now to the question about fear and running away. Assuming with William James (1884) that we meet a bear and run, should we feel more frightened or less frightened the faster we run away? The answer suggested by the foregoing discussion is that to the extent that running away is entirely data driven, dictated by the speed of the bear, the faster we run away the more fear we should experience. However, to the extent that we make an effort that goes beyond that required just to maintain a safe distance from the bear, the extra effort invested in running away should contribute toward reducing our feeling of fear. In general, variations in effort that are not accounted for entirely by data-driven effects should be correlated positively with variations in metacognitive feelings.

This is, in fact, the pattern that is suggested by the results of Experiments 5 and 7. These results illustrate the situation in which effort (study time or solution time) is both data driven and goal driven. Whereas increased data-driven effort reduced metacognitive feelings (JOLs or confidence), increased goal-driven effort enhanced these feelings. Both types of relations were observed within the same situation, suggesting that the MC and CM models are not mutually exclusive, as might seem to be implied by William James's discussion.

We also sketched a second mode in which the two models can be combined in everyday life—the cascaded mode (Koriat, 1998), in which monitoring-based control may lead to control-based monitoring. This mode is only implied by the results of Experiment 6, which suggest that a preliminary monitoring may drive increased investment in the study of the judged-easier items and that investment may then contribute to the higher JOLs associated with these items. More direct evidence for this mode has been presented elsewhere (e.g., Koriat & Levy-Sadot, 2001; Son & Metcalfe, 2005), as is discussed below.

In sum, the results presented in this article generally agree with the conceptual framework proposed. These results, however, were obtained within a restricted domain of metacognition. In what follows, we shall examine previous findings and discussions in metacognition in order to show that some of these share certain ideas advanced in this article.

Reflections on Story 1 in Metacognition Research

The assumptions of the MC model have their roots in social psychological approaches that stress the role of one's beliefs, perceptions, and attributions in mediating one's feelings and behavior (see Bandura, 1986; Bless & Forgas, 2000; Jost, Kruglanski, & Nelson, 1998; Schwarz, 2004). Within metacognition research, discussions that subscribe to the MC model generally

assume that metacognitive processes operate in the service of goal-oriented behavior.

Two features are common to several formulations embodying an MC model. The first is that self-regulation is hierarchically organized: At a superordinate level, decisions are made regarding the policy for the task as a whole, and that policy is then implemented at the subordinate level on the basis of online item-by-item monitoring. Thus, learners may plan to concentrate on the easier or on the more difficult items depending on such factors as time pressure (Metcalfe, 2002). This policy is then implemented in studying each item taking into account online item-by-item JOLs (Dunlosky & Thiede, 2004). Similarly, with regard to memory retrieval, it has been proposed that the general policy of spending more or less time searching for answers depends on the relative importance of speed versus accuracy (Barnes et al., 1999), but the amount of time spent searching for a particular answer before giving up is also influenced by the FOK associated with the respective question (Gruneberg et al., 1977; Nelson & Narens, 1990). Also, the tendency of rememberers (e.g., persons on the witness stand) to adopt a strict or lax criterion in deciding what to report and at which grain size depends on the relative utility of providing as complete and informative a report as possible versus as accurate a report as possible. At the subordinate level, however, the decision of whether to volunteer a particular piece of information and which grain size to use depends on the subjective confidence associated with it (Goldsmith, Koriat, & Pansky, 2005; Koriat & Goldsmith, 1996).

A second feature is that strategic regulation at the subordinate level is assumed to be guided by online iterative monitoring. Such has been assumed to be the case for JOLs, according to Dunlosky and Hertzog's (1998) discrepancy-reduction model. Barnes et al. (1999) proposed that memory search for a solicited target continues as long as FOK exceeds a certain level. Chen and Chaiken (1999) proposed that information processing and judgments are guided by the motivation to minimize the discrepancy between the actual and desired levels of confidence. Thus, the online regulation of behavior is assumed to be dynamically guided by online metacognitive feelings.

In sum, the dominant view in metacognition is consistent with Story 1 by assigning a critical role to metacognitive feelings in guiding and driving goal-oriented control operations (Brown, 1987). This view is most clearly seen in discussions that focus on the *function* of subjective experience. Such discussions also stress the critical contribution of accurate metacognitive judgments to effective cognitive performance. In contrast, discussions that focus on the *basis* of metacognitive judgments tend to lean toward Story 2, as we shall now show.

Reflections on Story 2 in Metacognition Research

We turn now to discussions that seem to endorse the assumptions underlying Story 2: first, that monitoring follows control, and second, that the accuracy of metacognitive judgments is mediated by their reliance on the feedback from control operations.

The basis of metacognitive feelings. A commonly held assumption is that sheer noetic feelings, such as the feeling of competence, the feeling of knowing associated with the tip-of-the-tongue state, or the subjective confidence in an answer, derive from the application of nonanalytic heuristics (see Jacoby & Brooks, 1984; Kelley & Jacoby, 1996) that operate unconsciously

to shape the subjective experience of knowing (see Koriat, 2000). Such experience can then serve as the basis of metacognitive judgments.

What are the cues for “intuitive” noetic feelings? Reviewing the work in metacognition, Koriat and Levy-Sadot (1999) concluded that these cues “lie in structural aspects of the information-processing system. This system, so to speak, engages in a self-reflective inspection of its own operation and uses the ensuing information as a basis for metacognitive judgments” (p. 496). This proposal incorporates the idea that noetic feelings monitor the feedback from one’s own cognitive processes and performance, and hence follow rather than precede control operations.

This idea resembles the notion of metamonitoring proposed by Carver and Scheier (1990, 1998) to underlie affective subjective experience. According to them, when a person engages in a goal-directed action, in parallel to the monitoring loop that evaluates the discrepancy between the actual state and the desired state (as implied by the CM model), a second, metamonitoring loop takes place that evaluates the *rate* at which this discrepancy is reduced. This rate is assumed to underlie the experience of positive or negative affect. Carver and Scheier’s notion of rate of discrepancy reduction has much in common with the notion of “processing fluency” that has been proposed to underlie noetic feelings (see Benjamin & Bjork, 1996; Kelley & Rhodes, 2002; Koriat & Ma’ayan, 2005). It is difficult to know whether Carver and Scheier’s model can be extended to noetic feelings. However, in line with this model, the position advanced by Koriat and Levy-Sadot (1999) assumes that sheer noetic feelings monitor characteristics of the *process* underlying various cognitive operations rather than their *outcome* (see also Schwarz, 2004; Winkelman et al., 2003).

What is the evidence for this generalization? Several researchers have proposed that JOLs are based on the ease with which to-be-remembered items are processed during encoding (Begg et al., 1989; Koriat, 1997; Matvey et al., 2001). In this article we assumed that study time reflects memorizing effort or memorizing fluency, but other indexes of fluency have also been explored. For example, Hertzog et al. (2003) found JOLs to increase with the success and speed of forming an interactive image between the cue and the target during paired-associate learning. Other researchers have emphasized retrieval fluency, arguing that JOLs are influenced by the ease and probability with which the to-be-remembered items are retrieved during learning (Benjamin & Bjork, 1996; Nelson et al., 1998). Benjamin et al. (1998), for example, observed that the faster it took participants to retrieve an answer, the higher was their estimate that they would be able to recall that answer at a later time. In reality, however, the opposite was the case. Matvey et al. (2001) found that JOLs increased with the fluency with which targets were generated to cues at study. Also, as indicated earlier, the superior accuracy of delayed JOLs over immediate JOLs was explained by assuming that JOLs monitor the ease and success of retrieval during study (Dunlosky & Nelson, 1994). Indeed, the recent findings of Nelson et al. (2004) and of Koriat and Ma’ayan (2005) support the claim that the basis of delayed JOLs lies in the feedback from the covert attempt to retrieve the to-be-remembered target from memory. Thus, these discussions imply that it is by attempting to memorize an item or by trying to retrieve it that learners monitor the likelihood of recalling the item at some later time.

Similarly, with regard to FOK judgments, it has been proposed that these judgments are based on the familiarity of the cue (e.g., question) that is used to probe memory. Indeed, advance priming of the cue that probes memory has been found to enhance FOK judgments (e.g., Reder, 1987, 1988; Schwartz & Metcalfe, 1992). Assuming that priming increases the familiarity of the cue by enhancing its fluent processing (Jacoby, 1991; Jacoby & Kelley, 1987; Jacoby, Woloshyn, & Kelley, 1989), this finding also implies that FOK monitors the feedback from the processing of the cue that prompts recall (Koriat & Levy-Sadot, 2001).

Another cue that was assumed to affect FOK judgments is the accessibility of partial information. As noted in the introduction, Koriat’s accessibility model of FOK (Koriat, 1993) actually assumes control-based monitoring: FOK judgments are based on the feedback from retrieval attempts, particularly the amount of partial information retrieved and its ease of retrieval (see Hicks & Marsh, 2002; Koriat, 1993, 1995; Schwartz & Smith, 1997). Thus, FOK judgments are assumed to follow rather than precede attempted retrieval.

Finally, it has been proposed that subjective confidence is also based in part on the feedback from controlled operations. Thus, explanations of the overconfidence phenomenon (for reviews, see McClelland & Bolger, 1994; Nickerson, 1998) in terms of a confirmation bias incorporate the notion that monitoring is retrospective in nature: When asked to make a decision and indicate their confidence, participants base their confidence judgments on a retrospective review of the arguments that influenced their decision, with a biased tendency to justify the decision reached (Koriat, Lichtenstein, & Fischhoff, 1980; McClelland & Bolger, 1994). Also, as noted earlier (see Experiment 7), the oft reported correlation between decision time and confidence (e.g., Costermans et al., 1992; Kelley & Lindsay, 1993; Nelson & Narens, 1990; Robinson et al., 1997) has been generally interpreted to imply control-based monitoring: Once an answer has been retrieved or selected, the amount of time or effort expended in retrieving or choosing it serves as a cue for the subjective correctness of the answer (Kelley & Lindsay, 1993).

In sum, several discussions and findings in metacognition imply that noetic feelings are retrospective in nature, being based on the feedback from behavior and performance. This view has been most explicitly voiced by Jacoby and Kelley, on the one hand, and by Whittlesea, on the other hand. As already noted, Kelley and Jacoby (1998) explicitly stated that their general theoretical position agrees with the James–Lange view that subjective experience follows rather than precedes performance. Indeed, the extensive research of Jacoby and his associates (see, e.g., Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989) suggests that subjective experience is shaped by one’s unconscious interpretation of one’s own performance. Whittlesea’s (1997, 2003) *selective construction and preservation of experiences* framework of memory also shares the assumption that monitoring follows performance. According to that framework, the interaction between memory and the environment consists of the construction of a mental model, and this construction has a production function and an evaluation function that monitors the integrity of the production. The evaluation function is assumed to result in several primitive perceptions (e.g., coherence, incongruity), and it is the interpretation of these perceptions that gives rise to a specific subjective feeling. Thus, subjective feelings follow production and are based on its quality. In sum, then, the James–Lange view seems to enjoy renewed

interest among students of metacognition who focus on the microgenesis of subjective experience.

The accuracy of experience-based metacognitive judgments. We turn next to the second assumption of Story 2, that the accuracy of metacognitive feelings derives from the diagnostic value of the feedback from one's own control operations. Thus, metacognitive accuracy should vary with the extent to which such feedback predicts actual memory performance (cue validity) and the extent to which that feedback is relied on as a basis for metacognitive judgments (cue utilization).

With regard to cue validity, an important question that suggests itself by our results is, why is the likelihood of recalling an item correlated negatively with the amount of time invested in studying that item? Why was the memory for the English translations of Spanish words inversely correlated with the number of trials needed to master these translations 8 years earlier in Bahrck and Phelps's study (1987)? More generally, why does the feedback from one's own cognitive processes predict future memory performance? We shall not discuss this question here except to note the need for its systematic investigation. However, assuming that the feedback from control processes is indeed diagnostic of memory performance, the accuracy of metacognitive judgments should increase with increased reliance on such feedback. Indeed, in Experiment 4 reliance on memorizing effort as a cue for JOLs increased with practice in parallel to the increase in the validity of memorizing effort in predicting recall. Koriat and Ma'ayan's results (2005) also suggest that with increased delay in soliciting JOLs, a shift in cue utilization occurs from reliance on encoding fluency toward greater reliance on retrieval fluency, and this shift parallels the change that occurs in the relative validity of these two cues with delay. The finding that the confidence–accuracy correlation is stronger for recall than for recognition (Koriat & Goldsmith, 1996; Robinson et al., 1997) also suggests that in recall testing people take advantage of an additional, generally valid cue for the correctness of their answers: the ease with which the answer comes to mind. Thus, metacognitive accuracy depends not only on cue validity but also on the extent to which the cue utilization mechanism is tuned to the relative validities of the various cues available and to the changes in these validities that occur with changes in different conditions.

In conclusion, the two assumptions underlying Story 2 are rarely explicitly endorsed in metacognition research. However, they are in fact implicit in many discussions.

The Bidirectional Links Between Monitoring and Control

We proposed that the MC and CM models are not mutually exclusive, and indeed, some of the discussions in the metacognition literature would seem to endorse that proposal. More generally, recent discussions of the cause-and-effect links between consciousness and behavior have emphasized the importance of developing interactional models that posit mutual causality as an alternative to traditional models that assume unidirectional causality (see Dent, 2003).

We proposed two general modes in which the monitoring and control functions can combine: a sequential and a simultaneous mode. The sequential mode has received some support in previous investigations. In that mode, initial monitoring informs control operations, and the feedback from these operations can serve then as the basis for monitoring, which can then guide subsequent

control operations, and so on. This mode is illustrated by the results of Koriat and Levy-Sadot (2001), which suggest that when one is presented with a question, the familiarity of that question may produce a preliminary positive FOK that can then induce memory search for the answer. FOK is then updated according to the accessibility of clues regarding the answer. Thus, cue familiarity, perhaps resulting from processing fluency, can drive memory search (i.e., monitoring-based control), and the feedback from that search can then affect later FOK judgments (i.e., control-based monitoring). Similarly, Vernon and Usher (2003), who examined the temporal course of metacognitive judgments during retrieval, showed that after the initial influence of cue familiarity, FOK judgments can actually increase or decrease over time depending on the information activated during the search for the target. Thus, preretrieval FOK can drive retrieval attempts, and the feedback from attempted retrieval can then be used to update the initial metacognitive judgments.

A similar two-stage model was advanced by Son and Metcalfe (2005) for JOLs. They proposed that JOLs involve a quick preretrieval stage based on cue familiarity, and the output of that stage may motivate the initiation of a subsequent retrieval stage. Once retrieval has been initiated, JOLs will then be based on the qualities of attempted retrieval.

With regard to the simultaneous mode in which the MC and CM models combine within the same situation, we could not find examples of this mode in previous research on metacognition. This mode is important because many real-life situations involve both goal-driven and data-driven regulation. In such situations top-down and bottom-up processes may affect metacognitive judgments in opposite directions, as was found to be the case in Experiments 5 and 7. Further investigations of the simultaneous operation of the CM and MC models are needed.

Some Issues and Questions for the Future

Admittedly, the conceptual framework that underlies the present work is far from being complete. Furthermore, in sketching that framework, we have deliberately avoided several important issues. In this final section we would like to comment briefly on several of these issues, which we think need to be addressed in future research.

The need for experimental support. We begin by noting a methodological weakness of our work: The major conclusions regarding the cause-and-effect relations between monitoring and control were based primarily on correlational results, which are open to alternative interpretations. Although we did use several experimental manipulations, the main purpose of these manipulations was to show that they modulate the relationship between the variables of interest. Thus, we manipulated incentives in Experiments 5–7 in order to show that variations in incentive produce a positive correlation between changes in study time (or in deliberation time) and changes in metacognitive judgments. The manipulation of practice in Experiment 4 was primarily intended to demonstrate that the study time–JOL correlation changes with repeated study–test cycles. The inclusion of delayed JOLs in Experiment 3 was aimed to test the hypothesis that the dependence of JOLs on study time decreases with JOL delay. The introduction of severe time pressure in Experiment 6 was intended to show that the correlation between study time and JOLs across items changes when severe constraints are placed on the overall study time

available. Clearly, it is not easy to find experimental manipulations that can directly test the cause-and-effect relations postulated in our theoretical framework. Therefore special effort must be made in the future to use additional experimental manipulations that can help produce converging evidence in support of the proposed conceptual framework.

The dynamics of data-driven regulation. We now turn to some of the substantive issues. One is the clarification of the dynamics of data-driven regulation. As far as MC models are concerned, these models are relatively clear about the dynamics of goal-driven regulation: This regulation is assumed to be modulated by metacognitive judgments, generally along the lines of the control-theory perspective. In contrast, the dynamics of data-driven regulation is far from being clear. If JOLs are based on study time rather than vice versa, what determines study time itself? How do learners decide when to stop studying an item?

We are currently exploring the possibility that the decision to continue studying an item or end study is based not on the perceived degree of mastery but on the monitoring of the mental effort expended in studying that item. Assuming, with Kahneman (1973), that the effort invested in a task is determined mainly by the intrinsic demands of the task, perhaps in self-paced study learners monitor the effort expended, stopping studying when no further increase in effort is detected. A somewhat similar proposal was advanced by Metcalfe and Kornell (2003): Learners continue studying an item until information uptake has plateaued so that there is diminished return. Unlike their proposal, however, we assume that it is the change in invested effort that is monitored rather than the change in degree of information uptake.

Feedback from outcome versus feedback from process. As noted earlier, Carver and Scheier (1990, 1998) distinguished between a monitoring function that operates in the service of reducing the discrepancy between actual and desired states and a meta-monitoring function that assesses the rate of discrepancy reduction. This implies a distinction between two types of cues. In a similar manner we have proposed that whereas MC models assign an important role to feedback from the outcome of goal-oriented operations, in the CM model, the feedback of concern is that pertaining to the process itself, for example, the effort needed to master a piece of information or to reach a decision.

A question arises, however, concerning the possible connection between information about the outcome and information about the process. Control-theory models, such as the discrepancy-reduction model, imply an iterative process that is controlled by its (monitored) outcome. The CM model, in contrast, was conceptualized to entail a retrospective monitoring that follows task completion. Clearly, however, as illustrated by the sequential mode, information about the process is also monitored online and may influence strategic decisions. For example, initial FOK, based on the feedback from processing a question, can drive a retrieval attempt, and the feedback from attempted retrieval can then be used to update the FOK (e.g., Vernon & Usher, 2003). Furthermore, even when retrieval fails, the feedback from the process can still be used to evaluate the likelihood that the solicited target will be retrieved given further effort (see Carver & Scheier, 1990). Therefore, a detailed model of the dynamics of monitoring and control requires an analysis of the online feedback from the outcome as well as that from the process, and a specification of their joint consequences.

Automatic and controlled processes. So far we have avoided reference to the commonly held distinction between automatic and

controlled processes, which seems relevant to our conceptual framework. Two questions suggest themselves. The first concerns the relation between this distinction and our distinction between data-driven and goal-driven effects. Researchers would possibly agree that goal-driven regulation should be considered a controlled process. However, would data-driven regulation—for example, the regulation of study time according to the intrinsic properties of items—be seen to represent an automatic process? Would the regulation of study time in accordance with differential incentives tax attentional resources to a greater extent than the regulation according to item difficulty? Such would be expected if goal-oriented regulation is assumed to be more “under the conscious control of the subject” (Posner & Snyder, 1975, p. 73; see Shallice, 1994) than data-driven regulation.

The second question concerns our analysis of the cause-and-effect relation between monitoring and control. In discussing control-based monitoring, we grouped together mnemonic cues that seem to stem from automatic processes (e.g., processing fluency and cue familiarity) with those that derive from more controlled processes (e.g., the amount of effort invested). A question that arises is whether these two types of cues should be distinguished, because it might be argued that only the latter constitute feedback from control operations and hence support the possibility that monitoring may follow control. We opted not to draw such a distinction. Indeed, in discussing the affinity between their view and the James–Lange position, Kelley and Jacoby (1998) used the more encompassing term *performance* in arguing that subjective experience may follow behavior. Also in sketching his position, James (1884) included examples of behaviors that can be said to differ along the continuum of automatic versus controlled actions, for example, “we lose our fortune, are sorry and weep” versus “we are insulted by a rival, are angry and strike” (p. 190).

The attribution process. The results of Experiments 5 and 7 support the simultaneous operation of top-down and bottom-up processes, which affect metacognitive judgments in opposite directions. As noted earlier, an attribution process must be postulated to account for this observation, which implies that the cognitive system can distinguish between the two sources of variation in study time: whether that variation is due to goal-driven or to data-driven effects. What is the process that permits such discrimination?

We might gain some insight into the underlying process from research on the control of action, which suggests a mechanism by which the cognitive system can distinguish between self-generated and externally generated movement (for a review, see Frith, Blake-more, & Wolpert, 2000). That mechanism assumes that when a movement is self-initiated, an efferent copy of the motor command is issued that predicts the outcome and allows compensation for it. Perhaps in a similar manner the cognitive system can distinguish between intention-driven effort and data-driven effort by specifying the proportion of variation that is due to each. The problem with this proposal is that we found experimenter-controlled variations in study time to have effects on JOLs and recall similar to those of self-regulated, goal-driven variation rather than those of data-driven variation (Koriat, 1997; Koriat & Ma’ayan, 2005). Thus, the hypothesized attribution process assumed to underlie the simultaneous effects of intention-driven and data-driven effort presents a challenge for future research.

A final word. The issue of the cause-and-effect relation between subjective experience and behavior continues to be a subject of intense debates among cognitive scientists in different disciplines. Clearly, this issue cannot be settled on the basis of empirical results. What we have offered in this article is, perhaps, not much more than a way to think about the issue. We believe that the experimental work that we presented illustrates the usefulness of the proposed conceptual framework and will hopefully generate further experiments and findings.

References

- Adams, J. K., & Adams, P. A. (1961). Realism of confidence judgments. *Psychological Review*, *68*, 33–45.
- Adelmann, P. K., & Zajonc, R. B. (1989). Facial efferece and the experience of emotion. *Annual Review of Psychology*, *40*, 249–280.
- Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In D. E. Meyer & S. M. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 183–218). Cambridge, MA: MIT Press.
- Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*, 126–131.
- Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.
- Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 344–349.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer Jr. (Ed.), *Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.
- Barnes, A. E., Nelson, T. O., Dunlosky, J., Mazzoni, G., & Narens, L. (1999). 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* (pp. 287–313). 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., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology*, *4*, 195–218.
- Bem, D. J. (1965). An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology*, *1*, 199–218.
- Bem, D. J. (1966). Inducing belief in false confessions. *Journal of Personality and Social Psychology*, *3*, 707–710.
- Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*, 183–200.
- Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55–68.
- Bless, H., & Forgas, J. P. (Eds.). (2000). *The message within: The role of subjective experience in social cognition and behavior*. Philadelphia: Psychology Press.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*, 227–287.
- Breznitz, S., & Ben-Dov, G. (1991). *Norms for word associations in Hebrew*. Unpublished manuscript, University of Haifa, Haifa, Israel.
- Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale, NJ: Erlbaum.
- Brunswick, E. (1956). *Perception and representative design in psychological experiments*. Berkeley: University of California Press.
- Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, *97*, 19–35.
- Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. New York: Cambridge University Press.
- Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*, 1078–1085.
- Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & Cognition*, *22*, 273–280.
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 73–96). New York: Guilford Press.
- Costermans, J., Lories, G., & Ansary, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 142–150.
- Dent, E. B. (2003). The interactional model: An alternative to the direct cause and effect construct for mutually causal organizational phenomena. *Foundations of Science*, *8*, 295–314.
- Duclos, S. E., Laird, J. D., Schneider, E., Sexter, M., Stern, L., & Van Lighten, O. (1989). Emotion-specific effects of facial expressions and postures on emotional experience. *Journal of Personality and Social Psychology*, *57*, 100–108.
- Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition*, *25*, 691–700.
- Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker (Ed.), *Metacognition in educational theory and practice* (pp. 249–275). Mahwah, NJ: Erlbaum.
- Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545–565.
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, *98*, 37–56.
- Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory & Cognition*, *32*, 779–788.
- Flanagan, O. (1992). *Consciousness reconsidered*. Cambridge, MA: MIT Press.
- Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*, *31*, 357–363.
- Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language*, *52*, 505–525.
- Gruneberg, M. M., & Monks, J. (1974). “Feeling of knowing” and cued recall. *Acta Psychologica*, *38*, 257–265.
- Gruneberg, M. M., Monks, J., & Sykes, R. N. (1977). Some methodological problems with feelings of knowing studies. *Acta Psychologica*, *41*, 365–371.
- Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*, 208–216.
- Hart, J. T. (1967). Memory and the memory-monitoring processes. *Journal of Verbal Learning and Verbal Behavior*, *6*, 685–691.
- Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of*

- Experimental Psychology: Learning, Memory, and Cognition*, 29, 22–34.
- Hicks, J. L., & Marsh, R. L. (2002). On predicting the future states of awareness for recognition of unrecalable items. *Memory & Cognition*, 30, 60–66.
- Hyland, M. E. (1988). Motivational control theory: An integrative framework. *Journal of Personality and Social Psychology*, 55, 642–651.
- 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. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 3–22.
- Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 1–47). San Diego, CA: Academic Press.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306–340.
- Jacoby, L. L., Jennings, J. M., & Hay, J. F. (1996). Dissociating automatic and consciously controlled processes: Implications for diagnosis and rehabilitation of memory deficits. In D. J. Herrmann, C. McEvoy, C. Hertzog, P. Hertel, & M. K. Johnson (Eds.), *Basic and applied memory research theory in context* (Vol. 1, pp. 161–193). Mahwah, NJ: Erlbaum.
- 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., Lindsay, D., & Toth, J. P. (1992). Unconscious influences revealed: Attention, awareness, and control. *American Psychologist*, 47, 802–809.
- Jacoby, L. L., Ste-Marie, D., & Toth, J. P. (1993). Redefined automaticity: Unconscious influences, awareness and control. In L. Weiskrantz & A. D. Baddeley (Eds.), *Attention: Selection, awareness and control: A tribute to Donald Broadbent* (pp. 261–282). New York: Clarendon Press.
- Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118, 126–135.
- Jacoby, L. L., Woloshyn, V., & Kelley, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, 118, 115–125.
- James, W. (1884). What is an emotion? *Mind*, 9, 188–205.
- Jost, J. T., Kruglanski, A. W., & Nelson, T. O. (1998). Social metacognition: An expansionist review. *Personality and Social Psychology Review*, 2, 137–154.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kelley, C. M. (1999). Subjective experience as a basis of “objective” judgments: Effects of past experience on judgments of difficulty. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 515–536). Cambridge, MA: MIT Press.
- Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, 35, 157–175.
- Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica*, 98, 127–140.
- 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.
- Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. In B. H. Ross (Ed.), *Psychology of learning and motivation: Advances in theory and research* (Vol. 41, pp. 293–320). New York: Academic Press.
- King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, 93, 329–343.
- Koriat, A. (1983). *The subjective monitoring of mastery during study*. Unpublished report, Israel Academy of Science.
- 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. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, 124, 311–333.
- Koriat, A. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370.
- Koriat, A. (1998). Metamemory: The feeling of knowing and its vagaries. In M. Sabourin, F. I. M. Craik, & M. Robert (Eds.), *Advances in psychological science* (Vol. 2, pp. 461–469). Hove, England: Psychology Press.
- Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, 9, 149–171.
- Koriat, A. (2002). Metacognition research: An interim report. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 261–286). Cambridge, England: Cambridge University Press.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490–517.
- Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one’s own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483–502). New York: Guilford Press.
- Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 34–53.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107–118.
- Koriat, A., & Ma’ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, 52, 478–492.
- Koriat, A., Sheffer, L., & Ma’ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, 131, 147–162.
- Laird, J. D., & Bresler, C. (1992). The process of emotional feeling: A self-perception theory. In M. Clark (Ed.), *Emotion: Review of personality and social psychology* (Vol. 13, pp. 223–234). Newbury Park, CA: Sage.
- Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York: McGraw-Hill.
- Leippe, M. R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. *Law and Human Behavior*, 4, 261–274.
- Le Ny, J. F., Denhiere, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarity in self-paced learning conditions. *Acta Psychologica*, 36, 280–289.
- Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 756–766.
- Mandler, G. (2002). *Consciousness recovered: Psychological functions and origins of conscious thought*. Amsterdam: John Benjamins.
- Marcel, A. J. (1983a). Conscious and unconscious perception: An approach

- to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, *15*, 238–300.
- Marcel, A. J. (1983b). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, *15*, 197–237.
- Marcel, A. J. (1986). Consciousness and processing: Choosing and testing a null hypothesis. *Behavioral and Brain Sciences*, *9*, 40–41.
- Marcel, A. J., & Bisiach, E. (Eds.). (1988). *Consciousness in contemporary science*. Oxford, England: Oxford University Press.
- Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytic basis for JOLs? *Memory & Cognition*, *29*, 222–233.
- 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.
- McClelland, A. G. R., & Bolger, F. (1994). The calibration of subjective probability: Theories and models 1980–94. In G. Wright (Ed.), *Subjective probability* (pp. 453–482). Chichester, England: Wiley.
- Metcalf, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality and Social Psychology Review*, *2*, 100–110.
- Metcalf, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, *131*, 349–363.
- Metcalf, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, *132*, 530–542.
- Metcalf, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 851–864.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston.
- 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). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. *Applied Cognitive Psychology*, *10*, 257–260.
- 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., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, *5*, 207–213.
- Nelson, T. O., Graf, A., Dunlosky, J., Marlatt, A., Walker, D., & Luce, K. (1998). Effect of acute alcohol intoxication on recall and on judgments of learning during the acquisition of new information. In G. Mazzoni (Ed.), *Metacognition and cognitive neuropsychology: Monitoring and control processes* (pp. 161–180). Mahwah, NJ: Erlbaum.
- 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., Leonesio, R. J., Landwehr, R., & Narens, L. (1986). A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 279–287.
- 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. 125–141). San Diego, CA: Academic Press.
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: MIT Press.
- Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods*, *9*, 53–69.
- Nhouyvanisvong, A., & Reder, L. M. (1998). Rapid feeling-of-knowing: A strategy selection mechanism. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 35–52). London: Sage.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.
- Pelegriña, S., Bajo, M. T., & Justicia, F. (1999). Allocation of time in self-paced memory tasks: The role of practice, instructions, and individual differences in optimizing performance. *Learning and Individual Differences*, *11*, 401–429.
- Pelegriña, S., Bajo, M. T., & Justicia, F. (2000). Differential allocation of study time: Incomplete compensation for the difficulty of the materials. *Memory*, *8*, 377–392.
- Perlmutter, L. C., & Monty, R. A. (1977). The importance of perceived control: Fact or fantasy? *American Scientist*, *65*, 759–765.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.
- Raven, J. C., Court, J. H., & Raven, J. (1979). *Manual for Raven's progressive matrices and vocabulary scales*. London: H. K. Lewis.
- Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, *19*, 90–138.
- Reder, L. M. (1988). Strategic control of retrieval strategies. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 227–259). San Diego, CA: Academic Press.
- 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.
- Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology*, *82*, 416–425.
- Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*, 379–399.
- Schacter, D. L. (1989). On the relations between memory and consciousness: Dissociable interactions and conscious experience. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 355–389). Hillsdale, NJ: Erlbaum.
- 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.
- Schwartz, B. L., & Smith, S. M. (1997). The retrieval of related information influences tip-of-the-tongue states. *Journal of Memory and Language*, *36*, 68–86.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*, 332–348.
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the

- availability heuristic. *Journal of Personality and Social Psychology*, 61, 195–202.
- 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.
- Shallice, T. (1994). Multiple levels of control processes. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and non-conscious cognitive processing* (pp. 396–420). Cambridge, MA: MIT Press.
- Sheffer, L. (2003). *The reliability and structure of metacognitive skills and their relationship to cognitive performance*. Unpublished doctoral thesis, University of Haifa, Haifa, Israel.
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 204–221.
- Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage model. *Memory & Cognition*, 33, 1116–1129.
- Son, L. K., & Schwartz, B. L. (2002). The relation between metacognitive monitoring and control. In T. J. Perfect & B. S. Schwartz (Eds.), *Applied metacognition* (pp. 15–38). Cambridge, England: Cambridge University Press.
- Stepper, S., & Strack, F. (1993). Proprioceptive determinants of emotional and nonemotional feelings. *Journal of Personality and Social Psychology*, 64, 211–220.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220–247.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95, 66–73.
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1024–1037.
- Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, 71, 673–679.
- Vernon, D., & Usher, M. (2003). Dynamics of metacognitive judgments: Pre- and postretrieval mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 339–346.
- Whittlesea, B. W. A. (1997). Production, evaluation, and preservation of experiences: Constructive processing in remembering and performance tasks. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 37, pp. 211–264). New York: Academic Press.
- Whittlesea, B. W. A. (2003). On the construction of behavior and subjective experience: The production and evaluation of performance. In C. J. Marsolek & J. S. Bowers (Eds.), *Rethinking implicit memory* (pp. 239–260). London: Oxford University Press.
- Winkielman, P., Schwarz, N., & Belli, R. F. (1998). The role of ease of retrieval and attribution in memory judgments: Judging your memory as worse despite recalling more events. *Psychological Science*, 9, 124–126.
- Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *Psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Mahwah, NJ: Erlbaum.
- Zacks, R. T. (1969). Invariance of total learning time under different conditions of practice. *Journal of Experimental Psychology*, 82, 441–447.
- Zajonc, R. B. (1985, April 5). Emotion and facial efference: A theory reclaimed. *Science*, 228, 15–21.
- 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 August 2, 2005

Revision received August 19, 2005

Accepted August 29, 2005 ■

Online Preview of *JEP: General* Articles

Are you an APA member or affiliate who subscribes to the *JEP: General*? If so, you now have on-line access to the most recently accepted articles before they appear in print. Articles accepted and scheduled for publication are available on the journal's Web site at least 2 months prior to print publication. Access to this feature is available at *no charge* via

<http://www.apa.org/journals/xge.html>

to APA members and affiliates who subscribe to *JEP: General*.