

# Are we aware of our ability to forget? Metacognitive predictions of directed forgetting

Michael C. Friedman · Alan D. Castel

© Psychonomic Society, Inc. 2011

**Abstract** It is often important to accurately predict not only what information we will later remember, but also what information we are likely to forget. The present research examined these abilities in the context of item-method directed forgetting, to determine whether people are aware of their strategic control of remembering and forgetting, as well as what cues are used when making metacognitive judgments. Participants studied words, each of which was followed by a cue to remember (R) or forget (F) the word for an upcoming test, and also made predictions of how likely they would be to later recall each word. When asked to recall all of the words, regardless of cue, both recall and predicted recall were fairly sensitive to the R or F instructions, despite some overconfidence. A similar and stronger pattern was found when words were assigned positive or negative point values as cues to remember or forget. These findings suggest that item-based cues to remember or forget information can be successfully utilized when making metacognitive judgments, and that people are fairly aware of the control they have over both remembering and forgetting information.

**Keywords** Metamemory · Directed forgetting · Memory · Forgetting · Judgments of learning

In many situations, it is important to forget previously learned material in favor of more important or updated

information. For example, imagine if you remembered every different location you have parked your car in a given parking lot, when really you are only concerned with where you parked today (Bjork, 1989). This adaptive quality of forgetting (cf. Anderson, Bjork, & Bjork, 1994; Bjork, Bjork, & Anderson, 1998) allows us not only to stay up to date with current information, but also allows us to forget information that will be irrelevant in the future. Such control over forgetting can be important in a number of contexts. For example, in court a judge may order that jurors forget particular information that has been deemed inadmissible. Such situations show the critical importance of the strategic control of forgetting, but also call into question the degree to which we are aware of the ability to selectively forget certain information. The goal of the present investigation is to examine the extent to which we are aware of our ability to forget specific information when explicitly instructed to do so.

Research comparing predicted and actual memory performance has emerged as an important field of study over the last 40 years (see, e.g., Arbuckle & Cuddy, 1969; Dunlosky & Nelson, 1997; Kelley & Jacoby, 1998), and while people are often capable of accurately predicting memory performance, important inaccuracies exist (e.g., Benjamin, Bjork, & Schwartz, 1998; Rhodes & Castel, 2008). Discrepancies between predicted and actual recall reveal important information regarding the cues that participants use when making judgments of learning (JOLs), in which participants report predictions of how likely they will be to remember information. For example, Rhodes and Castel (2008) observed that participants gave higher JOLs to words presented in a larger font size, relative to a smaller font size, despite actual recall being unrelated to size (see also Rhodes & Castel, 2009). This suggests that participants use certain characteristics of the

---

M. C. Friedman (✉) · A. D. Castel  
Department of Psychology, University of California, Los Angeles,  
Los Angeles, CA 90095, USA  
e-mail: m.friedman@ucla.edu

A. D. Castel  
e-mail: castel@ucla.edu

to-be-remembered information to predict their actual recall performance, such as perceptual information or processing fluency, but often are not aware of other factors that influence remembering.

In order to organize and illustrate the various factors that contribute to metacognitive judgments, Koriat (1997) outlined a cue-utilization approach to JOLs, which states that intrinsic and extrinsic cues can influence JOLs via different mechanisms (see also Dunlosky & Matvey, 2001). *Intrinsic cues* consist of the properties and characteristics of the studied items that are thought to indicate an item's ease or difficulty of learning. *Extrinsic cues* relate to the conditions of learning, such as the operations applied at encoding, serial position information, or the retention interval between study and test. Furthermore, participants can indirectly use both "theory-based" analytic inferences and more "experience-based" nonanalytic heuristics when deriving JOLs, and these two mechanisms can influence mnemonic factors that can then also impact JOLs (Koriat, 1997). This approach has provided a useful framework and also shows that certain variables are not adequately taken into account when making JOLs (Castel, 2008; Rhodes & Castel, 2008, 2009; Kornell, Rhodes, Castel, & Tauber, *in press*; Sungkhasettee, Friedman, & Castel, *in press*; Tauber & Rhodes, 2010). However, it remains unclear precisely how intrinsic and extrinsic cues are used in combination to generate JOLs.

While a great deal of research has examined metacognitive judgments about remembering, very little work has examined how well people are aware of forgetting, or more specifically, the control one has over remembering and forgetting. Rawson, Dunlosky, and McDonald (2002) showed that participants take into account differences in retention intervals when predicting later performance (in which retention interval was manipulated as a within-subjects variable), indicating that participants have some awareness of how forgetting may be related to the passage of time. However, Koriat, Bjork, Sheffer, and Bar (2004) found that participants were not aware of how retention intervals could influence forgetting, using a between-subjects design. When asked to predict how many words they would remember from a list for a test occurring 5 min later, participants gave accurate predictions relative to actual recall, but participants were overconfident when asked to predict their performance on the same test 2 days or a week later. In an attempt to elicit more accurate predictions, participants were asked to think about their predictions in terms of *forgetting* and not remembering, and this enhanced accuracy at predicting delayed recall performance (Koriat et al., 2004, Exp. 7). Finn (2008) also illustrated that framing predictions in terms of forgetting led to more information being selected for restudy than when predictions were framed in terms of remembering. Thus, it appears that thinking about forgetting allows participants to access their theory-based inferences and

give more accurate predictions, while taking into account general principles of forgetting.

To directly determine whether people are aware of the ability to forget, it is important to examine metamemory not only when instructed to remember information, but also when instructed to forget information. In the present study, we investigated whether it was possible for people to accurately predict their own memory performance when forgetting factors needed to be accounted for and, more specifically, when they were told to actually forget information. In order to examine this question, we employed a paradigm in which participants were explicitly instructed to forget specific information.

Bjork, LaBerge, and Legrand (1968) used a paradigm in which memory was modified through the instruction to forget. In the item-method directed forgetting task (see also Woodward & Bjork, 1971), items were presented one at a time and, after each item, a cue either to remember the item (R) or to drop or forget the item (F) was presented. When participants were asked to recall all of the items, regardless of the cues they had been presented with, participants remembered more of the to-be-remembered information, but also forgot a majority of the to-be-forgotten information (which the authors referred to as a *directed forgetting effect*). Subsequent studies have followed this procedure, explicitly instructing participants to forget specific information, usually during a list of words or word pairs (for a review, see MacLeod, 1998) and has yielded many different interpretations regarding how this forgetting occurs (see, e.g., Bjork, 1970; Sahakyan, 2004; Sahakyan & Delaney, 2005; Sahakyan, Delaney, & Kelley, 2004). Tekcan and Akturk (2001) examined the effect of item-based directed forgetting on the magnitude and accuracy of feeling-of-knowing (FOK) judgments at retrieval. They found that the directed forgetting manipulation influenced the magnitude or intensity (i.e., how certain their judgments were), but not the accuracy (i.e., how closely they matched with recall), of FOK judgments, suggesting that people may not be aware of how this manipulation influences the ability to retrieve items. However, to our knowledge no prior study has examined predictions of future memory performance for to-be-forgotten items.

The present study used an item-method directed forgetting paradigm to approach the specific question of whether participants would be able to predict their own forgetting when explicitly told to do so, on an item-by-item basis. Furthermore, it examined whether participants would use the intrinsic quality of the words to guide JOLs, or whether they would successfully incorporate more theory-guided information (i.e., the R or F cue) when making JOLs. The present investigation yields two competing predictions. First, if people feel that they can or will forget information that they have recently studied, then metacognitive judgments regarding

forgetting should be quite accurate. In fact, people might believe that they have such control over memory that if instructed to forget something, they would not recall this information on a later memory test (and thus perhaps even assign a JOL of or close to zero for F items). However, Koriat et al. (2004) have shown that people are not aware of how forgetting is related to retention intervals, suggesting that people are not explicitly aware of the principles of forgetting. This leads to a different prediction: Participants might believe that all studied information is encoded in some way, as if on a tape recorder (e.g., Neisser, 1985), and that it would be difficult to forget, even if told to do so. If this were the case, then people would be likely to overestimate recall of the forget items.

Across three experiments, we examined whether participants would accurately predict memory performance for items they were told to remember or to forget in an item-method directed forgetting paradigm. In addition, to make remembering and forgetting more salient to the participants, and to introduce some motivation to remember or forget certain words, we also extended this to situations in which point values dictated whether each item should be remembered or forgotten. In this case, words paired with positive point values (e.g., table, + 5) should be remembered, while words paired with negative point values (e.g., apple, - 5) should not be remembered or should be forgotten after initial encoding (see Castel, Farb, & Craik, 2007). While Koriat et al. (2004) were able to demonstrate that participants can accurately predict forgetting in some cases, we investigated whether it would be possible for participants to accurately predict their own forgetting, as well as the control that they might have regarding the forgetting of information.

## Experiment 1

In order to investigate whether or not people are aware of their ability to remember specific information, as well as to forget other information, participants made JOLs following each word in a standard item-method directed forgetting task, with explicit remember and forget cues. We selected JOLs as the metacognitive measure to use in this study (relative to the FOK judgments used in Tekcan & Akturk, 2001) since these would allow us to record predictions for all items at encoding, regardless of whether they were correctly recalled at retrieval. Additionally, prior research has reported that participants selectively rehearse to-be-remembered information during the time that other information is cued as to-be-forgotten in item-level directed forgetting paradigms (Sahakyan & Foster, 2009). This finding implies that, in item-level directed forgetting

paradigms, the “forgetting” that takes place occurs during encoding, rather than at retrieval (in contrast with list-level directed forgetting paradigms; see Sahakyan & Foster, 2009). If participants believe that they can forget information they have recently studied, we would expect them to give accurate JOLs for both to-be-remembered and -forgotten items. However, if participants feel that they have very little control over selective rehearsal and the forgetting process, we would expect participants to be overconfident in terms of JOLs for the to-be-forgotten items, but perhaps fairly accurate in terms of JOLs for the to-be-remembered items.

## Method

### *Participants*

A total of 32 undergraduate students from the University of California, Los Angeles, participated and received course credit. All participants were tested individually.

### *Materials and apparatus*

The experiment was conducted with Dell Dimension desktop computers with 19-in. monitors using Microsoft PowerPoint. A total of 24 words were selected from the English Lexicon Project database (Balota et al., 2007). The words selected had a frequency of 86.02 (mean KF word frequency; Kučera & Francis, 1967), were 4–7 letters long, and were presented in Arial 44-point font. The order in which the words were presented was block randomized such that within every block of four words presented, two R items and two F items were shown to the participant (six blocks in total). Words were counterbalanced such that they appeared equally often as F or R items across participants.

### *Procedure*

The procedure used was similar to other item-method directed forgetting experiments using item-level remember/forget cues (see Woodward & Bjork, 1971). Participants were instructed to recall only the “remember” items and not the “forget” items for a free-recall test that occurred after all items had been presented. After receiving the instructions, participants received two practice trials that were similar to the experimental trials, to acquaint themselves with the procedure, and were then presented the list containing the 24 experimental stimuli. Each individual trial was shown in the following sequence: item presentation (the word), cue presentation (R or F), and a prompt for prediction (the JOL).

During a trial’s item presentation, participants studied a word that appeared on the screen for 5 s. Participants were

instructed to study the word for a later memory test. Following the presentation of that word, participants viewed a cue (RRR or FFF) for 2 s. The cue instructed participants whether or not to remember (RRR) or forget (FFF) the item that had just been presented for the later test. Finally, participants were given 5 s to make a prediction of the likelihood that they would remember that item on the subsequent free recall memory test, using a scale from 0% to 100%. Participants were told a prediction (i.e., judgment of learning, JOL) of 0% indicated that they would not remember the item at all, whereas a prediction of 100% indicated that they would definitely recall the item. Participants were encouraged to use the entire scale (i.e., to use intermediate values) when making their predictions. Participants said their predictions aloud, and those predictions were recorded by the experimenter.

After studying and making predictions for all of the words, there was a 30-s distractor task, and then participants had 2 min to verbally recall items from the previous list. Participants were explicitly instructed to recall *all* of the items from the list, regardless of what cues had been associated with the items. The experimenter recorded participants' verbal responses. Following the 2 min of free recall, participants were debriefed.

## Results and discussion

Predicted and actual recall data are presented as percentages in Fig. 1. In general, both recall and JOLs were sensitive to remember or forget cues, although participants' JOLs were overconfident relative to actual performance. These data were analyzed in a 2 (measure: JOL, recall)  $\times$  2 (cue: remember, forget) repeated measures ANOVA. Main effects of measure,  $F(1, 31) = 25.92$ ,  $MSE = .033$ ,  $p < .001$ ,  $\eta_p^2 = .46$ , and cue,  $F(1, 31) = 45.30$ ,  $MSE = .020$ ,  $p < .001$ ,  $\eta_p^2 = .59$ , were present, as well as a marginally significant Measure  $\times$  Cue interaction,  $F(1, 31) =$

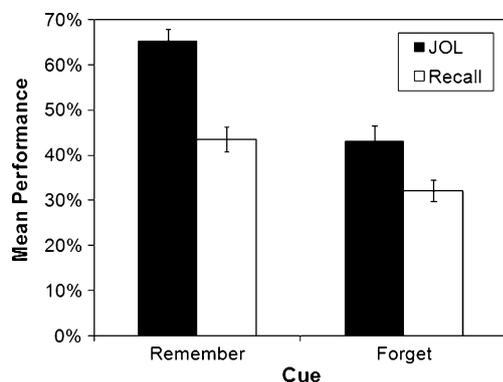
4.07,  $MSE = .022$ ,  $p = .052$ ,  $\eta_p^2 = .12$ . Post-hoc tests revealed that predictions for remember items ( $M = 65.18$ ,  $SE = 2.66$ ) overestimated actual performance ( $M = 43.49$ ,  $SE = 2.74$ ),  $t(31) = 5.04$ ,  $p < .001$ ,  $d = 1.43$ . Likewise, predictions for forget items ( $M = 43.12$ ,  $SE = 3.35$ ) exceeded recall performance ( $M = 32.03$ ,  $SE = 2.34$ ),  $t(31) = 2.77$ ,  $p = .009$ ,  $d = 0.69$ .

We also examined relative accuracy by calculating the gamma correlations for each participant. The mean correlation between JOLs and recall for R items ( $\gamma = .24$ ,  $SE = .08$ ) differed reliably from zero,  $t(31) = 3.27$ . Conversely, the mean correlation between JOLs and recall for F items ( $\gamma = .10$ ,  $SE = .09$ ) did not differ significantly from zero,  $t(31) = 1.16$ . In addition, there was no significant difference between the gamma correlations for R and F items,  $t(31) = 1.07$ ,  $p = .29$ .

These data indicate that participants' JOLs are in fact sensitive to the R and F cues, much like actual recall. In addition, in terms of calibration, JOLs were greater than actual recall for both R and F items, reflecting some overconfidence. However, there was a significant correlation between JOLs and recall of R items, suggesting that participants accurately gave higher JOLs for recalled items and lower JOLs for items that they failed to recall, yet this wasn't the case for F items. Overall, these findings appear to suggest that participants are aware of the control that they have regarding remembering and forgetting information, but some overconfidence was also evident for both the R and F items. In order to examine this issue in more detail, we conducted a follow-up experiment that strongly emphasized the importance of remembering and forgetting items during encoding, to see whether JOLs might be better calibrated when incentives and penalties are incorporated when participants are instructed to remember and forget information.

## Experiment 2

In order to investigate whether the remember/forget cues had the desired effect on participants, and to replicate and extend the main findings from Experiment 1, positive and negative point values were introduced in Experiment 2. Prior research has shown that point values cue people into encoding high-value items over others and inhibiting items with very low or negative values (Castel, Benjamin, Craik, & Watkins, 2002; Castel et al., 2007), and may therefore be a more salient cue than the RRR and FFF cues of Experiment 1. Words were paired with either positive or negative point values, which indicated how important it was to remember (or forget) the word (see Castel et al., 2007, for a similar paradigm). Participants were told to maximize their total score by



**Fig. 1** Mean percentages of predicted recall (JOLs) and of actual recall performance for “remember” and “forget” words from Experiment 1. Error bars represent standard errors of the means in all figures

remembering high-value items, whereas recall of any negative-value items would reduce their score. Thus, the negative values would not only prompt participants to try to forget the target item, but would also have undesirable consequences if they were incorrectly recalled at test. From the participants' perspective during encoding, recall of the negative-value words would be detrimental to performance because doing so would subtract points from their overall "score," thus making negative values more salient forget cues relative to the F cues used in Experiment 1. We hypothesized that the negative-point words might encourage participants to forget those words more than the "forget" cues had in Experiment 1, due to the added negative consequence if they were accidentally recalled, and JOLs might be more sensitive to these values than to the forget or remember cues. Under these conditions, participants might be more accurate at predicting their own remembering and forgetting at the time of encoding.

## Method

### Participants

A total of 36 undergraduates from the University of California, Los Angeles participated and received course credit. All participants were tested individually.

### Materials and apparatus

The words were the same as those in Experiment 1.

### Procedure

The procedure was nearly identical to that of Experiment 1. The experiment used a modified form of the item-level directed forgetting paradigm, but was different in that numerical point values (either + 5 or - 5) followed each word, as opposed to the letter form of the remember or forget cues. Specifically, the cues to remember (RRR) or forget (FFF) were replaced by the values + 5 and - 5 for Experiment 2. Participants were given instructions explaining that the points paired with each word would be awarded to them if they remembered the word during recall. For example, if a participant recalled four + 5 point items, their score would be 20. We also instructed participants not to recall items with a - 5 point value because those would be detrimental to their score, effectively making those to-be-forgotten items. For example, if a participant recalled two + 5 point items, but also two - 5 point items, their net score would be zero. After studying each word and its associated point value, participants made a JOL predic-

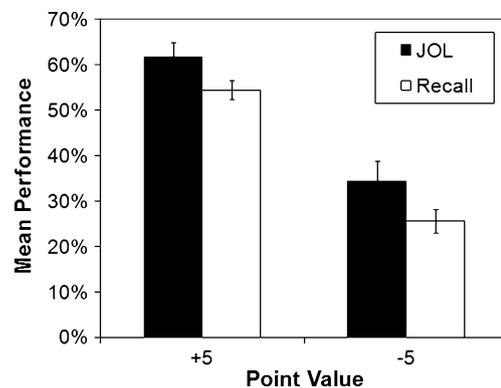
tion on a scale of 0%–100%, as in Experiment 1. Each individual study trial was also shortened in length from Experiment 1, such that each item presentation and the prompt to make a JOL were presented for only 3 s each (as opposed to 5 s in Exp. Exp. 1). However, the point value cue remained the same duration of 2 s. After the distractor task, participants were explicitly instructed to recall *all* of the words from the list, regardless of what value was associated with the item.

## Results and discussion

Predicted and actual recall data are presented as percentages in Fig. 2. In general, it appears that JOLs for the - 5 words overestimated recall, while JOLs also overestimated recall for + 5 words. These data were analyzed in a 2 (measure: JOL, recall) x 2 (Value: + 5, - 5) repeated measures ANOVA. A main effect of measure was found, such that JOLs were greater in value than recall,  $F(1, 31) = 21.50$ ,  $MSE = .031$ ,  $p = .014$ ,  $\eta_p^2 = .18$ . A main effect of value was also found, such that positively valued words were given larger JOLs and were recalled more than negatively valued words,  $F(1,31) = 72.10$ ,  $MSE = .035$ ,  $p < .001$ ,  $\eta_p^2 = .70$ . A Measure x Value interaction was not found,  $F < 1$ .

We examined relative accuracy by calculating the gamma correlations for each participant. The mean correlation between JOLs and recall for + 5 items ( $\gamma = .14$ ,  $SE = .08$ ) was marginally greater than zero,  $t(31) = 1.74$ ,  $p = .09$ , while the mean correlation between JOLs and recall for - 5 items ( $\gamma = .02$ ,  $SE = .11$ ) was not significantly greater than zero,  $t < 1$ . In addition, there was no significant difference between the gammas for R and F items,  $t = 1.32$ ,  $p = .20$ .

The results of Experiment 2 support some of the findings from Experiment 1, and also provide some novel results. Participants appeared to overestimate their recall for both the positive- and negative-value items to similar degrees,



**Fig. 2** Mean percentages of predicted recall (JOLs) and of actual recall performance for the to-be-remembered words (+ 5) and to-be-forgotten words (- 5) from Experiment 2

but the amount of overestimation (especially for R items) was reduced in [Experiment 2](#) relative to [Experiment 1](#). Thus, people do appear to be sensitive to forgetting, because both JOLs and recall were very sensitive to the point-value cues.

### Experiment 3

In [Experiment 2](#), we found that point values made participants' JOLs slightly better calibrated, although some overconfidence was still present for both the positively and negatively valued words. To investigate whether participants would make predictions that illustrated "graded" remembering or forgetting, we manipulated the values associated with each item such that certain items were more important to remember (or forget) than others. That is, we expected participants to give higher JOLs for positively valued items that were more valuable; however, we did not predict such differences in JOLs for high and low negatively valued items, because they should all be forgotten. [Experiment 3](#) again used positive and negative point values similar to those in [Experiment 2](#). However, the points values used in the present experiment had a larger range of values (i.e.,  $-10, -5, +5, +10$ ), unlike the prior experiment, which looked at value on a one-dimensional scale (i.e., the word was either positive or negative, but had the same numerical value of 5). The motivation for the present experiment was to determine whether greater negative values would prompt participants to access their theory-based judgments regarding forgetting, and to give lower JOLs for negatively valued words.

Furthermore, in order to investigate whether people think that they have control over forgetting in a more common, real-world scenario, a posttest questionnaire was included that asked participants about how well they could forget certain information in a courtroom setting. Participants were asked to imagine themselves as a juror in a criminal trial. Participants were told that some of the information they learned during the case was "stricken from the record" and that that information should not impact their decision for a verdict. These data were collected to determine whether the ratings of the ability to forget information in a courtroom setting would be at all related to JOLs for negative items in the directed forgetting experiment.

### Method

#### Participants

A total of 32 undergraduates from the University of California, Los Angeles, participated and received course credit. All participants were tested individually.

#### Materials and apparatus

The experiment used the same materials as [Experiment 2](#); however, additional point value cues were included, such that words could be paired with values of  $+10, +5, -5$ , or  $-10$ . All words and values were counterbalanced across different versions of the list, such that each of the 24 words was paired with all different point values across participants.

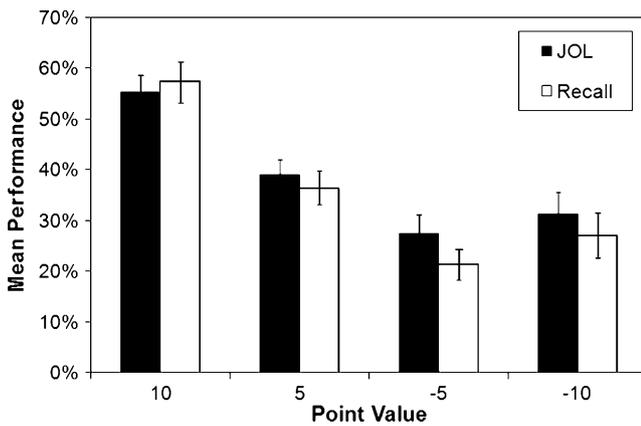
#### Procedure

The procedure was similar to that of [Experiment 2](#). The trial duration and instructions were the same as in [Experiment 2](#), except that participants were made aware of the fact that an item's point value could be  $+10, +5, -5$ , or  $-10$  and that an item's associated value would impact their "score" if recalled at test. At test, participants were explicitly instructed to recall *all* of the words from the list, regardless of what value had been associated with the item.

To examine whether or not to-be-forgotten information could have an impact on a natural, real-world scenario, and whether this would be related to participants' JOLs, following the recall test a postexperiment questionnaire instructed the participant to imagine him- or herself as a juror in a trial in which the accused had been accused of a crime. Participants were told that some of the information that they learned during the case had been "stricken from the record," and that that information should not impact their decision for a verdict. Participants were then asked on a scale of 0%–100% whether the to-be-forgotten information would impact their decision for a verdict in any way. A score of 0% indicated that there would be no influence on their decision (i.e., they could forget this information), whereas a score of 100% indicated that the information would be extremely influential in their decision (i.e., they could not exclude this information). Participants responded verbally and had as much time as they wanted to respond to the question.

### Results and discussion

Predicted and actual recall data are presented as percentages in [Fig. 3](#). Unlike in the prior experiments, participants accurately predicted their recall for the  $+10$  and  $+5$  items, but also accurately predicted their recall for the  $-5$  and  $-10$  items. A 2 (measure: JOL, recall)  $\times$  4 (value:  $+10, +5, -5, -10$ ) repeated measures ANOVA revealed only a main effect of value,  $F(3, 93) = 33.18$ ,  $MSE = .038$ ,  $p < .001$ ,  $\eta_p^2 = .52$ , such that highly positively valued words were predicted and recalled best overall, ( $+10$ :  $M = 56.3$ ,  $SE = 3.1$ ), followed by low positively valued words ( $+5$ :  $M = 37.7$ ,  $SE = 2.4$ ), then high and low negatively valued words ( $-10$ :  $M = 29.2$ ,  $SE =$



**Fig. 3** Mean percentages of predicted recall (JOLs) and of actual recall performance for the to-be-remembered words (+ 10, + 5) and to-be-forgotten words (− 5, − 10) from Experiment 3

3.7; −5:  $M = 24.3$ ,  $SE = 2.5$ ). Both the main effect of measure and the Measure  $\times$  Value interaction were not significant ( $F_s < 1$ ). These findings were also consistent when positive values (i.e., + 10 and + 5) and negative values (i.e., − 5 and − 10) were collapsed. Post-hoc  $t$ tests revealed that across every point value, JOLs did not differ from actual recall ( $t_s < 1$ ).

Examining participants' relative accuracy via gamma correlations revealed that the mean correlations between JOLs and recall for + 10 items ( $\gamma = .25$ ,  $SE = .13$ ) significantly differed from zero,  $t(29) = 1.97$ , while the correlation for +5 items ( $\gamma = .13$ ,  $SE = .11$ ) did not,  $t < 1$ . Additionally, the gamma correlations for − 5 items ( $\gamma = .20$ ,  $SE = .14$ ) and −10 items ( $\gamma = -.07$ ,  $SE = .16$ ) did not differ from zero, either. A converging pattern was found when the gamma correlations for positive ( $\gamma = .27$ ,  $SE = .07$ ) and negative point values ( $\gamma = -.01$ ,  $SE = .09$ ) were collapsed, such that the correlation between JOLs and recall for positively valued items (+ 10 and + 5) reliably differed from zero,  $t(29) = 3.79$ , while for negatively valued items (− 10 and − 5), the correlation did not,  $t < 1$ . Overall, there was a marginally significant difference between the gammas for positively and negatively valued items,  $t(29) = 1.92$ ,  $p = .07$ . This provides some evidence to suggest, in terms of relative accuracy, that participants were better at assigning higher JOLs for recalled positive-value items and lower JOLs for forgotten positive-value items, relative to the negative-value items. This suggests that participants were more accurate for positive-value items in terms of resolution, but this could have also resulted from using a more restricted range of JOLs for negative-value items.

In regard to the postexperiment question, participants gave a mean rating of 51.53 ( $SD = 19.27$ ), indicating that they did not feel certain as to whether they could or could not exclude information that they were told to forget or disregard in a courtroom setting. A correlation analysis

(Pearson's  $r$ ) was conducted between the posttest courtroom ratings and the collapsed JOL and recall scores for positive (5 and 10) and negative (− 5 and − 10) values. JOLs for neither the to-be-remembered items (positive values:  $r = .110$ ,  $p = .594$ ) nor the to-be-forgotten items (negative values:  $r = -.099$ ,  $p = .630$ ) yielded significant correlations with the courtroom ratings. This interesting disconnect between people's beliefs and their actual item-level JOLs has been shown in other contexts (e.g., Dunlosky & Hertzog, 2000; Kornell et al., *in press*), and it may suggest that participants' "global" beliefs are not always "guiding" beliefs and are not well incorporated into specific item-level JOLs.

The results of this experiment, like those of Experiment 2, support the idea that participants can effectively predict their memory for to-be-remembered as well as for to-be-forgotten information. Unlike Experiment 2, however, having differing values seems to have made remember and forget cues more salient, thereby making participants more sensitive in terms of providing accurate JOLs.

## General discussion

We examined whether or not it was possible for people to accurately predict their own remembering and forgetting when explicitly instructed to either remember or forget specific information. We generally found that participants were capable of accurately predicting their own recall performance for words they were told to remember, as well as those they were told to forget (especially in Exp. 3). Overall, both recall and JOLs were influenced by the R and F cues, suggesting that JOLs are sensitive to directed forgetting instructions. Thus, participants incorporate R and F cues when making JOLs, and this is even more the case when values or incentives are used to indicate which items should be remembered or forgotten.<sup>1</sup>

In terms of cue-utilization theory, in the present task, R and F cues could be considered as somewhat intrinsic, because they are closely tied to the manner in which words are processed and have a strong effect on both recall and JOLs. In other work that has examined metacognition and forgetting, retention intervals have been considered extrinsic factors (e.g.,

<sup>1</sup> We also conducted three experiments very similar to the ones reported here, except that participants were told to *only* recall the items that they had been told to remember (the R items) from the list of studied items. For the most part, this led to somewhat similar results, with the key exception being that while participants' JOLs were somewhat accurate for R items, there was overconfidence for F items (since participants recalled fewer F items, relative to the experiments reported here). However, this overconfidence for the F items could have resulted from participants potentially withholding F items from recall at test, leading to this observation of overconfidence for F items.

Koriat et al., 2004) because they are sometimes not well incorporated in participants' memory predictions (in between-subjects designs; but see Rawson et al., 2002). In addition, Tauber and Rhodes (2010) showed that participants do not always accurately incorporate list-length information when making JOLs, despite the fact that list length can strongly influence memory and forgetting. In the present research, JOLs may have been sensitive to forgetting due to the contrast in encoding strategies that participants used when studying R and F items, and this led to awareness of how one could indeed "forget" by engaging in selective rehearsal of R items at the cost of the F items. Thus, the item method of directed forgetting used in the present studies can be conceptualized as a within-subjects design that contained specific cues to both remember and forget certain information. Koriat et al. (2004) found that participants were not aware of forgetting dynamics when retention interval was manipulated in a between-subjects design (but see Rawson et al., 2002, for contrasting results with a within-subjects design). The results from the present series of experiments suggest that participants can indeed incorporate instructions to remember or forget items when making JOLs, indicating that they are aware of the control they have over remembering and forgetting information. This is also in line with other recent work that has shown that both younger and older adults are somewhat accurate when estimating the number of items they have forgotten after recalling previously studied categorized information (Halamish, McGillivray, & Castel, *in press*).

There was a tendency for the correlation between participants' JOLs and recall (or resolution, as measured by gamma correlations) to be slightly lower/poorer for F items, as compared to R items, which may suggest that people can't predict forgetting quite as well as remembering. This may occur because of a relative difficulty in conceptualizing forgetting (as opposed to remembering), when in the presence of the item in question. People usually think about and predict how well they will remember certain information, be it where they left their wallet, how well they will remember specific information for an upcoming test, or the date of their anniversary. However, people generally do not think about nor predict the information that they explicitly need to forget (e.g., inadmissible information in court, previous parking spaces) or are either free to forget or simply lack the need to retain (e.g., irrelevant knowledge for an upcoming test, an old phone number that is no longer used). In addition, participants may use a more restricted range for JOLs when judging F items, or may ignore subtle differences in specific words when applying a more general heuristic regarding F items, which might explain why the relative accuracy was poorer for to-be-forgotten than for to-be-remembered items. Despite these speculative explanations regarding potential

differences in resolution, participants do appear to be sensitive to the factors associated with forgetting, in terms of calibration.

In summary, in all of the experiments reported here, JOLs for to-be-forgotten information were lower than JOLs for to-be-remembered information, indicating a strong awareness of the dynamics of forgetting (i.e., to-be-forgotten items should not be rehearsed in favor of other items, and therefore should be recalled with lower likelihood at test). When point values were used as incentives to remember or forget, JOLs were highly sensitive to both remembering and forgetting. These findings have important implications for theories of both forgetting and metacognition, and possibly for education and clinical practice, given that people often need to accurately monitor the degree to which they can control both remembering and forgetting when learning new information. Future research could determine whether individual differences exist in the ability to monitor forgetting (e.g., Rawson et al., 2002) and whether any individual differences are related to participants' beliefs regarding forgetting and intelligence or to the difficulty associated with initial learning (Miele, Finn, & Molden, 2011). It would also be of interest to extend this research to situations in which forgetting can enhance new learning. For example, in list-method directed forgetting, when participants are instructed to forget previously studied information, there is often a benefit of forgetting previously studied information when learning a second list of new information. It would be important to know whether participants are aware of the potential benefits of forgetting that can enhance new learning, as this could help people become more efficient learners in certain contexts.

**Author note** We thank Bob Bjork, Elizabeth Ligon Bjork, Barbara Knowlton, Nate Kornell, Shannon McGillivray, Matt Rhodes, and Ben Storm and for their useful comments and insights. We also thank Victor Sungkhasettee, Michael Roberts, Greg Hanson, and Frances Javier for help with the data collection. Portions of this research were presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.

## References

- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063–1087.
- Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*, 126–131.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project.

- Behavior Research Methods*, 39, 445–459. doi:10.3758/BF03193014.
- 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. doi:10.1037/0096-3445.127.1.55.
- Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, 9, 255–268.
- Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.
- Bjork, E. L., Bjork, R. A., & Anderson, M. C. (1998). Varieties of goal directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 103–137). Mahwah, NJ: Erlbaum.
- Bjork, R. A., LaBerge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science*, 10, 55–56.
- Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition*, 36, 429–437. doi:10.3758/MC.36.2.429.
- 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. doi:10.3758/BF03194325.
- Castel, A. D., Farb, N. A. S., & Craik, F. I. M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition*, 35, 689–700. doi:10.3758/BF03193307.
- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about strategy effectiveness: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging*, 15, 462–474.
- Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1180–1191.
- Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language*, 36, 34–49.
- Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition*, 36, 813–821. doi:10.3758/MC.36.4.813.
- Halamish, V., McGillivray, S., & Castel, A. D. (in press). Monitoring one's own forgetting in younger and older adults. *Psychology and Aging*. doi:10.1037/a0022852
- Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica*, 98, 127–140.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370.
- Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133, 643–656.
- Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (in press). The ease of processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Mahwah, NJ: Erlbaum.
- Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered? It depends on your beliefs about intelligence. *Psychological Science*, 22, 320–324.
- Neisser, U. (1985). On the trail of the tape-recorder fallacy. *Social Action and the Law*, 11, 35–39.
- Rawson, K. A., Dunlosky, J., & McDonald, S. L. (2002). Influences of metamemory on performance predictions for text. *Quarterly Journal of Experimental Psychology*, 55A, 505–524. doi:10.1080/02724980143000352.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137, 615–625. doi:10.1037/a0013684.
- Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, 16, 550–554. doi:10.3758/PBR.16.3.550.
- Sahakyan, L. (2004). Destructive effects of “forget” instructions. *Psychonomic Bulletin & Review*, 11, 555–559.
- Sahakyan, L., & Delaney, P. F. (2005). Directed forgetting in incidental learning and recognition testing: Support for a two-factor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 789–801. doi:10.1037/0278-7393.31.4.789 (Correction at doi:10.1037/0278-7393.31.5.1164)
- Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review*, 11, 131–136. doi:10.3758/BF03206472.
- Sahakyan, L., & Foster, N. L. (2009). Intentional forgetting of actions: Comparison of list-method and item-method directed forgetting. *Journal of Memory and Language*, 61, 134–152.
- Sungkhassetee, V. W., Friedman, M. C., & Castel, A. D. (in press). *Memory and metamemory for inverted words: Illusions of competency and desirable difficulties*.
- Tauber, S. K., & Rhodes, M. G. (2010). Are judgments of learning (JOLs) sensitive to the amount of material to-be-remembered? *Memory*, 18, 351–362.
- Tekcan, A. I., & Akturk, M. (2001). Are you sure you forgot? Feeling of knowing in directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1487–1490.
- Woodward, A. E., & Bjork, R. A. (1971). Forgetting and remembering in free recall: Intentional and unintentional. *Journal of Experimental Psychology*, 89, 109–116.