Memory & Cognition
2008, 36 (2), 429-437
doi: 10.3758/MC.36.2.429

Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning

Alan D. Castel
University of California, Los Angeles, California

Although memory researchers know about primacy and recency effects, it is unclear whether students are aware of these effects and incorporate them when making judgments of learning (JOLs). The present research examined how participants use serial position information (extrinsic cues) when making JOLs after studying each item and showed that participants rely on the intrinsic qualities of the items and underestimate primacy and recency effects. However, when participants made JOLs *prior* to studying each item and engaged in multiple study–test sessions, their JOLs accurately reflected recall, as well as when serial position information was explicitly provided during the study phase. The findings are interpreted in a cue utilization framework and suggest that under certain conditions, participants can predict primacy and recency effects.

The ability to accurately assess one's own memory performance is a crucial skill that has important applied and theoretical implications. In order to better understand how people monitor memory performance, researchers have examined people's predictions of their future memory performance and how this relates to actual memory performance. This is often done by examining a judgment of learning (JOL) made at study, in which participants provide a rating of how well they think they have learned information, so that they can recall it at a later point in time. People are generally accurate when making these predictions, such that in the context of paired associate learning, participants exhibited strong correlations between a pair's associative strength and both predicted and later recall (e.g., Arbuckle & Cuddy, 1969), whereas Begg, Duft, Lalonde, Melnick, and Sanvito (1989) have demonstrated that memory predictions are typically based on ease of processing and fluency.

Although JOLs are often accurate predictors of later recall, several lines of research have shown that simply relying on encoding or retrieval fluency can lead to discrepancies between predicted and actual memory performance (e.g., Benjamin, Bjork, & Schwartz, 1998; Kelley & Jacoby, 1996; Koriat & Bjork, 2005; Koriat & Ma'ayan, 2005). This may occur because encoding and/or retrieval fluency represents salient, accessible, and compelling information; thus, the fluent processing of information is the salient and compelling experience that is then used to make JOLs, despite the fact that this sometimes leads to inaccuracies relative to later recall (e.g., Benjamin et al., 1998; Castel, McCabe, & Roediger, 2007; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Matvey, Dunlosky, & Guttentag, 2001; Zechmeister & Shaughnessy, 1980). Such discrepancies between predicted and actual performance provide important clues regarding how learning is assessed and, more generally, the insight people have about memory performance. There are situations in which participants are able to generate well-calibrated JOLs, such as when they have experience with both encoding and retrieval conditions. For example, King, Zechmeister, and Shaughnessy (1980) showed that JOLs become better calibrated with paired associated recall after several study–test trials. Dunlosky and Nelson (1994) have shown that JOLs are best calibrated with actual performance when participants can make *delayed* JOLs, which involve providing a JOL when asked to later retrieve an item that was previously studied. This suggests that when participants can incorporate both encoding and later retrieval dynamics to make metacognitive judgments, JOLs are more accurate.

In order to organize and illustrate the various factors and mechanisms that may 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. Intrinsic cues consist of the properties and characteristics of the studied items that are thought to disclose an item's ease or difficulty of learning. Extrinsic cues relate to the conditions of learning, such as the operations applied at encoding, the number of items to be remembered, and serial position information. Furthermore, participants can indirectly use both *theory-based* analytic inferences and more *experience-based* nonanalytic heuristics when de-

A. D. Castel, castel@ucla.edu

riving JOLs, and these two mechanisms can influence mnemonic factors that then give rise to JOLs. This account has provided a useful framework in which to understand how participants learn about the manner in which semantic relatedness can influence recall performance (e.g., Dunlosky & Matvey, 2001; Matvey, Dunlosky, & Schwartz, 2006), how participants are often inaccurate when predicting forgetting functions (Koriat, Bjork, Sheffer, & Bar, 2004), and how experience with several consecutive study–test cycles can lead to improvements in the calibration between recall and JOLs (e.g., King et al., 1980; Koriat & Bjork, 2006).

Understanding the robust effects of primacy and recency can have strong implications for maximizing memory performance, and one relatively neglected area of research in metacognition is the degree to which participants are aware of primacy and recency effects in free recall. Although memory researchers have long known about serial position effects in episodic memory (e.g., Glanzer & Cunitz, 1966; Murdock, 1962; Waugh & Norman, 1965), most notably that events or items at the beginning and end of a series (e.g., a list of words) will be best remembered, it remains unclear whether people (e.g., college students) are aware of these effects and incorporate these factors when making JOLs. Dunlosky and Matvey (2001) showed that JOLs were higher at the beginning of a list, relative to later JOLs, but, in general, JOLs were not very sensitive to serial position, relative to the semantic relatedness of the word pair. However, their investigation used word pairs and cued recall, and this often does not result in strong, observable primacy and recency effects, due to the manner in which items are encoded and later tested (but see Brooks, 1999), relative to studying individual items with an immediate free recall test (see also Matvey et al., 2006).

According to Koriat's (1997) cue utilization approach, list structure (e.g., the number of items in a to-be-remembered list) and the serial position of the items within a list represent extrinsic properties. Extrinsic properties are often not properly incorporated (or are overlooked) when JOLs are made, in comparison with more accessible and compelling intrinsic properties of the items (Koriat, 1997). Given the general reliance on the easily accessible intrinsic properties of items/words when JOLs are made, it is possible that participants might not properly incorporate serial position information when making JOLs, potentially leading to striking differences between perceived and actual recall performance.

The present study examined whether participants properly incorporate serial position information when making JOLs, in order to assess under what conditions primacy and recency effects can be captured by metacognitive judgments. Furthermore, it was of interest to determine whether participants could learn about primacy and recency effects if given multiple study–recall sessions, so that they could then incorporate this knowledge when making JOLs on later lists. It was thought that after several study–test cycles, participants might begin to assign higher JOLs for the first few and last few items in a list, in order to mirror actual recall performance. This would suggest that the use of experience-based learning leads to and informs theory-based concepts, possibly leading to better calibration between JOLs and recall performance. It was also thought that participants would rely primarily on intrinsic cues when making JOLs, and it was of interest to develop conditions in which participants would be more aware of extrinsic cues when making JOLs. One way to promote a shift from intrinsic to extrinsic cues when JOLs are made might be to ask participants to make JOLs prior to seeing or studying the item in question. Thus, in some of the present experiments, a novel approach was also incorporated in which the participants made *pre-JOL* prior to studying each word. In this case, the rationale was that making the JOL in the absence of the word would induce the participants to utilize extrinsic cues when making the JOL (given that no intrinsic information about the item was immediately available)—and thus become aware of serial position effects after several study–test cycles.

EXPERIMENT 1

In order to determine the degree to which JOLs are sensitive to primacy and recency effects in free recall, the participants studied a list of 15 words and, following each word, were asked to provide a JOL. After the final word and JOL rating, they were asked to recall as many words as they could remember. If the participants did not utilize or incorporate serial position information when making JOLs, one would expect that each word would receive a fairly similar JOL (given that the words were chosen in such a way that they were equally memorable in terms of word frequency, etc.). However, if the participants utilized serial position information (i.e., they were aware or became aware of primacy and recency effects), they might assign higher JOLs for the first and last few words in the list, in order to match recall performance. Finally, the participants might become more aware of the influence of serial position information as they studied subsequent lists (given that they engaged in five study–test sessions), so JOLs might best match recall performance for the final list.

Method

Participants. Twenty-four undergraduate students (18–24 years of age) participated and received course credit.

Procedure. The participants were told that they would study a list of words, presented 1 at a time, and that the list contained 15 words. They were told that there were five lists in total, with each list containing different words. The participants were instructed to remember the words and were informed that they would be asked to recall as many words as they could remember after the presentation of the last word on each list. They were also instructed that following the presentation of each word, they would be asked to make a JOL regarding how likely they would be to remember the word on the upcoming immediate recall test. The participants were then asked whether they had any questions and initiated the presentation of the first list by pressing the space bar.

During the study phase, the words were presented one at a time in the center of the computer screen for 2 sec. Following a 1-sec delay, the participants were prompted with the term "Prediction?" indicating that they should provide the JOL rating for the presented word. The participants were told that they should use a scale of 0 to 100, with 0 meaning they *definitely would not remember* and 100 meaning that they *definitely would remember* the word. The participants were instructed to use the entire range from 0 to 100 and gave their responses orally, and the experimenter recorded each response.

Following the study phase in each list, the participants were given 1 min to recall as many items as they could remember. They were told to make their response aloud for the experimenter to record. Following this, the participants were told that they would be presented with the next list and then initiated the presentation of the next list by pressing the space bar. After the recall session of the fifth and final list, the participants were debriefed and asked whether they had learned anything about which words might be most memorable in the list.

Materials and Apparatus. A group of 75 words were selected that were well matched on word length and frequency, and these words were then randomly assigned to five lists consisting of 15 words in each list. The words were selected so that they were of similar frequency and imageability, in order to reduce any possible differences in JOLs and memory based solely on these factors. These 75 words were unrelated items that were four to six letters long, two syllables, and of medium to high frequency (occurrence of at least 30 times per million according to the Thorndike–Lorge count; Thorndike & Lorge, 1944). Across participants, each word was presented in each list, and was then randomly assigned to a serial position for each participant. The stimuli were presented in the center of a computer screen (in white, with a black background) in Times New Roman, with a 32-point font size.

Results and Discussion

The main question of interest was how well JOLs would match actual recall performance—in particular, whether JOLs would be sensitive to serial position—and whether this changed after several study–test cycles. A summary of overall recall performance for the five lists, as well as the JOLs provided on List 1 and on List 5, are presented in Figure 1 (data regarding recall performance and JOLs for each list as a function of each serial position are available from the author). In general, a standard serial position effect was found for free recall (with primacy and recency effects), as was expected. However, JOLs were not highly sensitive to serial position on List 1 (a small downward linear trend, possibly reflecting primacy effects), and by List 5, JOLs appeared to be unrelated to serial position or recall performance.

In order to examine the relationship between JOLs and actual recall performance, a Pearson correlation was calculated on the overall JOL and recall data as shown in Figure 1. For List 1, this value was $r(15) = .19$, $p > .55$, indicating no significant relationship between the two measures. By List 5, the correlation remained positive but was not reliable [$r(15) = .31$, $p > .25$], suggesting that the participants were not systematically assigning JOLs in relation to actual recall across the various serial positions. In order to determine whether the serial position effects for recall and JOLs fit similar functions, trend analyses were conducted on the data, which showed that wheras recall was best fit by a quadratic function [$F(1,23) = 167.31$, $p < .0001$], JOLs (for all lists) were best fit by a linear function [$F(1,23) > 5.88$, $p < .05$].

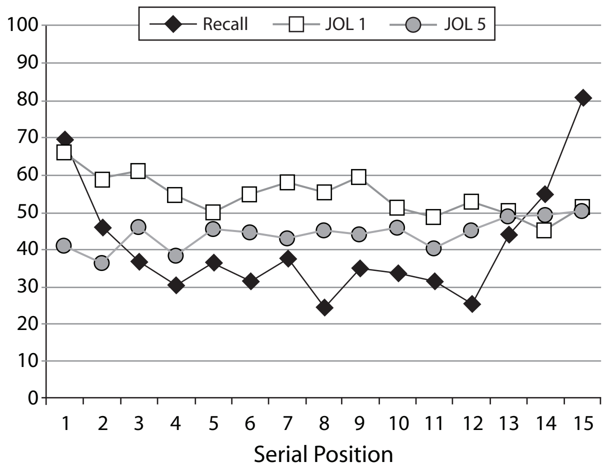

Figure 1. Mean predicted recall (judgments of learning, or JOLs) for List 1 (JOL 1) and for List 5 (JOL 5) and actual overall recall as a function of serial position in Experiment 1.

Given the specific interest in how well JOLs matched actual recall performance at the early and late serial positions, the data from Positions 1–3 were collapsed to reflect primacy effects and those from Positions 13–15 for recency effect and comparisons were made between JOLs and recall for these bins. In terms of the primacy effect in List 1, mean JOLs ($M_{\mathrm{JOL}} = 61.8$) did not significantly differ from actual recall ($M_{\mathrm{recall}} = 60.1$) [$t(23) = 0.011$, $p > .90$], suggesting that the participants accurately predicted performance for the first few words in the list (or perhaps simply begin using the JOL scale at the appropriate location for actual recall). In terms of the recency effect for List 1, mean JOLs ($M_{\mathrm{JOL}} = 48.9$) were significantly lower than actual recall ($M_{\mathrm{recall}} = 65.3$) [$t(23) = 3.31$, $p < .01$], suggesting that the participants underestimated the recency effect.

In order to examine whether participants learn about these effects and whether this is reflected in their JOLs after several study–test cycles, similar analyses were conducted on the data from List 5. In terms of the primacy effect in List 5, mean JOLs ($M_{\mathrm{JOL}} = 41.0$) were significantly lower than actual recall ($M_{\mathrm{recall}} = 54.2$) [$t(23) = 2.21$, $p < .05$]; participants were not quite as accurate as in List 1 for the primacy items. In terms of the recency effect for List 5, mean JOLs ($M_{\mathrm{JOL}} = 49.5$) were significantly lower than actual recall ($M_{\mathrm{recall}} = 63.9$) [$t(23) = 3.32$, $p < .01$], again suggesting that the participants underestimated the recency effect, as in List 1.

Overall, these data suggest that the participants' JOLs were not highly sensitive to serial position information, although in List 1 they were somewhat consistent with the primacy effect (as was also shown by Dunlosky & Matvey, 2001). What is somewhat surprising is that even after multiple study–test cycles, the participants did not adjust their JOLs in order to accurately reflect the recall advantage for the first and last positions in the list. One possibility (given the consistency of the averaged JOLs across serial position) is that the participants used the intrinsic characteristics of each word, as opposed to utilizing information about the list structure (i.e., serial position), such as which words they often recalled first from the list (i.e., typically the last few words that were presented). In order to further investigate conditions in which participants may or may not utilize extrinsic properties, a second experiment was conducted in which participants might be more inclined to utilize extrinsic cues (i.e., serial position information) and to monitor retrieval processes (i.e., which words are typically recalled with respect to serial position) in order to provide JOLs that accurately reflected primacy and recency effects.

EXPERIMENT 2

The results from Experiment 1 suggest that participants do not take into account serial position information when making JOLs. It was of specific interest in the present investigation to develop conditions in which participants would be more likely to consider serial position information and the effects of primacy and recency, so that JOLs would better match recall performance. Koriat (1997) stated that intrinsic properties of to-be-remembered items are often given priority when JOLs are made, relative to extrinsic cues, possibly because of the ease of accessibility of intrinsic cues. In order to prevent participants from being "captured" by the intrinsic properties of the word themselves and to promote the incorporation of extrinsic factors (such as serial position), in Experiment 2 the participants were asked to make their JOL rating *prior* to studying the upcoming word. Thus, this JOL procedure (a prediction, or pre-JOL) is a novel method of assessing metacognitive ability, in which the participants must use extrinsic information, given that they have yet to view the actual stimulus.

In Experiment 2, for each item, the participants were asked to provide their pre-JOL prediction and were then presented the word. It was possible that under these conditions, participants would assign similar JOLs for each word, or just an overall decrease in JOLs as the list progressed. However, after several study–test sessions, the participants might become aware that the pre-JOL allowed them to only use extrinsic properties (i.e., list structure, serial position), as opposed to the item's specific intrinsic properties (or their own idiosyncratic assessment of why the item may or may not be memorable). As a result, the participants might shift to using extrinsic information when making pre-JOLs. This might cause the participants to monitor which words they recalled and where they fell in terms of serial position, so that for the next list they could better assign a JOL, given that they could not use the actual word to make this prediction. Thus, JOLs might best match recall performance under these conditions when the participants were forced to utilize serial position information in the absence of information about the specific item.

Method

Participants. Twenty-four undergraduate students (18–24 years of age) participated and received course credit. None of these participants participated in any of the other experiments.

Procedure. The procedure was somewhat similar to that in Experiment 1, with one important difference. In Experiment 2, the participants were told that *prior* to seeing the upcoming word, they would be asked to make a JOL, or prediction, regarding how likely it was that they would remember the upcoming word on an immediate memory test that would follow the presentation of all 15 words. The participants were then asked whether they had any questions and initiated the presentation of the first list by pressing the space bar.

During the study phase, the words were presented one at a time in the center of the computer screen for 2 sec. Prior to the presentation of each word, the participants were prompted with the term "Prediction?" indicating that they should provide the JOL rating for the upcoming word. The participants were told that they should use a scale of 0 to 100, with 0 meaning that they *definitely would not remember* and 100 meaning that they *definitely would remember* the word, and they made their rating and then pressed the space bar to advance to the word. The participants were instructed to use the entire range from 0 to 100 and gave their responses orally (the experimenter recorded each response). The procedures used for the test of immediate recall, the materials, and the apparatus were identical to those used in Experiment 1.

Results and Discussion

The main results are presented in Figure 2, in terms of overall recall performance for all of the lists and JOLs for List 1 and List 5 as a function of serial position. As was expected, a standard serial position effect was found for free recall and, unlike in Experiment 1, JOLs appeared to be sensitive to serial position, particularly by List 5 for the recency portion of the list. The correlation between JOLs and actual recall performance in List 1 was $r(15) = .59$, $p < .05$, indicating a significant positive relationship between the two measures. For List 5, this correlation was slightly stronger and significant [$r(15) = .78$, $p < .0001$] and is easily observable from Figure 2, suggesting that the participants were now incorporating serial position information when making pre-JOLs. What is most striking is the degree to which JOLs began to match recall performance by List 5 in terms of both primacy and recency effects. Trend analyses shows that although JOLs for List 1 were best fit by a linear function [$F(1,23) = 52.92$, $p < .0001$], both recall and JOLs for List 5 were best fit by a quadratic function [$F(1,23) > 24.79$, $p < .0001$].

In order to examine how JOLs matched actual recall performance at these early and late serial positions, data were again collapsed from Positions 1–3 to reflect primacy effects and from Positions 13–15 for recency effects. In terms of the primacy effect in List 1, mean JOLs ($M_{\mathrm{JOL}} = 73.1$) did not significantly differ from actual recall ($M_{\mathrm{recall}} = 77.8$) [$t(23) = 0.95$, $p > .35$]. Thus, consistent with the results of Experiment 1, the participants accurately predicted

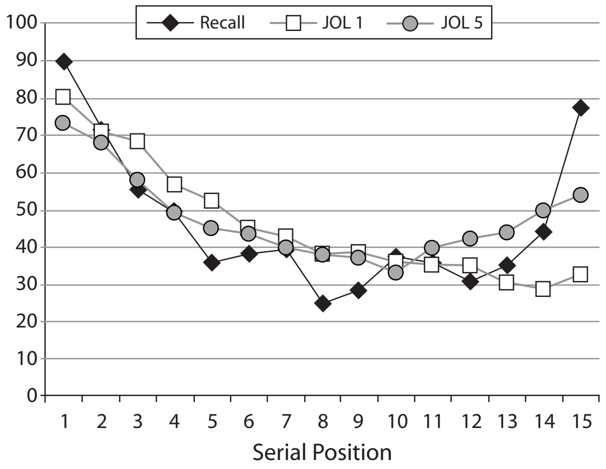

Figure 2. Mean predicted recall (judgments of learning, or JOLs) for List 1 (JOL 1) and for List 5 (JOL 5) and actual overall recall as a function of serial position in Experiment 2.

performance for the first few words in the list (although it is unclear whether they were simply anchoring at an appropriate value for actual recall, as opposed to being aware of primacy effects). In terms of the recency effect for List 1, mean JOLs ($M_{JOL} = 30.7$) were significantly lower than actual recall ($M_{recall} = 48.6$) [$t(23) = 3.04, p < .01$]. These data are consistent with those in Experiment 1 on List 1, again suggesting that the participants underestimated the recency effect.

It was of specific interest to determine whether participants learn to assign higher JOLs to primacy and recency items after several study test cycles, suggesting that they may monitor recall performance and then incorporate this knowledge when making JOLs. For the primacy items in List 5, mean JOLs ($M_{JOL} = 66.5$) were very similar to actual recall ($M_{recall} = 69.4$) [$t(23) = 0.46, p > .65$], and in terms of the recency effect for List 5, mean JOLs ($M_{JOL} = 49.6$) did not significantly differ from actual recall ($M_{recall} = 61.1$) [$t(23) = 1.965, p < .062$; this approached conventional levels of significance]. This trend suggests (as in Experiment 1) that the participants still underestimated the recency effect to some extent, but perhaps not to the same degree as in Experiment 1.

These results suggest that when participants make JOLs prior to studying the actual item, they begin to incorporate serial position information into their metacognitive judgments. It may also allow for more rehearsal of each item, which can enhance recall but also make participants more aware of how well they will recall the items. This might be because, without any other specific information to use, participants attend to recall output and then notice that primacy and recency items typically are recalled more frequently than items from the middle of the list. Thus, under these conditions, participants begin to incorporate extrinsic cues when making JOLs, leading to a stronger relationship between JOLs and actual recall.

EXPERIMENT 3

One potential reason the participants have had difficulty assigning accurate JOLs for the final few items in the list may have been that they were not aware of either the precise serial position of the item or the point at which the list would end. In fact, several participants reported that they would begin to give higher JOLs when they felt the list was about to end but that, often, they were unaware of exactly where they were in the list. To better understand whether specific serial position information can be incorporated when JOLs are made, in Experiment 3 serial position information was given to the participants during study. Thus, the participants were presented with serial position information about the upcoming item (e.g., 1), reported their pre-JOL (e.g., Prediction? e.g., 75), and were then presented with the word (e.g., *table*). It was thought that under these conditions, the participants' JOLs would best match recall, especially for the last list, since the participants could develop and incorporate theory-based inferences about how to assign JOLs to the words. Thus, this presented a situation in which the participants might best utilize the provided serial position information and were not captured by the intrinsic qualities of the word (since they needed to make their rating prior to seeing the word). It was expected that JOLs would best match recall performance after several lists—an important result, given that the JOL was made prior to the word's being studied and was based more on the extrinsic properties of the list.

Method

Participants. Twenty-four undergraduate students (18–24 years of age) participated and received course credit. None of these participants participated in any of the other experiments.

Procedure, Materials, and Apparatus. As in the previous experiments, the participants were told that they would study a list of words, presented one at a time, and that the list contained 15 words. They were told that there were five lists in total, with each list containing different words. They were instructed to remember the words and were told that they would be asked to recall as many words as they could remember after the presentation of the last word on each list. Much as in Experiment 2, they were told that prior to seeing the upcoming word, they would be asked to make a JOL, or prediction, regarding how likely it was that they would remember the upcoming word on an immediate memory test that would follow the presentation of all 15 words. However, unlike in Experiment 2, the participants were also given the serial position of the upcoming word, coupled with the prompt to make the prediction, in the following manner, for example, "1. Prediction?" "2. Prediction?" . . . "15. Prediction?" Thus, the serial position information and the JOL prompt were presented simultaneously, and the participants made their response prior to seeing the upcoming word. The participants were then asked whether they had any questions and initiated the presentation of the first list by pressing the space bar. All other parts of the procedure, materials, and apparatus were similar to those in Experiment 2.

Results and Discussion

The main results are presented in Figure 3, in terms of overall recall performance and JOLs for List 1 and List 5. A standard serial position effect was found for free recall, and consistent with the results of Experiment 2, JOLs appear to have been somewhat sensitive to serial position. This was apparent in List 1 and especially in List 5, where JOLs and recall appear to have been in synchrony. The

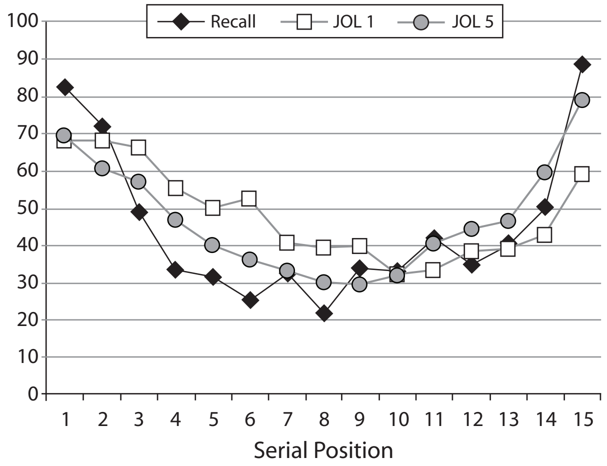

Figure 3. Mean predicted recall (judgments of learning, or JOLs) for List 1 (JOL 1) and for List 5 (JOL 5) and actual overall recall as a function of serial position in Experiment 3.

correlation between JOLs and actual recall performance for List 1 was significant [$r(15) = .69, p < .01$], as was the correlation in List 5 [$r(15) = .81, p < .001$], as is evident from Figure 3. Trend analyses showed that the recall and JOL data for Lists 1 and 5 were best fit by quadric functions [$F\text{s}(1,23) > 35.14, p < .0001$]. Thus, when participants make pre-JOLs and are given serial position information, they are very well calibrated in terms of assigning appropriate and accurate JOLs to items at almost all levels of serial position, taking into account the primacy and recency effects.

To examine how well JOLs matched actual recall performance at the early and late serial positions, data were again collapsed into bins to reflect primacy and recency positions. In terms of the primacy effect in List 1, mean JOLs ($M_{\text{JOL}} = 68.6$) did not significantly differ from actual recall ($M_{\text{recall}} = 72.2$) [$t(23) = 0.56, p > .55$], suggesting an accurate match between JOLs and recall for the primacy items, as was observed in the previous experiment. However, unlike in Experiments 1 and 2, in terms of the recency effect for List 1, mean JOLs ($M_{\text{JOL}} = 46.8$) did not significantly differ from actual recall ($M_{\text{recall}} = 52.8$) [$t(23) = 0.88, p > .35$]. Thus, when the participants were given serial position information and had to make a prediction about the items, accurate JOLs could be produced. This may have occurred because serial position was made accessible and salient (and the participants could not rely on the intrinsic properties of the words). By List 5, this trend was even more apparent: For the primacy items in List 5, mean JOLs ($M_{\text{JOL}} = 62.6$) were again very similar to actual recall ($M_{\text{recall}} = 60.3$) [$t(23) = 0.29, p < .75$], and in terms of the recency effect for List 5, mean JOLs ($M_{\text{JOL}} = 62.2$) did not significantly differ from actual recall ($M_{\text{recall}} = 65.1$) [$t(23) = 0.37, p > .70$].

These results suggest that when participants make JOLs prior to studying the actual item, and are informed about the serial position of the upcoming item, they can successfully use this information to generate a highly accurate JOL. As in Experiment 2, this is likely due to participants' focusing on extrinsic properties, and in the present study serial position was made especially salient. Thus, when participants make predictions about their learning of upcoming items and are provided with serial position information, this leads to a strong relationship between perceived and actual recall performance. One remaining issue is the degree to which participants will utilize serial position information once they have already studied the word (i.e., under the more standard *post-JOL* condition). This issue was examined in Experiment 4, in which conditions were created in which both intrinsic and extrinsic information were readily available and accessible when JOLs were made.

EXPERIMENT 4

Given that participants can incorporate serial position information when making pre-JOLs (presumably, by being forced to rely solely on extrinsic factors) and this leads to fairly good calibration between perceived and actual recall performance, it was of interest to determine whether participants would incorporate serial position information when making more standard post-JOLs. Experiment 1 suggests that participants will not spontaneously use serial position information when making JOLs, possibly due to a reliance on intrinsic properties of the word. However, if presented with the serial position prior to the presentation of the word, participants may be more inclined to utilize this information when making a JOL. Thus, the participants in Experiment 4 were first presented with the serial position of the word (e.g., 1) and then with the word itself (e.g., *table*) and were then asked to make a JOL (Prediction?). Under these circumstances, participants can incorporate both extrinsic properties of the list, such as serial position information, and intrinsic properties of the word, and it was thought that this might eventually (after several study–test cycles) lead to good calibration between JOLs and recall.

Method

Participants. Twenty-four undergraduate students (18–24 years of age) participated and received course credit. None of these participants participated in any of the other experiments.

Procedure, Materials, and Apparatus. As in the previous experiments, the participants were told that they would study a list of words, presented one at a time, and that the list contained 15 words. They were told that there were five lists in total, with each list containing different words. They were instructed to remember the words and were told that they would be asked to recall as many words as they could remember after the presentation of the last word on each list. However, unlike in the previous experiments, the participants were first presented with the serial position information for 2 sec and then with the word for 2 sec and were then asked to make a JOL rating. Thus, the serial position information was presented first, then the word, and then the JOL prompt. All other parts of the procedure, materials, and apparatus were similar to those in Experiments 1, 2, and 3.

Results and Discussion

The main results are presented in Figure 4, in terms of overall recall performance across all lists and JOLs for List 1 and List 5. As in all of the previous experiments, a standard serial position effect was found for free recall

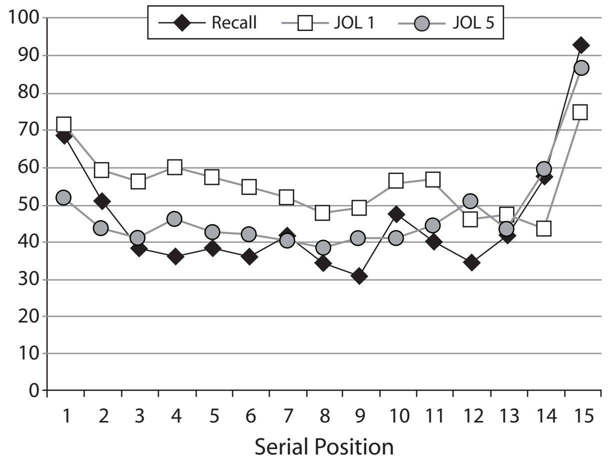

Figure 4. Mean predicted recall (judgments of learning or JOLs) for List 1 (JOL 1) and for List 5 (JOL 5) and actual overall recall as a function of serial position in Experiment 4.

and, as was observed in Experiments 2 and 3, JOLs appear to have been somewhat sensitive to serial position. In the present experiment, this was particularly true for the final word that was presented. Again, the correlation between JOLs and actual recall performance for List 1 was significant [$r(15) = .78$, $p < .0001$], as was the correlations in List 5 [$r(15) = .79$, $p < .0001$], and this is illustrated in Figure 4. Trend analyses showed that the recall and JOL data for Lists 1 and 5 were best fit by quadric functions [$Fs(1,23) > 11.82$, $p < .002$]. In this experiment, both intrinsic and extrinsic factors were likely incorporated into the JOL, and this resulted in extremely accurate predictions of later recall, despite the participants' having made the JOL after observing the word. Thus, in contrast to the findings from Experiment 1, when serial position information is presented and made salient, participants can use this information when making standard JOLs, taking into account the primacy and recency effects that are evident in free recall.

To examine how well JOLs matched actual recall performance at the early and late serial positions, the data were collapsed into the previously described bins to reflect primacy and recency positions. In terms of the primacy effect in List 1, mean JOLs ($M_{\text{JOL}} = 64.7$) did not significantly differ from actual recall ($M_{\text{recall}} = 54.2$) [$t(23) = 1.48$, $p > .15$], suggesting a somewhat accurate match between JOLs and recall for the primacy items, as was observed in the previous experiments (although a trend may exist to suggest that JOLs differed from recall). Also, in terms of the recency effect for List 1, mean JOLs ($M_{\text{JOL}} = 57.2$) did not significantly differ from actual recall ($M_{\text{recall}} = 56.9$) [$t(23) = 0.051$, $p > .95$]; in fact, they were strikingly similar, suggesting that when the participants were given serial position information prior to making a JOL, accurate JOLs could be produced, possibly because serial position was made accessible and salient. This was especially evident in List 5, in terms of the primacy effect in which JOLs ($M_{\text{JOL}} = 48.2$) were again very similar to actual recall ($M_{\text{recall}} = 52.8$) [$t(23) = 0.67$, $p > .50$]. The same pattern was apparent for the recency effect for List 5, since mean JOLs ($M_{\text{JOL}} = 65.8$) did not significantly differ from actual recall ($M_{\text{recall}} = 65.3$) [$t(23) = 0.087$, $p > .90$]. In fact, JOLs and recall were strikingly similar for the last serial position, since the participants were likely aware that they could simply immediately recall this item (from experience with previous lists) and, accordingly, assigned it an extremely high JOL in the final list.

The participants in Experiment 4 made standard post-JOLs, but serial position was highlighted by presenting this information prior to their studying the item and making the JOL. This created conditions in which intrinsic and extrinsic information were made available, and the participants seem to have capitalized on the enhanced availability of serial position information and were likely made more aware of the role that serial position can play in memory performance. Thus, participants can utilize extrinsic variables when making JOLs, but the availability and accessibly of these extrinsic cues needs to be made salient when information is studied, and also reinforced by monitoring the output of later recall.

GENERAL DISCUSSION

The goal of the present study was to examine the degree to which participants use intrinsic and extrinsic cues when making JOLs, in order to predict primacy and recency effects. Experiment 1 demonstrated that the participants did not utilize extrinsic cues, and JOLs were not well calibrated with performance in terms of primacy and recency effects. This tendency to discount extrinsic cues when metacognitive judgments are made has been found in other areas (e.g., Dunlosky & Nelson, 1994; Koriat et al., 2004; Zechmeister & Shaughnessy, 1980). In Experiments 2–4, the participants were more inclined to use extrinsic cues, possibly due to the lack of other information on which to base the JOLs (in Experiment 2) and when the serial position information was made accessible and salient prior to the word's being studied (Experiment 3). Finally, when both extrinsic and intrinsic information were available (Experiment 4), the participants were successful at incorporating this information to make accurate JOLs in terms of the primacy and recency effects.

The need to incorporate retrieval processes (and in the present experiments, the contents and serial output of recall) when JOLs are made has been a critical issue in metacognitive research. Koriat and Bjork (2006) found that with multiple study test cycles, participants became better at predicting differences in performance, as they had more experience with the retrieval conditions. This is also in line with the memory-for-past-test heuristic (Finn & Metcalfe, 2007), in which participants recall and make use of knowledge about previous tests when making JOLs in later lists. In the present experiments, the participants gained useful experience with retrieval operations and also learned about which items were more likely to be recalled on the basis of the serial position. This likely occurred with experience, in terms of noticing that items at the beginning and end of the list figured prominently in the contents of recall output. The participants then incorporated this knowledge (i.e., a theory-based heuristic) when making JOLs on subsequent lists, leading to better calibration between JOLs and recall for primacy and recency items.

Although the precise mechanism(s) that gives rise to more accurate JOLs by List 5 are debatable, it seems likely that participants begin to monitor the contents of recall in terms of serial position (cf. Castel & Logan, 2007; Finn & Metcalfe, 2007), and this is especially so when they are making pre-JOLs (given the lack of intrinsic or word-specific information to guide these judgments). However, Castel and Logan found that JOLs were not sensitive to the spacing effect with multiple lists, even when participants scored their own recall performance. In the present study, it seems plausible that the participants did not specifically remember the exact serial position of the items that were recalled but knew that they were recalling items that were recently presented first (as has been shown in research on clustering analyses for primacy items in the free recall task; e.g., Glanzer & Cunitz, 1966; Welch & Burnett, 1924). Although the present study did not specifically address whether participants are aware that they are recalling recency items first, they may monitor retrieval output and then successfully apply this knowledge on later lists when making JOLs or when trying to strategically maximize recall performance (e.g., Castel, 2007; Castel, Benjamin, Craik, & Watkins, 2002; Castel, Farb, & Craik, 2007; Dunlosky & Hertzog, 2000).

Although Koriat's (1997) framework regarding intrinsic and extrinsic cues can be used to explain the findings, it is also conceivable that in Experiments 3 and 4, when serial position information was explicitly provided, this information then became an intrinsic cue. This would blur the distinction between when cues are intrinsic and extrinsic in nature and how mnemonic cues are used in the present task, and this is a challenge for theories of metacognition. The prejudgment recall and monitoring model (Nelson, Narens, & Dunlosky, 2004) also suggests that subjects use retrieval principles when making JOLs, and this would account for the more accurate JOLs on later lists in the present set of experiments. In general, the incorporation of serial position information on later lists is impressive (given that it actually relies on memory for past performance) and is useful in terms of the generation of more accurate JOLs, providing an illustration of how participants can use experience with retrieval to inform JOLs for new items on a subsequent list.

Learning about and incorporating extrinsic properties of to-be-remembered material can greatly enhance metacognitive awareness, but it should be noted that in the present study, the results were largely correlational in nature. Other studies have manipulated level of difficulty or item characteristics and have shown that participants can learn about intrinsic properties of word pairs (via experience with multiple study–test cycles; e.g., Koriat & Bjork, 2006) and that this change can occur if participants are asked to make predictions or "postdictions" about words with different word frequency. Matvey, Dunlosky, and Schwartz (2006) observed that participants were somewhat sensitive to serial position effects for lists that contained sets of semantically related items, in that an increase in JOL magnitude across the words of a related set was found, suggesting that the participants could use serial position information in some way. Guttentag and Carroll (1998) found that participants were better at making "postdictions" in a recognition memory test regarding the memorability of the items on a subsequent test, and Benjamin (2003) found that participants then used this knowledge when making future predictions at encoding on subsequent memory tests. Finally, this may also be related to the findings that JOLs are more accurate when they are made after an initial delay between study and the final test (Nelson & Dunlosky, 1991), possibly because, under these conditions, participants also consider retrieval factors when making JOLs (see also Kimball & Metcalfe, 2003). Thus, participants will incorporate results from memory tests (e.g., Finn & Metcalfe, 2007) when making future JOLs, and in the present study this was extended to extrinsic cues, leading to successful use of knowledge about primacy and recency effects.

The awareness of list structure and primacy and recency effects can play an important role in metacognitive judgments, but also in terms of supplementing actual memory performance. For example, Watkins and Watkins (1974) and Crowder (1969) showed that if participants are told the length of a to-be-remembered list of words (relative to conditions in which the lists were of variable unknown lengths), this generally leads to marginally better immediate free recall of the items (by enhancing either the primacy or the recency effect). In a related vein, deWinstanley and Bjork (2004) found that participants became aware of the effects of generating answers with experience with reading and generating information for a later test. Thus, the present findings—and, in general, making people aware of primacy and recency effects—can have pedagogical implications if applied in an appropriate manner that emphasizes the subjective experience of learning (e.g., Ghodsian, Bjork, & Benjamin, 1997). For example, potentially, students could learn to direct study time to items in relation to JOLs (e.g., Dunlosky & Thiede, 1998; Kornell & Metcalfe, 2006; Nelson, 1993; Nelson, Dunlosky, Graf, & Narens, 1994), and in the present context, it is possible that with the awareness of primacy and recency effects, participants could learn to practice information that does not benefit from primacy and recency effects (i.e., items in the middle of a list or, perhaps, chapters in the middle of a text book).

The present findings highlight the degree to which metacognitive awareness can inform theories regarding human memory. For instance, the primacy and recency effects may have been first discovered quite incidentally, via simple awareness of what information is typically recalled when a series of similar items is studied. Stigler (1978; see also Roediger, Gallo, & Dolan, 2001) reported that the physicist Francis Nipher (in the year 1876) may have been the first to document the serial position effect when he noted that he was better able to recall the first few and last few graphs and figures that were presented in a physics paper (and he later tested this notion). Thus, initial observations of the serial position effect might have resulted from metacognitive judgments about the recallability of items in a series. Noticing what information is well remembered and which areas require more effort can lead to important selection of methods and strategies for students who are studying for an upcoming test, as for well as researchers who are trying to understand the function and properties of human memory (see Castel, 2007).

In summary, the present research shows that although participants typically do not predict primacy and recency effects via JOL ratings, under certain conditions participants can successfully synchronize JOLs with recall in terms of the primacy and recency effects. The present findings can be interpreted in the context of a cue utilization approach (Koriat, 1997), in which extrinsic and intrinsic cues can give rise to JOLs. The findings that participants will use serial position information only under certain conditions suggests that intrinsic cues will typically be given priority (e.g., Koriat et al., 2004) and that serial position information needs to be made salient both at encoding and at retrieval (via monitoring the content of recall), and participants need to be aware of how recall is influenced by primacy and recency effects via experience with multiple study–test cycles. The understanding of how experienced-based and theory-based learning can influence JOLs can be of help in terms of a better awareness of robust memory phenomena, as well as critical observations regarding what we know about our own memory.

AUTHOR NOTE

Many people provided valuable suggestions and insight regarding this research at various points; specifically, I thank David McCabe, Matthew Rhodes, Nate Kornell, Bob Bjork, Elizabeth Bjork, Aaron Benjamin, Lisa Son, Asher Koriat, and John Dunlosky for helpful comments and discussions and Andrew Addessi, Rachel Gartner, and Sarah Moynan for help with data collection. Please address all correspondence to A. D. Castel, Department of Psychology, University of California, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563 (e-mail: castel@ucla.edu).

REFERENCES

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

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

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

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

Brooks, B. M. (1999). Primacy and recency in primed free association and associative cued recall. *Psychonomic Bulletin & Review*, **6**, 479-485.

Castel, A. D. (2007). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 48, pp. 225-270). San Diego: Academic 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.

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.

Castel, A. D., & Logan, J. M. (2007, November). *Learning about the spacing effect: The role of experience and feedback*. Poster presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA.

Castel, A. D., McCabe, D. P., & Roediger, H. L., III (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, **14**, 107-111.

Crowder, R. G. (1969). Behavioral strategies in immediate memory. *Journal of Verbal Learning & Verbal Behavior*, **8**, 524-528.

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, **32**, 945-955.

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology & 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, & Cognition*, **27**, 1180-1191.

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

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect people's self-paced study and yield "labor-and-gain" effects. *Acta Psychologica*, **98**, 37-56.

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 238-244.

Ghodsian, D., Bjork, R. A., & Benjamin, A. S. (1997). Evaluating training during training: Obstacles and opportunities. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 63-88). Washington, DC: American Psychological Association.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior*, **5**, 351-360.

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

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

Kelley, C. M., & Jacoby, L. L. (1996). Adult geocentricism: Subjective experience versus analytical bases for judgment. *Journal of Memory & Language*, **35**, 157-175.

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

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

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

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition*, **34**, 959-972.

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

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

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 609-622.

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.

Matvey, G., Dunlosky, J., & Schwartz, B. (2006). The effects of categorical relatedness on judgments of learning (JOLs). *Memory*, **14**, 253-261.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, **64**, 482-488.

Nelson, T. O. (1993). Judgments of learning and the allocation of study time. *Journal of Experimental Psychology: General*, **122**, 269-273.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning 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., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods*, **9**, 53-69.

Roediger, H. L., III, Gallo, D. A., & Dolan, P. O. (2001, November). *Francis E. Nipher: The first memory researcher*. Paper presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL.

Stigler, S. M. (1978). Some forgotten work on memory. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 1-4.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College Press.

Watkins, M. J., & Watkins, O. C. (1974). Processing of recency items for free recall. *Journal of Experimental Psychology*, **102**, 488-493.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89-104.

Welch, G. B., & Burnett, C. T. (1924). Is primacy a factor in association formation? *American Journal of Psychology*, **35**, 396-401.

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.

(Manuscript received January 4, 2007;
revision accepted for publication August 10, 2007.)