

Are awareness questionnaires valid? Investigating the use of posttest questionnaires for assessing awareness in implicit memory tests

TERRENCE M. BARNHARDT AND LISA GERACI
Texas A&M University, College Station, Texas

Two experiments—one employing a perceptual implicit memory test and the other a conceptual implicit memory test—investigated the validity of posttest questionnaires for determining the incidence of awareness in implicit memory tests. In both experiments, a condition in which none of the studied words could be used as test responses (i.e., the none-studied condition) was compared with a standard implicit test condition. Results showed that reports of awareness on the posttest questionnaire were much less frequent in the none-studied condition than in the standard condition. This was especially true after deep processing at study. In both experiments, 83% of the participants in the none-studied condition stated they were unaware even though there were strong demands for claiming awareness. Although there was a small bias in the questionnaire (i.e., 17% of the participants in the none-studied condition stated they were aware), overall, there was strong support for the validity of awareness questionnaires.

The distinction between explicit and implicit memory (Graf & Schacter, 1985) has intrigued researchers for over 3 decades (e.g., Warrington & Weiskrantz, 1970, 1974). In explicit memory tests, participants are asked to retrieve information to which they had previously been exposed. For example, in stem-cued recall tests, participants are given the first three letters of words (i.e., stems) and are instructed to respond with studied words that begin with those letters. In contrast, in implicit memory tests, participants are given a task on which performance can be enhanced by prior exposure to information, but they are not asked to retrieve the prior information. For example, in stem completion tests, the stimuli are the same as those in the stem-cued recall test, but participants are instructed to respond with the first word that comes to mind. Usually, instructions for stem completion tests simply do not mention the prior portions of the experiment. In such tests, participants display priming in that they will more often respond with a word if they have studied it than if they have not. A number of researchers have argued that the distinction between explicit and implicit memory tests can be used to reveal differences between different types of memory retrieval mechanisms, variously characterized as intentional versus unintentional (e.g., Schacter, 1987), voluntary versus involuntary (e.g., Richardson-Klavehn, Gardiner, & Java, 1994), controlled versus automatic (e.g., Jacoby, 1991), or conscious versus nonconscious (Tulving & Schacter, 1990).

However, it should be noted that there is nothing to prevent participants with intact memory, while they are

completing a typical implicit memory test, from becoming aware that they are using information to which they have been previously exposed. For example, imagine that the stem *fla* was presented in a stem completion test and the first word that came to mind was a word that had been presented earlier, such as *flash*. Participants may become aware that they have produced a studied word only *after* they have produced the word. This has been referred to as involuntary explicit memory (e.g., Schacter, 1987; Schacter, Bowers, & Booker, 1989) or involuntary aware memory (e.g., Kinoshita, 2001; Richardson-Klavehn et al., 1994). Often, participants displaying involuntary aware memory are simply described as *test aware* (Bowers & Schacter, 1990).

Moreover, once participants become aware that they are retrieving studied words, there is nothing—besides being instructed to use the first word that comes to mind—to prevent them from intentionally retrieving those words. That is, once participants become aware that they are using previously studied words such as *flash* to complete the stems, they may engage in a variety of strategies that constitute a shift away from saying the first word that comes to mind and toward intentionally responding with studied words. For example, participants may adopt a mental set in which only the words from the study portion of the experiment can serve as potential responses. Alternatively, when faced with a number of possible completions (e.g., *flab*, *flag*, *flake*, *flame*, *flash*, *flat*, *flavor*, etc.) for a particular stem (i.e., *fla*), participants may consciously favor a studied word (such as *flash*). The presence of intentional retrieval

T. M. Barnhardt, tbarnhardt@tamu.edu

strategies in implicit memory tests has been referred to as voluntary explicit memory (e.g., Schacter, 1987) or voluntary aware memory (e.g., Richardson-Klavehn et al., 1994). Both the possibility of test awareness and the possibility of intentional retrieval fall under the rubric of explicit memory contamination.

Explicit memory contamination of implicit memory tests is worrisome because such contamination could alter the pattern of performance on an implicit memory test and could lead researchers to mistakenly attribute properties of explicit memory to implicit memory. For example, researchers have used explicit contamination of implicit memory tests to explain age differences in implicit memory (i.e., elderly populations display less priming because they have less explicit contamination than do young populations; for a discussion, see Mitchell & Bruss, 2003) and the presence of cross-modality effects in perceptual-priming tests (i.e., without explicit contamination, there would be no priming on a perceptual test when the materials were studied in a different modality; see, e.g., Jacoby, Toth, & Yonelinas, 1993; for a discussion, see Schacter & Badgaiyan, 2001).

Several methods for reducing explicit contamination in implicit memory tests have been described, including the use of elaborate cover stories, filler tasks, foil explicit memory tests, and a small proportion of studied items in the implicit test (e.g., Roediger & Geraci, 2005; Roediger & McDermott, 1993). However, the use of such methods does not guarantee that contamination will be eliminated. Therefore, a number of methods exist for assessing the presence of explicit contamination, including (but not limited to) the process dissociation procedure (Jacoby, 1991), the method of opposition (Richardson-Klavehn et al., 1994), and the retrieval intentionality criterion (Schacter et al., 1989).

One easy and often-used method for assessing the presence of explicit contamination has been the administration of posttest awareness questionnaires. This method was the focus of the present research. The awareness questionnaire method consists of asking participants, after they have completed the implicit test, about the experience they just had. The first researchers to report this method were Bowers and Schacter (1990). Their experiments used a typical study-test paradigm, in which participants were exposed to words during the first (study) phase of the experiment and were subsequently (after a filler task) administered an implicit memory test during the test phase of the experiment. After the participants had been administered the implicit memory test, they were asked four questions: "What did you think was the purpose of the stem completion task that you just finished?" "What was your general strategy in completing the word stems?" "Did you notice any relation between the words I showed you earlier and the words produced on the stem completion test?" and "While doing the stem completion test, did you notice whether you completed some of the stems with the words studied in the earlier list?" The participants were classified as test aware if they responded positively—that is, indicated that they were aware of having responded with studied words—on at least one of the four questions.

From the time the posttest awareness questionnaire was first introduced, one obvious concern has been whether it is a valid measure of test awareness (for a discussion, see Reingold & Toth, 1996). That is, does classifying participants on the basis of their questionnaire responses actually capture the distinction between unaware, unintentionally retrieving participants, on the one hand, and aware (either unintentionally or intentionally retrieving) participants, on the other?

At first glance, it seems that the use of awareness questionnaires after implicit memory tests is fraught with methodological problems. Some of these methodological problems may occur because the questionnaire is administered after the implicit test. For example, participants may have forgotten the nature of the awareness they had experienced during the implicit memory test by the time the questionnaire is administered. In addition, participants may report the nature of the awareness that they are experiencing while completing the questionnaire, even though they are asked to report the nature of the awareness that they had experienced during the implicit test. Alternatively, some of these methodological problems may occur because of the typical study-test paradigm that is used to test implicit memory. For example, in repetition-priming experiments, participants study some type of stimuli in a first phase of the experiment and then are later given a task involving the same type of stimuli. This instates quite high experimental demands for participants to state that they were aware of having used the studied stimuli during the test task. In addition, experimenters often go through a number of contortions in order to disguise the fact that the implicit test is a memory test. These contortions can instate an experimental demand in participants to state that they were unaware that they had produced studied words on the implicit test. Finally, some of these problems may occur because of the nature of the awareness questionnaire itself. The questionnaire poses a number of leading questions that, in effect, funnel participants' responses toward an endorsement of test awareness.¹ For example, Question 3 asks, "Did you notice any relation between the words I showed you earlier and the words produced on the stem completion test?" Participants may respond positively to this question not only if they noticed that they were repeating the stimuli that were presented earlier, but also if they feel they've noticed any relation (e.g., semantic? grammatical?) between their test responses and the study stimuli.

Nevertheless, there is some evidence indicating that the use of posttest questionnaires is a valid approach to assessing the presence of awareness in implicit memory tests. This evidence comes in two forms. First, the magnitude of priming in implicit memory tests is often dependent on awareness classification, with greater priming obtained in aware participants than in unaware participants (e.g., Barnhardt, 2004; Bowers & Schacter, 1990). This kind of result can be interpreted as evidence for the validity of the awareness questionnaire. The logic behind this interpretation goes something like this: If (1) the absolute magnitude of explicit memory performance is greater than that for implicit memory performance, (2) there is

explicit contamination of an implicit memory test, and (3) the awareness questionnaire is sensitive to that contamination, implicit memory test performance should increase as reported test awareness increases. Again, greater priming for aware participants than for unaware participants can be taken as support for the validity of posttest questionnaires.

Second, the pattern of priming effects observed in implicit memory tests is often dependent on awareness classification (e.g., Barnhardt, 2004; Bowers & Schacter, 1990; Camp, Pecher, & Schmidt, 2005; Geraci & Rajaram, 2002; Mace, 2003a, 2003b; Mulligan & Hartman, 1996; Pilotti, Chodorow, & Tan, 2004; Richardson-Klavehn, Lee, Joubert, & Bjork, 1994). Most important, the direction of the effects in aware participants is often similar to the direction of the effects obtained in explicit memory tests (although the magnitude of the effect is typically smaller in the implicit test), whereas such effects are absent in unaware participants. For example, Geraci and Rajaram (2002) obtained an orthographic distinctiveness effect (superior memory for unusual-looking words such as *sequoia*) on explicit memory tests. In an implicit memory test of word fragment completion (e.g., *s _ q u _ _ a*), the orthographic distinctiveness effect was obtained in those participants who reported test awareness (see also Kinoshita & Miller, 2000), but not in those who were unaware. Similar patterns of results have been observed when the implicit memory tests were conceptual in nature. For example, Mulligan, Guyer, and Breland (1999) found that level-of-processing (LOP) effects (e.g., Craik & Lockhart, 1972) in the category exemplar generation task depended on awareness (see also Mace, 2003a). These types of results can also be interpreted as support for the validity of the awareness questionnaire. The logic behind this claim is a slight variation on the rationale described earlier. In this case, if (1) explicit memory is more sensitive to a particular manipulation than is implicit memory, (2) there is explicit contamination of an implicit memory test, and (3) the awareness questionnaire is sensitive to that contamination, the effect in question should increase as test awareness increases. In sum, both greater priming as test awareness increases and a greater *explicit memory effect* as test awareness increases constitute evidence for the validity of the awareness questionnaire.

Although there is clear evidence for the validity of posttest questionnaires, there is also some evidence that is less clear with respect to this claim. For example, there are studies in which some participants reported test awareness but such awareness did not mediate either the degree of priming or the presence of some effect typically associated with explicit memory (e.g., Baques, Saiz, & Bowers, 2004; Mulligan, 2003). For example, Lustig and Hasher (2001) found that awareness did not mediate interference effects in a fragment completion test. In addition, Mulligan and colleagues found that awareness did not mediate divided attention effects (Mulligan & Hartman, 1996) or generation effects (Mulligan, 2002) in category exemplar generation tests. The interpretation of these results with respect to the validity of the awareness questionnaire is unclear. One possibility is that the awareness question-

naire accurately tracked the presence of awareness but that priming and test awareness were functionally independent in those experiments. The other possibility is that the awareness questionnaire did not provide an accurate measure of awareness and, as a result, the observation of any relation between awareness and priming was precluded. Given that instances in which priming and awareness are independent may be interpreted as evidence against the validity of the questionnaire, it may be useful to have an experimental approach that does not rely exclusively on the relationship between awareness and priming performance.

One alternative approach is to examine the effects of certain manipulations on the frequency with which participants report awareness, rather than relying on the relationship between awareness and priming magnitude or typical explicit memory effects. For example, using oral stem completion tests, Barnhardt (2004) showed that certain experimental manipulations affected the number of aware participants in a theoretically defensible manner. In one manipulation, intertrial interval (i.e., the time elapsed between the onset of two consecutive stimuli) was varied at test, and fewer participants reported awareness when the interval was short than when it was long. The interpretation offered for this effect was that the shorter intervals reduced the opportunity for the relatively slow advent of awareness to occur. In another manipulation, the number of solutions for stems was varied, and fewer participants reported awareness when there were many solutions than when there were fewer solutions. The interpretation offered for this effect was that stems with fewer solutions presented many more opportunities for awareness to occur because there was a much greater probability that participants would state a studied word in the case of stems with fewer solutions (even if it was just by chance) than in the case of stems with many solutions. In sum, it appears that the validity of awareness questionnaires can also be investigated by examining the effects of certain variables on the frequency of awareness.

In the experiments reported here, a simple but novel manipulation was used to investigate the validity of the awareness questionnaire, once again by assessing changes in the frequency of awareness reports across certain conditions. In one condition, none of the studied words could be used as legitimate test responses (i.e., the none-studied condition). In the other condition, studied words could be used as legitimate test responses (i.e., a standard implicit memory test condition). The prediction was very straightforward: If the awareness questionnaire is valid, fewer participants should state that they were aware of having produced studied words in the none-studied condition than in the standard condition. Such a finding would converge with other, similar findings (e.g., Barnhardt, 2004) in supporting the validity of awareness questionnaires.

The inclusion of the none-studied condition also provided another test of the validity of the awareness questionnaire. In the none-studied condition, studied words cannot be used as legitimate test responses. As a result, in the none-studied condition, the frequency of awareness is known prior to the administration of the aware-

ness questionnaire. As such, the none-studied condition provides a known standard by which the validity of the awareness questionnaire can be measured. If the awareness questionnaire is perfectly valid, no participants in that condition should state that they are aware of having said studied words. In short, the closer the frequency of unaware reports in the none-studied condition is to 100%, the more confident researchers can be in the validity of the questionnaire.

In addition, the relationship between priming and awareness in the standard implicit test condition was evaluated. As was mentioned earlier, an association between priming and awareness would also constitute evidence for the validity of awareness questionnaires.

With these ideas in mind, two experiments were conducted in order to investigate the validity of awareness questionnaires. In Experiment 1, a standard perceptual implicit memory test (i.e., stem completion) was used. In Experiment 2, a standard conceptual implicit memory test (i.e., category production) was used. The nature of the implicit test was varied across the two experiments in order to examine whether the validity of the awareness questionnaire would be altered by using different types of implicit memory tests.

EXPERIMENT 1

In the introduction, experiments were described in which priming and/or explicit memory effects in an implicit memory test did not increase as awareness increased (e.g., Mulligan, 2002). As was noted earlier, when faced with this type of result, it is difficult to know whether priming and awareness were functionally independent or whether the posttest questionnaire failed to accurately track awareness. In order to disentangle these two interpretations, LOP (Craik & Lockhart, 1972) was manipulated between subjects during the study phase. Prior research has revealed a much stronger association between test awareness and deep processing than between test awareness and shallow processing (e.g., Graf, Mandler, & Haden, 1982; Mace, 2003a, 2003b; Richardson-Klavehn et al., 1994; Toth, Reingold, & Jacoby, 1994). The predictions and rationale were as follows. Priming would be obtained in the standard condition, but not in the none-studied condition, after both deep and shallow processing. However, frequency of awareness would vary much more across the standard and none-studied conditions after deep processing than after shallow processing. This pattern of results in the deep condition would constitute evidence for the validity of the awareness questionnaire. Usually, however, the pattern of results in the shallow condition, when taken alone, would be ambiguous with respect to the validity of the questionnaire. But in this case, because the validity of the questionnaire would be supported under the very same circumstances in the deep condition, the interpretation of this result in the shallow condition would be heavily tilted in favor of the independence of priming and awareness, rather than a lack of validity in the questionnaire.

To sum up, there were three ways in which the validity of the awareness questionnaire could be supported: first, if frequency of awareness was much less in the none-studied condition than in the standard condition after deep processing; second, if the frequency of unaware reports in the none-studied condition was close to 100%; and third, if a positive association between awareness and priming was observed in the standard condition. In addition, if the predicted contrast between the deep and the shallow processing conditions was obtained, this would imply that observations of statistical independence between awareness and priming, as reported in Mulligan (2002), may be due to the functional independence of awareness and priming, rather than to a lack of validity in the awareness questionnaire itself.

Method

Design. The design of this experiment included two between-subjects factors: test type (standard vs. none-studied) and LOP (deep vs. shallow). For the test type manipulation, half of the participants received a word stem completion test that included some stems that could be completed with studied words (i.e., the standard condition), and half of the participants received a word stem completion test that included only stems that could not be completed with any studied words (i.e., the none-studied condition). For the LOP manipulation, participants studied the items either deeply (i.e., they rated the pleasantness of the meaning of each word) or shallowly (i.e., they counted the number of letters with enclosed spaces, such as *g* but not *n*).

Participants. There were 144 participants. They received credit toward the research participation component of their introductory psychology course.

Materials. Three study lists of 40 items each were derived from a pool of 120 low- to medium-frequency words. These three lists were used to counterbalance for study status across participants. Each study list included 4 filler words at the beginning and end of the lists to reduce possible primacy and recency effects. For each study list, two test lists were created. One list served as a typical implicit test in the standard condition. This test list contained the stems from the 40 studied items as well as 40 stems from nonstudied items (counterbalanced across the other two study lists), for a total of 80 test stems. The other test list contained stems from the two other study lists, also for a total of 80 test stems. This test list served as the control implicit test in the none-studied condition.

Procedure. During the incidental study task, the participants were told that they would be presented with a list of words and would be asked to make judgments about these words. For each word, the participants were asked to make either deep judgments or shallow judgments about the items in the study list, using a 5-point scale. Words were presented for 5 sec, and the participants were required to make their judgments within this amount of time. After the study list had been presented, the participants were given a 5-min distractor task in which they were asked to mentally rotate objects. Next, the participants were given the stem completion test. They were told that the test was designed to measure word knowledge. For this test, they were told that they would be given a list of stems containing only the first three letters of words and that they should try to complete the stem with the first word that came to mind. They were informed that speeded responding was of utmost importance and that if they could not complete an item within approximately 5 sec, they should proceed to the next item. Lastly, they were told not to use any proper nouns or words with four or fewer letters to complete the stems. In total, this test took approximately 7 min to complete.

After the word stem completion test, the participants were given a slightly modified version of the posttest questionnaire used by Bowers and Schacter (1990). In addition to the four questions that they used, three other questions were included. The fifth question was

designed to determine whether test-aware participants continued to use the first word that came to mind or whether they changed their retrieval strategy to one in which they were intentionally trying to respond with studied words (see also Barnhardt, 2004; Mace, 2003a, 2003b). The sixth question asked test-aware participants to estimate the point in the test at which they had become aware, and the seventh question asked test-aware participants to state the particular word, if they could, on which they had become aware of saying studied words. The questionnaire is presented in the Appendix. The questions were printed on both sides of a single piece of paper, with the first three questions on one side and the last four questions on the other side. The participants wrote their own answers in response to the questions. They were told, "Please take a few minutes to briefly answer the questions on this sheet, starting with number one. Look up when you have finished and I will collect your questionnaire." Only the first five questions were used to classify the participants. If the participants answered all five of these questions negatively, they were classified as *unaware*; if they answered at least one of the first four questions positively, they were classified as *aware*; and if they answered Question 5 in a way that indicated that they were intentionally retrieving the studied words, they were classified as *intentional*, regardless of their other responses.

Results and Discussion

The results were broken into four subsections. The significance level for all statistical tests was set at $p < .05$.

Priming. First, it was important to determine whether priming was obtained in the standard condition, since the presence of priming may have triggered test awareness. Priming was calculated for each participant by subtracting the proportion of target responses in the baseline portion of the standard condition from the proportion of target responses in the studied portion of the standard condition. In the deep standard condition, $M = .31$, $SD = .09$, for the studied items, and $M = .20$, $SD = .07$, for the baseline items. In the shallow standard condition, $M = .30$, $SD = .09$, for the studied items, and $M = .23$, $SD = .06$, for the baseline items. A significant amount of priming was obtained in both the deep condition [$M = .11$; $t(35) = 7.39$, $SEM = .015$] and the shallow condition [$M = .07$; $t(35) = 4.24$, $SEM = .017$]. Although priming was numerically greater in the deep condition than in the shallow condition (i.e., the LOP effect), the difference was not significant ($p = .11$).

The influence of test condition and LOP on the frequency of awareness. Next, the overall accuracy of the posttest questionnaire was examined by comparing reports of awareness across the none-studied and standard conditions. In order to do this, the participants were classified into one of the three categories described above. The number of participants in each awareness category, as a function of LOP and test type, is displayed in Table 1. In general, reports of awareness were much more frequent in the deep standard condition than in the deep none-studied condition but were relatively infrequent in both the shallow standard and the shallow none-studied conditions. In the deep condition, the distribution of participants across the awareness categories was significantly different in the standard and the none-studied conditions [$\chi^2(2) = 30.45$]. In contrast, in the shallow condition, this difference was not significant [$\chi^2(2) = 3.19$, $p = .074$]. However, a $2 \times 2 \times 3$ hierarchical log linear analysis of the

three-way dependence of LOP, test type, and awareness classification was not significant (difference $G^2 = 3.914$, $p = .14$). An inspection of Table 1 revealed that three of the four intentional cells had an observed frequency of zero and the fourth had an observed frequency of four. Classifying the intentional cases as rare cases eliminated only 4 of the 144 (2.7%) participants. The subsequent $2 \times 2 \times 2$ hierarchical log linear analysis was significant (difference $G^2 = 3.916$, $p = .0478$). This finding verified the prediction that the difference in awareness across the deep standard and deep none-studied conditions would be greater than that difference in the shallow condition. As such, it supported the notion that instances in which awareness and priming are statistically independent are due to their functional independence, rather than to a lack of validity in the awareness questionnaire.

It was also argued that the validity of the awareness questionnaire would be supported if the percentage of unaware participants in the none-studied condition was close to 100%. In this experiment, 60 of 72 participants (83%) in the none-studied condition described themselves as test unaware. Although not perfect, a number of reasons will be given in the General Discussion section for why this accuracy rate can be considered in a positive light.

Questionnaire response patterns. Although the classification of participants in the none-studied condition as unaware was highly accurate, it was not perfect. Twelve of 72 participants (17%) in the none-studied condition mistakenly described themselves as test aware even though they did not respond with any studied words. We will use the phrase *mistaken awareness* to describe this phenomenon. Mistaken awareness may be attributable to experimental demands in the study-test paradigm or the questionnaire. It may also be triggered by false memories (e.g., Roediger & McDermott, 1995). We will have more to say about mistaken awareness in the General Discussion section.

For present purposes, we were interested in whether the nature of awareness in mistakenly aware participants was similar to the awareness in aware participants in the studied condition. To remind the reader, the participants were classified as aware if they endorsed any one or more of the first four questions. Perhaps the number of questions endorsed by mistakenly aware participants was different than that for aware participants.

Table 2 displays the frequencies with which both aware and mistakenly aware participants endorsed awareness

Table 1
Number of Participants (With Percentage Conversions in Parentheses) in Each Awareness Category As a Function of Level of Processing and Test Condition in Experiment 1

Awareness Category	Deep		Shallow	
	Standard	None Studied	Standard	None Studied
Unaware	5 (14%)	28 (78%)	26 (72%)	32 (89%)
Aware	27 (75%)	8 (22%)	10 (28%)	4 (11%)
Intentional	4 (11%)	0	0	0

Note—Unaware, no questions endorsed; aware, any of Questions 1–4 endorsed; intentional, Question 5 endorsed.

Table 2
Number of Aware Participants (With Percentage Conversions
in Parentheses) Endorsing Either One or at Least Two Questions
As a Function of Level of Processing and Test Condition in Experiment 1

Number of Questions Endorsed	Deep		Shallow	
	Standard	None Studied	Standard	None Studied
1 question	3 (11%)	8 (100%)	6 (60%)	4 (100%)
At least 2 questions	24 (89%)	0 (0%)	4 (40%)	0 (0%)
Total	27	8	10	4

Note—"Total" refers to the number of aware participants in each condition.

either on only one question or on at least two questions. In the deep condition, a high percentage of aware participants in the standard condition (88%) endorsed two or more of the first four questions, whereas this was true for none of the participants in the none-studied condition; that is, all of the mistakenly aware participants endorsed only a single question. The difference in the pattern of questionnaire responses between the aware participants in the deep standard and deep none-studied conditions was significant ($p < .001$, Fisher's exact test). In contrast, there was relatively little difference between aware participants in the shallow standard and none-studied conditions ($p = .21$, Fisher's exact test). In the shallow condition, a relatively high percentage of participants in both conditions (standard, 60%; none-studied, 100%) endorsed a single question. It was not surprising to find a larger difference in the deep condition than in the shallow condition for the number of questions endorsed. It was also in the deep condition that there was a large discrepancy across the standard and none-studied conditions in the frequency of aware participants.

These results, especially those in the deep condition, showed (1) that the awareness reported in the standard condition was not necessarily like the awareness reported in the none-studied condition and (2) that the awareness questionnaire was sensitive to potentially important, perhaps subtle, differences within the awareness category. These results also showed that the traditional awareness questionnaire classification scheme that we have used up to this point may be biased in that it is too liberal in categorizing participants as aware. In turn, this suggested that another type of classification scheme might have more accurately reflected the participants' state of awareness during this particular implicit memory test. For example, if the participants in the none-studied condition who responded positively to only a single question were counted as unaware, every participant in the none-studied condition would have been correctly classified as unaware. If this revised classification scheme were applied to the standard condition, this would imply that the participants in the standard condition who responded positively to only a single question might also have been mistakenly aware and might have been better classified as unaware, rather than aware. We will return to this point in the following discussion.

Awareness and priming. As was mentioned in the introduction, the dependency of priming magnitude (or priming effects) on awareness is an important piece of evidence for evaluating the validity of the questionnaire. Here, in the

first set of analyses, priming was examined as a function of the traditional awareness classification scheme that has been used throughout these analyses. When examined as a function of this scheme, priming did not differ, regardless of whether the deep and the shallow standard conditions were examined separately, whether the interaction of LOP and awareness was examined, or whether the two standard conditions were collapsed across LOP (all F s < 1.55). Collapsing across the two standard conditions, the mean proportion of priming (and the number of participants) for each awareness category was $M = .08$ ($n = 31$) for unaware, $M = .10$ ($n = 37$) for aware, and $M = .16$ ($n = 4$) for intentional. The correlation of priming and this awareness classification was $r = .187$, $p = .116$.

One possible reason for the lack of a significant relationship between priming and awareness may lie in the nature of the implicit tests themselves. Buchner and Wipich (2000) have shown that implicit memory tests are relatively unreliable measures of priming. In turn, the low reliability of a measure limits its validity. If the amount of priming observed in the implicit test is not representative of the amount of priming actually present, clearly this could put a cap on the ability to observe a relationship between magnitude of priming and test awareness.

However, another possible reason for the lack of a significant relationship between priming and awareness may lie in the awareness classification procedure. As was noted earlier, all mistakenly aware participants endorsed only a single question, and it seemed possible that the participants endorsing a single question in the standard condition might also be better classified as unaware. Given this, in a second set of analyses, the participants in the standard condition were classified as unaware even when they endorsed one question (other than Question 5). When priming was examined as a function of this revised awareness classification scheme, priming again did not differ as a function of awareness classification in either the deep ($F < 1$) or the shallow ($F < 2$) standard conditions, nor was the interaction significant ($F < 1$), although the means were in the predicted direction. However, when the two standard conditions were collapsed across LOP, there was a strong trend for priming to increase as awareness increased [$F(2,69) = 2.88$, $MS_e = 0.009$, $p = .063$]. The mean proportion of priming (and number of participants) for each awareness category was $M = .07$ ($n = 40$) for unaware, $M = .11$ ($n = 28$) for aware, and $M = .16$ ($n = 4$) for intentional. In addition, the correlation of priming and the revised awareness classification was significant ($r = .277$).

To sum up, there were several pieces of evidence for the validity of the awareness questionnaire. First, in the deep study condition, reports of awareness were much less frequent in the none-studied condition than in the studied condition. Second, when the traditional classification scheme was used, the questionnaire correctly classified the vast majority of participants (83%) in the none-studied condition as unaware. Third, when the revised classification scheme was used, there was some evidence that priming increased as awareness increased in the standard condition. Finally, the three-way interaction of awareness, test type, and LOP was significant, signaling that the difference in awareness across the standard and none-studied conditions after deep processing was greater than that same difference in the shallow condition. As was outlined earlier, the contrast between the deep and the shallow conditions supported the notion that instances in which priming and awareness are empirically independent are due to functional independence between priming and awareness, rather than to a lack of validity in the awareness questionnaire. In Experiment 2, these same issues were examined when a conceptual, rather than a perceptual, implicit memory test was used.

EXPERIMENT 2

The conceptual implicit memory test used in this experiment was the category production test. In this test, participants were presented with a category label and were asked to generate as many exemplars as they could within a limited time period. Again, LOP and test type (standard vs. none-studied) were manipulated in this experiment. As before, the main questions of interest were whether the awareness questionnaire would (1) accurately reflect the difference between the none-studied and the standard conditions after deep study, (2) accurately reflect the similarity between the none-studied and the standard conditions after shallow study, (3) accurately classify participants in the none-studied condition as unaware, and (4) accurately classify participants in the standard condition so that priming would increase as awareness increased.

Method

Design. The design for this experiment was the same as that used in Experiment 1: a 2×2 between-subjects design in which the principal factors were test type (standard vs. none-studied) and LOP (deep vs. shallow).

Participants. There were 120 participants. They received credit toward the research participation component of their introductory psychology course.

Materials. Three study lists consisting of 10 eight-item groups of categorized words was randomly selected from a pool of 30 different groups of 8-item categories. In addition to these 80 items, each study list included four filler words at the beginning and end of the lists to reduce possible primacy and recency effects. The study words were presented in a mixed order (not blocked) with respect to their category membership. For each study list, two test lists were created. One list served as a typical implicit test. This test list contained the 10 studied category labels, as well as 10 nonstudied category labels (counterbalanced across the two other study lists), for a total of 20 test categories. The other test list contained all nonstudied category labels taken from the two other study lists, for a total of 20 nonstudied test categories.

Procedure. The procedure for the study task and the filler task was the same as that in Experiment 1. For the category production task, each category label (e.g., a type of bird) was presented on a separate page, and the participants were given 30 sec to write down as many exemplars of the category as they could in that amount of time. At the end of the 30 sec, the participants heard a beep sounded by an audio cassette that prompted them to proceed to the next category label. This test took approximately 10 min. After the category production test, all the participants were given the same awareness questionnaire as that in Experiment 1.

Results and Discussion

The results were again broken into four subsections. The significance level for all statistical tests was set at $p < .05$.

Priming. First, it was important to determine whether priming was observed in the standard condition, since the presence of awareness may be dependent upon the presence of priming. As in Experiment 1, priming was calculated for each participant by subtracting the proportion of target responses (in this case, category exemplars) in the baseline portion of the standard condition from the proportion of target responses in the studied portion of the standard condition. In the deep standard condition, $M = .33$, $SD = .07$, for the studied items, and $M = .20$, $SD = .08$, for the baseline items. In the shallow standard condition, $M = .22$, $SD = .06$, for the studied items, and $M = .21$, $SD = .06$, for the baseline items. A significant amount of priming was observed in the deep condition [$M = .12$; $t(29) = 8.44$, $SEM = .015$], but not in the shallow condition [$M = .01$; $t(29) = 1.07$, $SEM = .013$, $p = .29$]. The difference between the conditions was significant [$t(58) = 5.74$, $SEM = .019$].

The absence of a statistically significant amount of priming in a category generation test after shallow processing was a bit unusual. Typically, conceptual priming after shallow processing is less than after deep processing (e.g., Mulligan et al., 1999), but not altogether absent. However, the absence of priming was not totally unexpected. From the perspective of transfer-appropriate processing theories (e.g., Roediger, 1990), the magnitude of priming is a function of the overlap between the type of processing engaged at study and the type of processing engaged at test. Since processing surface features of letters during the shallow study task would seem to have very little in common with processing conceptual information during the category generation task, it was not surprising that almost no priming was obtained in the shallow standard condition.

The influence of test condition and LOP on the frequency of awareness. Next, the overall accuracy of the posttest questionnaire was examined by comparing frequency of awareness across the different LOP \times test type combinations. The traditional classification scheme was used to categorize participants on the basis of their questionnaire responses. The results are displayed in Table 3. As can be seen, reports of awareness and intentional retrieval were much more frequent in the deep standard condition than in the deep none-studied condition [$\chi^2(2) = 39.77$]. In contrast, reports of awareness across the shallow standard and none-studied conditions were relatively infrequent and not statistically different

Table 3
Number of Participants (With Percentage Conversions in Parentheses) in Each Awareness Category As a Function of Level of Processing and Test Condition in Experiment 2

Awareness Category	Deep		Shallow	
	Standard	None Studied	Standard	None Studied
Unaware	2 (7%)	26 (87%)	18 (60%)	24 (80%)
Aware	16 (53%)	4 (13%)	11 (37%)	6 (20%)
Intentional	12 (40%)	0	1 (3%)	0

Note—Unaware, no questions endorsed; aware, any of Questions 1–4 endorsed; intentional, Question 5 endorsed.

$[\chi^2(2) = 3.33, p = .189]$. When the deep and the shallow standard conditions were compared, awareness was again much more prevalent in the deep condition [$\chi^2(2) = 23.03$]. The hierarchical log linear analysis of the three-way dependence of LOP, test type, and awareness classification was significant (difference $G^2 = 8.904$). As in Experiment 1, this result supported the notion that instances in which awareness and priming are statistically independent are not necessarily due to a lack of validity in the awareness questionnaire.

As was noted earlier, the validity of the awareness questionnaire would be supported if the percentage of unaware participants in the none-studied conditions was close to 100%. In this experiment, 50 of 60 participants (83%) in the none-studied condition described themselves as test unaware. This was the same percentage as in Experiment 1. Again, a number of reasons will be given in the General Discussion section for why this accuracy rate can be considered in a positive light.

Finally, it is interesting to note that fully 40% of the participants in the shallow standard condition described themselves as aware or intentionally retrieving despite the fact that there was virtually no priming in that condition. We will use the phrase *awareness without priming* to describe this phenomenon. The topic of awareness without priming will be postponed until the General Discussion section.

Questionnaire response patterns. As in Experiment 1, the classification of participants in the none-studied condition as unaware was very good, but it was not perfect. Ten (17%) of the 60 participants in the none-studied condition were classified as aware even though none of the studied category exemplars could have served as legitimate responses. As in Experiment 1, we were interested in whether the nature of awareness in the aware standard participants was similar to that in the mistakenly

aware none-studied participants. Again, in order to answer this question, we looked at the number of awareness questions that these participants endorsed (see Table 4).

There were distinct patterns of question endorsement across the deep standard and deep none-studied conditions: 100% of the aware participants in the standard condition endorsed at least two questions, whereas only 50% of the mistakenly aware participants endorsed two questions in the none-studied condition ($p = .03$, Fisher's exact test). The same was true in the shallow condition, where 81% of the aware participants in the standard condition endorsed two or more of the first four questions, whereas only 13% of the mistakenly aware participants endorsed two questions in the none-studied condition ($p = .02$, Fisher's exact test). These results, in both LOP conditions, showed once again that (1) the awareness reported in the standard condition was not necessarily like the awareness reported in the none-studied condition and (2) the awareness questionnaire was sensitive to potentially important, perhaps subtle, differences within the awareness category.

Awareness and priming. As was noted in the Results and Discussion section of Experiment 1, an examination of the relationship between awareness and priming is important because it can provide converging evidence for the validity of the awareness questionnaire. In the deep, standard condition, priming varied as a function of awareness classification [$F(2,27) = 8.63, MS_e = 0.004$]. The mean proportion of priming (and number of participants) for each awareness category was $M = .01$ ($n = 2$) for unaware, $M = .10$ ($n = 16$) for aware, and $M = .18$ ($n = 12$) for intentional. In the shallow, standard condition, there was no overall priming, and as might be expected on the basis of that observation, neither was there any significant association between priming and awareness classification in that condition ($F < 1$).

In summary, there were several pieces of evidence that supported the validity of the awareness questionnaire when it was used after a conceptual priming test. First, in the deep study condition, there was a large difference between the standard and the none-studied conditions in the distribution of participants across the awareness categories. Second, the questionnaire correctly classified the vast majority of participants (again, 83%) in the none-studied conditions as unaware. Third, priming increased as awareness increased in the deep standard condition. Finally, the three-way interaction of awareness, test type, and LOP was significant, indicating a strong contrast between the deep and the shallow processing conditions with regard to the

Table 4
Number of Aware Participants (With Percentage Conversions in Parentheses) Endorsing Either One or at Least Two Questions As a Function of Level of Processing and Test Condition in Experiment 2

Number of Questions Endorsed	Deep		Shallow	
	Standard	None Studied	Standard	None Studied
1 question	0 (0%)	2 (50%)	2 (18%)	5 (83%)
At least 2 questions	16 (100%)	2 (50%)	9 (82%)	1 (17%)
Total	16	4	11	6

Note—"Total" refers to the number of aware participants in each condition.

manner in which frequency of awareness varied across the none-studied and standard conditions. Again, the contrast between the deep and the shallow conditions supported the notion that instances in which priming and awareness are empirically independent are due to a functional independence between priming and awareness, rather than to a lack of validity in the awareness questionnaire.

In many ways, the results of Experiment 2 were similar to those of Experiment 1, indicating that the awareness questionnaire was at least as appropriate for use after a conceptual implicit memory test as after a perceptual implicit memory test. Other similarities and differences between the two experiments and some more general issues will be discussed next.

GENERAL DISCUSSION

“This thorny issue of awareness will exercise the field for some time to come” (Roediger & McDermott, 1993, p. 70). Indeed, since the introduction of the explicit/implicit memory distinction, the measurement and control of awareness (and its consequences) in implicit memory tests has generated a great deal of experimental attention. Posttest awareness questionnaires have been a popular choice for addressing the issue of awareness in implicit memory tests, even though their validity has not been well established.

In the present experiments, the validity of awareness questionnaires was investigated by comparing the frequency of awareness across two conditions: one in which the studied words could *not* serve as legitimate test responses (i.e., the none-studied condition) and one in which studied words could serve as legitimate test responses (i.e., a standard implicit test condition). In both experiments, there was a large difference in the frequency of awareness reports across the none-studied and standard conditions after the participants had studied words deeply. Thus, the awareness questionnaire was highly proficient at tracking predicted changes in awareness across certain conditions.

There were other pieces of evidence that supported the validity of the awareness questionnaire. For example, in Experiment 2, priming increased as awareness increased. This was also true in Experiment 1, but the results were somewhat more complicated and will be considered further below. In general, the relationship between priming and awareness in the present experiments supported the validity of awareness questionnaires.

Another piece of evidence for the validity of the awareness questionnaire was the accuracy with which the participants in the none-studied condition were classified as unaware. How close to a perfect classification rate did the awareness questionnaire achieve in the none-studied condition? At an empirical level, the answer to this question is a simple one. In both experiments, 83% of the participants in the none-studied conditions did not endorse any of the questions on the awareness questionnaire. At an evaluative level, there are reasons to be encouraged by this degree of classification accuracy. For example, despite a number of potential methodological shortcomings in the aware-

ness questionnaire, these results far exceeded the minimal expectations that some researchers have had for the capacity of awareness questionnaires to track meaningful changes in awareness during implicit memory tests (Reingold & Toth, 1996). In addition, the 83% accuracy rate is probably an underestimate of the accuracy that awareness questionnaires can ultimately obtain. None of the methodological weaknesses in the awareness questionnaire approach that have been described—whether in the study–test procedure, in the nature of the questions in the awareness questionnaire, in the method of administering the questionnaire, or in the scheme for classifying participants on the basis of their questionnaire responses—have received much systematic attention with an eye toward improving the questionnaire’s accuracy. Thus, with some work, an even more accurate awareness questionnaire seems readily attainable.

Finally, the contrast between the deep and the shallow processing conditions provided evidence that some instances in which awareness increases but priming does not are attributable to the functional independence of awareness and priming, rather than to poor validity of the awareness questionnaire. In both of the present experiments, at the same time that there was a large shift in frequency of awareness across none-studied and standard conditions after deep processing, there was a relatively small shift in the frequency of awareness across the none-studied and standard conditions after shallow processing. In Experiment 1, the relatively small difference in awareness across the none-studied and standard conditions after shallow processing occurred despite a relatively large difference in priming across those conditions. Simply stated, in Experiment 1, the questionnaire was not sensitive to awareness in the shallow standard condition because there was very little awareness in that condition (despite the presence of priming), and not because the questionnaire was invalid. This result was reminiscent of other findings in which awareness—as measured by a posttest questionnaire—and priming were statistically independent (e.g., Mulligan, 2002).

The shallow standard condition was important for another reason. Awareness was very likely to occur in the deep standard condition and very unlikely to occur in the none-studied condition. Some researchers might find it hardly surprising that the awareness questionnaire was sensitive to the extreme difference between those conditions. However, the shallow condition can be viewed as a kind of midpoint between these extremes. As such, comparisons between the shallow standard and the none-studied conditions might be more compelling in that they might be viewed as a test of the capacity of the questionnaire to detect less extreme differences in awareness. In Experiment 1, the difference between the shallow standard and the none-studied conditions was not significant ($p = .071$), but an effect size of small to medium magnitude was observed (Cramer’s $\phi = .21$). In Experiment 2, the difference again was not significant ($p = .189$), but again an effect size of small to medium magnitude was observed (Cramer’s $\phi = .24$). In short, it appears that the awareness questionnaire not only was sensitive to differ-

ences in awareness across extremely different conditions, such as the deep standard and none-studied conditions, but also was sensitive to more subtle differences in awareness, such as those found across the shallow standard and none-studied conditions.

There were a couple of other findings that deserve further attention. One of these was the fact that, in both experiments, 17% of the participants in the none-studied condition described themselves as test aware. Earlier, we used the term *mistaken awareness* to describe this phenomenon. It was also noted that this phenomenon could be attributable either to a kind of false memory (e.g., Roediger & McDermott, 1995) or to the experimental demands present in both the study–test paradigm and the questionnaire itself. Hypothetically, it would not be surprising that at least some of these instances of mistaken awareness were due to false memory. Memory illusions of this type frequently occur both in recognition tests when participants produce false alarms to lures and in free recall tests when participants produce intrusions. In the present experiments, the seventh question in the posttest questionnaire provided some direct evidence for the contribution of false memory to mistaken awareness. To remind the reader, this question asked, “If you noticed that you were writing words that had been presented earlier, did you notice while you were responding with a particular word? If so, what was that word?” Across the two experiments, of the 22 aware participants in the none-studied condition, 3 (14%) indicated a particular word on which they had become aware. In contrast, of the 81 aware participants in the standard condition, 30 (37%) indicated a particular word on which they had become aware. Thus, even though the percentage of mistakenly aware participants who cited a particular word was not as high as that for aware participants, the fact that some participants in the none-studied condition were willing to indicate a particular word supports the idea of a memory-related mistaken awareness in at least some instances.

There was one other finding that was of interest. In the shallow standard condition of Experiment 2, there was no evidence of priming, yet fully 40% of the participants described themselves as aware. We used the phrase *awareness without priming* to describe this phenomenon. How is awareness without priming possible? To remind the reader, the fact that there was no priming in the shallow standard condition did not mean that these participants never responded with a studied critical exemplar. It just meant that the rate at which they produced studied words (to “studied” category labels) was not any greater than the rate at which they produced nonstudied critical exemplars (to baseline, or “nonstudied,” category labels). The production of a studied critical exemplar, even when it was not retrieved from recent memory, could still have served to remind the participants that they had studied that word earlier in the experiment and, hence, given rise to test awareness. One might even expect that reported test awareness should have increased as more critical studied items were produced, even if that rate was no greater than that for baseline. To examine this possibility, we investigated the relationship between the number of studied critical items produced (without subtracting

out the baseline) and reports of awareness. The correlation between these two variables was not significant in either the deep standard condition ($r = .27, p = .15; n = 30$) or the shallow standard condition ($r = .24, p = .21; n = 30$), but it was significant when the standard conditions were collapsed across LOP ($r = .54, n = 60$). Thus, there was some support for the idea that awareness in the shallow standard condition was generated by the production of words at a baseline rate (i.e., from general knowledge), that these words just happened to be those that were studied earlier, and that the participants were reminded of having studied those words after they had produced them.

Finally, even though these experiments provided strong evidence for the validity of awareness questionnaires, they also showed that the awareness questionnaire (and the traditional classification scheme associated with it) was not perfect. In both experiments, 17% of the participants in the none-studied conditions were classified as aware even though none of the studied words could be used as legitimate responses in the implicit memory tests. As it stands, the awareness questionnaire has a liberal bias in that it tends to overestimate the number of aware participants, at least in the none-studied condition. This liberal bias is attributable to both the nature of the questionnaire and the nature of the classification system. First, the composition of the awareness questionnaire encouraged the participants to set a very liberal criterion for subjectively evaluating the presence of awareness. It did this by posing questions that became increasingly direct, to the point that, in Question 4, the participants were told the correct answer (it was a memory test) and all they needed to do was to agree to this in order to indicate that they were aware. Second, the nature of the categorization scheme also set a very liberal criterion for classifying the participants as aware. It did this by categorizing the participants as aware if they indicated awareness on a single question, even if it was on Question 4.

Moreover, there is no reason to think that this liberal bias would be absent in the standard conditions. This is welcome news for researchers who use the awareness questionnaire for eliminating participants who are test aware. It seems highly unlikely that a procedure such as the present awareness questionnaire, which has a liberal bias in that it tends to classify too many unaware participants as aware, could at the same time, mistakenly, classify aware participants as unaware. To put it another way, it seems highly unlikely that this procedure, given its liberal bias, would “miss” any aware participants, classifying them as unaware instead.

However, the fact that the awareness questionnaire has a liberal bias may be more problematic for researchers who use the questionnaire to study the relationship between awareness, priming, and priming effects. This was not the case in Experiment 2, where an association between priming and awareness was observed, even though the liberal bias was present. In Experiment 1, however, a significant relationship between awareness and priming was not observed when the traditional classification scheme was used to categorize the participants. This type of result undermines confidence in the awareness questionnaire and

the traditional classification scheme, inasmuch as it appears that even a relatively minor misclassification error can lead to potentially inaccurate conclusions regarding the association of priming and awareness.

The important point here is that the none-studied condition was also used to resolve this bias problem. When the patterns of question endorsement for the mistakenly aware participants in the none-studied condition were examined, the specific nature of the liberal bias was very clear: All the mistakenly aware participants had endorsed only a single question. This suggested that the participants in the standard condition who had endorsed only a single question may also have been mistakenly aware. A correction for this liberal bias in the standard condition was then undertaken, so that the 9 participants in the standard condition who had endorsed only a single question (and had been previously classified as aware) were reclassified as unaware. After this correction procedure was undertaken, a significant relationship between awareness and priming was observed. Using this same correction procedure, Barnhardt (2004) also observed a much stronger association between awareness and priming in an oral stem completion test. Thus, it appears that the inclusion of a none-studied condition not only can provide evidence on the validity of the questionnaire and reveal possible biases in the classification of participants, but also can be used to correct for those biases.

In conclusion, the evidence presented here suggests that the awareness questionnaire, even in its present state, yields a fairly valid assessment of test awareness after both perceptual implicit memory tests (in this case, stem completion) and conceptual implicit memory tests (in this case, category production). Several pieces of evidence supported this claim. First, the frequency distributions of the participants across the awareness categories shifted as the conditions changed from none-studied to standard and from deep to shallow. Second, 83% of the participants in the none-studied conditions were correctly classified as unaware. Third, priming increased as awareness increased both in Experiment 1 (after correcting for a liberal bias in the questionnaire) and in Experiment 2. Although our overarching assessment of the awareness questionnaire is a positive one, future research that includes an appropriate control condition (i.e., a none-studied condition followed by an awareness questionnaire) will allow the field to improve the validity of awareness questionnaires.

AUTHOR NOTE

We thank Liz Adams for conducting all of the data collection and data coding for the two experiments reported here. Correspondence concerning this article should be addressed to T. M. Barnhardt, Department of Psychology, Texas A&M University, 230 Psychology Building, College Station, TX 77843 (e-mail: tbarnhardt@tamu.edu).

REFERENCES

- BAQUES, J., SAIZ, D., & BOWERS, J. S. (2004). Effects of working memory load on long-term word priming. *Memory*, *12*, 301-313.
- BARNHARDT, T. M. (2004). Different involuntary mechanisms underlie priming and LOP effects in stem completion tests. *Memory*, *12*, 614-636.
- BOWERS, J. S., & SCHACTER, D. L. (1990). Implicit memory and test awareness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 404-416.
- BUCHNER, A., & WIPPICH, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology*, *40*, 227-259.
- CAMP, G., PECHER, D., & SCHMIDT, H. G. (2005). Retrieval-induced forgetting in implicit memory tests: The role of test awareness. *Psychonomic Bulletin & Review*, *12*, 490-494.
- CRAIK, F. I. M., & LOCKHART, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671-684.
- GERACI, L., & RAJARAM, S. (2002). The orthographic distinctiveness effect on direct and indirect tests of memory: Delineating the awareness and processing requirements. *Journal of Memory & Language*, *47*, 273-291.
- GRAF, P., MANDLER, G., & HADEN, P. E. (1982). Simulating amnesic symptoms in normal subjects. *Science*, *218*, 1243-1244.
- GRAF, P., & SCHACTER, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 501-518.
- HAMILTON, M., & GERACI, L. (2006). The picture superiority effect in conceptual implicit memory: A conceptual distinctiveness hypothesis. *American Journal of Psychology*, *119*, 1-20.
- JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, *30*, 513-541.
- JACOBY, L. L., TOTH, J. P., & YONELINAS, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*, 139-154.
- KINOSHITA, S. (2001). The role of involuntary aware memory in the implicit stem completion and fragment completion tasks: A selective review. *Psychonomic Bulletin & Review*, *8*, 58-69.
- KINOSHITA, S., & MILLER, M. (2000). The orthographic distinctiveness effect on fragment completion: Not implicit. *Australian Journal of Psychology*, *52*, 63-68.
- LUSTIG, C., & HASHER, L. (2001). Implicit memory is vulnerable to proactive interference. *Psychological Science*, *12*, 408-412.
- MACE, J. H. (2003a). Involuntary aware memory enhances priming on a conceptual implicit memory task. *American Journal of Psychology*, *116*, 281-290.
- MACE, J. H. (2003b). Study-test awareness can enhance priming on an implicit memory task: Evidence from a word completion task. *American Journal of Psychology*, *116*, 257-279.
- MITCHELL, D. B., & BRUSS, P. J. (2003). Age difference in implicit memory: Conceptual, perceptual, or methodological? *Psychology & Aging*, *18*, 807-822.
- MULLIGAN, N. W. (2002). The effects of generation on conceptual implicit memory. *Journal of Memory & Language*, *47*, 327-342.
- MULLIGAN, N. W. (2003). Effects of cross-modal and intramodal division of attention on perceptual implicit memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 262-276.
- MULLIGAN, N. W., GUYER, P. S., & BRELAND, A. (1999). The effects of levels-of-processing and organization on conceptual implicit memory in the category exemplar production test. *Memory & Cognition*, *27*, 633-647.
- MULLIGAN, N. W., & HARTMAN, M. (1996). Divided attention and indirect memory tests. *Memory & Cognition*, *24*, 453-465.
- PILOTTI, M., CHODOROW, M., & TAN, J. (2004). Asymmetric interference in implicit memory: Effects of awareness and stimulus order. *American Journal of Psychology*, *117*, 43-64.
- REINGOLD, E. M., & TOTH, J. P. (1996). Process dissociations versus task dissociations: A controversy in progress. In G. Underwood (Ed.), *Implicit cognition* (pp. 159-202). Oxford: Oxford University Press.
- RICHARDSON-KLAVEHN, A., GARDINER, J. M., & JAVA, R. I. (1994). Involuntary conscious memory and the method of opposition. *Memory*, *2*, 1-29.
- RICHARDSON-KLAVEHN, A., LEE, M. G., JOUBRAN, R., & BJORK, R. A. (1994). Intention and awareness in perceptual identification priming. *Memory & Cognition*, *22*, 293-312.
- ROEDIGER, H. L., III (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*, 1043-1056.
- ROEDIGER, H. L., III, & GERACI, L. (2005). Implicit memory tasks in cognitive research. In A. Wenzel & D. C. Rubin (Eds.), *Cognitive*

- methods and their application to clinical research* (pp. 129-151). Washington, DC: American Psychological Association.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Graham (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63-131). Amsterdam: Elsevier.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.
- SCHACTER, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 501-518.
- SCHACTER, D. L., & BADGAIYAN, R. D. (2001). Neuroimaging of priming: New perspectives on implicit and explicit memory. *Current Directions in Psychological Science*, **10**, 1-4.
- SCHACTER, D. L., BOWERS, J., & BOOKER, J. (1989). Intention, awareness, and implicit memory: The retrieval intentionality criterion. In S. Lewandowsky, J. C. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 47-65). Hillsdale, NJ: Erlbaum.
- TOTH, J. P., REINGOLD, E. M., & JACOBY, L. L. (1994). Toward a redefinition of implicit memory: Process dissociations following elaborative processing and self-generation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 290-303.
- TULVING, E., & SCHACTER, D. L. (1990). Priming and human memory systems. *Science*, **247**, 301-306.
- WARRINGTON, E. K., & WEISKRANTZ, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature*, **217**, 972-974.
- WARRINGTON, E. K., & WEISKRANTZ, L. (1970). Amnesic syndrome: Consolidation or retrieval? *Nature*, **228**, 628-630.

NOTE

1. Indeed, some researchers have modified the standard awareness questionnaire in an attempt to avoid the use of leading questions (e.g., Geraci & Rajaram, 2002; Hamilton & Geraci, 2006).

APPENDIX

Exact Text of the Posttest Questionnaire Used in Experiment 1

The exact wording of the seven questions in the posttest questionnaire used in Experiment 1 is given below. Of course, for Experiment 2, references to the word stem completion task were replaced with references to the category generation task.

1. What do you think was the purpose of the word stem completion task that you just finished?
2. What was your general strategy in completing the word stems?
3. While you were doing the word stem completion task, did you notice any relation between the words that were presented on the screen at the beginning of the experiment and the words you wrote in your booklet?
4. While you were doing the word stem completion task, did you notice whether some of the words you wrote were the same as the words that had been displayed on the screen?
5. If you noticed that you were writing words that had been displayed on the screen, did you simply continue to use the first word that came to your mind or did you try to complete the stems with the words that had been displayed on the screen?
6. If you noticed that you were completing stems with words that had been displayed on the screen, how soon into the word stem completion task did you notice that you were doing so?
7. If you noticed that you were writing words that had been displayed on the screen, did you notice while you were responding with a particular word? If so, what was that word?

(Manuscript received July 11, 2006;
revision accepted for publication May 4, 2007.)