

The effects of domain knowledge on metacomprehension accuracy

THOMAS D. GRIFFIN, BENJAMIN D. JEE, AND JENNIFER WILEY

University of Illinois, Chicago, Illinois

In the present research, we examined the relationship between readers' domain knowledge and their ability to judge their comprehension of novel domain-related material. Participants with varying degrees of baseball knowledge read five texts on baseball-related topics and five texts on non-baseball-related topics, predicted their performance, and completed tests for each text. Baseball knowledge was positively related to absolute accuracy within the baseball domain but was unrelated to relative accuracy within the baseball domain. Also, the readers showed a general underconfidence bias, but the bias was less extreme for higher knowledge readers. The results challenge common assumptions that experts' metacognitive judgments are less accurate than novices'. Results involving topic familiarity ratings and a no-reading control group suggest that higher knowledge readers are not more likely to ignore text-specific cues in favor of a domain familiarity heuristic, but they do appear to make more effective use of domain familiarity in predicting absolute performance levels.

A great deal of formal instruction occurs via reading, and the ability to accurately judge one's level of comprehension from reading—known as *metacomprehension accuracy*—has important consequences for learning from text (Maki & Berry, 1984; Wiley, Griffin, & Thiede, 2005; Winne & Hadwin, 1998). Metacomprehension accuracy can affect learning by contributing to whether one engages in effective study strategies and devotes one's limited attention to where it is most needed (Thiede, Anderson, & Theriault, 2003). Inaccurate metacomprehension can mean overlearning material that is already understood while neglecting other material. Several decades of research reveal that readers are generally poor at self-assessing what they have understood from a text, as is evidenced by low intraindividual correlations between predicted and actual test performance (e.g., Dunlosky & Lipko, 2007; Maki, 1998b; Thiede, Griffin, Wiley, & Redford, 2009). Furthermore, there is typically a great deal of variance around these low averages, with some readers showing near perfect accuracy and others showing the opposite (Griffin, Wiley, & Thiede, 2008; Thiede, Griffin, Wiley, & Anderson, in press). Identifying variables that distinguish readers with high and low accuracy could help explain and improve generally poor levels of accuracy.

One source of variance in metacomprehension accuracy could be domain expertise. Expertise positively predicts performance on many tasks, including text comprehension (e.g., Chi, Glaser, & Farr, 1988; Ericsson & Kintsch, 1995; Feltovich, Prietula, & Ericsson, 2006; Spilich, Vesonder, Chiesi, & Voss, 1979). Also, domain knowledge has been suggested to be positively related to awareness and control of cognitive activities and to engag-

ing in more effective strategies during learning (Baker, 1989; Brown & DeLoache, 1978; Glaser & Chi, 1988; Schneider, 2002). If the effective use of study strategies depends partly on accurate comprehension monitoring, these findings might indirectly reflect a positive impact of expertise on metacomprehension accuracy. Interestingly, a prevailing assumption and explicit claim about metacomprehension accuracy and expertise is that they are negatively related (see Jacoby, Bjork, & Kelley, 1994). The conflict between this claim and the implications of the expertise literature warrants further investigation into the knowledge–metacomprehension relationship.

To begin to address this question, a basic model is outlined of the judgment process and its relation to observed metacomprehension accuracy, along with evidence that informs and supports the model. This is followed by a discussion of the various types and measures of metacomprehension and their relations to the issues raised by the model of judgments. The focus then turns to the empirical evidence for effects of domain knowledge, beginning with a review of prior research on individual differences in test performance (a proxy for knowledge) and their relation to metacomprehension accuracy, followed by an examination of the two studies that have more directly examined the expertise–metacomprehension relationship. The most well known of these latter studies relies on a popular theory that experts experience poorer accuracy than novices, because of a reliance on simple domain familiarity judgments. Thus, the explication of the *domain familiarity hypothesis* receives special attention as a basis for the prediction that expertise may undermine metacomprehension accuracy. Finally, an experiment that provides

for an examination of the expertise–metacomprehension relationship is reported.

A Basic Model of Metacomprehension Judgments and Their Predictive Accuracy

Metacomprehension is a process of inferring one's level of understanding of text via the evaluation of available cues. Accuracy is determined by the extent to which readers can reliably predict their own comprehension of texts, usually operationalized by examining the relation between predictions and actual test performance. A *cue utilization* framework (e.g., Koriat, 1997) posits that people can use a variety of judgment cues for monitoring and that accuracy will depend on whether the cues utilized are directly and strongly tied to the quality of the representation that determines future performance. The cue utilization approach was formulated in reference to memory for lists of paired associates. Such learning, and therefore the monitoring of that learning, differs in important ways from text comprehension (Wiley et al., 2005). However, the basic distinction between cue types and their relation to mental representations and thus to performance is still useful for understanding possible sources of variability in metacomprehension accuracy.

A basic model of the metacomprehension process is outlined in Figure 1. The readers have many possible cues on which to base their judgments of comprehension or predictions of future test performance. Some cues (e.g., topic familiarity or interest, prior knowledge, mood, etc.) are available before, during, or after actually reading the text. Because these cues are available whether or not a text has been read and do not relate to the construction of a particular text representation, we refer to the use of these cues as the *heuristic route* to judgment in Figure 1. In contrast, other cues (e.g., accessibility or coherence of text representation, ability to summarize or explain) that arise from processing or constructing a particular text representation become available only during or after reading. When a reader relies on these cues that reflect a particular text representation, the reader has taken a *representation-based route* to judgment, as is shown in Figure 1.

Readers may take a heuristic route for many reasons; for example, the heuristic route is less effortful and involves salient available cues that can be modestly predictive of test performance. However, the predictive validity of heuristic cues is limited, because they reflect only a subset of influences on text comprehension. Comprehension entails constructing a representation of the text in memory (Kintsch, 1998), so representation-based cues will be more valid indices of actual comprehension. Furthermore, because judgment accuracy is defined as judgment–performance correspondence, judgments based on cues more proximally tied to comprehension performance will be more accurate.

The third basis for comprehension judgments depicted in Figure 1 is the *postdiction route*, which contrasts with the other routes, which constitute predictions. For postdictions, readers can simply use information about their test performance to make their judgments. Even if the post-test judgment prompt asks readers to predict performance on the next test, the readers are apparently still making a postdiction and basing their judgments on prior test performance (Finn & Metcalfe, 2007). If cues that are more proximal to the determinants of performance yield more accurate judgments, postdiction judgments should be the most accurate. Indeed, this is supported by the observed postdiction superiority effect (Glenberg, Sanocki, Epstein, & Morris, 1987; Maki & Serra, 1992; Pierce & Smith, 2001). However, such judgments may not reflect metacognitive monitoring and reflection, because additional external cues become available for postdictions, such as feedback on test performance and the frequency of unanswered test questions. Because of this qualitatively different nature of postdictive judgments, the present focus is on the accuracy of predictive judgments.

The present distinction between heuristic and representation-based routes to judgment is both rooted in and supported by a growing number of recent studies that have shown large improvements in metacomprehension accuracy using manipulations designed to make representation-based cues more salient or available for readers (for a review, see Thiede et al., 2009). For example, accuracy can be substantially improved through tasks

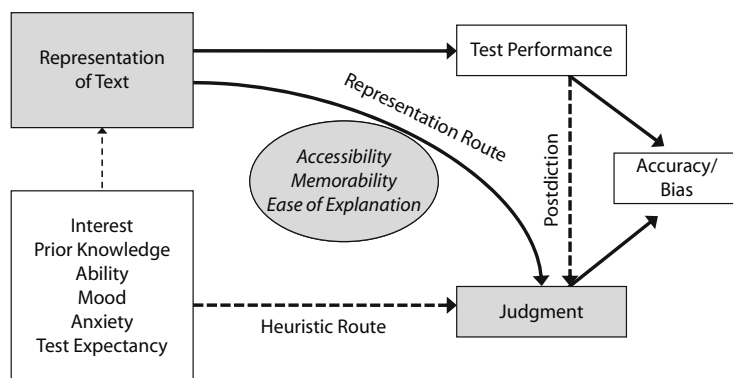


Figure 1. Potential routes to judgment.

that afford or even require readers to access or utilize their text representations, such as rereading, self-explanation, concept mapping, or delayed summarization (e.g., Dunlosky & Rawson, 2005; Griffin et al., 2008; Rawson, Dunlosky, & Thiede, 2000; Thiede, Dunlosky, Griffin, & Wiley, 2005; Thiede et al., in press). Although discussion of these effects has focused specifically on the situation-model level of text representation, the findings are consistent with the more general idea that the more valid cues for judgments are those tied to the text representation that will be assessed at testing. All of these findings suggesting the importance of a representation-based route to judgment have used the same measure of accuracy—namely, relative accuracy. Because expertise could plausibly relate differently to the various measures of metacomprehension, a detailed discussion of these measures follows.

Common Measures of Metacomprehension

Various measures of metacomprehension compare predictions with actual performance, but they do so in different ways. Three independent measures of judgment–performance relations have been employed: absolute accuracy, confidence bias, and relative accuracy (Maki, 1998a). *Absolute accuracy*, also referred to as *calibration*, is the mean absolute (or mean squared) deviation between judged and actual performance. A related measure, *confidence bias*, concerns the directionality of these deviations, computed as the signed difference between mean judgments and mean performance. This measure, sometimes referred to as *over-/underconfidence*, does not reflect the level of accuracy in terms of the number or magnitude of judgment errors but, rather, of whether those judgment errors are systematically biased in one direction (Yates, 1990).¹

Absolute accuracy and confidence bias are statistically dependent on marginal mean performance levels, which can allow for nonmetacognitive influences of performance levels on accuracy measures (for a longer discussion, see Nelson, 1984). Therefore, a third measure, *relative accuracy*, or *resolution*, has become more common. Relative accuracy quantifies a participant's accuracy in predicting performance on one text relative to other texts in terms of an intraindividual correlation between performance predictions and test performance, such as Pearson's r or the Goodman–Kruskal γ (Nelson, 1996). Such a measure is not dependent on mean performance but can still be artificially constrained by low within-person variance in performance.

Figure 1 illustrates the point that test performance can have a direct impact on computed measures of monitoring accuracy, independent from the effects of the judgment process itself. Performance effects may underlie the repeatedly observed underconfidence-with-practice effect, whereby people increasingly underestimate their performance after successive trials of study and testing (Koriat & Bjork, 2006). Close examination of the data shows that learners actually increase in their confidence judgments with more practice. However, because performance scores increase at a faster rate, the deviation between judgments and performance (i.e., underconfidence) also increases (see Koriat & Bjork, 2006, Experiment 1; Koriat, Shef-

fer, & Ma'ayan, 2002, especially Figures 1–4, 7, and 8). Such a divergence could be observed without any actual impact of practice on the metacognitive processes underlying judgments. Because domain knowledge can affect test performance, metacognitive effects of expertise cannot be inferred from simple effects on computed metacomprehension accuracy unless performance effects are controlled for.

The three measures of metacomprehension are conceptually and statistically distinct and can give conflicting impressions of metacomprehension accuracy (e.g., Kelemen, Frost, & Weaver, 2000; Maki, Shields, Wheeler, & Zaccchilli, 2005). People can have perfect relative accuracy in their belief that they understand their physics course better than their biology course but could still have low absolute accuracy in believing that they understand almost everything in both courses when they actually understand very little in either. Alternatively, having superior absolute accuracy does not guarantee superior relative accuracy. Different accuracy measures could also be differentially affected by different cue use. Heuristics like general domain familiarity could provide a modestly useful basis for predicting the absolute level of performance in the domain but would be useless in predicting relative differences in performance within the same domain. Thus, any variable that affects cue use can produce very different effects on absolute and relative accuracy.

Possible Effects of Domain Knowledge on Metacomprehension

The main question to be considered here is whether individual differences in domain knowledge affect the metacomprehension process. There are several theoretical reasons to expect a positive effect of domain knowledge. In problem solving contexts, experts show better estimates of progress toward solution and more efficient learning strategies (e.g., question asking, answer checking), speculated to result from the knowledge of successful solutions in the domain that can be used as a reference point (Brown & DeLoache, 1978; Glaser & Chi, 1988). Wiley et al. (2005) argued that having a grasp of what it means to fully comprehend a text (i.e., a valid point of reference) could be critical for metacomprehension accuracy. Experts might have a better sense of what text comprehension is within their domain, giving them a valid standard and absolute reference point that might be especially critical for the absolute accuracy of judgments.

A concurrent processing framework for metacomprehension provides an additional basis to predict a positive effect of expertise. The metacognitive monitoring construct is traditionally defined in terms of two levels of information processing: The *object-level* processing of the text or object of study and the *meta-level* processing of the cues that reflect the quality of the resulting representation at the object level (Fischer & Mandl, 1984; Nelson & Narens, 1990). Griffin et al. (2008) recently emphasized that effective monitoring of text comprehension may require that readers attend to both the object and meta levels *concurrently*, encoding monitoring cues while actively constructing a text representation.

Griffin et al. (2008) reinterpreted the rereading effect using this concurrent processing framework in a manner directly related to the representation versus heuristic cues distinction. After a single reading, readers with lower reading ability or working memory capacity had significantly worse accuracy, but their accuracy was especially improved after rereading. Griffin et al. argued that this reading \times ability interaction arises because lower ability readers do not have access to certain valid cues related to their text representation (i.e., representation cues), because they lack the resources to attend to these cues during the reading process. This forces lower ability readers to rely more heavily on cues that are available independent of the reading process (i.e., heuristic cues). During a second reading, the attention demands of basic text processing are reduced, so lower working memory capacity/ability readers can concurrently attend to more of the available representation cues.

This concurrent processing perspective is especially relevant in light of previous work showing that domain knowledge can improve working memory capacity for domain-related material (Fincher-Kiefer, Post, Greene, & Voss, 1988) and can compensate for low reading ability (Adams, Bell, & Perfetti, 1995; Recht & Leslie, 1988) or deficits in general abilities (Hambrick & Engle, 2002; Schneider, K rkel, & Weinert, 1989). Prior knowledge might reduce the attention resources needed for basic text processing (Miller, Cohen, & Wingfield, 2006), thus increasing attention to representation-related cues. However, this prediction depends on experts actually directing any added resources toward the task of monitoring.

Since domain-related comprehension generally improves with expertise (e.g., Ericsson & Kintsch, 1995; Miller, Stine-Morrow, Kirkorian, & Conroy, 2004; Spilich et al., 1979; Wiley, 2005), existing empirical findings relating levels of test performance and metacomprehension accuracy could be used to predict the effects of domain knowledge. However, only studies involving performance within a particular domain are relevant to the issue of domain knowledge. Hacker and colleagues repeatedly found that students with better exam performance also more accurately predict their absolute magnitude of performance (Bol & Hacker, 2001; Bol, Hacker, O'Shea, & Allen, 2005; Hacker, Bol, & Bahbahani, 2008; Hacker, Bol, Horgan, & Rakow, 2000). Glover (1989) and Pressley, Snyder, Levin, Murray, and Ghatala (1987) observed findings consistent with these results. Several of these researchers reported that better performing students also show less of a confidence bias. The results for relative accuracy are less clear, with findings equally split between a positive relationship with test performance and no relationship (Maki & Berry, 1984; Maki, Jonas, & Kallod, 1994). Individual differences in test performance are only partially constrained by prior knowledge, so these findings are only suggestive of the potential knowledge–metacomprehension relationship.

Tests of the Domain Knowledge–Metacomprehension Relationship

There appear to be only two studies in which the researchers attempted to directly examine the relation be-

tween prior knowledge and metacognitive accuracy for learning from text. In one study, Schneider et al. (1989) found that soccer knowledge positively predicted the absolute accuracy of judgments about learning soccer-related text. However, only absolute accuracy was measured, and the effects of performance differences were not controlled for.

In a second study, Glenberg and Epstein (1987) examined the relative accuracy of comprehension judgments among readers with varying degrees of knowledge about the topics of two sets of texts. The participants read several texts on music and several texts on physics, provided an estimate of their confidence in their ability to answer a text-related question after reading each text, and then took a single-item test about each text. Each participant's expertise was assessed by determining the number of college courses that they had taken in music and physics. Glenberg and Epstein concluded that the relationship between domain knowledge and metacomprehension accuracy was negative.

There are also a number of issues with Glenberg and Epstein's (1987) study, including mixed results across domains and several methodological issues that limit the general conclusions that can be drawn from it. Three main issues are noted here. First, although a weak negative relationship between metacomprehension accuracy and expertise was found for the physics domain, the relationship in the music domain was nonexistent. Second, Glenberg and Epstein did not actually report the key test of the negative relation—that is, whether the coefficient for expertise on metacomprehension of the physics texts was significantly lower than zero. Instead, they tested an interaction concerning whether the physics expertise coefficient for the physics texts differed from the slightly positive coefficient for the music texts. The reported analysis tells us that physics expertise related to metacomprehension differently depending on the domain of the text, but it does not directly test whether experts had worse metacomprehension than nonexperts for the physics texts. An additional problem for interpretation was the failure to control for differences in variance in test performance.²

A third issue is that Glenberg and Epstein's (1987) measure of comprehension was a single true/false test item and the overall metacomprehension accuracy in this study was not significantly different from zero. Reliability problems due to assessing comprehension with a single test question have been previously discussed in the metacomprehension literature (Glenberg et al., 1987; Maki, 1998b; Maki & Serra, 1992; Weaver, 1990; Wiley et al., 2005). Poor reliability is a particular problem when attempting to find relations with individual differences (Kelemen et al., 2000; Maki et al., 2005).

Methodological problems aside, Glenberg and Epstein (1987) proposed a self-classification account to explain the reported result. They suggested that higher knowledge individuals experienced poorer accuracy because, rather than assessing their actual comprehension of the specific texts within the domain, they relied on simple familiarity with the domain in general. Relating this to the present framework, experts supposedly employ a familiarity heu-

ristic route to judgment rather than a text-representation route, which prevents them from making accurate relative judgments among texts within the same domain. The two key assumptions of this domain familiarity hypothesis are that domain familiarity is a major determinant of comprehension judgments and that domain experts are especially prone to rely on domain familiarity heuristics and will thus suffer in their within-domain accuracy.

The first assumption—that readers rely on topic familiarity as a heuristic for metacomprehension judgments—has been explored, yielding mixed results. Some studies have supported the hypothesis, showing that judgments can be highly correlated ($r_s = .62-.70$) with ratings of topic interest (Lin, Zabrocky, & Moore, 1997) or familiarity (Glenberg et al., 1987). However, Maki and Serra (1992) found that postreading predictions of test performance had higher relative accuracy than did prereading judgments based only on the topics, indicating that readers do take some text-specific information into account. In another study, Maki (1998b) found that topic familiarity ratings and comprehension judgments were related at only $r = .15$, suggesting that familiarity was not a common or strong influence on judgments. Regardless, the most critical assumption—that experts are especially reliant on a familiarity heuristic—has not been previously examined, and it is not theoretically obvious why experts would rely on their topic familiarity but novices would not rely on their unfamiliarity.

In summary, this often cited study does not provide adequate evidence for claims of experts' poorer metacomprehension or reliance on domain familiarity. Yet this result is often relied on in support of these conclusions. In fact, a widely distributed report by the National Academies (Jacoby et al., 1994, p. 70) cites only this one article to support the claim that "because of reliance on prior knowledge when judging comprehension, an expert is even more subject to illusions of comprehension than is a novice."

Overview of the Present Study

There appear to be some theoretical reasons and some indirect empirical support for expecting a positive rather than negative relationship between expertise and metacomprehension accuracy, but no clear conclusions about this relationship can be drawn from the existing literature. In this study, the particular topic used to explore the effects of domain knowledge on metacomprehension was baseball. Baseball (like other sports-related topics) is often chosen as a domain for examining the effects of knowledge on cognitive processing, because baseball expertise is less likely to correlate strongly with general cognitive abilities than is expertise in academic domains (Voss, Vesonder, & Spilich, 1980). Moreover, the baseball domain is likely to evidence a desirable range and distribution of expertise across ability levels in normal U.S. samples (Hambrick & Engle, 2002).

In the present study, participants were presented with five baseball-related texts. After reading each text, the participants judged how many items they would correctly answer on a five-item test and then took the tests. Predictive judgments and performance were on the same scale,

which allowed for the computation of both absolute accuracy and confidence bias. In addition, topic familiarity ratings were collected to more directly evaluate the assumptions underlying the domain familiarity hypothesis and its relation to expertise.

The participants also completed a standard metacomprehension paradigm using a set of texts on various non-baseball-related topics. These data were collected so that any spurious relations between baseball knowledge (BK) and general metacomprehension skill could be removed from regression analyses and so that the unique effects of knowledge on domain-related metacomprehension processes could be investigated. Similarly, the direct influences of test performance and variance were removed, so that the effects of domain knowledge on the judgment process could be isolated.

As is detailed above, previous findings related to individual differences in test performance, along with the traditionally positive effects of expertise on performance, suggest that comprehension monitoring (especially absolute accuracy) could be enhanced by domain knowledge. Also, if prior knowledge operates in the same fashion as general reading ability or working memory capacity, the concurrent processing framework predicts that experts may have better relative accuracy because of greater attention to representation-based cues. However, if the domain familiarity hypothesis is correct, experts should be *more* likely to use a heuristic route to judgment, leading to worse relative accuracy within a domain.

METHOD

Participants

In exchange for course credit, 131 undergraduates at the University of Illinois at Chicago participated in this experiment as part of an introductory psychology participant pool. Relative accuracy cannot be computed for participants who lack text-to-text variance in either their judgments or performance. Thus, data had to be eliminated for 20 participants who lacked variability on either of these measures. Data for an additional 7 participants were lost because of incomplete data, as were those for 3 other participants because of equipment failures, leaving usable data from 101 participants (53 male and 48 female).

Materials

Five texts on baseball topics were written for this study (about 400 words each). The baseball texts were on bunting strategy, corking bats, a new batting average statistic, the crack of the bat, and curveballs. For each text, five short-answer test items were constructed, with the intention that at least one would be relatively easy for novices (a gist question or a question about a specific fact mentioned in the text), whereas several questions required inferences from the text. It is because of these inference questions that accurate monitoring of performance in this paradigm is referred to as *metacomprehension accuracy*. At the same time, several questions were designed to test text-specific information, so that high-knowledge participants could not correctly answer all of the baseball-related items without comprehending the texts. An example text and test is included in the Appendix.

Five texts on non-baseball-related topics were written that matched the baseball texts in length and format. The purpose of including a set of non-baseball texts was to obtain a measure of general metacomprehension skill using the standard paradigm that could be entered in regressions to ensure that any observed relation-

ships with BK were not the result of spurious correlations between BK and general skill. The non-baseball texts created for this study were on the Irish potato famine, cell division, dinosaur extinction, electric cars, and heart disease. Five short-answer test items were constructed that also included some detail questions, as well as questions that required inferences from each text.

In addition, a practice text and test on stalactites and stalagmites was created in similar length and format to the experimental texts and tests.

Text comprehension measures. A scoring key was devised by the experimenters, which specified acceptable responses that were equivalent in meaning for each short answer question on the baseball and non-baseball comprehension tests. Each question was worth a point and was scored as a 1 or 0. Two independent coders scored a subset of 1,000 responses to establish intercoder reliability. Using Cohen's κ as an index, agreement between coders was found to be acceptable ($\kappa = .82, p < .001$). The two coders resolved scoring discrepancies through discussion and refinement of the scoring key. The remaining responses were divided between the two scorers. The total scores from the baseball and non-baseball tests were used as measures of baseball comprehension and non-baseball comprehension.

BK. BK was assessed with a 45-item BK questionnaire (Spilich et al., 1979). The questionnaire items did not overlap with the text content. Scores ranged from 1 to 42 (out of 45 possible), and the reliability of this assessment was very high (Cronbach's $\alpha = .96$). BK did differ by gender, with males scoring significantly higher ($M = 24.9, SD = 11.4$) than females ($M = 10.6, SD = 10.5$) [$t(99) = 6.51, p < .05$]. To rule out gender as a factor, all regression analyses reported below were also run entering gender as a control variable and the results did not change. Also, BK significantly predicted comprehension of the baseball texts ($r = .62, p < .05$) but not comprehension of the non-baseball texts ($r = .18, n.s.$), which demonstrates the validity of the measure.

Procedure

The participants were tested in small groups (≤ 10) on individual computers. In the general instructions, the participants were informed that they would read several texts at their own pace as if studying for an exam. Rereading was allowed, but once the readers finished a text and advanced the screen, they could not go back. The participants were informed that they would have to make a judgment about the number of questions (0–5) that they would be able to answer about each text and that they would answer questions about the texts after reading them.

After reading the general instructions, the participants were given a practice text, judged their comprehension, and then completed a five-item test. The baseball-related and non-baseball-related texts were then presented in different blocks and in two different text orders. The block and text orders were counterbalanced across participants. The judgment task was presented immediately after each text. After reading and making judgments for all five texts, the participant completed the comprehension tests in the order of text presentation. Finally, the participants rated their prior familiarity with each text topic (prompted by the text titles) on a scale from 1 (*not familiar at all*) to 6 (*extremely familiar*) and completed the paper-and-pencil BK questionnaire.

Additional No-Reading Control Condition

The self-classification account essentially presumes that it is merely the familiarity of the domain and not consideration of the content of the texts that drives higher knowledge individuals' judgments and thus affects their accuracy. To further examine this hypothesis, we had 66 undergraduates participate in a no-reading control condition for the baseball texts following the logic of Maki and Serra (1992). The data from 8 participants were eliminated because relative measures could not be computed, leaving usable data from 58 participants. In this condition, the participants judged their expected performance on the baseball-related texts *without actually*

reading the texts and then completed the same tests as in the reading condition. They completed a paper-and-pencil BK questionnaire at the end of the session. In the judgment phase, the participants were given only the topic of the text (e.g., bunting strategy, corking bats, etc.). This no-reading condition allowed us to test whether the participants' judgments, performance, and metacomprehension accuracy were influenced by actually reading the texts. Mean prior knowledge scores for the reading condition ($M = 18.1, SD = 13.1$) and the no-reading condition ($M = 16.0, SD = 11.7$) did not significantly differ [$t(99) < 1.0, n.s.$].

RESULTS

Two main questions are addressed here. First, the effects of BK are examined in relation to the three measures of within-domain metacomprehension for baseball-related texts: absolute accuracy, confidence bias, and relative accuracy. Second, several tests of the domain familiarity hypothesis and whether it provides a compelling account for the judgment process are reported.

Preliminary analyses revealed that the two measures used to compute metacomprehension—judgments and performance—had adequate range and variance and did not suffer from ceiling or floor effects.³ Correlations among these measures and the measures used in the regressions are reported in Table 1. As was noted previously, test performance can directly constrain computed accuracy. To isolate the specific effects of knowledge on domain-specific metacomprehension judgment processes, the effects of mean test performance, test variance, and metacomprehension accuracy on the non-baseball texts were entered first as covariates in regressions, so any unique effects of domain knowledge on judgments could be assessed.

Effects of BK on Baseball Metacomprehension

The three metacomprehension measures (absolute accuracy, confidence bias, and relative accuracy) were computed using the standard procedures previously discussed. The reliability and validity of these measures was established using procedures adapted from Maki et al. (1994), in which an intersubsets alpha (rather than interitem) is computed.⁴ For all three metacomprehension measures, $\alpha > .91$.

Absolute metacomprehension accuracy. The mean absolute deviation between predictions and actual performance on each test was 1.22 ($SD = 0.54$). In order to create a metric in which greater values would indicate greater accuracy, the mean absolute deviations were subtracted from zero. Table 2A presents the results of the hierarchical regression analysis predicting absolute accuracy on the baseball-related texts. BK was a significant predictor of absolute accuracy, accounting for 13% of the variance. The main result of this analysis was that prior knowledge led to less absolute deviation and more accurate predictive absolute accuracy, not less.

The control variables of average magnitude and variance in baseball test performance were both significant negative predictors of absolute accuracy, but the beta weight for BK was still significant, if not greater, after controlling for these influences. There was no significant

Table 1
Pearson Correlations Among Baseball Knowledge (BK), Comprehension Judgments, Mean Test Performance, Metacomprehension Measures, and Familiarity Ratings for Baseball Tests ($N = 101$)

Measure	1	2	3	4	5	6	7	8	9
1. BK	—	.60*	-.03	.62*	-.23*	.32*	.18	-.07	.80*
2. Mean judgment		—	-.06	.55*	-.19	.56*	.71*	-.14	.57*
3. Judge variance†			—	-.16	-.05	.08	.07	.25*	-.06
4. Mean test score				—	-.21*	.05	-.21*	-.14	.51*
5. Test variance†					—	.39*	-.04	.20*	-.31*
6. Absolute accuracy						—	.61*	.06	.34*
7. Confidence bias							—	-.04	.26*
8. Relative accuracy								—	-.09
9. Mean familiarity									—

†These measures of variance represent within-person variability across the five baseball texts. * $p < .05$.

effect for the control variable of absolute accuracy on the non-baseball texts. This result replicates the results of Kelemen et al. (2000), who found little general correlation for absolute accuracy across different metacognitive tasks and sets of materials.

Confidence bias. The mean judgment minus performance difference score on the baseball texts was significantly lower than zero [$M = -0.62$, $SD = 0.94$; $t(100) = 6.62$, $p < .05$], indicating that the participants tended to underestimate their performance on the baseball tests. A hierarchical regression analysis indicated that BK was a significant predictor of confidence bias, accounting for 18% of the variance. The results are presented in Table 3A. The positive beta for BK must be interpreted in relation to the overall underestimation and negative average bias score of the sample. As BK increased, the computed bias score increased, meaning that the negative scores got smaller and closer to zero. This indicates less underestimation (less of a confidence bias) with increasing domain knowledge.

The results for the control variables showed that the participants who performed better on the baseball tests were more likely to underestimate their performance (more negative bias scores). As with absolute accuracy, this reflects the unique and direct effect of performance levels on the monitoring measure that is independent of the judgment process and how that process is influenced by domain knowledge. Also, bias on the non-baseball texts was a positive predictor of confidence bias on the baseball texts, accounting for about 18% of the variance. This result is similar to that of Kelemen et al. (2000), who also found that individual differences in confidence bias correlated across tasks and materials.

Relative metacomprehension accuracy. The mean intraindividual judgment–performance coefficient was significantly greater than zero [$M = .13$; $t(101) = 2.46$, $p < .05$]. As is presented in Table 4A, the results of the hierarchical regression analysis predicting relative accuracy on the baseball-related texts reveal that BK was not a significant predictor. In fact, the coefficient is virtually zero and slightly positive after entering the control variables. Relative accuracy for non-baseball texts did not predict relative accuracy on the baseball-related texts. This is again consistent with Kelemen et al. (2000), who did not

find significant correlations in relative metacomprehension accuracy across different sets of texts.

Contrary to widespread presumption, we found no evidence of a negative relationship between domain knowledge and relative metacomprehension accuracy within the domain.

Tests of the Domain Familiarity Hypothesis

There are two basic assumptions of the domain familiarity hypothesis as originally proposed by Glenberg and Epstein (1987). The first is that participants may rely on a topic familiarity heuristic when computing test predictions and fail to consider their actual comprehension for a particular text. The second is that experts will be more likely to use such a strategy than will nonexperts. These assumptions are tested first by examining whether reading the specific texts influenced judgments and accuracy, and second by examining whether readers' familiarity ratings for the various topics strongly correlate with their comprehension judgments for the specific texts (Glenberg & Epstein, 1987; Glenberg et al., 1987; Lin et al., 1997; Maki, 1998b; Maki & Serra, 1992).

Effects of reading on comprehension judgments.

The domain familiarity hypothesis implies that cues derived from actually reading a text are not taken into account, especially by experts. As a test of this hypothesis, the previously analyzed judgments of the reading group were compared with the no-reading control group, who judged and were tested on the baseball topics without reading the texts. The readers had better test performance ($M = 3.25$, $SD = 0.80$) than did the nonreaders ($M = 2.52$, $SD = 0.61$) [$t(157) = 5.97$, $p < .01$]. More important, reading affected judgments, and the readers gave significantly higher judgments ($M = 2.63$, $SD = 1.11$) than did the nonreaders ($M = 1.90$, $SD = 1.22$) [$t(157) = 3.80$, $p < .01$].

Two hierarchical regressions that controlled for performance effects and used BK, reading condition, and their interaction (see Tables 2B and 4B) as predictors showed that reading the texts increased both the absolute and relative accuracy of judgments, but it did not interact with BK. Compared with the nonreaders, the readers more accurately predicted their average test performance and the variability in their test performance. Thus, contrary to the

Table 2
Regression Analyses Predicting Absolute
Metacomprehension Accuracy on Baseball Texts

Criterion Variable	Inc. R^2	F Value	β	t Value	pr^2
(A) Read Condition					
Step 1	.10	11.47*			
BK			.32	3.39*	.10
Step 2	.16	6.84*			
BK			.43	3.77*	.13
Non-baseball text absolute accuracy			-.06	-0.67	.00
Average baseball test performance			-.285	-2.53*	.06
Variance on baseball tests			-.362	-3.96*	.14
(B) Reading Effects on Absolute Accuracy					
Step 1	.25	12.85*			
BK			.34	3.68*	.08
Reading condition			.29	3.63*	.08
Average baseball test performance			-.26	2.56*	.04
Variance on baseball tests			-.31	-4.22*	.01
Step 2	.01	1.51			
BK \times reading condition interaction			.15	1.23	.01

Note—Inc. R^2 , increment in variance accounted for; β , standardized regression coefficient; pr^2 , squared partial correlation; BK, baseball knowledge. * $p < .05$.

domain familiarity hypothesis, the participants made use of cues tied to reading the specific texts, and the lack of significant interaction terms means that high-knowledge readers did so to an equal degree. The lack of an interaction for absolute accuracy also shows that the greater accuracy of higher knowledge readers would have existed even without reading the texts.

A third hierarchical regression predicted confidence bias using BK, reading condition, and their interaction. As is shown in Table 3B, although reading condition was not itself a significant predictor of confidence bias, there was a significant BK \times reading condition interaction. Reading the text reduced confidence bias scores more for the higher knowledge than for the lower knowledge participants. Reading did not have a notable effect on the lowest knowledge participants (BK score below the 25th percentile), who were highly underconfident regardless of having read the texts ($M = -0.76$, $SD = 1.12$) or not ($M = -0.82$, $SD = 0.70$). In contrast, reading had a sizable effect on the highest knowledge participants (above the 75th percentile), who were moderately underconfident if they read the texts ($M = -0.38$, $SD = 0.66$) but slightly overconfident without reading ($M = 0.20$, $SD = 0.72$). Because test performance effects were controlled for, this result means that reading the texts impacted the judgment process of the higher knowledge readers and resulted in nearly a full standard deviation reduction in confidence bias.

Correlations of familiarity ratings with predictive judgments. Table 1 shows that the average magnitude of comprehension judgments was significantly correlated with average familiarity ratings for the five baseball texts, suggesting that the readers may have used familiarity as a partial basis for their comprehension judgments. However, average baseball test performance (an index of text comprehension) was also significantly correlated with both familiarity ratings and average comprehension judgments. When domain familiarity and average test performance

were simultaneously regressed onto judgments, they both remained significant predictors (β s = .38 and .39, respectively) [$ts(95) > 4.20$, $ps < .0001$]. This suggests that the readers may have relied partly on domain-related classifications to anchor their judgment magnitudes, but they were also sensitive to other cues tied to their actual comprehension.

These interindividual correlations may merely reflect that people who generally have more knowledge have generally greater confidence. If the participants actually used variability in their prior familiarity with various topics as the basis for the judgments for each text, there should be a significant intraindividual correlation between their topic familiarity ratings and corresponding comprehension judgments. Contrary to this hypothesis, this mean intraindividual correlation was not significantly different from zero ($M = .07$, $SD = .48$) [$t(78) = 1.39$, n.s.], showing that the use of familiarity as a basis for comprehension judgments was generally limited.

Of course, there was variability among the readers' reliance on topic familiarity, and each participant's intraindividual familiarity-judgment coefficient is an index of that reliance on a familiarity heuristic. This index was positively related to absolute accuracy ($r = .25$, $p < .03$) but was unrelated to relative accuracy within the domain ($r = .09$, n.s.). This finding supports the speculation that the use of heuristic cues like topic familiarity is more likely to be effective for absolute than for relative accuracy. Critically, this index of reliance on topic familiarity for the baseball texts was unrelated to BK ($r = .04$, n.s.).

These results converge with the results of the no-reading analyses in suggesting that experts and novices make similarly limited use of prior familiarity during judgment and are sensitive to some text-specific cues that arise from reading the text. Contrary to the domain familiarity hypothesis, the experts may have been *more*, not less, sensitive to representation-based cues, as is indicated by

Table 3
Regression Analyses Predicting Confidence Bias on Baseball Texts

Criterion Variable	Inc. R^2	F Value	β	t Value	pr^2
(A) Read Condition					
Step 1	.03	3.27			
BK			.18	1.81	.03
Step 2	.30	14.60*			
BK			.50	4.69*	.18
Non-baseball text bias			.37	4.48*	.18
Average baseball test performance			-.51	4.80*	.19
Variance on baseball tests			-.04	0.49	.00
(B) Reading Effects on Confidence Bias					
Step 1	.19	8.91*			
BK			.56	5.84*	.18
Reading condition			.17	2.00	.03
Average baseball test performance			-.48	4.60*	.12
Variance on baseball tests			.03	0.44	.00
Step 2	.02	3.82*			
BK \times reading condition interaction			-.25	1.96*	.03

Note—Inc. R^2 , increment in variance accounted for; β , standardized regression coefficient; pr^2 , squared partial correlation; BK, baseball knowledge. * $p < .05$.

the greater effect of reading on their level of confidence bias. The experts did not make more use of a familiarity heuristic, and yet they did have greater absolute accuracy, even without reading the texts, when only heuristic cues were available. This could suggest that the experts made equally limited but more effective use of heuristic cues.

DISCUSSION

The purpose of this study was to reexamine the relationship between domain knowledge and metacomprehension accuracy, and to assess the validity of the common presumption that greater knowledge in a domain harms one's ability to judge their comprehension of domain-related information. Contrary to this presumption, expertise was unrelated to the accuracy of relative judgments within a domain. Furthermore, expertise was *positively* related to absolute accuracy, meaning that higher knowledge readers were better able to predict their average performance on the set of domain-related texts.

Claims about a negative relation between expertise and metacomprehension accuracy are usually rooted in the assumptions of the domain familiarity hypothesis (Glenberg & Epstein, 1987; Jacoby et al., 1994)—namely, that readers rely on domain familiarity as the basis for judgments of individual texts and that this is acutely true among domain experts. However, the numerous analyses in the present study found only limited support for the first assumption and no support for the second. Judgment variability both between and within readers was only weakly related to the variability in ratings of prior familiarity. Also, comparison with a no-reading condition showed that the readers were sensitive to text-specific cues and used these cues to improve both their relative and absolute accuracy. The greater effect of reading on the confidence bias of experts suggests that they may have been even more sensitive than novices to representation-based cues. The evidence fails to support and sometimes contradicts the idea that readers

with more domain knowledge rely on a domain familiarity heuristic more than novices do.

The results suggest that experts made similarly limited but perhaps more effective use of cues related to domain familiarity than did novices, which accounts for why their judgments were more accurate in absolute magnitude without being any less accurate in predicting relative performance within a domain. Furthermore, the failure to observe a negative expertise–metacomprehension relation cannot be attributed to mere issues of measurement, reliability, or sampling error, because the reliability of the key measures was directly assessed, and the measures showed the expected relationships to other variables.

The present analyses also highlight the methodological importance of controlling for direct performance effects on accuracy measures. Variance in performance was related to both absolute and relative accuracy. Also, average test performance had an independent negative effect on absolute accuracy (as is revealed in the regression results), but this relationship was hidden in the nonexistent bivariate correlation (Table 1), because of the positive correlation of performance with domain knowledge, which was in turn positively related to absolute accuracy. Had the real positive effect of domain knowledge been weaker, it might have been masked by the opposing negative effect of test performance. In addition, because in the present analyses, we controlled for the direct effects of test performance and variance on metacomprehension measures, used an independent measure of domain knowledge, and controlled for general metacomprehension skills, this is the first study that clearly speaks to the unique effects of prior domain knowledge on metacomprehension judgments.

The present findings do more than just undermine a popular claim about the negative effects of expertise on metacomprehension and the associated domain familiarity hypothesis. On a more positive note, the superior absolute and across-domain accuracy for high-knowledge readers supports speculations in the classic expertise literature

Table 4
Regression Analyses Predicting Relative
Metacomprehension Accuracy (Pearson's r)

Criterion Variable	Inc. R^2	F Value	β	t Value	pr^2
(A) Within-Domain Relative Accuracy: Read Condition					
Step 1	.01	0.45			
BK			-.07	0.67	.00
Step 2	.04	2.60			
BK			.03	0.26	.00
Non-baseball relative accuracy			.17	1.69	.03
Average baseball test performance			-.16	1.26	.02
Variance on baseball tests			.19	1.89	.04
(B) Reading Effects on Within-Domain Relative Accuracy					
Step 1	.11	4.54*			
BK			.06	0.63	.00
Reading condition			.35	3.93*	.09
Average baseball test performance			-.18	1.60	.02
Variance on baseball tests			.15	1.89	.02
Step 2	.00	0.02			
BK \times reading condition interaction			.02	0.89	.00

Note—Inc. R^2 , increment in variance accounted for; β , standardized coefficient; pr^2 , squared partial correlation; BK, baseball knowledge. * $p < .05$.

that domain knowledge promotes better monitoring (Glaeser & Chi, 1988) and recent claims in the metacomprehension literature that accuracy may improve with greater knowledge about what it means to fully comprehend a text (Wiley et al., 2005). A valid reference point for the domain would not only aid absolute judgments, but would also aid judgments of the domain relative to others. Unlike the domain familiarity hypothesis, this interpretation does not make the unsupported prediction of a negative effect for within-domain relative accuracy.

The present findings are also consistent with current assumptions about cue validity and differing routes to judgment presented in Figure 1. Only the participants who actually read the texts had access to cues related to their text representation, whereas the nonreaders were forced to rely solely on heuristic cues. The observed superior judgment accuracy of the readers over the nonreaders supports the notion that a representation route gives readers access to more valid and predictive cues than a heuristic route does. This support for the importance of the representation route converges with the results of an increasing number of studies showing that relative metacomprehension accuracy improves with increased access to cues tied to one's situation-model representation (Dunlosky & Rawson, 2005; Griffin et al., 2008; Rawson et al., 2000; Thiede et al., 2005). In addition, improved accuracy without improvements in test performance was found in all of these prior studies. A similar independence was found in the present results, with expertise improving test performance without affecting relative accuracy. Because test performance reflects the quality of the text representation that was constructed, these various results converge to support a recently emphasized distinction (Griffin et al., 2008) that access to valid cues about a text representation is not dependent on the quality of the representation that a reader constructs, but rather on the extent to which the reader merely attempts to construct or access a representation.

On a related note, in the present study, attention to the meta level did not seem to vary as a function of topic knowledge in the same way that it was found in a previous study to vary as a function of ability differences (Griffin et al., 2008). Of course, domain knowledge and general ability are distinct constructs, and it is likely that some features that they do not share (perhaps interest in the topic of the text) could lead high-knowledge and high-ability readers to utilize their additional processing resources differently. To the extent that better test performance reflects higher quality text representations, it is possible that these high-knowledge readers may have reinvested any freed resources afforded by their prior knowledge into focusing more deeply on the text ideas (i.e., the object level) rather than into monitoring representation cues during reading (i.e., the meta level). This would reconcile these findings with the combined implications of the concurrent-processing framework of monitoring (Griffin et al., 2008) and the previously observed relationships between prior knowledge and text processing efficiency (Adams et al., 1995; Fincher-Kiefer et al., 1988).

A final note is that these results are consistent with many previous reports of dissociations between the effects of manipulations or individual differences on measures of absolute and relative accuracy (Kelemen et al., 2000; Maki, 1998a; Maki et al., 2005). Going beyond these prior observations, the present findings show that cue use can impact these conceptually distinct types of accuracy differently with heuristic familiarity cues being more useful for absolute accuracy. Of course, the practical goal is to find contexts that improve monitoring accuracy in both respects, which should lead to the most effective studying behaviors, where students can correctly judge not only which subject, but also which topics within that subject, are most critical for them to review. From this perspective, further investigation of how prior knowledge may be leveraged to improve both absolute and relative accuracy

and discovery of mechanisms common to the accuracy of both types would be quite interesting.

In conclusion, there is no convincing evidence that higher knowledge individuals become subject to an illusion of comprehension as a function of their expertise. Higher knowledge individuals not only achieve better comprehension of domain-related texts, but they also estimate their performance on domain-related tests more precisely and are less biased in their estimates. Our findings are inconsistent with the notion that higher knowledge individuals rely solely on simplistic self-classifications as experts in judging their comprehension, and to the extent that their prior knowledge does inform their judgments, it seems to do more good than harm.

AUTHOR NOTE

The authors thank Keith Thiede for his discussion on this project, Melinda Jensen, Robert Ariel, and Megan Engleman for their assistance with data collection, and Gregory Colflesh, Pat Cushen, Travis Ricks, Christopher Sanchez, and James Voss for consultation on the baseball-related materials. This research was supported by Grants R305H03170 and R305B07460 from the Institute for Education Sciences Cognition and Student Learning program to T.D.G., Keith W. Thiede, and J.W. Writing of the manuscript was supported by a Humboldt Research Fellowship for Experienced Researchers to the third author. Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect those of the funding agency. Correspondence concerning this article should be addressed to J. Wiley, Department of Psychology (M/C 285), University of Illinois, Chicago, IL 60607 (e-mail: jwiley@uic.edu).

REFERENCES

- ADAMS, B. C., BELL, L. C., & PERFETTI, C. A. (1995). A trading relationship between reading skill and domain knowledge in children's text comprehension. *Discourse Processes*, **20**, 307-323.
- BAKER, L. (1989). Metacognition, comprehension monitoring, and the adult reader. *Educational Psychology Review*, **1**, 3-38. doi:10.1007/BF01326548
- BOL, L., & HACKER, D. J. (2001). The effect of practice tests on students' calibration and performance. *Journal of Experimental Education*, **69**, 133-151.
- BOL, L., HACKER, D. J., O'SHEA, P., & ALLEN, D. (2005). The influence of practice and achievement level on calibration accuracy. *Journal of Experimental Education*, **73**, 269-290. doi:10.3200/JEXE.73.4.269-290
- BROWN, A. L., & DELOACHE, J. S. (1978). Skills, plans, and self-regulation. In R. S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, NJ: Erlbaum.
- CHI, M. T. H., GLASER, R., & FARR, M. (Eds.) (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- DUNLOSKY, J., & LIPKO, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, **16**, 228-232. doi:10.1111/j.1467-8721.2007.00509.x
- DUNLOSKY, J., & RAWSON, K. A. (2005). Why does rereading improve metacomprehension accuracy? Evaluating the levels-of-disruption hypothesis for the rereading effect. *Discourse Processes*, **40**, 37-55. doi:10.1207/s15326950dp4001_2
- ERICSSON, K. A., & KINTSCH, W. (1995). Long-term working memory. *Psychological Review*, **102**, 211-245. doi:10.1037/0033-295X.102.2.211
- FELTOVICH, P. J., PRIETULA, M. J., & ERICSSON, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 39-68). Cambridge: Cambridge University Press.
- FINCHER-KIEFER, R., POST, T. A., GREENE, T. R., & VOSS, J. F. (1988). On the role of prior knowledge and task demands in the processing of text. *Journal of Memory & Language*, **27**, 416-428. doi:10.1016/0749-596X(88)90065-4
- 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. doi:10.1037/0278-7393.33.1.238
- FISCHER, P. M., & MANDL, H. (1984). Learner, text variables, and control of text comprehension and recall. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 213-254). London: Erlbaum.
- GLASER, R., & CHI, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. xv-xvii). Hillsdale, NJ: Erlbaum.
- GLENBERG, A. M., & EPSTEIN, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, **15**, 84-93.
- GLENBERG, A. M., SANOCKI, T., EPSTEIN, W., & MORRIS, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, **116**, 119-136.
- GLOVER, J. (1989). Reading ability and the calibrator of comprehension. *Educational Research Quarterly*, **13**, 7-11.
- GRIFFIN, T. D., WILEY, J., & THIEDE, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, **36**, 93-103. doi:10.3758/MC.36.1.93
- HACKER, D. J., BOL, L., & BAHBAHANI, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition & Learning*, **3**, 101-121. doi:10.1007/s11409-008-9021-5
- HACKER, D. J., BOL, L., HORGAN, D. D., & RAKOW, E. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, **92**, 160-170. doi:10.1037/0022-0663.92.1.160
- HAMBRICK, D. Z., & ENGLE, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, **44**, 339-387. doi:10.1006/cogp.2001.0769
- JACOBY, L. L., BJORK, R. A., & KELLEY, C. M. (1994). Illusions of comprehension, competence, and remembering. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, and believing: Enhancing human performance* (pp. 57-80). Washington, DC: National Academy Press.
- KELEMEN, W. L., FROST, P. J., & WEAVER, C. A., III (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, **28**, 92-107.
- KINTSCH, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- KORIAT, A. (1997). Monitoring one's knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, **126**, 349-370. doi:10.1037/0096-3445.126.4.349
- 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., SHEFFER, L., & MA'AYAN, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, **131**, 147-162.
- LIN, L., ZABRUCKY, K., & MOORE, D. (1997). The relations among interest, self-assessed comprehension, and comprehension performance in young adults. *Reading Research & Instruction*, **36**, 127-139.
- MAKI, R. H. (1998a). Metacomprehension of text: Influence of absolute confidence level on bias and accuracy. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 38, pp. 223-248). San Diego: Academic Press.
- MAKI, R. H. (1998b). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117-144). Mahwah, NJ: Erlbaum.
- MAKI, R. H., & BERRY, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 663-679. doi:10.1037/0278-7393.10.4.663
- MAKI, R. H., JONAS, D., & KALLOD, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin & Review*, **1**, 126-129.

- MAKI, R. H., & SERRA, M. (1992). Role of practice tests in the accuracy of test predictions on text material. *Journal of Educational Psychology*, **84**, 200-210. doi:10.1037/0022-0663.84.2.200
- MAKI, R. H., SHIELDS, M., WHEELER, A. E., & ZACCHILLI, T. L. (2005). Individual differences in absolute and relative metamemory accuracy. *Journal of Educational Psychology*, **97**, 723-731. doi:10.1037/0022-0663.97.4.723
- MILLER, L. M. S., COHEN, J. A., & WINGFIELD, A. (2006). Contextual knowledge reduces demands on working memory during reading. *Memory & Cognition*, **34**, 1355-1367.
- MILLER, L. M. S., STINE-MORROW, E. A. L., KIRKORIAN, H. L., & CONROY, M. L. (2004). Age differences in knowledge-driven reading. *Journal of Educational Psychology*, **96**, 811-821. doi:10.1037/0022-0663.96.4.811
- NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133. doi:10.1037/0033-2909.95.1.109
- NELSON, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. *Applied Cognitive Psychology*, **10**, 257-260. doi:10.1002/(SICI)1099-0720(199606)10:3<257::AID-ACP400>3.0.CO;2-9
- NELSON, T. O., & NARENS, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-173). New York: Academic Press.
- PIERCE, B. H., & SMITH, S. M. (2001). The postdiction superiority effect in metamemory of text. *Memory & Cognition*, **29**, 62-67.
- PRESSLEY, M., SNYDER, B. L., LEVIN, J. R., MURRAY, H. G., & GHATA, E. S. (1987). Perceived readiness for examination performance (PREP) produced by initial reading of text and text containing adjunct questions. *Reading Research Quarterly*, **22**, 219-236. doi:10.2307/747666
- RAWSON, K. A., DUNLOSKY, J., & THIEDE, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, **28**, 1004-1010.
- RECHT, D. R., & LESLIE, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, **80**, 16-20. doi:10.1037/0022-0663.80.1.16
- SCHNEIDER, W. (2002). Memory development in childhood. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 236-256). Oxford: Blackwell.
- SCHNEIDER, W., KÖRKE, J., & WEINERT, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, **81**, 306-312. doi:10.1037/0022-0663.81.3.306
- SPILICH, G. J., VESONDER, G. T., CHIESI, H. L., & VOSS, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior*, **18**, 275-290. doi:10.1016/S0022-5371(79)90155-5
- THIEDE, K. W., ANDERSON, M. C. M., & THERIAULT, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, **95**, 66-73. doi:10.1037/0022-0663.95.1.66
- THIEDE, K. W., DUNLOSKY, J., GRIFFIN, T. D., & WILEY, J. (2005). Understanding the delayed keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 1267-1280. doi:10.1037/0278-7393.31.6.1267
- THIEDE, K. W., GRIFFIN, T. D., WILEY, J., & ANDERSON, M. C. M. (in press). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*.
- THIEDE, K. W., GRIFFIN, T. D., WILEY, J., & REDFORD, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85-106). New York: Routledge.
- VOSS, J. F., VESONDER, G. T., & SPILICH, G. J. (1980). Text generation and recall by high-knowledge and low-knowledge individuals. *Journal of Verbal Learning & Verbal Behavior*, **19**, 651-667. doi:10.1016/S0022-5371(80)90343-6
- WEAVER, C. A., III (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 214-222. doi:10.1037/0278-7393.16.2.214
- WILEY, J. (2005). A fair and balanced look at the news: What affects memory for controversial arguments? *Journal of Memory & Language*, **53**, 95-109. doi:10.1016/j.jml.2005.02.001
- WILEY, J., GRIFFIN, T. D., & THIEDE, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology*, **132**, 408-428. doi:10.3200/GENP.132.4.408-428
- WINNE, P. H., & HADWIN, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ: Erlbaum.
- YATES, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice Hall.

NOTES

1. Highly inaccurate judgments that fail to predict performance in either absolute or relative terms will still result in zero confidence bias if the directions of the judgment errors are randomly distributed around actual performance, because errors in opposite directions will cancel out in averaging. Given this, confidence bias effects are best interpreted alongside measures of absolute accuracy.

2. The lack of a key inferential test is even more problematic given the low number of participants ($N = 50$) and the fact that the expertise measure suffered from a strong positive skew with extreme outliers nearly 4 SD above the mean ($M = 2.56$, $SD = 2.38$, range = 0-11; see Table 1).

3. The mean average comprehension judgment for baseball texts was 2.63 ($SD = 1.11$). On average, the participants answered 3.25 out of 5.0 questions correctly for each of the five baseball texts ($SD = 0.80$). For the non-baseball texts, the mean judgment was 2.55 ($SD = 0.82$) and the mean test performance was 2.17 ($SD = 0.82$).

4. We examined the estimated intraindividual correlations for each subset of four texts. The range of average correlations for the five possible subsets was .11-.17 with Cronbach's $\alpha = .93$, indicating a high level of consistency. Furthermore, the level of relative accuracy observed here ($\gamma = .13$) is similar to that observed by Maki (1998b; $\gamma = .14$) when the readers made judgments on texts about a single topic. Although relative accuracy on the baseball texts was low ($M = .12$, $SD = .53$), it was significantly greater than zero [$t(101) = 2.24$, $p < .05$].

The mean relative metacomprehension accuracy for the set of non-baseball texts was .32, which was also significantly greater than zero [$t(100) = 6.82$, $p < .001$]. Using the same reliability procedure, we found the estimated correlations for the five possible subsets of texts to range from .25 to .38 ($\alpha = .91$). Furthermore, the observed level of accuracy ($\gamma = .38$) is similar to typical levels of metacomprehension in the literature ($\gamma = .27$; Dunlosky & Lipko, 2007; Maki, 1998b), which supports the validity of this as a metric of general metacomprehension skill.

The reliability of the absolute accuracy measures was assessed using the same subsets approach. For the baseball texts, estimates of error for the five possible subsets ranged from 1.17 to 1.25 ($\alpha = .96$). For non-baseball texts, the estimates of error for the five possible subsets ranged from 1.11 to 1.13 ($\alpha = .96$).

The same subsets approach was used for the bias measure. For baseball texts, the estimates of bias for the five possible subsets ranged from -.55 to -.66 ($\alpha = .98$). For non-baseball texts, the estimates of bias for the five possible subsets ranged from .26 to .53 ($\alpha = .98$). Confidence bias was also correlated across the two text sets ($r = .37$), which is similar to the modest correlations in confidence bias across contexts obtained by Kelemen et al. (2000; Spearman $r_s = .25-.44$).

APPENDIX**Example Baseball-Related Text: Corked Bats**

A corked bat is one in which a cavity has been drilled into the barrel of a wood bat. The cavity is usually filled with some substance, such as compressed cork, so that the bat sounds normal. Because wood has been removed from the bat and replaced by some substance with less density, the bat is lighter by about 1–2 oz. Not only is the bat lighter, but the center of gravity or balance point of the bat moves closer to the hands. This means that the “swing weight” of the bat is also reduced. It is much easier to swing something when the weight is concentrated closer to your hands than when it is concentrated far from your hands. Therefore, the argument is that a batter can get higher bat speed with a corked bat than with a comparable bat that has not been corked.

The distance a baseball travels after being hit depends on three primary factors: the angle at which the ball leaves the bat, how fast the ball is hit, and the mass of the bat. If the bat is standing still and the ball hits it, the ball will bounce off the bat with most, but not all, of the pitch speed. If the ball is standing still and is hit by the bat, it is given a good portion of the bat’s speed. Combine the two and you can see that a pitched ball hitting a swinging bat gains a good portion of the sum of both the pitch and the bat speed. So, if your bat is swinging faster the ball should go further. There is a second advantage of faster swing. A hitter can wait a few milliseconds longer before committing to a swing. This means he can watch the pitched ball travel about 5 or 6 more feet, which can help him figure out its movement and make contact.

On the other hand, in his book *The Physics of Baseball*, Robert K. Adair notes that the mass of the bat also affects how far the ball will go. In general, a heavier bat will hit the ball farther. So the advantages of a faster swing may be cancelled out by the loss of mass. He also notes that some effects achieved by corking can be obtained legally by just using a lighter bat. Indeed, the rules of Major League Baseball merely specify that a bat be no longer than 42 inches and no wider than 2.75 inches. There’s no maximum or minimum bat weight, so it’s not clear to Adair why players go through the trouble of creating illegal bats. Perhaps, he claims, it is simply superstition. The key factor to the corked bat may be that if a player thinks it will make a difference in his game then it very likely might make a difference.

Test Questions for Corked Bats

1. How would swinging speed be affected if the batter choked up on the bat (i.e., if the batter were to move their hands up the handle)?
 2. When would it take more energy to swing a golf club, if you hold the club or the handle?
 3. According to the article, why do batters cork their bats?
 4. How would swinging speed be affected by using a shorter bat of the same weight?
 5. Is there a maximum or minimum bat weight in Major League Baseball?
-

(Manuscript received June 28, 2007;
revision accepted for publication April 19, 2009.)