1-1-2005

Putting the Comprehension in Metacomprehension

Jennifer Wiley
University of Illinois at Chicago

Thomas D. Griffin
University of Illinois at Chicago

Keith Thiede
Boise State University

Putting the Comprehension in Metacomprehension

Jennifer Wiley & Thomas D. Griffin
University of Illinois at Chicago

Keith W. Thiede
Boise State University

Abstract

The purpose of the present piece is to integrate some current theories of text comprehension with the body of work on metacomprehension, and especially the calibration of comprehension monitoring. This paper explores some important methodological and conceptual issues, inspired by current theories in the text comprehension literature, which suggest that the nature of the texts used for metacomprehension studies may be a critical, and currently unrecognized, factor that should be considered. First, we need to re-examine what we mean by "comprehension," and how we should measure it. There are important differences between memory for text and comprehension of text that need to be considered. Second, to fully deal with these concerns, we need to pay more attention to the kinds of expository text that are being used, the different ways that readers may understand these texts, and how readers may interpret the concept of "understanding" as they make their judgments.

Research on self-regulated learning has been increasing in recent years, and with it an interest in metacognition or monitoring ability (see Hacker, Dunlosky, & Graesser, 1998; Zimmerman & Schunk, 2001). Many models of self-regulated learning describe learning as an interaction between metacognitive monitoring and control -- also called regulation of study (e.g., Nelson & Narens, 1990; Thiede & Dunlosky, 1999; Winne & Butler, 1997). For instance, consider a student preparing for an exam. As the student studies, she monitors her progress toward the goal of mastering the material. If her monitoring indicates that she has not yet mastered the material, she will likely restudy the material. If her monitoring indicates mastery has been accomplished, then that material is not selected for restudy. Therefore, accurate metacognitive monitoring is critical to effective regulation of study (Winne & Perry, 2000). If a person is not able to accurately differentiate well-learned material from less-learned material, he or she could waste time studying material that is already well learned, or worse, fail to restudy material that has not yet been adequately learned (e.g., Dunlosky & Thiede, in press; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999).

Since expository text is so central to educational practice (especially in science and social studies instruction), the ability of students to monitor their own understanding of these texts seems very important. The general conclusion that has been reached, however, is that readers lack an ability to track their own comprehension, especially on expository texts. They are not very accurate when asked to predict their performance on a test (Glenberg, Wilkinson & Epstein, 1982; Maki & Berry, 1984; Weaver, 1990). They do not select the least understood texts for re-study (Thiede, Anderson, & Therriault, 2003). Findings such as these have led to the conclusion that readers lack metacomprehension skills. But, are readers really this poor at monitoring their own level of comprehension? We cannot confidently answer this question because a number of important factors in text comprehension have been largely ignored in research on monitoring understanding of text. Thus, this paper explores some methodological and conceptual issues, inspired by current theories of text comprehension, which suggest that the nature of the texts and tests used for metacomprehension studies may be critical factors to consider.

Metacomprehension studies inspired by metamemory research

The tradition of metacomprehension research that has focused on the "calibration of comprehension" or the accuracy of metacognitive monitoring during text comprehension tasks has sprung from the metamemory literature. To date, the majority of investigations exploring factors that affect the accuracy of metacognitive monitoring have been done using associative learning tasks (e.g., learning word pairs). This research has shown that monitoring accuracy improves when a person monitors learning (a) after a delay rather than immediately after studying an item (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991), (b) after practice monitoring his or her learning (Vesonder & Voss, 1985), (c) after items are actively generated during study rather than passively read during study (Mazzoni & Nelson, 1993; but for failed replication see Matvey, Dunlosky, & Guttentag, 1997), and (d) following a practice test of the material (King, Zechmeister, & Shaughnessy, 1980; Lovelace, 1984; Shaughnessy & Zechmeister, 1992). Furthermore, under certain circumstances, people attain near perfect levels of monitoring accuracy. For example, monitoring accuracy, operationalized as the correlation between predicted performance and actual performance, was +.93 in Nelson and Dunlosky (1991).

Monitoring accuracy is far less impressive when the task involves learning from written texts. Glenberg and Epstein (1985) developed the original metacomprehension paradigm that is the model that research has followed to date. To assess the accuracy of comprehension monitoring, mirroring the paradigm used in research on associative learning, they had participants read 15 texts. Participants then rated their comprehension of each text and completed a test of reading comprehension for each one. A person's monitoring accuracy was operationalized as the intra-individual correlation between a person's comprehension ratings and his or her test performance; thus, greater monitoring accuracy was indexed by a stronger correlation. A standard term for this predictive accuracy is metacomprehension accuracy (Maki, 1998a).

In sharp contrast to accuracy of monitoring during associative learning tasks, metacomprehension accuracy has typically been quite low. For instance, Glenberg et al. (1987) concluded the following, "Data from our laboratory has almost uniformly demonstrated poor calibration" (p. 120). Likewise, Maki (1998a) reported that the mean intra-individual correlation between comprehension ratings and test performance across 25 studies from her lab was only +.27. More importantly, many of the factors that have improved monitoring accuracy in associative learning tasks have failed to produce similar affects on metacomprehension accuracy. Some studies have found improvements in metacomprehension skills, but results have been inconsistent. For example, the effect of delaying one's judgments of learning (versus making an immediate judgment) in associative learning tasks has produced the largest improvement in accuracy reported in the literature. Sometimes introducing a delay between reading and judging improves performance (Maki & Berry, 1982). However, Maki (1998b) failed to replicate the *delayed judgment of learning effect* for a reading task. Comprehension ratings made after a delay were less accurate than those made immediately after reading. Similarly, practice tests have been shown to improve monitoring accuracy in associative learning tasks, but have had at best modest effects on the accuracy of monitoring comprehension during reading (Maki & Serra, 1992b). Sometimes reading instructions (Magliano, Little & Graesser, 1993; Schommer & Surber, 1986) and adjunct questions can have an effect (Glover, 1989; Pressley, Snyder, Levin, Murray & Ghatala, 1987; Walczyk & Hall, 1989). However, more striking than these isolated findings is the recognition that manipulations often do not work across contexts or studies (c.f. Maki, 1998).

If we assume that memory for presented stimuli and comprehension of texts' conceptual meaning are distinct psychological phenomena, then the effective study of metacomprehension will entail more than replacing paired associates with texts and will require the careful consideration of methodological issues that have not arisen in metamemory paradigms. This paper explores some issues inspired by current theories of the text comprehension literature suggesting that the nature of the texts and tests used for these studies may be critical factors that should be considered. Consistent with the points made by Weaver (1990), we need to re-examine what we mean by "comprehension," and how we should measure it. However, an additional insight, is that to fully deal with these concerns, we need to pay more attention to the kinds of expository text that are being used.

The analyses we present in this paper suggest that metacomprehension research has paid insufficient attention to the nature of the texts and test items employed. One problem is that test items often measure something other than comprehension, such as surface memory. As will be discussed in more detail below, part of the reason for this is the use of texts that lack sufficient structural complexity to allow for comprehension tests that are not heavily influenced by surface memory. Thus, some studies may be measuring metamemory for text rather than metacomprehension. A

second problem has been variability in the kinds of expository texts that have been used. Inconsistencies in expository text structures across studies make comparisons difficult. Within-study inconsistencies in expository text are likely to impede readers' ability to gauge the type and level of understanding they are being asked to judge and monitor. In sum, a failure to closely consider the nature of the expository texts and test questions that have been used may result in measurements of metacomprehension that are misleading or not measures of metacomprehension at all.

What do we mean by "comprehension"?

As Weaver (1990) and others have noted (Rawson, Dunlosky & Thiede, 2000; Weaver, Bryant, & Burns, 1995) a bridge needs to be made between the literature on text comprehension and studies on metacomprehension. Current theories of text comprehension view the act of comprehension as occurring on multiple levels. The most prominent of such models is Kintsch's theory of comprehension (1998). On one level, a representation of the text is developed in terms of the surface form or the exact words that appear. A second level, the text-base, is a slightly more processed version where the information that was read is now represented in terms of syntactic or semantic constituents. Essentially, the text-base can be seen as an abstraction of the exact words in propositional form, so that the meaning is still what was presented with only minor inferences, such as anaphor resolution. The elements or links in the text-base have a direct correspondence to what was in the text. On the highest level of representation, termed by Kintsch as the situation model, there is integration of the ideas that were presented with prior knowledge. Texts usually describe situations that can exist, and the reader builds a representation or model of these situations based on the text, their prior experience, and what might be probably true based on those propositions.

In terms of these levels, it is the creation of the situation model that can be seen as the process of deeply understanding a text, or comprehension, and acquiring the information so that it can be used in novel contexts (Kintsch, 1994). While the surface and text-base levels can be seen as memory-based representations of the exact words and ideas that appeared in a text, the situation model represents the instantiation of both explicit and implicit relations between ideas. Other theories of text comprehension describe this final level of representation as the construction of a causal model or causal chain (Trabasso & van den Broek, 1985) or as scenario building (Sanford & Garrod, 1998). The critical component in all these theories is that the reader establishes causal, logical or explanatory relations between the propositions given in the text. So if the student is given a text about blood circulating through the heart, then the situation model should capture the process of blood flow, and a reader with a good situation model should be able to recreate and explain the dynamic process of blood flow that was described by the text.

The structural complexity of the expository text is quite important for determining the extent to which the text-base and situation model will differ. Simple linear texts with few implicit relations will have very similar text-base and situation model representations. As a result, memory for the text will be indiscriminable from comprehension. However, in more complex text, with more implicit relations, the text-base and situation model will be quite different representations. In these cases, measures of memory will not necessarily be good measures of comprehension. Instead, test items that directly tap into the implicit or causal relations implied by the text will be a more direct measure of comprehension.

In other words, only complex texts provide researchers the opportunity to test for situation model level inferences that cannot be simply recalled from the text-base. Such inference questions must be about relations that are implied by the text, but not explicitly stated. Simple descriptions lack the complexity of relations to produce such implied associations. In contrast, causally complex texts generally convey numerous implicit logical and causal relations among ideas. When a text states that 'A' causes 'B', this conveys an implied relationship between 'A' and other ideas with a causal relation to 'B', and vice versa. The implied relations conveyed by complex causal structures are fruitful ground for test questions that require a situation model level understanding. Also, texts of insufficient complexity will constrain tests to be strongly affected by surface memory, resulting in monitoring accuracy scores that are more reflective of metamemory than metacomprehension.

These theories of comprehension suggest that text comprehension can occur in several different ways, ranging from the generation of a memory for what was read, to the creation of explanatory models, depending on the nature of the expository text. Thus, the first lesson from the text comprehension literature, is that we need to attend to the nature of both the texts and test questions that have been used in research on metacomprehension. It is quite probable that

some studies have tested readers' memory while others have tested readers' comprehension. Failed replications or seemingly conflicting results may occur from the use of materials that assess different types of learning. Importantly, the only condition under which we can truly study metacomprehension is when tests measure comprehension. Thus, any attempt to make sense of findings across the literature must carefully consider the type of learning they assessed.

A second contribution from the text comprehension literature compounds the problem. Several lines of investigation have found that the conditions that improve memory for text (or memory for facts or details presented in the text) are often different than the conditions that improve understanding of a text (i.e. the creation of a causal or explanatory model). Surface memory and deeper comprehension do not necessarily correlate and are affected by different factors. This means that if one considers only metacomprehension studies using measures of learning based on memory for facts, one could reach misleading conclusions. For example, readers may better predict their test performance by monitoring their memory for the text, rather than their comprehension. And, particular interventions may improve judgment-performance covariance by improving meta-memory for text, but fail to improve or even harm comprehension monitoring. Therefore, it becomes especially critical that we align our learning measures to be clear whether better memory for the text, or understanding of the text, is the basis for the study. Indeed, the lack of discrimination among learning measures could be part of the reason for inconsistent findings about metacomprehension skills and their improvement.

The third and primary contribution from the text processing literature that will be discussed here is a methodology for examining the nature of the texts that are used to study the phenomenon of metacomprehension accuracy. A great deal of research on text processing has shown that it is the causal structure underlying a text that dictates its patterns of memory and understanding (Trabasso, van den Broek & Suh, 1989; Myers, Shinjo, & Duffy, 1983; van den Broek, 1990). Employing causal analysis of texts, we can show which expository texts have a large number of implicit relations. Explanatory texts (texts which have an explanatory structure embedded within them) would be especially good texts with which to investigate metacomprehension accuracy as the situation models that should be developed, and the casual inferences that can be derived from them, are clear. On the other hand, some expository texts are more descriptive. Without the presence of a causal model underlying a text, comprehension approximates simple memory for facts.

Structural complexity of texts has important implications for metacomprehension research. Readers are sensitive to text structure and it affects the way they process and represent a text (e.g., Meyer and Freedle, 1984). Further, understanding improves when readers are explicitly aware of these structures (Kintsch & Yarrough, 1982). It is quite possible that text structure impacts a reader's notion of what it means to understand a particular text. This means that variability in research materials could artificially lower reader's monitoring accuracy in two ways. First, reader's judgments of understanding will fail to predict performance if the tests do not assess the same type and level of understanding implied to readers by the text structure. If the texts prompt readers to monitor their representation of idea relations, but the tests assess memory, then judgment-performance predictions will be poor. Second, even if the tests align with the text structures, if the text structures vary across texts, then readers will be forced to alter their notion of "understanding" from text to text resulting in confusion and misapplication of earlier notions of "understanding" to later texts.

Thus, a primary contribution of this paper is to examine the structures of texts and nature of tests that have been used in metacomprehension research. In the original Glenberg studies, for example, the stimuli that were used were single paragraph expository texts, and the tests were single questions about the main topic of the passage. Previous researchers have already commented on the use of a single test question (Glenberg, Sanocki, Epstein & Morris, 1987; Lin & Zabrucky, 1998; Maki & Serra, 1992b; Weaver, 1990) and documented the improvement in estimates of metacomprehension accuracy that occur when multiple items per text are used. Although multiple questions may lead to improvements in test reliability, there is still the issue of whether or not questions truly require understanding to be answered correctly.

The test questions from the early studies were intended to be inference questions related to the central proposition of the text. In practice they might have been inferences that were available to the student without having read the text. A problem, especially with main idea questions, is that they can be answered strategically using prior knowledge or keywords from the passage (i.e. without having really understood or even read the text). Admittedly, the creation of inference questions is a very difficult task. It requires a delicate balance between what is stated and what is unstated

in the text, and what can be assumed to be common knowledge and what is unlikely to already be known by a reader. As long as one tests for information that is directly given in the text, then one is measuring the text-base. On the other extreme, if one tests for an idea that is not available from integrating ideas in the text, then one is testing for prior knowledge. If we are interested in studying situation-model level comprehension, and how sensitive readers can be to its creation, then what is needed are inference questions that require the construction of a causal, situation model in order to be answered. However, as noted previously, this requires texts that actually have a causal model embedded in them, so that these inferences are available to readers. Hence, before we can examine the nature of the test questions, we also must investigate the nature of the texts that have been traditionally used in metacomprehension calibration studies.

What kinds of expository texts are these?

Most studies of metacomprehension accuracy have used expository texts as their stimuli. Although these texts may all fall into an expository genre when contrasted against more purely narrative texts, it is an interesting question to ask whether they are all the same kinds of expository texts. This section presents an analysis of several of the one paragraph passages that have been used in so many studies on metacomprehension, originally introduced by Glenberg and Epstein (1985). The original and revised texts have been used in a large portion of the studies in this literature (Glenberg, Sanocki, Epstein, & Morris, 1987; Lin, Moore & Zabrucky, 2001; Lin, Zabrucky & Moore, 1997; Maki & Serra, 1992a, b; Magliano, Little & Graesser, 1993; Morris, 1990; Rawson, Dunlosky & McDonald, 2002; Rawson, Dunlosky & Thiede, 2000; Thiede & Anderson, 2003; Weaver, 1990). While to an outsider it is striking that such a large portion of a literature has utilized the same set of texts, it is understandable when investigators were concerned with aligning the effects of new manipulations with previous results.

Previous reviews of the literature have discussed the possible role of the text itself in metacomprehension accuracy. These reviews have focused almost exclusively on the simple distinction between narrative and expository genres. These differences have generally been discussed in terms of difficulty, as expository texts usually deal with more technical and less familiar subject matter than narrative texts (Lin & Zabrucky, 1998; Weaver, 1990). Relatedly, expository text has been proposed to be (a) processed in a more item-specific as opposed to global, relational manner (Einstein, McDaniel, Bowers & Stevens, 1984), (b) less coherent to readers, (c) require or assume more specific background knowledge to be understood (Graesser & Bertus, 1998; van den Broek, Virtue, Everson, Tzeng, & Sung, 2002), and (d) require more effort to generate even within-text inferences (Wiley & Myers, 2003).

However, there are different kinds of text that fall into the catchall generic category of "expository". Earlier attempts to classify subtypes of expository texts have used the structure of texts as the basis for delineation (c.f. Kintsch, 1982; Meyer & Freedle, 1984). For example, Meyer and Freedle (1984) identified 4 types of expository text that differed in the complexity of the interrelations among their idea units (i.e., collection, comparison, causation, and problem/solution). Collection texts have the simplest structure consisting of lists of details with no logical or causal associations between them. The details can be a set of descriptive facts or examples that are unrelated to each other, but linked by the common topic or principle they describe or exemplify. Alternatively, a collection can be a chronology of events grouped by their sequence in time. Comparison texts are slightly more complex with ideas that are interlinked through comparing and contrasting them against each other. Causal and problem/solution texts are the most structurally complex and contain ideas that are interrelated both in terms of time sequence and causal or quasi-causal associations.

As argued previously, the type of text structure is an important factor that could affect metacomprehension accuracy and constrains whether the tests of comprehension are assessments of comprehension or memory of facts. For our present purposes, we can combine causal and problem/solution texts into a single category of *explanatory*, resulting in the three categories of *collection*, *comparison*, and *explanatory,* which vary in structural complexity from least to greatest. Of course, any text longer than a single proposition can contain more than one of these text types, but often a single text type will capture a majority of the idea relations within a text. Our purpose here is to examine a set of texts that are central in the metacomprehension literature in light of these structural distinctions.

Seven sample texts (originally from Glenberg & Epstein, 1985) were available from appendicies of the original and follow-up articles (*Detection of Black Holes*, *Control of Eating by Blood Sugar*, and *A Good Hanging Never Hurt Anyone* in Glenberg & Epstein, 1985; *Rules for Setting Bail* in Lin, Zabrucky & Moore, 1997; *Global Temperature and Flooding* in Maki & Serra, 1992a; *Freedom and the Fear of Genetic Intervention* in Maki & Serra, 1992b;

Evolutionary Adaptations for New Biological Functions in Morris, 1990).   Seven more that were in the original set, and used in Weaver (1990), are included in an appendix of this article, *Correlation and Causation*, *Rules for Borrowing*, *Viruses and Cancer*, *Listening Space*, *Scientific Theories*, *Earth Homes*, and *Nuclear Disarmament*.

Five of these fourteen texts conveyed a clear causal model or process to the reader. For example, in *Detection of Black Holes* (Glenberg & Epstein, 1985), the paragraph details the causal relation between a collapsing star, the mass of the star, the ability to see light, and the effect of the collapsing star on other objects. The inference item is based on the proposition that a collapsing star has an effect on other objects, and therefore can only be detected when there are other objects around it. Similarly, in the *Control of Eating by Blood Sugar* (Glenberg & Epstein, 1985), the paragraph details the causal relation between hunger and glucose levels in the bloodstream.   The inference item probes for understanding of the relation between glucose and hunger by introducing new concept of insulin, which suppresses glucose. *Global Temperature and Flooding* (Maki & Serra, 1992a) details the relationship between carbon dioxide emissions, the greenhouse effect, global temperatures, polar icecap melting, and ocean levels. (Unfortunately the original inference item is not included). *Thermal Characteristics of Earth Sheltered Homes* introduces causal relations between air temperature, ground temperature, warming effects of the Sun and cooling effects of wind. (No inference item available.) In *Evolutionary Adaptations for New Biological Functions* (Morris, 1990), the paragraph explains the relation between several evolutionary principles: how natural selection leads to the gradual emergence of new structures, and how adaptation occurs under one set of selection pressures, then structure can be used for new purpose once evolved. The inference item for this passage is based on a macroproposition that can be constructed from the text: the relation between history of a structure and its current use.

A sixth text *Viruses and Cancer*, is essentially a logical argument. At its base is a syllogism that if cancer is caused by viruses in animals, and biological phenomena are constant across species, then viruses may cause cancer in humans. Although this is not a causal structure in a physical sense, it is explanatory in the logical sense. The logical argument is a common structure in expository texts that does depend on the creation of relations between the premises, and as such can be seen as a model that needs to be constructed by the reader (c.f., Wiley & Myers, 2003). In all, six out of the fourteen texts examined possess a predominantly explanatory structure.

A second set of five passages state or provide examples of a principle for the reader, and thus are more collective in structure. Although some of the principles contain a causal relation, the descriptive principle-example relation is the over-arching structure. In *Rules for Setting Bail* (Lin, Zabrucky & Moore, 1997), the paragraph describes a key principle of the Supreme Court: that their purpose is to detect conflict with constitutional rights. The passage then elaborates on this through an example. (Unfortunately, original inference questions were not included for any of these texts). *Enhancing the Apparent Size of the Listening Space* introduces the principle behind Bose speakers, and then illustrates it with an analogy.   *When to Borrow* gives readers the principle that one should borrow when an investment yields higher interest than the interest on the loan. *The Correlational Method and Causation* states that causal conclusions cannot be drawn from correlational data, but does not really explain why. An example of a correlation is given. Similarly, *Determining the Truth and Falsity of Scientific Theories*, again states a principle of the scientific method, that theories can only be proven false, but does not really explain the principle in a way that a novice reader could understand.

Although none of the available examples from the original Glenberg passages seem like collections of sequenced events, sample expository texts from other studies have been. For example, the sample passage in Thiede, Anderson and Therriault (2003) on *Norse Settlements* is based on an encyclopedia entry and is basically a chronological account of facts about populations in Norway. Examples of other kinds of descriptive expository texts are found in Gillstrom and Ronnberg (1995) where readers were given single paragraph texts such as *Banana*, which offers facts about the historical emergence of bananas in Europe and the Americas.

Three final examples from the Glenberg passages are comparison texts that argue in favor of some position or action against some alternative. *Freedom and the Fear of Genetic Intervention* (Maki & Serra, 1992b) offers reasons for and against genetic design. *The Dilemma of Nuclear Disarmament* does exactly as its title suggests, as it offers reasons for and against the elimination of nuclear weapons. Similarly, *A Good Hanging Never Hurt Anyone* (Glenberg & Epstein, 1985), advances the claim that "we should stall the rush toward capital punishment" and offers several reasons to support that claim, including that the judicial process is imperfect, that errors cannot be corrected or compensated for because death is permanent, and errors that cannot be later compensated are unacceptable. The

inference item for this passage required readers to verify as true or false, the statement: *The availability of prosthetic devices, that is artificial substitutes, is a consideration that will affect the uncertainty associated with surgical amputation in cases of conflicting medical opinion.* The logic behind constructing this statement was that readers who have understood the principle that *irreversible actions should not be taken unless there are compensations that can be made,* should find the test statement to be true.

These texts exemplify a specific type of comparative text, an evaluative comparison that is especially problematic for valid assessment of comprehension. The central thesis of such texts is a subjective, value-based judgment that one action is superior or preferable to others. What distinguishes these texts is that they are centered around contestable personal values that are integral to the ideas provided for the main thesis. Evaluative texts can contain more or less descriptive or explanatory relations. For example, the Glenberg capital punishment text presents a descriptive list of reasons and personal values against the practice, whereas the nuclear disarmament text explains how an overreaching effort to quickly eliminate all nuclear arms could cause some undesirable outcomes. There can be logical explanatory relations in value-based arguments such as the logical relationships between a set of asserted facts (e.g., there are errors in judgments of guilt and no compensation for wrongful execution), presumed values (e.g., erroneous acts without compensation should not occur), and the main thesis (e.g., capital punishment should not occur). Comprehension of such a logical argument would entail recognition of how the main thesis is necessitated by the combination of the asserted facts and presumed values.

Evaluative texts do not necessarily lack explanatory structure or implied idea relations, but they are problematic because of the centrality of the author's personal values. Inference questions about such texts will often fail to validly assess understanding, because they require personal acceptance of the presumed values of the text's author. Even the inference question itself may introduce additional presumed values that require acceptance in order to generate a "correct" response. As Griffin and Ohlsson (2001) argued, there is a critical but largely ignored distinction in the comprehension literature between understanding of ideas and acceptance of or belief in those ideas and assertions. In the case of the Glenberg and Epstein (1985) capital punishment text, the reader could understand the text and yet fail to "correctly" verify the test statement either because they do not personally agree with the author's value-based presumption that uncorrectable errors should never occur, or because they do not agree with the evaluative premise in the inference statement that prosthetics provide compensation for a lost limb. Likewise, a reader who personally agrees with the values inherent in the test statement may "correctly" endorse it as true, even though they fail to understand the abstract principle that connects the statement to the central idea from the text. Test items that are so readily influenced by factors orthogonal to understanding (e.g., personal values) provide a misleading basis for calculating metacomprehension accuracy.

So, what would be a valid inference the reader should construct and be asked about after reading an evaluative text? Rather than asking the reader whether or not a new statement is true, it might be more fitting to ask whether a new statement supports or opposes the author's position, or the claim that was advanced. However, more research is surely needed before one could say what would be the best method of assessing comprehension of these kinds of evaluative text in a way that is not affected by the readers' personal values. Comprehension and metacomprehension of evaluative, value-laden explanatory arguments is certainly an interesting avenue of research in its own right. However, evaluative texts pose serious validity concerns when they are used in the context of studying metacomprehension at a more basic and general level, pariticularly when explicit attention is not paid to the comprehension vs. agreement distinction when constructing inference items for such texts.

Summary of Analysis of Example Expository Texts

Even in this subset of the original texts, there is variability in the kinds of expository text that are used. Several fit into a causal/explanatory model category. Some are descriptive or collective texts where a principle is stated and often illustrated with an example. A few others are comparative texts that vary in degree of explanation but have an additional quality of being evaluative texts where value-based assertions are made and supported. These types of expository text seem quite different, and how one might learn each of them may also vary (e.g. Kintsch & Yarbrough, 1982; Weaver & Kintsch, 1991). For the explanatory texts, the process that is described can be represented and explicit and implicit connections can be created between ideas. The implicit connections represent aspects of the situation model that go beyond and are empirically separable from mere memory for the text. The goal of descriptive/collection text, is to convey to the reader a novel principle that may be well captured by a single summary statement. Other collection texts have the goal of conveying lists of facts or details about a topic.

Comparison texts fall somewhere in between in terms of representational complexity. Evaluative texts can contain more or less explanation, but share the feature of comparing and contrasting some event or practice to a value-based ideal in the aim of convincing the reader to adopt a position about the practice.

The present text classifications are based largely upon Meyer and colleagues (e.g., Meyer & Freedle, 1984; Meyer, 1985), but have some obvious connections to Kintsch's (1982) types of rhetorical structure. Kintsch's (1982) analysis yielded three main types of rhetorical schemas that map onto the types identified above: General-particular relations, object-part relations, and object-object relations. General-particular relations are schemas that identify, define, classify and/or illustrate phenomena. This pattern seems similar to those of the descriptive/collection texts identified in the above analysis. A second schema is the object-part relation, where the text engages in a technical analysis of phenomena (in either functional or causal terms). This seems similar to the causal/explanatory structures noted above. A third schema is the object-object relation, where the text juxtaposes two entities for the purposes of comparison or evaluation. One could view the evaluative texts as possessing this rhetorical structure, where a state of affairs is compared with what ought to be, or two sides of an issue are compared to a value or ideal. The correspondence between the present analysis and Kintsch's (1982) rhetorical analysis is
particularly relevant because when readers are given texts of these different subtypes, their understanding is improved when the specific rhetorical structures are signaled as opposed to when they are not (Kintsch & Yarbrough, 1982). This suggests that modes for learning from these different types of texts are indeed different and that reader awareness of text structure affects text processing.

The point of the present analysis is not to single out these texts for criticism, but to illustrate more general concerns about insufficient attention in metacomprehension research to these text processing issues. There are implications and complications that may be the consequence of the texts and test items that are used. Certainly descriptive texts that are merely collections of facts about a topic would require similar kinds of learning as paired-associates or lists of ideas. But, how are evaluative texts processed? Are these kinds of texts understood and represented in the same way as other expository texts? This is currently an unanswered question in the text processing literature. Nevertheless it is an important question and it seems problematic to run studies using different kinds of expository text if they are not all processed, represented or understood in similar ways.

When one examines the statements that were used for the inference verification tests, variety exists there as well. As a measure of understanding of the written material, students were instructed to verify whether a statement was true or false based on "what you have learned from the passage." As already discussed, some test items were as much a test of personal agreement as of understanding. However, even among the other items, Glenberg and his colleagues explicitly recognized the range of difficulty. Some statements were taken directly from the texts (a sentence representing the central proposition). Some statements were macropropositions (abstractions or restatements of a central point in more general terms). Some statements required problem solving with new information (based on the application of a central principle advanced in the text). All questions were designed to tap into the main idea of each text. However, given that some items were taken directly from the text, and some required novel application of ideas, these items may differ in more than difficulty. In particular, only the later items may require the construction of a situation model to answer correctly. Of course, having generated an accurate situation model could help you to verify correct statements and summaries of a text, but this could also be accomplished via other means. Therefore, the critical variety here may be in the kinds of understanding that these different item types are tapping.

A major implication of all this variety, and perhaps the most important level of concern, is the effect that this variability may have on the reader who is attempting to judge their understanding of these texts. What does the reader infer the meaning of "understanding" to be? How should they gauge their learning? Should they base their judgments on their memory for the texts? Or on their ability to summarize? Or on their ability to apply principles included in the text in new circumstances?

A reader's ability to predict test performance will depend heavily upon whether their interpretation of "understanding" corresponds to the type and level (e.g. text-base vs. situation model) of the understanding that the texts and tests require. Even when a reader interprets "understanding" in terms of developing a good situation model, if within the study, the level of understanding being assessed is inconsistent, then correspondence between performance and judgments will necessarily be quite low. Also, since readers appear to modify their processing based upon cues about text structure (Kintsch & Yarbrough, 1982), they could very well adjust their interpretation of understanding in response to text structure. Thus, inconsistency in text structure could mislead readers to misapply

their interpretation of understanding produced by the structure of early texts to other texts later in the protocol. Even if text structure is consistent throughout the study, readers may be misled to monitor the wrong type of understanding across all texts, if the type of understanding implied by the text structure does not match what the tests measure. Thus, a hypothesis to consider is that poor calibration findings in metacomprehension studies may be partly arising from the use of texts and tests that are inconsistent in the kinds of understanding they require.

The present analysis suggests a future goal for metacomprehension research: to examine reader's metacomprehension judgements with a more uniform and explanatory set of expository texts, so that the understanding or situation model that underlies each text, as well as the inferences to be derived from it, are clear to both researchers and readers. When the content of expository texts can be delineated with a causal or structural analysis, then that analysis can be used to create a "gold standard" for understanding – an explicit causal model of the text. One can then create of sets of test questions that tap into the causal model and represent necessary causal or explanatory inferences that extend beyond the explicitly stated text propositions. The result is a direct measure of understanding of the process described by the text.

Several researchers interested in contexts that improve learning from text have taken such an approach. For example, Mannes and Kintsch (1987) performed an analysis of a passage about fermentation to develop test questions based on inferences needed in order to understand the text. Rereading and the presence of a matched outline improved memory for the details in the text, while the presentation of a mismatched outline improved performance on inference questions. Presumably, readers in the mismatched outline condition were more active in constructing a situation model, whereas readers in the former condition could rely on their surface or textbase models. In a similar study, Chi, de Leeuw, Chiu and LaVancher (1994) investigated learning from a text about the circulatory system and examined when readers generated explanations that made relationships between parts of the text explicit. When readers were prompted to self-explain, more connections were generated, and readers were more likely to recognize correct inferences about the circulatory system. The mental model and key inferences were defined by an analysis of the text and the 11 key functions of the anatomy involved in the circulatory system, which were unstated by the text (See also Chi, 2000). Similarly, Wiley (2001) found that when readers of a text on plate tectonics were asked to generate a causal argument of why volcanoes erupt, they were more likely to understand the process. Again, this investigation of learning from text was done by deriving the causal model underlying volcanic eruptions, and testing for understanding of relationships in that model using an inference verification task (see also Wiley & Voss, 1999). Using a set of explanatory texts such as these, where inferences are directly derived from a causal model, we can see whether metacomprehension accuracy might improve when readers have a more consistent expectation of what "understanding a text" might mean.

There are many alternative sources of confusion and noise in the standard metacomprehension paradigm. Readers are asked to judge their own understanding of texts, but no definition of understanding is given. Even if it were, the kinds of text they are given, their goals for reading and learning, the scales we give them, the number of comprehension questions we give them, the nature of the comprehension questions we give them, and even the way we compute accuracy, are inconstant factors that may obscure any reliable comprehension monitoring skill that readers may actually possess. The admonition presented here is that we need to be sure to attend to what it means to learn from the expository texts that we present. Metamemory and metacomprehension are likely to be distinct psychological phenomena impacted by different constraints and requiring different methodological considerations for study. More attention must be given to this distinction, particularly to ways in which metacomprehension gives rise to methodological issues that have not been considered within metamemory paradigms. If we are interested in studying readers' ability to monitor their own understanding of text, then we need to be clear what it means to understand. To the extent that text comprehension is a process of developing explanatory, causal representations, then we need to be sure that the expository texts we present have such models available, and that test items are based on the successful development of such models.

References

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology, Vol. 5* (161-238). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., DeLeeuw, N., Chiu, M., & Lavancher, C. (1994). Elicit self-explanation improves understanding. *Cognitive Science, 18*, 439-477.

Dunlosky, J. & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 374-380.

Dunlosky, J. & Thiede, K. W. (in press). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory & Cognition*.

Einstein, G. O., McDaniel, M. A., Bowers, C. A., & Stevens, D. T. (1984). Memory for prose: The influence of relational and proposition-specific processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 133-143.

Gillstrom, A, & Ronnberg, J. (1995). Comprehension calibration and recall prediction accuracy of texts: Reading skills, reading strategies, and effort. *Journal of Educational Psychology, 87*, 545-558.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 702-718.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension*. Journal of Experimental Psychology: General, 116*, 119-136.

Glover, J. A. (1989) Improving reader's estimates of learning from text: The role of inserted questions. *Reading Research Quarterly, 28*, 68-75.

Graesser, A.C., & Bertus, E.L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading, 2*, 247-269.

Griffin, T. D., & Ohlsson, S. (2001). Knowledge versus beliefs: A necessary distinction for predicting, explaining, and assessing conceptual change. In L. R. Gleitman and A. K. Joshi (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (1998). *Metacognition in Educational Theory and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

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

Kintsch, W. (1982). Text Representations. In W. Otto & S. White (Eds.), *Reading Expository Material* (pp. 87-102). New York, NY: Academic Press.

Kintsch, W. (1994). Learning from text. *American Psychologist, 49*, 294-303.

Kintsch, W. (1998). Comprehension: A paradigm for cognition. New York: Cambridge University Press.

Kintsch, W. & Yarbrough, C. J. (1982) Role of rhetorical structure in text comprehension. *Journal of Educational Psychology, 74*, 828-834.

Lin, L., Moore, D., & Zabrucky, K. M. (2001). An assessment of students' calibration of comprehension and calibration of performance using multiple measures. *Reading Psychology,2*, 111-128.

Lin, L. & Zabrucky, K. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*, 345-391.

Lin, L., Zabruck:y, K., & Moore, D. (1997). The relations among interest, self-assessed comprehension, and comprehension performance in young adults. *Reading Research and Instruction, 36*, 127-139.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10*, 756-766.

Magliano, J.P., Little, L.D., & Graesser, A.C. (1993). The impact of comprehension instruction on the calibration of comprehension. *Reading Research and Instruction, 32*, 49-63.

Maki, R. H. (1998a). Test predictions over text material. In Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). *Metacognition in Educational Theory and Practice*. (pp. 117-144). Hillsdale, NJ: LEA.

Maki, R. H. (1998b). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition, 26*, 959-964.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10*, 663-679.

Maki, RH & Serra, M. (1992a). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 116-126

Maki, R. H., & Serra, M. (1992b). Role of practice tests in the accuracy of test predictions on text material. *Journal of Educational Psychology, 84*, 200-210.

Mannes, S. & Kintsch, W. (1987). Knowledge Organization and Text Organization. *Cognition and Instruction, 4*, 91-115.

Mazzoni, G. & Nelson, T. O. (1993). Metacognitive monitoring after different kinds of monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1263-1274.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.

Meyer, B. J. F. (1985). The organization of prose and its effect on memory. Amsterdam: North-Holland.

Meyer, B. J. F. & Freedle, R. O. (1984) Effects of discourse type on recall. *American Educational Research Journal, 21*, 121-143.

Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 223-232.

Myers, J. L., Shinjo, M., & Duffy, S. A., (1987) Degree of causal relatedness and memory. *Journal of Memory and Language, 26*, 453-465.

Nelson, T. O. & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed JOL effect." *Psychological Science, 2*, 267-270.

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, 26,* pp. 125-141. New York: Academic Press, Inc.

Pressley, M., Snyder, B. L., Levin, J. R., Murray, H. G., & Ghatala, 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.

Rawson, K., & Dunlosky, J., & McDonald, S. L. (2002). Influences of metamemory on performance predictions for text. *Quarterly Journal of Experimental Psychology, 55*, 505-524.

Rawson, K., & Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition, 28*, 1004-1010.

Sanford, A.J. & Garrod, S.(1998) The role of scenario mapping in text comprehension. *Discourse Processes, 26*, 159-190.

Schommer, M. & Surber, J. (1986). Comprehension-monitoring failure in skilled adult readers. *Journal of Educational Psychology, 82*, 498-504.

Shaughnessy, J. J. & Zechmeister, E. B. (1992). Memory monitoring accuracy as influenced by the distribution of retrieval practice. *Bulletin of the Psychonomic Society, 30*, 125-128.

Son, L. K. & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204-221.

Thiede, K. W. & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology. 28*, 129-160.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology , 95*, 66-73.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024-1037.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24*, 595-611.

Trabasso, T., van den Broek, P., & Suh, S. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes, 12*, 1-25.

van den Broek, P. W. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 423-446). Hillsdale, NJ: Erlbaum.

van den Broek, P., Virtue, S., Everson, M., Sung, Y., & Tzeng, Y. (2002). Comprehension and memory of science texts: Inferential processes and the construction of a mental representation. In J. Otero, J. Leon, & A.C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 131-154). Mahwah, NJ: Erlbaum.

Vesonder, G. T. & Voss, J. F. (1985). On the ability to predict one's own responses while learning. *Journal of Memory and Language, 24*, 363-376.

Walczyk, J. J., & Hall, V. C. (1989). Effects of examples and embedded questions on the accuracy of comprehension self-assessments. Journal *of Educational Psychology, 81*, 435-437.

Weaver, C. A. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning , Memory, & Cognition, 16*, 214-222.

Weaver, C. A., Bryant, D. S. & Burns, K. D. (1995). Comprehension monitoring: Extensions of the Kintsch and van Dijk model. In C. A. Weaver, Mannes, S., & Fletcher, C. (Ed.), *Discourse comprehension. Essays in honour of Walter Kintsch* (pp. 177-193). Hillsdale, NJ: Lawrence Erlbaum Associates.

Weaver, C. A. & Kintsch, W. (1991) Expository Text. In R. Barr, M. Kamil, P. Mosenthal & P. D. Pearson (Eds) *The Handbook of Reading Research* (Vol 2, pp. 230-245). White Plains, NY: Longman.

Wiley, J. (2001) Supporting understanding through task and browser design. . In L. R. Gleitman and A. K. Joshi (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Wiley, J. & Myers, J. L. (2003) Availability and Accessibility of Information and Causal Inferences from Scientific Text. *Discourse Processes. 36*, 109-129.

Wiley, J. & Voss, J. F. (1999) Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology,91*, 301-311.

Winne, P. H. , & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts and P. Pintrich (Eds.), *Handbook of Self-Regulation, pp. 531-566.* New York: Academic Press, Inc.

Zimmerman, B. J. & Schunk, D. H. (2001). *Self-Regulated Learning and Academic Achievement*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Author Notes

Appendix: Glenberg Passages from Weaver (1990)

Enhancing the Apparent Size of the Listening Space

Recognizing the fact that most of the sound heard at a live performance reaches the listener not directly from the sound source--the musicians' instruments-- but by way of reflections from the walls and ceiling, Dr. Bose fashioned a loudspeaker to disperse the sound in a similar way. In consequence, the Bose 901 radiates most of its sound not toward the listener but toward the wall in back of the speaker and toward the sides of the room. Where the rearward and sideways sound hits the walls, it reflects as from a mirror. Surprisingly, this changes the apparent size of the listening room. Compared to hearing sounds directly from the source, hearing sounds reflected from the walls of a room increases the apparent size of the room. This curious effect is more easily understood by analogy to light reflections. Suppose you look at yourself in a mirror from a distance of five feet. Your reflection then appears not in the plane of the mirror itself; rather, your image seems to be standing five feet behind the mirror. Likewise, the sound reflections produced by the Bose 901 seem to be coming not from the wall but from behind the wall. This gives rise to the impression of sitting in a larger space than the actual dimensions of the listening room. Such spatial enlargement contributes to the illusion of hearing music in the kind of acoustic ambiance for which most music was intended.

The Dilemma of Nuclear Disarmament

Advocates of Universal Nuclear Disarmament need to realize that the risks of nonnuclear war may be generated by total removal of nuclear arms. The question of whether global nuclear disarmament would increase the likelihood of non-nuclear war must be asked, especially since some forms of nonnuclear war may be as threatening as the danger averted by universal nuclear disarmament. Now, when all superpowers have nuclear weapons, war between those powers may become a nuclear confrontation in which all participants will suffer terrible losses. For this reason, the existence of nuclear weapons acts to deter war. However, the existence of nuclear weapons also is dangerous. However, in advocating steps to deflect an obvious danger, we need to be sure that the risks of other dangers had not been increased. Would it be better if elimination of the nuclear danger increased the risk of nonnuclear war? Note

that believing this does not require one to believe that nuclear arsenals should continue to grow. One might hope that an adequate and reasonably cheap nuclear deterrent could be established at a level well below current levels. The nuclear deterrent is not an ideal solution to the problem of avoiding war between the superpowers. It is certainly dangerous, but all other solutions, except being sensible, which we seem to be unwilling to try, are more dangerous.

When to Borrow?

It has been observed that in the lives of many persons, acquisition of a car is the second most costly single purchase. It is sensible, therefore, to consider the means of payment carefully. Assuming that the potential buyer has the financial means to pay the full cost at the time of purchase, would this be the prudent action? Or should the purchaser finance the acquisition with the help of funds borrowed from a lending institution? For some people, the answer to this question will be determined by personal style. For example, there are people who are very uncomfortable when in debt. But let us set aside subjective considerations, and deal with the question purely as a financial matter. From the financial perspective, two factors must be considered: (a) the interest charged by the lender, that is, the cost of the borrowed funds; (b) the opportunities for earning a return if a sum equal to the contemplated loan is invested. If one could be certain that the amount earned would exceed the interest paid, then borrowing would be advisable. But, if as is often the case, the return on investment is uncertain, whether to borrow or to pay with available funds cannot be decided easily.

Thermal Characteristics of Earth Sheltered Homes

Because of the rising price of energy, consumers and engineers are investigating methods of containing inflation of home heating and cooling costs. One of the more intriguing ideas uses earth sheltered (usually three sides and the roof of the structure) or earth bermed (three sides) houses. Thermal characteristics of earth sheltered structures provide three benefits. First, building material of the earth sheltered house are typically massive, usually concrete. These structural members act as a thermal mass to store heat (from the sun, for example) for long periods of time. Second, the earth protects the house from winds and associated effects such as wind chills and draft of cold air into the house. The most important benefit comes from the fact that the energy needed to change the temperature of a closed space depends on the difference between the desired temperature inside and the temperature of the environment immediately surrounding the enclosed space. Even at depths as
modest as a few feet, the ground maintains an almost constant temperature of about 40-50 degrees. Thus in winter, the inside temperature of the earth sheltered house need only be raised 20-30 degrees from that of the surrounding environment compared to 70 or more degrees for the above-grade house. In the summer, the earth environment provides a natural cooling system for the earth-sheltered house.

The Correlational Method and Causation

The correlational method involves careful observation of two or more variables to determine if they are related, but neither variable is imposed on the organisms being observed. Two examples of the correlational method are (1) a medical researcher who observes frequency of drinking alcohol (one
variable) and digestive tract cancer rates (a second variable); and (2) a consumer advocate who observes automobile types and frequency of breakdowns. No analysis applied to data collected by correlational methods can determine if one of the variables causes the observed changes in the other variable, however. Simply observing a relationship between alcohol consumption and cancer rate, for example, cannot justify the claim that alcohol causes cancer. It might be the case that cancer causes (either physiologically or psychologically) people to drink. Also, it might be the case that a third variable (e.g. stress) that may not have been observed causes similar changes in both variables.

Determining the Truth and Falsity of Scientific Theories

Scientists work in a curious, seemingly backward way, to develop and test theories. The layman might suspect that it is possible to prove a good scientific theory true. Nonetheless, the collection of facts and observations cannot be used to guarantee the absolute truth of a theory. Therefore, data (observations) consistent with a scientific theory do not prove the theory true; however, data that are inconsistent with a theory can certainly prove it false. Theories make predictions about what will occur when a given set of conditions arise. For example, a theory can be used to predict the outcome of an experiment before it is actually conducted. Once the data from the experiment are

collected, they can be used to evaluate the theory. There are two possibilities. If the prediction and the data are in agreement, then the theory is supported, but not necessarily proven true (other theories could also have predicted the outcome of the experiment and they can't all be true). If the theory and the data do not agree, the theory is logically proven false. Since the theory made an incorrect prediction, something is wrong with the theory. Thus, although scientific theories cannot be proven true, they can be proven false. This has led some philosophers of science to argue that scientists should attempt to prove their theories false, rather than attempting to prove them true.

Viruses and Cancer

Despite the fact that today viruses are known to cause cancer in animals and in certain plants, there exists great reluctance to accept viruses as being of importance in human cancer. Since there is no evidence that human cancer is infectious, many persons contend that because viruses are infectious agents they cannot possibly be significant causal agents in the development of human cancer. However, viruses can mutate and examples are known in which a virus that never kills its host can mutate to form a new strain of virus that always kills its host. It does not seem unreasonable to assume that an innocuous latent virus might mutate to form a strain that causes cancer. Add to this argument the general observation that biological phenomena generally do not differ strikingly as one goes from one species to another. If viruses have been implicated in the development of cancer in lower species it is plausible to consider viral infection as an explanation of human cancer.