

An evaluation of empirical measures of source identification

KEVIN MURNANE

Pennsylvania State University, University Park, Pennsylvania

and

UTE J. BAYEN

University of Memphis, Memphis, Tennessee

Source identification refers to memory for the origin of information. A consistent nomenclature is introduced for empirical measures of source identification which are then mathematically analyzed and evaluated. The ability of the measures to assess source identification independently of identification of an item as old or new depends on assumptions made about how inconsistencies between the item and source components of a source-monitoring task may be resolved. In most circumstances, the empirical measure that is used most often when source identification is measured by collapsing across pairs of sources (sometimes called "the identification-of-origin score") confounds item identification with source identification. Alternative empirical measures are identified that do not confound item and source identification in specified circumstances. None of the empirical measures examined provides a valid measure of source identification in all circumstances.

If a few days from now you remember reading an article about empirical measures of source identification and then try to remember in which of the many journals on your shelf you read it, you are facing a task that requires source memory. *Source memory*, or source identification, refers to memory for the origin of information. Questions regarding source identification have attracted increased attention in many fields of psychology, following work by Johnson and her colleagues, which has made use of reality- and source-monitoring experimental paradigms (see, e.g., Foley, Johnson, & Raye, 1983; Johnson, Kahan, & Raye, 1984; Johnson & Raye, 1981; Johnson, Raye, Foley, & Foley, 1981). This widespread interest in source identification highlights the need for appropriate measures of memory for the source of information. The ideal measure of source identification would be sensitive to differences in the ability to identify sources of information without being sensitive to differences in the ability to remember the information itself. In this paper we will sketch a computational-level theory (cf. Marr, 1982) of the memory and decision processes involved in the performance of a source-monitoring task. This theory will then be used to analyze several of the empirical measures that have been used to assess source memory. Whether

individual measures provide adequate measures of source identification depends on assumptions made about how inconsistencies produced by the item and source processes involved in source monitoring are resolved. We will show that most of the measures that are commonly used confound memory for the source with memory for the item in most circumstances. We will also identify a set of assumptions under which two rarely used measures of source memory are not affected by memory for the item. Finally, we will identify a different set of assumptions under which all of the analyzed measures confound item memory and source memory. We begin with a brief survey of the source-monitoring literature. A comprehensive review can be found in Johnson, Hashtroudi, and Lindsay (1993).

Source memory is usually tested with a source-monitoring paradigm in which subjects are asked if items of information have either been presented previously by one of several sources or have not been presented before. Three basic types of source-monitoring paradigm have received attention (Johnson et al., 1993). In an external source-monitoring paradigm, the task is to remember which one of several sources outside of oneself (e.g., different people or different journal articles) provided certain information. Internal source monitoring refers to the discrimination of different kinds of self-generated information (e.g., discrimination of thoughts and dreams). In a reality-monitoring paradigm, one must determine whether remembered information came from an external source or was self-generated. The domains in which these three types of source-monitoring paradigm have been used are very diverse and include basic and applied memory research; cognitive, developmental, social, and personality psychology; neuropsychology; psychopathol-

Preparation of this article was supported by NSF Grant SBR-9319549, an Alumni Faculty Research Fellowship, and a Research Initiation Grant from The Pennsylvania State University to K.M. and a fellowship from the German National Fellowship Foundation to U.J.B. We thank Robert Belli, Rich Carlson, Ron Kinchla, Matt Phelps, Bob Seibel, and two anonymous reviewers for helpful comments on an earlier version of this manuscript. Correspondence concerning this article should be addressed to K. Murnane, Department of Psychology, University of Maryland, College Park, MD 20742-4411.

ogy; and psycholinguistics. Source memory paradigms have been used to address questions regarding the relationship between the source and the credibility of information (see, e.g., Begg, Anas, & Farinacci, 1992; Gruder et al., 1978; Hannah & Sternthal, 1984), eyewitness testimony (Lindsay, 1990; Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989), cryptomnesia (Brown & Halliday, 1991; Marsh & Bower, 1993), and the relationship between source-monitoring ability and certain personality traits (Durso, Reardon, & Jolly, 1985). Source monitoring has also been examined in different subject populations, including children (Foley & Johnson, 1985; Gopnik & Graf, 1988; Lindsay, Johnson, & Kwon, 1991), older adults (Dywan & Jacoby, 1990; Ferguson, Hashtroudi, & Johnson, 1992; Hashtroudi, Johnson, & Chrosniak, 1990), and various clinical populations such as patients with Alzheimer's disease (Mitchell, Hunt, & Schmitt, 1986), Parkinson's disease (Taylor, Saint-Cyr, & Lang, 1990), amnesia (Janowsky, Shimamura, & Squire, 1989; Schacter, Harbluk, & McLachlan, 1984; Shimamura & Squire, 1987), and schizophrenia (Harvey, 1985).

Source-monitoring experiments have also proven useful in the study of memory for attributes of information. In this type of study, sources differ in attributes such as input modality or language. At test, the subject is asked to name the attribute that characterized the targets at the time of presentation. Areas of interest have included memory for modality (Kelley, Jacoby, & Hollingshead, 1989; Light, LaVoie, Valencia-Laver, Albertson Owens, & Mead, 1992), memory for voice attributes (Palmeri, Goldinger, & Pisoni, 1993), and psycholinguistic questions regarding bilingual coding of information (for a review, see Gerard & Scarborough, 1989).

In this paper our concern is with the empirical measures that have traditionally been used in the investigation of source memory. Empirical measures are calculated directly from performance data without specification of a formal model of either the underlying memory or decision processes that give rise to performance in source-monitoring tasks. The field has recently witnessed the development of multinomial models that specify underlying decision models for source-monitoring tasks (Batchelder, Hu, & Riefer, 1993; Batchelder & Riefer, 1990; Bayen, Murnane, & Erdfelder, 1996). Multinomial models have the distinct advantage of including separate and independent parameters for measuring item identification, source identification, and several types of response bias. However, as is often the case when a new class of models is introduced into a field, controversy exists over the adequacy of current models (Batchelder, Riefer, & Hu, 1994; Bayen et al., 1996; Kinchla, 1994) and new models are being introduced (Batchelder et al., 1994; Bayen et al., 1996; Riefer, Hu, & Batchelder, 1994). We discuss this interesting class of models elsewhere (Bayen et al., 1996).

In the section that follows we will sketch a computational-level theory of the memory and decision processes involved in source monitoring and show how these processes are related to the data from a typical source-monitoring

task. In the third section of the paper we discuss the difference between experimental designs in which source memory is measured separately for each of two individual sources and designs in which source memory is measured by collapsing across pairs of sources. This is an important distinction in considering which measure of source memory to use. In the fourth section of the paper we analyze several empirical measures of source identification in terms of whether each is successful at separating item memory and source memory. Finally, the results are summarized and conclusions are drawn.

Memory and Decision Processes

Although the use of empirical measures of source memory necessitates assumptions about the memory and decision processes involved in performing a source-monitoring task, these assumptions are rarely examined in detail. In this section we will sketch what Marr (1982) called a computational-level theory of the memory and decision processes involved in source monitoring. Our goal is to provide a framework for the analysis of empirical measures of source memory that is sufficiently general to be consistent with a wide range of algorithmic-level theories that specify memory and decision processes in detail. We will first describe a typical source-monitoring task and stipulate some very general assumptions about the memory and decision processes involved in the performance of this task. We will then show that these assumptions regarding decision processes are consistent with theories of source monitoring that are based on either signal detection or threshold theories of decision. Finally, we will show how the data from a typical source-monitoring task are related to the memory and decision processes we describe.

In a typical source-monitoring experiment, a number of items (such as words, sentences, etc.) are presented to subjects by at least two different sources. These sources can be people, study lists, input modalities (e.g., auditory vs. visual) or, as in the reality-monitoring paradigm, a mixture of external and internal sources (e.g., perceived words vs. self-generated words). Although experiments with more than two sources have been carried out (Riefer et al., 1994), most studies have used a two-source paradigm. During learning, information from the two sources (which we will label A and B) may be blocked by source, partially blocked by source, alternated between sources, or randomly intermixed. After learning takes place, the subject is given a three-alternative recognition test. Test items (targets and distractors) are presented one at a time and the subject must respond whether the item (a) was originally presented by Source A, (b) was originally presented by Source B, or (c) is a new item that was not experienced during learning. The data that are gathered from a typical source-monitoring experiment can be summarized conveniently with the set of response frequencies shown in Table 1 (slightly modified from Batchelder & Riefer, 1990), where the Y_{ij} values are the frequencies of type j responses to type i items. The empirical measures we will examine are computed from the re-

Table 1
Data From a Typical Source-Monitoring Experiment

Source	Response		
	A	B	N
A	Y_{AA}	Y_{AB}	Y_{AN}
B	Y_{BA}	Y_{BB}	Y_{BN}
N	Y_{NA}	Y_{NB}	Y_{NN}

Note—A, Source A; B, Source B; N, distractor item; Y_{ij} , frequency of responses of type j to items of type i .

sponse frequencies shown in Table 1. Justification for the validity of the measures rests on assumptions concerning the memory and decision processes that are involved in performance of the source-monitoring task.

Marr (1982) introduced the idea that an information-processing system can be described at three levels. A description at the computational level specifies both the general nature of the information that is input to and output from the system, and the goal of the computation that the system completes. A description at the algorithmic level specifies both the form of the inputs and outputs and the processing mechanisms that carry out the computation. Most general memory theories and smaller grained source-monitoring theories are constructed at the algorithmic level. A description at the implementation level specifies how the processing mechanisms that carry out the computation are physically implemented.

At the computational level, source monitoring can be viewed as a complex memory task that is composed of two subtasks, memory for the item and memory for the source, as depicted in Figure 1. These two subtasks are described as separate memory tasks because each can be defined in terms of different information that must be retrieved from memory and each can be carried out separately from the other. The source task demands the retrieval of information about the origin of information. At the algorithmic level, the source task may be modeled as a local access process through which source information is recollected from the memory representation of the learning episode. The item task demands retrieval of information that indicates whether the test item was presented during the learning episode. At the algorithmic level, the item task may be modeled as either a local access recollective process or a local or global access familiarity process. Although the item and source tasks are combined in the typical source-monitoring task, either could be carried out in isolation from the other given an appropriately designed experimental paradigm. For example, the standard *old-new* recognition paradigm exemplifies the item task being carried out without the source task. Likewise, an experimental paradigm in which subjects were given only learned items at test and were simply asked to name the source of each test item would exemplify the source task without the item task.

Input on each test trial in a source-monitoring task is a retrieval cue that includes information about the test item, along with context information that defines the relevant learning episode (Humphreys, Wiles, & Dennis,

1994; Murnane & Phelps, 1994). The same cue may be used or different cues may be generated for the item and source components of the task. Output from memory in response to the retrieval cue(s) is used to reach a decision in the source-monitoring task. This output can be thought of as evidence that indicates whether the test item was experienced during the learning phase of the experiment (item evidence) and whether the test item was originally presented by Source A or Source B (source evidence). Both item and source evidence are assumed to vary in strength along continuous dimensions.

The memory processes that produce levels of item and source evidence must be supplemented by decision processes to produce a response in a source-monitoring task. Output from memory serves as input to the decision processes. As a component of each subtask, decisions must be reached about the status (old or new) and the source of the test item. The output from the memory and decision processes concerning the status and the source of the test item must then be combined to produce a single response that corresponds to one of the available response categories in the typical source-monitoring task. The production of this response is the goal of the source-monitoring task. We will refer to the separate decision processes that correspond to the item and source components of the task as Stage 1 or identification decisions and the decision processes that combine the two Stage 1 decisions into a response as a Stage 2 or response deci-

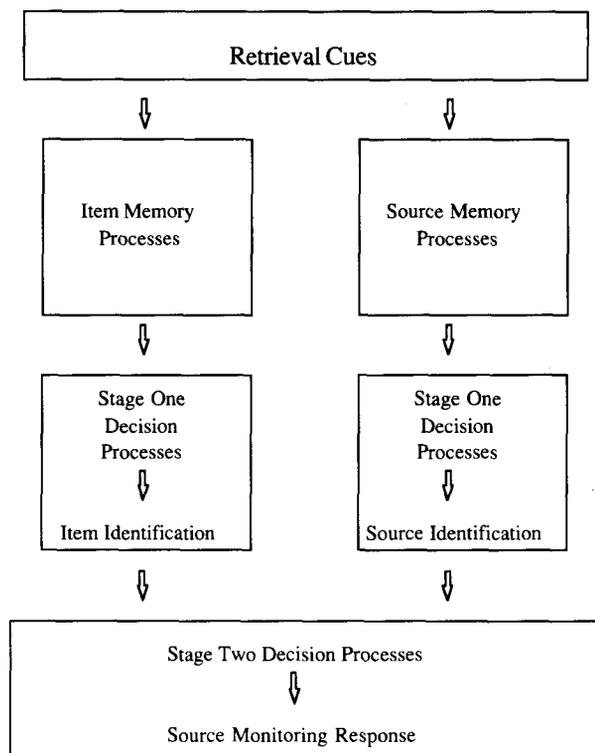


Figure 1. Computational-level sketch of the item and source processes involved in the performance of a source-monitoring task.

sion. Note that if the Stage 1 item decision is that the test item is identified as new, the subject may respond "neither" without completing the Stage 1 source processes.

It is important to remember that a theoretical sketch at the computational level does not imply specific processing mechanisms that accomplish the goal of the computation. Thus, our characterization of separate item and source subtasks and our graphical depiction of these tasks in Figure 1 should not be taken to imply that an algorithmic-level theory of source monitoring should include separate item and source processes that operate in parallel. An algorithmic-level theory of source monitoring may include separate and parallel item and source processes, but this is only one possibility among many. The analyses of empirical measures of source identification that follow are based on the computational-level theory just described and do not depend on how specific memory and decision processes are implemented at the algorithmic level.

There are four possible outcomes after Stage 1 decisions have been reached for both the item and source components of a source-monitoring task. These outcomes are shown in Table 2. Taking items that were originally presented by Source A as an example, let $p(O|S_A)$ represent the probability that the outcome of the Stage 1 item subtask indicates that the stimulus item is old, given that it is a target that was originally presented by Source A. We will refer to this probability as the probability that the stimulus item is identified as old given that it is an item from Source A. Likewise, $p(N|S_A)$ represents the probability that a stimulus item is identified as new, given that it was originally presented by Source A, and $p(A|S_A)$ and $p(B|S_A)$ represent the probabilities that a stimulus item originally presented by Source A is identified as having been presented by Source A and Source B, respectively. Critical factors that determine whether a test item is identified as old (or as from Source A) may affect memory processes, decision processes, or both.

Multinomial models of source monitoring are variants of the general class of threshold models that have been developed for simple detection and recognition paradigms (Batchelder & Riefer, 1990; Batchelder et al., 1994; Bayen et al., 1996). The development of these multinomial models has generated a debate over whether source monitoring is best modeled using threshold or signal detection theories of decision (Batchelder et al., 1994; Kinchla, 1994). Either type of theory could be used to model Stage 1 decision processes. To illustrate this point we will show how both types of decision theory could be involved in

the production of the probability that the test item is identified as old. Equivalent arguments can be given for the production of the probability that the test item is identified as having been presented by Source A.

In theories of item recognition that incorporate signal detection theories of decision, a test item is identified as old if the output from memory is greater than a criterion set by the subject. Criterion setting is usually assumed to be a function of the overlap between target and distractor distributions and any bias the subject might have to respond that the test item is old or new. In signal detection terms, $p(O|S_A)$ is the probability that the output from memory is greater than the response criterion when the test item is a target. The probability $p(O|S_A)$ is measured as the area of the target distribution to the right of the response criterion. Note that $p(O|S_A)$ does not differentiate between memory factors (the overlap of the target and distractor distributions) and decision factors (response bias) that may affect criterion placement.

In threshold theories of item detection, the output from memory is used to map stimuli into discrete cognitive states with some input probability, and cognitive states are mapped into the set of available response categories with some output probability (see, e.g., Macmillan & Creelman, 1991). For the case in which the stimulus is a target item, two distinct cognitive states are proposed in several threshold theories; test items are mapped into the detect state with probability q and are mapped into the guessing state with probability $1-q$. Items that are mapped into the detect state are identified as targets with probability 1.0; items mapped into the guessing state are identified as targets with a guessing probability, r . In threshold terms, $p(O|S_A)$ is equal to the sum of these two probabilities. That is,

$$p(O|S_A) = q + (1-q)r. \quad (1)$$

Note that in threshold theories, the probabilities of detecting and of guessing the status of the test item are estimated separately. The probability $p(O|S_A)$ combines these separate probabilities into a single value.

Once the status and source of the test item have been identified, a Stage 2 decision must be made to combine the output of the two identification processes into a single response category. The Stage 2 decision is straightforward when the test item is identified as old. If the test item is identified as old and the source is identified as A (or B), an "old item from Source A (or B)" response is made. The case is less clear when the test item is identified as new (and the source identification processes are completed), because the responses "new item from Source A (or B)" are not sensible in the context of the typical source-monitoring task and are not made available to the subject. When the test item is identified as new there is some probability that a "new" response is made and some probability that an "old item from Source A (or B)" response is made.¹ Thus the probabilities that an "old" response is given for a stimulus item of type i are given by

Table 2
Possible Outcomes After Stage 1 Decisions Have Been Reached in a Typical Source-Monitoring Task

	Source	Item Identification	
		Old	New
Source	A	OA	NA
Identification	B	OB	NB

Note—O, old; N, new; A, Source A; B, Source B.

$$p(R_A|S_i) = p(O|S_i) p(A|OS_i) + p(N|S_i) p(A|NS_i) \quad (2)$$

$$p(R_B|S_i) = p(O|S_i) p(B|OS_i) + p(N|S_i) p(B|NS_i), \quad (3)$$

where $p(R_A|S_i)$ and $p(R_B|S_i)$ are the probabilities that an "old item from Source A" and an "old item from Source B" response is made to a test item originally presented by Source i , $p(A|OS_i)$ and $p(B|OS_i)$ are the probabilities that a Source i item is identified as originating from Source A and Source B given that it has been identified as old, and $p(A|NS_i)$ and $p(B|NS_i)$ are the probabilities that a Source i item is identified as originating from Source A and Source B given that it has been identified as new.²

It may seem counterintuitive that an "old" response may be made when the test item is identified as new (i.e., that $p[A|NS_i] > 0$ and $p[B|NS_i] > 0$). This situation may occur, for example, when the evidence that the test item was originally presented by one of the sources is very strong and the evidence that the test item is new is very weak. The important consideration is that the two Stage 1 decisions are inconsistent when either source is identified and the item is identified as new. Although the Stage 2 decision may produce a "new" response in the majority of these cases, there is no a priori reason to assume that the inconsistency is always resolved with the decision to respond "new" at Stage 2. That is, there is no a priori reason to assume that $p(A|NS_i) = p(B|NS_i) = 0$ must be true.

The response frequencies in Table 1 can be treated as empirical estimates of the response probabilities given by Equations 2 and 3. The empirical measures we analyze are all derived from the four cases in which a target is presented and subjects correctly recognize it as having been present on the learning list. The equations for the four cases of interest are

$$p(R_A|S_A) = p(O|S_A) p(A|OS_A) + p(N|S_A) p(A|NS_A) \quad (4)$$

$$p(R_A|S_B) = p(O|S_B) p(A|OS_B) + p(N|S_B) p(A|NS_B) \quad (5)$$

$$p(R_B|S_B) = p(O|S_B) p(B|OS_B) + p(N|S_B) p(B|NS_B) \quad (6)$$

$$p(R_B|S_A) = p(O|S_A) p(B|OS_A) + p(N|S_A) p(B|NS_A). \quad (7)$$

We can now clearly see one of the problems involved in measuring source memory. The probability of correctly identifying the source of a previously studied item— $p(R_A|S_A)$, for example—is a function of both the item identification probabilities $p(O|S_A)$ and $p(N|S_A)$ and the conditional source identification probabilities $p(A|OS_A)$ and $p(A|NS_A)$. That is, performance in a typical source-monitoring task is clearly affected by both memory for the item and memory for the source. However, a great deal of the interest in source monitoring as a topic of research stems from an interest in memory for the source independent of memory for the item. Investigators have used a source-monitoring task that combines memory for the source and memory for the item as a means to examine source memory. To fulfill the needs of researchers who are interested in source memory, a measure of source identification must be computed from observed frequen-

cies and yet must not depend on the relevant item identification probabilities.

Comparing Performance Between Single Sources and Between Pairs of Sources

Independent variables in source-monitoring studies may be manipulated so that performance is compared between single sources or between pairs of sources. As we will show in the following section, this distinction is important for determining which measure of source memory is appropriate. Comparisons between single sources may be carried out using either within- or between-subjects designs. Single-source comparisons using a within-subjects design are often used when interest is focused on the effects of individual source attributes on source memory. In a typical experiment of this type, subjects learn items from two Sources, A_1 and B_1 , and source identification for items from Source A_1 is compared with source identification for items from Source B_1 . One value of the manipulated source variable is operationalized as an attribute of Source A_1 , and a different value of the manipulated source variable is operationalized as an attribute of Source B_1 . Main effects of source in an analysis of variance (ANOVA) are found by comparing source identification for items from Source A_1 with source identification for items from Source B_1 . For example, in a study of the effects of sensory modality on source memory (Light et al., 1992) subjects learned items from a visual source and an auditory source, and source identification was compared between the two sources.

Single-source comparisons using a between-subjects design are often used when interest is focused on the effects of between-subjects variables on memory for individual sources with specific attributes. In a typical experiment, items are learned from two Sources, A_1 and B_1 , performance for items from Source A_1 is compared across subject groups, and performance for Source B_1 is compared across subject groups. As is the case for single-source comparisons that use a within-subjects design, one value of the manipulated source variable is operationalized as an attribute of Source A_1 and a different value of the manipulated source variable is operationalized as an attribute of Source B_1 . Main effects of source in an ANOVA are found by comparing source identification for items from Source A_1 collapsed across the two subject groups with source identification for items from Source B_1 collapsed across the two subject groups. An example can be found in a study by Rabinowitz (1990) concerned with age differences in source memory. Groups of younger and older adults read words and generated words from word fragments. The effects of age on source memory were analyzed separately for the read and generate items.

Paired-source comparisons generally use a between-subjects design, and are often used when interest is focused on the effects on source memory of relational characteristics between sources. In a typical experiment subjects in one group learn items from two Sources, A_1 and B_1 ,

and subjects in a second group learn items from Sources A_2 and B_2 . Performance is measured by combining performance across the pair of sources in each group and source memory is examined by comparing performance in the A_1B_1 group with performance in the A_2B_2 group. One value of the manipulated source variable is operationalized as an attribute of the relationship between Sources A_1 and B_1 and a different value of the manipulated source variable is operationalized as an attribute of the relationship between Sources A_2 and B_2 . Main effects of source in an ANOVA are found by comparing source identification for items from Source A_1 collapsed with Source B_1 with source identification for items from Source A_2 collapsed with Source B_2 . For example, Ferguson et al. (1992, Experiment 1) examined the effects of the perceptual similarity of two sources on source memory. In the high-perceptual-similarity condition, items were presented by two female sources; in the low-perceptual-similarity condition, items were presented by a male source and a female source. Performance was compared across conditions by collapsing across the two sources in each condition.

Comparison of performance between single sources and between pairs of sources may be combined in a single study. However, care must be taken in how performance is measured for the two types of comparisons. As we will show in the analyses that follow, the single-source version of the most commonly used measure is an appropriate measure of source memory under one plausible set of assumptions, but the paired-source version of this measure is flawed in almost all circumstances.

SEPARATION OF SOURCE IDENTIFICATION AND ITEM IDENTIFICATION

In the analyses that follow we will examine a number of measures of source memory in terms of whether they successfully separate source identification and item identification. First, we examine the simplified cases that result when it is assumed that the subject always responds "new" when the Stage 1 item process identifies the test item as new. In this case $p(A|NS_A) = p(B|NS_B) = 0$ and Equations 4 through 7 reduce to

$$p(R_A|S_A) = p(O|S_A) p(A|OS_A) \quad (8)$$

$$p(R_A|S_B) = p(O|S_B) p(A|OS_B) \quad (9)$$

$$p(R_B|S_B) = p(O|S_B) p(B|OS_B) \quad (10)$$

$$p(R_B|S_A) = p(O|S_A) p(B|OS_A). \quad (11)$$

Analysis of this simplified case is sufficient to show that almost all of the commonly used measures of source identification confound item and source memory. We will then explore the more complex case in which $p(A|NS_A) > 0$ and $p(B|NS_B) > 0$.

We will introduce a related set of names for the measures we examine in an attempt to bring some order to a confusing situation. Many different measures are currently used to assess source memory and these measures

go by many different names. Often the same measure is known by several different names, some of which are descriptive of the nature of the measure and some of which are not. In addition, the same name is sometimes used to describe different measures. The difference between single-source and paired-source variants of measures is generally ignored even when both are used in analyzing data from a single study. Identification of the measure used in a particular study may necessitate trying to infer the measure from an ambiguous description or calculating possible measures until the measure used in the study is discovered. The terms *source identification measure* (SIM), *conditional source identification measure* (CSIM), and *average conditional source identification measure* (ACSIM) were chosen because each term describes the measure and indicates its relationship to the others.

The analysis for each measure will follow the same form. The general formula for the measure will be given, followed by an appropriate computational formula based on the observed frequencies from Table 1. We will treat the observed frequencies as estimates of response probabilities, and the underlying probabilities given in either Equations 4–7 or Equations 8–11 will be substituted; the resulting formulae will be algebraically reduced to determine whether the item identification probabilities can be eliminated. Although the analyses are based on the two-source case, the conclusions hold for any number of sources.

Source Identification Measure, False Source Identification Measure, and Frequency Counts

In this section we examine the source identification measure (SIM), the false source identification measure (FSIM), and frequency counts of the number of correct source identifications and the number of incorrect source identifications as potential measures of source memory. The analysis focuses on SIM but we will show that the basic conclusions hold for all of the measures. SIM is calculated as the number of correct source identifications divided by the total number of targets.

The general formula for SIM is

$$\text{SIM} = \frac{\sum_{i=1}^m Y_{ii}}{MN}, \quad (12)$$

where M equals the number of sources and N equals the number of items presented by each source. The computational formula for single-source SIM for Source A is

$$\text{single-source SIM} = \frac{Y_{AA}}{N}. \quad (13)$$

Treating the observed frequencies as estimates of response probabilities and substituting from Equation 8 gives

$$\text{single-source SIM} = \frac{p(O|S_A)p(A|OS_A)}{p(A|S_A) + p(B|S_A) + p(N|S_A)}, \quad (14)$$

which is an equation of the form

$$\text{single-source SIM} = \frac{xa}{1} = xa, \quad (15)$$

where x equals $p(O|S_A)$ and a equals $p(A|OS_A)$. Notice that x is a factor representing item identification and a is a factor representing source identification. It is apparent upon examination that x cannot be algebraically eliminated from Equation 15 and hence that source memory as measured by single-source SIM varies with changes in the probability that the test item is identified as old. In conditions in which the proportion of correctly recognized items for which the source is correctly discriminated remains unchanged, single-source SIM will change with changes in the probability of correctly identifying the status of the test item as old or new.

The computational formula for the paired-source version of SIM is,

$$\text{paired-source SIM} = \frac{Y_{AA} + Y_{BB}}{2N}. \quad (16)$$

Substituting from Equations 8 and 10 gives

$$\text{paired-source SIM} = \frac{p(O|S_A)p(A|OS_A) + p(O|S_B)p(B|OS_B)}{p(A|S_A) + p(B|S_A) + p(N|S_A) + p(A|S_B) + p(B|S_B) + p(N|S_B)}, \quad (17)$$

which has the form

$$\text{paired-source SIM} = \frac{xa + yb}{2}, \quad (18)$$

where x again equals $p(O|S_A)$, y equals $p(O|S_B)$, a equals $p(A|OS_A)$, and b equals $p(B|OS_B)$. In Equation 18 the x and y factors represent the item identification probabilities and the a and b factors represent the conditional source identification probabilities. It is again apparent upon examination that x and y cannot be algebraically eliminated from Equation 18 and hence that source memory as measured by the paired-source version of SIM varies with changes in the probability of item identification. Paired-source SIM will change with changes in the probability of correctly identifying either A or B items as old independent of changes in source identification.

There are several special cases in which both versions of SIM are not dependent on item identification, which we illustrate with the paired-source measure. If $x = y = 1.0$, then Equation 18 reduces to

$$\text{paired-source SIM} = \frac{(a + b)}{2}. \quad (19)$$

Thus, if item identification is perfect, it does not affect SIM. It is also the case that if either $x = y = 0$ or $a = b = 0$, then Equation 18 reduces to zero. None of these special cases are sufficiently interesting to warrant further discussion.

Several measures related to SIM have been used to measure source memory. We will discuss these measures

in relation to the paired-source variant of SIM. It is clear that all conclusions apply equally to single-source versions of the measures. FSIM is calculated as the number of incorrect source identifications (i.e., calling Source A items "B" and Source B items "A") divided by the total number of targets. The analysis of FSIM is identical to that of SIM, substituting Y_{AB} and Y_{BA} for Y_{AA} and Y_{BB} in Equation 16 and making substitutions based on Equations 9 and 11 in place of Equations 8 and 10 in Equation 17. Algebraic reduction gives an equation of the form

$$\text{paired-source FSIM} = \frac{xc + yd}{2}, \quad (20)$$

where c equals $p(B|OS_A)$, and d equals $p(A|OS_B)$. The conclusion is identical to that for paired-source SIM; the item identification probabilities x and y cannot be algebraically eliminated from Equation 20 (except in the special circumstances noted above), and thus source memory as measured by FSIM confounds item and source identification.

Some authors have used frequency counts of the number of correct source identifications (e.g., Anderson, 1984) or the number of incorrect source identifications (e.g., Brown & Halliday, 1991) as a measure of source memory. Marsh and Bower (1993) have criticized the latter measure on the basis of Monte Carlo simulations that demonstrated a negative correlation between the number of incorrect source identifications and item misses. The frequency measures for number of correct and incorrect source identifications are given by the numerators of Equations 18 and 20, respectively. It is obvious that the criticism leveled against SIM and FSIM applies to frequency counts as well; except under the special circumstances noted above, frequency counts also confound item and source identification.

On the basis of the foregoing analysis, we suggest that both the single-source and the paired-source variants of SIM, FSIM, and frequency counts of either the number of correct or the number of incorrect source identifications are all flawed as measures of source memory. With few exceptions, all of these measures vary with changes in item identification when the proportion of correctly recognized items for which the source is correctly discriminated remains constant. The exceptions are not of sufficient interest to justify use of any of these measures in source-monitoring studies.

Conditional Source Identification Measure and Conditional False Source Identification Measure

Marsh and Bower (1993) and Durso et al. (1985) argued that measuring source memory conditional upon correct item identification can avoid confounding source and item identification. The conditional measure most frequently used to measure source memory is CSIM, which is defined as the proportion of correctly recognized items for which a correct source judgment is made. However, it has also been suggested (Batchelder & Riefer, 1990; Johnson, Foley, & Leach, 1988; Rabinowitz, 1990)

that source memory as measured by CSIM may be confounded with item identification when recognition differs over the conditions in an experiment. The analysis that follows will clarify the circumstances under which both claims about the ability of CSIM to separate item and source identification are accurate.

The general formula for CSIM is

$$CSIM = \frac{\sum_{i=1}^m Y_{ii}}{\sum_{i=1}^m \sum_{j=1}^m Y_{ij}}, \quad (21)$$

where m indexes source. In a number of studies, CSIM calculated for single sources has been used to measure source memory (e.g., Anderson, 1984; Cohen & Faulkner, 1989; Light et al., 1992; Rabinowitz, 1989; Voss, Vesonder, Post, & Ney, 1987). If we use Source A as an example, the computational formula for single-source CSIM is

$$\text{single-source CSIM} = \frac{Y_{AA}}{Y_{AA} + Y_{AB}}. \quad (22)$$

The sum of the frequencies in the denominator of Equation 22 serves as an estimate of the probability of correctly identifying an item from Source A as old, that is, $p(O|S_A)$. Treating the observed frequencies as estimates of response probabilities and substituting from Equation 8 gives

$$\text{single-source CSIM} = \frac{p(O|S_A)p(A|OS_A)}{p(O|S_A)}, \quad (23)$$

which reduces to an equation of the form

$$\text{single-source CSIM} = \frac{xa}{x} = a, \quad (24)$$

where, as previously, x is the item probability $p(O|S_A)$ and a is the conditional source probability $p(A|OS_A)$. The elimination of x from Equation 24 means that changes in item identification have no effect on single-source CSIM and thus that single-source CSIM is an appropriate measure of source memory when comparisons are made between individual sources.

CSIM is also commonly used to measure source memory in studies that involve paired-source comparisons (e.g., Ferguson et al., 1992; Foley, Aman, & Gutch, 1987; Foley & Johnson, 1985; Johnson, Raye, Foley, & Kim, 1982; Markham, 1991; Palmeri et al., 1993). The paired-source version of CSIM is sometimes called an identification-of-origin score (e.g., Batchelder & Riefer, 1990; Finke, Johnson, & Shyi, 1988; Johnson et al., 1988) and is calculated as the sum of the A and B items for which the source is correctly identified divided by the number of correctly recognized old items. The computational formula for paired-source CSIM is

$$\text{paired-source CSIM} = \frac{Y_{AA} + Y_{BB}}{Y_{AA} + Y_{AB} + Y_{BB} + Y_{BA}}. \quad (25)$$

Treating the observed frequencies as estimates of response probabilities and substituting the probabilities from Equations 8 and 10 in the numerator and $p(O|S_A)$ and $p(O|S_B)$ in the denominator of Equation 25 gives

$$\text{paired-source CSIM} = \frac{p(O|S_A)p(A|OS_A) + p(O|S_B)p(B|OS_B)}{p(O|S_A) + p(O|S_B)}, \quad (26)$$

which is an equation of the form

$$\text{paired-source CSIM} = \frac{xa + yb}{x + y}, \quad (27)$$

where, as previously, x and y are the item identification probabilities $p(O|S_A)$ and $p(O|S_B)$, respectively, and a and b are the conditional source identification probabilities $p(A|OS_A)$ and $p(B|OS_B)$. It is clear that the item identification factors x and y cannot be algebraically eliminated from Equation 27 and thus, with the exceptions noted below, paired-source CSIM confounds item and source identification.

Paired-source CSIM provides a measure of source memory that is independent of item identification in two special cases that are important for empirical studies of source memory. When the probabilities of correctly identifying the source of correctly recognized items are identical, that is, when $a = b$,

$$\text{paired-source CSIM} = \frac{xa + yb}{x + y} = \frac{a(x + y)}{x + y} = a. \quad (28)$$

In this case, paired-source CSIM is equivalent to single-source CSIM (see Equation 24). This is a sensible measure of source memory and it does not vary with differences in the probabilities of correctly identifying A items, B items, or both A and B items as old. The second special case exists when the probabilities of correctly identifying A items and B items as old are equal. That is, when $x = y$,

$$\text{paired-source CSIM} = \frac{xa + yb}{x + y} = \frac{x(a + b)}{2x} = \frac{a + b}{2}. \quad (29)$$

This is also a sensible measure of source memory that does not vary with changes in the probabilities of correctly identifying A items, B items, or both A and B items as old.

As noted in the introduction to this section, Marsh and Bower (1993) and Durso et al. (1985) claimed that it is possible to avoid confounding memory for the item with memory for the source by measuring source memory conditional upon item identification. For CSIM, this claim is correct if single-source CSIM is used when comparisons are drawn between individual sources or if paired-source CSIM is used when comparisons are drawn between pairs of sources and either $a = b$ or $x = y$. Batchelder and Riefer (1990), Johnson et al. (1988), and Rabinowitz (1990) claimed that CSIM may confound item and source identification when item identification differs over the con-

ditions in an experiment. This claim is correct if paired-source CSIM is used and $a \neq b$ or $x \neq y$. In the latter case it is important to notice, however, that the critical comparison is not between levels of item identification across experimental conditions, but is rather between levels of item identification for the two sources within an experimental condition. If $a = b$ or $x = y$ in all conditions of an experiment, the levels of item identification over conditions in a paired-source experiment do not affect paired-source CSIM.

It is also important to notice that the two cases in which paired-source CSIM provides an acceptable measure of source memory (Equations 28–29) describe circumstances that may be instantiated in a source-monitoring experiment. Published data analyses from paired-source experiments tend to focus on between-condition comparisons, whereas within-condition examinations of either item identification for each source (i.e., x and y) or source identification for each source (i.e., a and b) are often ignored or at least not addressed with data analyses. However, the interest in between-condition comparisons leads to a tendency to try to equate some within-condition factors. Although published studies must be addressed on a case-by-case basis, practices such as sampling Source A and Source B items from a common pool would tend to produce conditions in which item identification is roughly equivalent for both sources within a condition (other factors being equal), thereby raising the possibility that paired-source CSIM may be used as a valid measure of source memory.

Marsh and Bower (1993) used a conditional false source identification measure (CFSIM) to measure source memory, defined as the proportion of correctly recognized items for which an incorrect source judgment is made. As was the case for FSIM and SIM, the analysis of CFSIM is identical to that of CSIM. We illustrate with the paired-source variant of CFSIM. Substituting Y_{AB} and Y_{BA} for Y_{AA} and Y_{BB} in Equation 25 and making substitutions based on Equations 9 and 11 in place of Equations 8 and 10 in the numerator of Equation 26 leads to an equation of the form

$$CFSIM = \frac{xc + yd}{x + y}, \tag{30}$$

where, once again, c equals $p(B|OS_A)$, and d equals $p(A|OS_B)$. It is clear that CFSIM has the identical strengths and weaknesses of CSIM; except under the special circumstances discussed above, CFSIM varies with changes in item identification independently of changes in source identification.

When comparisons are drawn between individual sources, single-source CSIM and CFSIM provide measures of source memory that are unaffected by the level of item identification for each source. However, when comparisons are drawn between pairs of sources, paired-source CSIM and CFSIM are flawed because they vary with changes in item identification in conditions in which the proportion of correctly identified sources does not change. If it can be established that either $x = y$ for all

sources in all experimental conditions or that $a = b$ for all sources in all experimental conditions, paired-source CSIM and CFSIM provide valid measures of source memory. Establishing the validity of paired-source CSIM and CFSIM complicates experimental design and involves carrying out additional data analyses to demonstrate that the conditions needed to justify using the measures were achieved in the experiment. It is fortunate that these extra complications can be eliminated through the use of an alternative measure that does not confound item and source identification under any of the circumstances we have examined thus far. We now turn to an analysis of that measure.

Average Conditional Source Identification Measure

ACSIM is the arithmetic mean of several single-source CSIMs. Anderson (1984) and Lindsay et al. (1991, Experiment 2) computed ACSIM, although these authors did not explicitly identify the measure, compare CSIM and ACSIM, or address the problem of sensitivity of the measures to changes in item identification. Although ACSIM has been occasionally used as a performance measure, to the best of our knowledge its properties as a measure of source memory have not been examined and its status as an empirical measure that is insensitive to differential levels of item identification under some conditions has not been pointed out (although see Bayen et al., 1996, for a use of ACSIM based on the arguments presented here).

The general formula for ACSIM is

$$ACSIM = \frac{\sum_{i=1}^m \left(\frac{Y_{ii}}{\sum_{j=1}^m Y_{ij}} \right)}{M}, \tag{31}$$

where M equals the number of sources. When there are two sources,

$$ACSIM = \frac{\frac{Y_{AA}}{Y_{AA} + Y_{AB}} + \frac{Y_{BB}}{Y_{BB} + Y_{BA}}}{2}. \tag{32}$$

Treating the observed frequencies as estimates of response probabilities and substituting the appropriate underlying probabilities gives

$$ACSIM = \frac{\frac{p(O|S_A)p(A|OS_A)}{p(O|S_A)} + \frac{p(O|S_B)p(B|OS_B)}{p(O|S_B)}}{2}, \tag{33}$$

which is an equation of the form

$$ACSIM = \frac{\frac{xa}{x} + \frac{yb}{y}}{2} = \frac{a+b}{2}, \tag{34}$$

where a and b are the conditional source identification probabilities $p(A|OS_A)$ and $p(B|OS_B)$, respectively. The item identification probabilities x and y can be algebraically eliminated from Equation 34 and hence ACSIM does not vary with changes in the level of item identification.

Comparison of Equations 29 and 34 shows that when the item identification probabilities for items from both sources are equal (i.e., $x = y$), paired-source CSIM is equivalent to ACSIM. As noted, comparison of Equations 24 and 28 shows that when the probabilities of correctly identifying the source of correctly recognized items are equal (i.e., $a = b$) paired-source CSIM is equivalent to single-source CSIM. Thus, the conclusions drawn from these analyses of CSIM and ACSIM under the assumptions that $p(A|NS_A) = p(B|NS_B) = 0$ are simple and straightforward. When comparisons are drawn between pairs of sources, ACSIM should be used to measure source memory; when comparisons are drawn between individual sources, single-source CSIM is the preferred measure. Both measures are sensitive to changes in the proportion of items for which the source is correctly identified and are not affected by changes in the proportion of items that are correctly identified as old or new.

Responding “Old” When the Test Item Is Identified as New

The foregoing analyses were based on the assumption that inconsistencies between Stage 1 item and source identifications are always resolved in favor of the decision that the test item is new, that is, that $p(A|NS_A) = p(B|NS_B) = 0$. We will now explore the consequences of the assumption that Stage 1 inconsistencies may be resolved in favor of the decision that the test item is old, that is, that $p(A|NS_A) > 0$ and $p(B|NS_B) > 0$.

Compare Equations 4 and 8:

$$p(R_A|S_A) = p(O|S_A)p(A|OS_A) + p(N|S_A)p(A|NS_A) \quad (4)$$

$$p(R_A|S_A) = p(O|S_A)p(A|OS_A). \quad (8)$$

When it is assumed that $p(A|NS_A) = 0$, it can be seen from Equation 8 that the response probability, $p(R_A|S_A)$, is a function of a single item probability, $p(O|S_A)$, and a single conditional source probability, $p(A|OS_A)$. In this case ACSIM and single-source CSIM provide unconfounded measures of source memory because they estimate $p(A|OS_A)$ independently of $p(O|S_A)$. However, when it is assumed that $p(A|NS_A) > 0$, it can be seen from Equation 4 that $p(R_A|S_A)$ is a function of two conditional source identification probabilities, $p(A|OS_A)$ and $p(A|NS_A)$. The contributions made by each of these conditional source identification probabilities to the probability of a “Source A” response is a function of the probability that the test item is identified as old. As $p(O|S_A)$ increases, $p(A|OS_A)$ makes a proportionally larger contribution to the response probability $p(R_A|S_A)$, and, because $p(N|S_A) = 1 - p(O|S_A)$, $p(A|NS_A)$ makes a proportionally smaller contribution to $p(R_A|S_A)$. The ideal empirical measure of source identification in a source-monitoring task would provide separate and unconfounded

estimates of each of these conditional source identification probabilities.

Let $a'_A = p(A|NS_A)$, $b'_A = p(B|NS_A)$, $a'_B = p(A|NS_B)$; and $b'_B = p(B|NS_B)$. Then, substituting response probabilities from Equations 4 and 6 and using x , y , a , and b as before, the empirical measures (single-source measures are given for Source A) have equations of the following forms:

$$\text{single-source SIM} = xa + (1 - x)a'_A, \quad (35)$$

$$\text{paired-source SIM} = \frac{xa + (1 - x)a'_A + yb + (1 - y)b'_B}{2}, \quad (36)$$

$$\text{single-source CSIM} = \frac{xa + (1 - x)a'_A}{x + (1 - x)(a'_A + b'_A)}, \quad (37)$$

$$\text{paired-source CSIM} = \frac{xa + (1 - x)a'_A + yb + (1 - y)b'_B}{x + (1 - x)(a'_A + b'_A) + y + (1 - y)(a'_B + b'_B)}, \quad (38)$$

$$\text{ACSIM} = \frac{xa + (1 - x)a'_A}{x + (1 - x)(a'_A + b'_A)} + \frac{yb + (1 - y)b'_B}{y + (1 - y)(a'_B + b'_B)}. \quad (39)$$

It is apparent that the item identification probabilities, x and y , that is, $p(O|S_A)$ and $p(O|S_B)$, cannot be eliminated from any of these equations; hence all of these empirical measures confound item identification with source identification. Although the formulae are not given, the same conclusion applies to the FSIM and CFSIM variants discussed earlier. Thus, under the assumption that Stage 1 inconsistencies may be resolved in favor of the decision that the test item is old, none of these empirical measures are capable of adequately measuring source memory in a typical source-monitoring task.

DISCUSSION

Many different empirical measures have been used to assess source memory and these measures have been known by many different names. We have introduced a consistent nomenclature for source memory measures that indicates both the nature of each individual measure and its relationship to other measures. In addition, we have pointed out that an often-ignored aspect of source-monitoring experimental designs is important for determining which measure should be used to assess performance in a source-monitoring task. Measures that may be appropriate when source memory is measured separately for each source in an experiment may be flawed when source memory is measured by collapsing across pairs of sources.

We analyzed a series of empirical measures of source memory with the help of a sketch of a computational-level theory of source monitoring. Each measure was analyzed

in terms of whether item identification is confounded with source identification in measuring source memory. The results of the analyses are straightforward. The most commonly used measure of source memory in paired-source designs, paired-source CSIM (also known as the identification-of-origin score), confounds item and source identification under almost all circumstances. If it is assumed that a "new" response is made whenever the result of Stage 1 item processes is that the test item is identified as new—that is, when $p(A|NS_A) = p(B|NS_B) = 0$ —then single-source CSIM and ACSIM provide acceptable empirical measures of source memory that do not confound item and source identification. If it is assumed that inconsistencies between Stage 1 item and source identifications can be resolved in favor of a Stage 2 decision to respond that the test item is old—that is, when $p(A|NS_A) > 0$ and $p(B|NS_B) > 0$ —then all of the empirical measures confound item and source identification.

How should source memory be measured in light of these findings? One strategy is to avoid the use of all empirical measures and analyze data using the parameters of a formal theory of source monitoring such as the one-high threshold multinomial model presented by Batchelder and Riefer (1990) or the two-high threshold multinomial model presented by Bayen et al. (1996). Another possible strategy is to rely on a fully developed and empirically supported theory of source monitoring that specifies how inconsistencies between item and source identification processes are resolved. Unfortunately, current theory development has not reached a point where there is agreement on this issue. In any case, the analyses presented in this article indicate that the theoretical implications of using a particular empirical measure of source memory should be given careful attention before that measure is used to measure performance in a source-monitoring task.

REFERENCES

- ANDERSON, R. E. (1984). Did I do it or did I only imagine doing it? *Journal of Experimental Psychology: General*, **113**, 594-613.
- BATCHELDER, W. H., HU, X., & RIEFER, D. M. (1993). Analysis of a model for source monitoring. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology*. New York: Springer-Verlag.
- BATCHELDER, W. H., & RIEFER, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, **97**, 548-564.
- BATCHELDER, W. H., RIEFER, D. M., & HU, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review*, **101**, 172-176.
- BAYEN, U. J., MURNANE, K., & ERDFELDER, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 197-215.
- BEGG, I. M., ANAS, A., & FARINACCI, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, **121**, 446-458.
- BROWN, A. S., & HALLIDAY, H. E. (1991). Cryptomnesia and source memory difficulties. *American Journal of Psychology*, **104**, 475-490.
- COHEN, G., & FAULKNER, D. (1989). Age differences in source forgetting: Effects on reality monitoring and on eyewitness testimony. *Psychology & Aging*, **4**, 10-17.
- DURSO, F. T., REARDON, R., & JOLLY, E. J. (1985). Self-nonsel-segregation and reality monitoring. *Journal of Personality & Social Psychology*, **48**, 447-455.
- DYWAN, J., & JACOBY, L. L. (1990). Effects of aging on source monitoring: Differences in susceptibility to false fame. *Psychology & Aging*, **5**, 379-387.
- FERGUSON, S. A., HASHTRUDI, S., & JOHNSON, M. K. (1992). Age differences in using source-relevant cues. *Psychology & Aging*, **7**, 443-452.
- FINKE, R. A., JOHNSON, M. K., & SHYI, G. C.-W. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, **16**, 133-137.
- FOLEY, M. A., AMAN, C., & GUTCH, D. (1987). Discriminating between action memories: Children's use of kinesthetic cues and visible consequences. *Journal of Experimental Child Psychology*, **44**, 335-347.
- FOLEY, M. A., & JOHNSON, M. K. (1985). Confusions between performed and imagined actions: A developmental comparison. *Child Development*, **56**, 1145-1155.
- FOLEY, M. A., JOHNSON, M. K., & RAYE, C. L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Development*, **54**, 51-60.
- GERARD, L. D., & SCARBOROUGH, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 305-315.
- GOPNIK, A., & GRAF, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, **59**, 1366-1371.
- GRUDER, C. L., COOK, T. D., HENNIGAN, K. M., FLAY, B. R., ALESSIS, C., & HALAMAJ, J. (1978). Empirical tests of the absolute sleeper effect predicted from the discounting cue hypothesis. *Journal of Personality & Social Psychology*, **36**, 1061-1074.
- HANNAH, D. B., & STERNTHAL, B. (1984). Detecting and explaining the sleeper effect. *Journal of Consumer Research*, **11**, 632-642.
- HARVEY, P. D. (1985). Reality monitoring in mania and schizophrenia. *Journal of Nervous & Mental Disease*, **173**, 67-72.
- HASHTRUDI, S., JOHNSON, M. K., & CHROSNIAK, L. D. (1990). Aging and source monitoring. *Psychology & Aging*, **4**, 106-112.
- HUMPHREYS, M. S., WILES, J., & DENNIS, S. (1994). Toward a theory of human memory: Data structures and access processes. *Behavioral & Brain Sciences*, **17**, 655-692.
- JANOWSKY, J. S., SHIMAMURA, A. P., & SQUIRE, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia*, **27**, 1043-1056.
- JOHNSON, M. K., FOLEY, M. A., & LEACH, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, **16**, 337-342.
- JOHNSON, M. K., HASHTRUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- JOHNSON, M. K., KAHAN, T. L., & RAYE, C. L. (1984). Dreams and reality monitoring. *Journal of Experimental Psychology: General*, **113**, 329-344.
- JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67-85.
- JOHNSON, M. K., RAYE, C. L., FOLEY, H. J., & FOLEY, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, **94**, 37-64.
- JOHNSON, M. K., RAYE, C. L., FOLEY, M. A., & KIM, J. K. (1982). Pictures and images: Spatial and temporal information compared. *Bulletin of the Psychonomic Society*, **19**, 23-26.
- KELLEY, C. M., JACOBY, L. L., & HOLLINGSHEAD, A. (1989). Direct versus indirect tests of memory for source: Judgments of modality. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 1101-1108.
- KINCHLA, R. A. (1994). Comments on Batchelder and Riefer's multinomial model of source monitoring. *Psychological Review*, **101**, 166-171.
- LIGHT, L. L., LA VOIE, D., VALENCIA-LAVER, D., ALBERTSON OWENS, S. A., & MEAD, G. (1992). Direct and indirect measures of memory for modality in young and older adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1284-1297.

- LINDSAY, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 1077-1083.
- LINDSAY, D. S., & JOHNSON, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, **17**, 349-358.
- LINDSAY, D. S., JOHNSON, M. K., KWON, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, **52**, 297-318.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MARKHAM, R. (1991). Development of reality monitoring for performed and imagined actions. *Perceptual & Motor Skills*, **72**, 1347-1354.
- MARR, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- MARSH, R. L., & BOWER, G. H. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 673-688.
- MITCHELL, D. B., HUNT, R. R., & SCHMITT, F. A. (1986). The generation effect and reality monitoring: Evidence from dementia and normal aging. *Journal of Gerontology*, **41**, 79-84.
- MURNANE, K., & PHELPS, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition*, **22**, 584-590.
- PALMERI, T. J., GOLDINGER, S. D., & PISONI, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 309-328.
- RABINOWITZ, J. C. (1989). Judgments of origin and generation effects: Comparisons between young and elderly adults. *Psychology & Aging*, **4**, 259-268.
- RABINOWITZ, J. C. (1990). Effects of repetition of mental operations on memory for occurrence and origin. *Memory & Cognition*, **18**, 72-82.
- RIEFER, D. M., HU, X., & BATCHELDER, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 680-693.
- SCHACTER, D. L., HARBLUK, J. L., & MCLACHLAN, D. R. (1984). Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning & Verbal Behavior*, **23**, 593-611.
- SHIMAMURA, A. P., & SQUIRE, L. R. (1987). A neuropsychological study of fact memory and source amnesia. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 464-473.
- TAYLOR, A. E., SAINT-CYR, J. A., & LANG, A. E. (1990). Memory and learning in early Parkinson's disease: Evidence for a "frontal lobe syndrome." *Brain & Cognition*, **13**, 211-232.
- VOSS, J. F., VESONDER, G. T., POST, T. A., & NEY, L. G. (1987). Was the item recalled and if so by whom? *Journal of Memory & Language*, **26**, 466-479.
- ZARAGOZA, M. S., & KOSHMIDER, J. W., III (1989). Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 246-255.

NOTES

1. We thank an anonymous reviewer for focusing our attention on this interesting and important case.
2. Response probabilities are given in this form rather than in the more familiar $p(R_A|S_i) = p(S_i) p(O|S_i) p(A|OS_i) + p(S_i) p(N|S_i) p(A|NS_i)$ because the source is always known when data analyses are being conducted. That is, $p(S_i) = 1$ for all data analyses.

(Manuscript received February 24, 1995;
revision accepted for publication June 7, 1995.)