On the Recollection of Specific- and Partial-Source Information

Chad S. Dodson, Paul W. Holland, and Arthur P. Shimamura
University of California, Berkeley

Memory judgments can be based on information that is more or less specific with respect to
the source of an item. The authors introduce a procedure and multinomial model for measuring
specific- and partial-source information. In 2 experiments, participants heard words spoken by
4 different voices: 2 male voices and 2 female voices. During the test, participants were
required to remember who spoke the test items (e.g., Male 1, Male 2, Female 1, Female 2, or
new word). Participants often remembered information about the gender of the source (i.e.,
partial-source information) when they did not remember information that identified the source
itself (i.e., specific-source information). Dividing attention during retrieval impaired partici-
pants' memory for specific-source information (i.e., voice information) but did not affect
memory for partial-source information (i.e., gender information).

A central goal of memory research has been to understand
and characterize the experience of recollection. For instance,
the exceptional aspect of Luria's patient, S, is that S seems
capable of remembering nearly all past events with amazing
clarity and vividness (Luria, 1968). An equally surprising
aspect of S, though, is the apparent absence of variability in
the vividness of his memories. In contrast, for most people,
memories often contain only fragments of information rather
than being complete records of our personal past. That is,
past experiences are remembered with varying degrees of
precision. Sometimes people remember exactly where, when,
and who said something. On other occasions, people recol-
lect a range of partial information, such as remembering the
approximate location of a parking spot, the approximate
temporal period of a desired article, or only the gender of
the person who said something. For instance, Huttenlocher and
colleagues have shown that when participants do not remem-
ber the precise date of an earlier event, their incorrect
responses are likely to cluster around the correct date
(Huttenlocher, Hedges, & Prohaska, 1992; see Nairne, 1991,
for related findings with spatial memory). In short, many
studies have shown that most people, unlike S, remember a
range of specific information about earlier events.

Memory for source information is a prominent aspect of
the experience of recollection (e.g., Johnson, 1992). The
term source refers to the memorial features, such as spatial,
temporal, and perceptual information, that together identify
the origin of a memory (Johnson, Hashtroudi, & Lindsay,
1993). Following the source-monitoring framework of
Johnson and colleagues, we believe that the specificity of
source information varies along a continuum such that some
events are remembered vaguely and other events are remem-
bered vividly (e.g., Johnson et al., 1993). We introduce a
procedure for measuring the specificity of information that
can contribute to memory performance. For the ease of
communication, we use the terms specific-source memory
and partial-source memory to refer to memorial information
that characterizes the origin of an event more and less
precisely, respectively. These terms are meant to describe
memory in a relative way, because specific and partial are
defined within the context of our experiment. We further
elaborate on these terms when we describe our experiments.

Although people recollect past events with varying de-
grees of precision, current methods of measuring the process
of recollection primarily focus on the overall amount of
source information that answers a particular discrimination.
For example, the process dissociation procedure of Jacoby
and colleagues investigates the contributions of recollection
and familiarity to performance (e.g., Jacoby, 1991). In the
two-list version of this procedure, participants study items
from two different lists, such as solving anagrams for List 1
items and solving word fragments (e.g., sw__d) for List 2
items (e.g., Dodson & Johnson, 1996). Recollection and
familiarity are measured with two different tests: On the
inclusion test, participants are instructed to call both List 1
and List 2 items "old" and to call nonstudied items "new"
(i.e., an old--new recognition test); on the exclusion test,
participants are instructed to call only the List 2 items "old"
and to call List 1 items and nonstudied items "new." The
recollection parameter measures the overall amount of
information that allows participants to correctly call List 1
items "old" on the inclusion test and to call List 1 items
"new" on the exclusion test. In the foregoing example,
either recollecting that the test item was earlier seen as an

Chad S. Dodson and Arthur P. Shimamura, Department of
Psychology, University of California, Berkeley; Paul W. Holland,
School of Education and the Department of Statistics, University of
California, Berkeley.

This research was supported in part by National Research
Service Award MH11135 F32 to Chad S. Dodson from the National
Institute of Mental Health and by National Institute of Health
Grants MH48757 and NS17778. We are grateful for the helpful
comments of Mike Ciranni, Lawrie Balfour, Mike Anderson and
Eliot Hirshman, and Steve Lindsay.

Correspondence concerning this article should be addressed to
Chad S. Dodson, Department of Psychology, University of Cal-
ifornia, Berkeley, 3210 Tolman Hall #1650, Berkeley, Cal-
ifornia 94720-1650. Electronic mail may be sent to chadd@
socrates.berkeley.edu.
anagram in List 1 or the test item’s familiarity can be used as a basis for calling the test item “old” on the inclusion test. On the exclusion test, however, only recollecting information about seeing the test item as an anagram in List 1 is sufficient for a correct response of “new.”

Although participants can recollect varying degrees of information about List 1 and List 2 items, only some of this information is measured by the recollection parameter. For instance, recollecting that the test item was earlier seen on the computer and not remembering whether the item was seen as an anagram in List 1 or as a word fragment in List 2 would not be measured by the recollection parameter. This partial-source information would contribute to the familiarity parameter. Recently, Gruppuso, Lindsay, and Kelley (1997), Mulligan and Hirshman (1997), and Yonelinas and Jacoby (1996) have argued for a distinction between memorial information that identifies the source of test items and memorial information that does not accomplish this task. The terms diagnostic and non-diagnostic recollection have been used to refer to these different kinds of memorial information (Mulligan & Hirshman, 1997). In short, the recollection parameter only measures the overall amount of diagnostic source information that is contributing to performance.

Another dominant approach to studying recollection has been to examine the processes that are involved in identifying the source of earlier studied items (e.g., Johnson et al., 1993). As embodied by the source-monitoring framework of Johnson and colleagues, studies of source monitoring have provided important insights into the role of different memory characteristics, different decision processes, aging, brain damage, and so forth on the ability to identify the origin of memories (e.g., Dodson & Johnson, 1993; Janowsky, Shimamura, & Squire, 1989; Johnson, Foley, & Leach, 1988; Multhaup, 1995; Raye, Johnson, & Taylor, 1980; Shimamura & Squire, 1991; see Johnson et al., 1993, for a review). With this approach, for example, participants might be asked to judge whether a test item was earlier spoken by a man or a woman or that it is a new item (e.g., Ferguson, Hashtroudi, & Johnson, 1992; Johnson, De Leonardis, Hashtroudi, & Ferguson, 1995). The source identification score, typically used to measure source memory performance in these studies, measures how often participants identify the correct source of test items given that they know the item is not a new item (e.g., Johnson et al., 1988). However, like the recollection parameter from the process-dissociation procedure, the source identification score measures the total amount of memorial information that identifies the source of past events (see Shimamura, Dodson, & Prinzmetal, 1997, for a fuller discussion of the similarity between the recollection parameter and measures of source memory from source tests). Neither of these measures can disentangle the separate contributions of memories of varying specificity. For instance, when judging whether a man or woman earlier spoke an item, participants may remember specific information about the idiosyncratic quality of the speaker’s voice or they may remember less specific information about the gender of the speaker. Both voice and gender information in this example would contribute to a correct response on the source test.

We desire a richer profile of people’s memory for past events than is provided by current methods. This article has two goals. First, we introduce an approach to studying source memory that systematically measures specific- and partial-source memory. We also introduce an extension of the multinomial models of Batchelder and Riefer (1990) that measures these different kinds of source memories. Our second goal is to demonstrate the value of this approach to examining memory by investigating the attentional demands of retrieving specific- and partial-source information. We show that dividing attention during retrieval impairs memory for specific-source information, but has no effect on the recollection of partial-source information.

Multiple-Source Similarity Paradigm

Our method of measuring memory for specific- and partial-source information uses multiple sources that vary in similarity to each other. For example, in our experiments participants hear words at study that are spoken by two different male voices (i.e., Anthony and Eric) and two different female voices (i.e., Julie and Lucy). Each source is more similar to the remaining same-gender source than to the other two different-gender sources (e.g., Anthony is more similar to Eric than to Julie or Lucy). At test, we present participants with old and new words and ask them to make a five-alternative forced choice: Is the test word new or was it spoken earlier by one of the four voices? This test contains three increasingly specific discriminations: (a) discriminating old words from new words, (b) identifying the correct gender of the speaker, and (c) identifying the correct speaker of an item. The last two discriminations divide source memory into two parts: (a) the amount of activated memorial information that successfully identifies the correct gender of the speaker but not the speaker and (b) the amount of memorial information that identifies the speaker. Although these thresholds imply that there are only two kinds of source memories, we believe (as we stated earlier) that memory for source information varies along a continuum of specificity. In fact, instead of dividing source memory into these two parts—gender and voice information—one could use stimuli that allow for more divisions and thus characterize source memory more finely.

In sum, we introduce an approach to studying source memory that separately measures memories that contain specific and partial information about the source of an item. The key element of our multiple-source procedure is that the sources are related to each other in an increasingly similar manner. These similarity relationships create multiple discriminations that allow us to measure memories of varying specificity. Because of the potential difficulty of interpreting scores from a source test, we have developed an extension of the multinomial models of Batchelder and Riefer (1990) to measure specific- and partial-source information. In the next section, we discuss the importance of multinomial models in general, and our model in particular.
Multinomial Modeling of Source Monitoring

Responses on a source test can reflect some combination of the influences of both the quality of the memorial information and the kind of decision processes used to evaluate this information (e.g., Johnson et al., 1993). Because of the difficulty of separating the influence of memory from the influence of certain decision processes like response bias, Batchelder, Riefer, and colleagues have advocated the use of multinomial models to disentangle the contributions of these two factors to performance (Batchelder, Hu, & Riefer, 1994; Batchelder & Riefer, 1990; Bayen, Murnane, & Erdfelder, 1996; Riefer & Batchelder, 1988; Riefer, Hu, & Batchelder, 1994). In the next section, we outline the standard multinomial model of source monitoring and then describe our multinomial model for measuring specific- and partial-source information.

In a typical source-monitoring experiment, participants learn information from different sources and later are required to remember the source of the test items. For instance, in one of our experiments (described in detail later), participants heard words spoken by four different sources, identified as Anthony, Eric, Julie, and Lucy. On the source test, words from all four sources were mixed with new distractor words. For each test word, participants made a five-alternative forced-choice: Is it a new word that was not heard earlier or was the word spoken by one of the four sources, and if so, by whom? In this example, the data from the source recognition test can be summarized in the confusion matrix in Table 1, in which the row headings correspond to the actual source of the test item and the column headings correspond to the participant’s response to the test item (for other analyses using confusion matrices, see Conrad, 1964; Homa & Viera, 1988; Zechmeister & McKillip, 1972). In Table 1, A, E, J, and L refer to the sources Anthony, Eric, Julie, and Lucy, respectively. The summary data for responses to the words from each source, such as Anthony, can be represented by the following probabilities: (a) the probability of correctly stating that a word spoken by Anthony came from that source (i.e., $P(\text{"A"} | \text{"A"})$), where "A" refers to the participant’s response of “Anthony”; (b) the probability of incorrectly stating that a word spoken by Anthony came from another source (i.e., $P(\text{"E"} | \text{"A"})$ or $P(\text{"J"} | \text{"A"})$ or $P(\text{"L"} | \text{"A"})$); and (c) the probability of incorrectly stating that a word spoken by Anthony was a new word (i.e., $P(\text{"N"} | \text{"A"})$).

<table>
<thead>
<tr>
<th>True source</th>
<th>&quot;Anthony&quot;</th>
<th>&quot;Eric&quot;</th>
<th>&quot;Julie&quot;</th>
<th>&quot;Lucy&quot;</th>
<th>&quot;New&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthony</td>
<td>$P(\text{&quot;A&quot;}</td>
<td>\text{&quot;A&quot;})$</td>
<td>$P(\text{&quot;E&quot;}</td>
<td>\text{&quot;A&quot;})$</td>
<td>$P(\text{&quot;J&quot;}</td>
</tr>
<tr>
<td>Eric</td>
<td>$P(\text{&quot;A&quot;}</td>
<td>\text{&quot;E&quot;})$</td>
<td>$P(\text{&quot;E&quot;}</td>
<td>\text{&quot;E&quot;})$</td>
<td>$P(\text{&quot;J&quot;}</td>
</tr>
<tr>
<td>Julie</td>
<td>$P(\text{&quot;A&quot;}</td>
<td>\text{&quot;J&quot;})$</td>
<td>$P(\text{&quot;E&quot;}</td>
<td>\text{&quot;J&quot;})$</td>
<td>$P(\text{&quot;J&quot;}</td>
</tr>
<tr>
<td>Lucy</td>
<td>$P(\text{&quot;A&quot;}</td>
<td>\text{&quot;L&quot;})$</td>
<td>$P(\text{&quot;E&quot;}</td>
<td>\text{&quot;L&quot;})$</td>
<td>$P(\text{&quot;J&quot;}</td>
</tr>
<tr>
<td>New</td>
<td>$P(\text{&quot;A&quot;}</td>
<td>\text{&quot;N&quot;})$</td>
<td>$P(\text{&quot;E&quot;}</td>
<td>\text{&quot;N&quot;})$</td>
<td>$P(\text{&quot;J&quot;}</td>
</tr>
</tbody>
</table>

Table 1
Summary of Response Outcomes From a Five-Alternative Forced-Choice Source Test

Note. $A =$ Anthony, $E =$ Eric, $J =$ Julie, $L =$ Lucy, and $N =$ new.
Figure 1. Tree diagrams for the standard-source multinomial model, with separate trees for items said by the Anthony source and for new items. \( D_1 \) = probability of detecting Anthony items as old; \( d_1 \) = probability of identifying the source of Anthony items; \( a_i \) = probability of guessing that a detected item is from source \( i \); \( b \) = probability of guessing an item is old; \( g_i \) = probability of guessing that an undetected item is from source \( i \). From "Multinomial Processing Models of Source Monitoring," by W. H. Batchelder and D. M. Riefer (1990), Psychological Review, 97, p. 551. Copyright 1990 by the American Psychological Association. Adapted with permission of the authors.

The \( a \) and \( g \) probabilities must satisfy the following constraint: \( a_1 + a_2 + a_3 + a_4 = 1 \) and \( g_1 + g_2 + g_3 + g_4 = 1 \).

In sum, from our perspective, the important characteristic of this model is that source memory is measured in an all-or-none manner. That is, like the source identification score and the process-dissociation procedure's measure of recollection, the \( d \) parameter only measures the overall amount of diagnostic source information that contributes to performance.

Partial-Source Multinomial Model

Figure 2 illustrates our partial-source model, which builds on the standard source memory models of Batchelder, Riefer, and Hu (e.g., Batchelder & Riefer, 1990; Riefer et al., 1994). The major difference between our partial-source model and the standard-source model described earlier is...
that we include a parameter that specifically measures memory for partial-source information. That is, we add an extra component to the source identification process, represented by the P-branch of the model tree (seen in Figure 2). The partial-source model reduces to the standard-source model when the value of P is 0 (i.e., when there is no discriminative partial-source information; compare Figures 1 and 2). The item detection process and the response biases are the same as indicated in Points 1 and 3 of the standard-source model just described.

In our partial-source model, if an old item is detected as "old," then its source, i, is either identified with probability $d_i$, or not, with probability $1 - d_i$. If not identified, partial information about source $i$ (in our experiment, the gender of source $i$) is recollected with probability $P_i$. When in this state of partial recollection of source $i$, the participant guesses the correct identity of source $i$, with probability $e_i$, and guesses (incorrectly) the other same-gendered source with probability $(1 - e_i)$. In our model, we make the natural assumption that responses made in the partial recollection state are governed by the same response bias process that operates when the participant neither identifies the source nor recalls the source's gender (i.e., it is governed by the $a$ parameters). The $e_i$s are related to the $a_i$s by being the relative values of the corresponding $a_i$s given that the participant is guessing between only two of the sources. For example, in our experiment Sources 1 and 2 are both men, so $e_1 = a_1/(a_1 + a_2)$ and $e_2 = 1 - e_1 = a_2/(a_1 + a_2)$. Similarly,
Sources 3 and 4 are both women, so \( e_3 = a_3/(a_3 + a_4) \) and \( e_4 = 1 - e_3 = a_4/(a_3 + a_4) \).

**Experiment 1**

Participants initially heard words spoken by four different voices. Later, participants completed a surprise source-memory test on which they tried to identify words as old or new, and if "old," to identify who had said the word during the study phase. To examine memory for partial-source information, we compared two different conditions. In the first condition, we expected that participants would remember partial-source information about the studied items when they could not remember specific-source information about the item. In the second condition, we reduced the effectiveness of such partial information. In the different-gender condition, participants heard two men and two women say the words; in the same-gender condition, participants heard four different men say the words. In the different-gender condition, when participants remember partial-source information about studied items, gender differences should emerge in the probabilities of their responses. For instance, if participants cannot remember the source of the studied word (i.e., the speaker), then recollecting partial-source information about the gender of the speaker will result in the selection of the same-gender incorrect response on the test. We expected the results from the same-gender condition to illustrate an all-or-none source identification process. In other words, when participants cannot remember the speaker of the word, there is no salient partial-source information in the same-gender condition that can discriminate between any of the alternative responses. In fact, this pattern was the case in our experiments.

**Method**

Participants. The participants were 42 paid volunteers who were students at the University of California, Berkeley. These participants were randomly assigned to either the different-gender or the same-gender condition.

Materials. The target materials consisted of 105 nouns. The words were divided into seven sets of 15 that were matched for length (5 letters) and frequency (\( M = 75 \)) (Kučera & Francis, 1967). A recording was made of six speakers, four men and two women, saying each of the 105 words. The words were recorded into an Apple Macintosh computer by using the SoundEdit (1992) program. In the experiment, the computer replayed these words to participants at volumes corresponding to normal conversational levels.

The sets of words were rotated in the experimental design so that each set was spoken by each of the six speakers and was also presented as new items on the forced-choice recognition test. Seven counterbalancing conditions were created to rotate the sets of words through the seven different sources (i.e., six speakers and new items). In the different-gender condition, participants heard words spoken by two men (Anthony and Eric) and by two women (Julie and Lucy), whereas in the same-gender condition, participants heard words spoken by four men (Anthony, Eric, Mike, and Tom).

The study list presented to each participant contained 70 words. However, the words in the first five and in the last five positions were buffers. The remaining 60 words on the study list were the target materials, with 15 words spoken by each voice. The words were randomly intermixed, with the constraint that no more than four words from one voice appeared consecutively. After all 60 words were presented, they were repeated in a different random order with the same constraint. On each study trial, participants first heard the spoken word, which was accompanied by an 8-cm \( \times \) 5.5-cm black-and-white schematic picture of the speaker, and then after a 1-s delay, the visual form of the word appeared on the screen. The schematic pictures were constructed with the Mac-a-Mug (1987) program.

The test list consisted of 75 words (60 old words randomly mixed with 15 new words) and was constructed so that no more than four words from each source appeared consecutively. In addition to the 75 words, there were another four words in the beginning of the test that were practice items and were not scored.

Procedure. We randomly assigned participants to either the different- or same-gender conditions, and tested them individually. At the beginning of the experiment, participants were told that we were interested in the processes that are involved in perceiving and imagining voices. No mention was made of a later memory test. Participants were told that they would hear words spoken by four different voices and that they were to imagine hearing the person say the word again.

To help participants imagine the voice of each speaker, participants initially listened to each speaker read a 30-s passage from a story. Participants were instructed to concentrate on the particular sound of each voice. They were also directed to examine a schematic picture of the speaker while listening to the passage.

After participants heard the passages, we informed them that they would then hear words spoken by each of the voices. After hearing each word, participants were directed to imagine hearing that word spoken by the speaker and to rate how easy or difficult it was to imagine the word being said by that voice. Participants were instructed to use the picture of the person as an aid when attempting to imagine the word being said by that person. The first five words were practice trials in order to demonstrate the task to the participants. All participants appeared to understand the instructions.

Finally, after the study phase, all participants were given a surprise memory test. They were informed that the test would contain both new words as well as old words that had been spoken by the four speakers. Participants were told to identify the speaker of the test word if it were one that had been heard earlier or to decide that the test word was new (i.e., the test phase was a five-alternative, forced-choice test). Each test word appeared in the center of the screen. In the different-gender condition, the response options of "Anthony," "Eric," "Julie," "Lucy," or "new" appeared two cm below the test word. In the same-gender condition, the response options were "Anthony," "Eric," "Mike," "Tom," or "new." In both conditions, a schematic picture of the source appeared underneath each voice's response option. The picture was 4 cm \( \times \) 3 cm and (except for its size) was identical to the schematic picture seen during the study phase of the experiment. The keyboard keys a, j, f, j, and n were labeled with the first letter of the name of the response option "Anthony," "Eric," "Julie," "Mike," "Lucy," "Tom," and "new," respectively. Participants were instructed to push the labeled keys corresponding to the alternative sources. After each response, the screen cleared and was then followed by a 1-s delay before the presentation of the next word.

**Results and Discussion**

The 75 responses for each participant were sorted into a 5 \( \times \) 5 table in which the rows corresponded to the true
source (Anthony, Eric, Julie/Mike, Lucy/Tom, or new) and the columns corresponded to the participant's response ("Anthony," "Eric," "Julie"/"Mike," "Lucy"/"Tom," or "new"). Appendix A reports the frequencies from these 5 × 5 tables summed across all of the participants for both the different-gender and same-gender conditions. Table 2 presents the corresponding response percentages for the five sources in the different- and same-gender conditions.

**Participant-based analyses.** For the purposes of comparability with some earlier source-monitoring experiments, we first analyzed the data using standard participant-level analyses (e.g., Johnson et al., 1988). That is, on the basis of the response frequencies in the 5 × 5 table for each participant, we formed contrast-like quantities to measure old–new recognition, source identification, and partial-source identification. These “contrasts” were conducted only on the responses to the Anthony and Eric stimuli because these stimuli were the only ones that were the same across the different- and same-gender conditions. Table 3 presents each of these scores as well as the false alarm rate to the new words. Appendix B contains the formulas for computing each of these scores.

Old–new recognition (the ability to distinguish studied items from new items) was assessed by examining the proportion of studied words that were attributed to a speaker regardless of source accuracy. As seen in Table 3, there was a higher old–new recognition rate of the words spoken by Anthony and Eric in the different-gender condition (.85) than in the same-gender condition (.75), t(40) = 2.58, p < .01. The false alarm rate to the new words was lower, but not significantly, in the different-gender condition (.14) than in the same-gender condition (.19), t(40) = 0.78.

We next examined participants’ ability to identify the source of previously heard words. Previous studies have relied on the identification of origin score as a measure of correct source identification (e.g., Finke, Johnson, & Shyi, 1988). This score refers to the proportion of items identified as “old” that were also attributed to the correct source (see Appendix B). As seen in Table 3, participants in the different-gender condition were better able to remember the speaker of words spoken by Anthony and Eric than were participants in the same-gender condition (.59 vs .36, respectively), t(40) = 3.70, p < .01.

Finally, we measured participants’ ability to remember partial-source information about words spoken by Anthony and Eric. As presented in Appendix B, we define partial-source information (in this paradigm) as the probability of correctly identifying the gender of the speaker of a studied word when participants selected the incorrect source. The partial source score measures how much more likely participants are to choose a correct gender response instead of an incorrect gender response when they know the item is “old” and they cannot remember the correct speaker. For instance, a score of 0 would mean that participants are equally likely to choose a response from either gender when they cannot remember the speaker.

As seen in Table 3, in the absence of memory for the correct source, participants were much more likely to select “Anthony” and “Eric” in the different-gender condition (.44) than in the same-gender condition (.11), t(39) = 3.85, p < .01. One participant was excluded from this analysis because she performed perfectly on the test. Because there were no mistakes for this participant, there was no way to measure memory for partial-source information. The score of .11 in the same-gender condition indicates there was a slight bias to respond “Anthony” and “Eric” when participants selected the incorrect source, although this score was not significantly different from chance responding (i.e., a score of 0), t(40) = 1.94, p > .05.

In short, participants were able to remember partial-source information about the gender of the speaker of previously studied items when they did not remember the speaker. Gender information guided responses in the different-gender condition but not in the same-gender condition, because all the words were spoken by men. In addition, increasing intralist similarity by having four male voices, as compared with two male and two female voices, reduced both the old–new recognition rate and the source identification rate.

**Model-based analyses.** Before outlining our model-analysis strategy, it is important to emphasize that the standard-source model is a special case of the partial-source model. When the P parameters of the partial-source model are set equal to 0, we obtain the standard-source model.
other words, the standard-source model is nested within the partial-source model. We exploit this fact in our use of these models to test hypotheses involving the parameters of both models. In the rest of this article, following Batchelder and Riefer (1990), we estimate all of the model parameters, using the method of maximum likelihood, and use differences in likelihood-ratio chi-squares (i.e., $G^2$) to obtain chi-square tests of null hypotheses. An example of such a null hypothesis is that the $D$ parameters for a common source across the two experimental conditions are equal.

There are two steps in our model-analysis strategy (see Dodson, Prinzmetal, & Shimamura, in press, for a fuller discussion of this strategy). In the initial "goodness-of-model-fit phase," we are interested in finding submodels of the partial-source model with the fewest number of parameters that fit the data in each experimental condition. Goodness of fit was measured using the log-likelihood statistic $G^2$, which assesses the degree to which the model's predicted response probabilities match the actual response probabilities. The value of $G^2$ is directly related to the difference between the predicted and actual probabilities. For instance, a small difference results in a small value of $G^2$ (see Bishop, Fienberg, & Holland, 1975; Riefer & Batchelder, 1988). In this initial phase we want to find values of $G^2$ whose corresponding $p$ values are large, and we require the $p$ values to be higher than .05. This would indicate that any differences between the model's predicted response probabilities and the actual response probabilities are nonsignificant.

In the second "parameter-testing phase," we have a good-fitting model for each experimental condition and now want to see whether the model can be further simplified by assuming equality between some of the parameters that appear in the different- and same-gender conditions, such as assuming that the $d$ (voice identification) parameters can vary independently of each other; in the other model both parameters are set equal to each other. If the model that allows both parameters to vary independently fits the data significantly better than does the model in which both parameters are set equal, then it is possible to conclude that the values of the parameters are significantly different. Conversely, if the fits of both models are not significantly different, then the parameters are not significantly different. This general procedure can be used to compare any nested models (i.e., any situation in which one model is a special case of another).

In the following analyses we used the standard-source model for the same-gender data because we deliberately constructed this condition so that there could not be any discriminative partial-source information that would guide participants' responses when they were unable to remember specific-source information about the test item. We used the partial-source model for the different-gender data.

Although our data set has enough degrees of freedom (i.e., 20) to use the complete 15-parameter standard-source model and the 19-parameter partial-source model, we used submodels of these models with the fewest number of parameters that fit the data (see Appendix C for the parameter values of the complete 15- and 19-parameter models). In the same-gender condition, we used a 10-parameter submodel of the standard-source model that was constructed by setting the parameters $D$ and $g$ to be equal for each voice (i.e., $D_{Anthony} = D_{Eric} = D_{Mike} = D_{Tom}; g_{Anthony} = g_{Eric} = g_{Mike} = g_{Tom}$). The parameters $d$ and $a$ were free to vary for each of the four voices (in addition to the parameter $b$). This submodel fit the data well, $G^2(10) = 8.05, p = .62$. In the different-gender condition, we used a 6-parameter submodel of the partial-source model that fit the data, $G^2(14) = 20.61, p = .11$. This submodel was formed by setting the parameters $D, P, d, a,$ and $g$ to be equal for the four different voices (along with the parameter $b$). See Appendix D for further thoughts on our partial-source model.

Table 4 presents the parameter estimates from the submodels for the different- and same-gender voices conditions. Overall, the parameter values for the two conditions parallel the participant-based analyses. Specifically, the value of $D$, which corresponds to old—new detectability, was significantly higher in the different-gender (.84) than in the same-gender condition (.69), $G^2(1) = 54.31, p < .01$. There was no significant difference in the values of $b$ (i.e., the false alarm rate to the new items) between the different- and same-gender conditions (.14 vs. .19, respectively), $G^2(1) = 3.40, p > .05$.

Similarly, the value of $d$ (i.e., source identification) was lower in the same-gender than in the different-gender condition. Because of the separate values of $d$ for each voice in the model of the same-gender condition, we analyzed the parameters for only the Anthony and Eric conditions.
voices in the two conditions (because these stimuli were identical in both conditions). Participants were better able to identify the source of the words spoken by Anthony and Eric in the different-gender than in the same-gender conditions, \( G^2(2) = 31.67, p < .01 \).

Of importance is that our partial-source model provides a formal method for measuring memory for partial-source information (i.e., gender information). In the same-gender condition, the value of 0 for the \( P \) parameter indicates that there was no discriminative partial-source information. Not surprisingly, the estimates of \( P \) were higher in the different-gender condition (.47) than in the same-gender condition (0), \( G^2(1) = 52.46, p < .01 \). Even when participants failed to remember the speaker of a word, they nevertheless remembered information that identified the gender of the speaker 47% of the time.

Finally, Table 4 indicates that there were different biases (i.e., the \( a \) parameters) for responding “Anthony” and “Eric” in the different- and same-gender conditions when participants were unable to remember any source information about the item. Specifically, participants had a higher tendency to respond “Anthony” in the different (.27) than in the same-gender (.21) conditions, \( G^2(1) = 6.63, p < .01 \). However, there was a lower tendency to say “Eric” in the different (.27) than in the same-gender (.36) conditions, \( G^2(1) = 13.33, p < .01 \). There was no difference in the biases for responding to any one of the voices when participants made a false alarm to a new item (i.e., the values of \( g \) between the two conditions). However, in both the different- and same-gender conditions, there was a bias to choose the two middle keys of the keyboard (i.e., “Eric” and “Julie” in the different-gender condition and “Eric” and “Mike” in the same-gender condition) rather than the two outer keys. This is shown in Appendix A by the higher frequency of responses to the sources associated with the two middle keys than to the two outer keys when participants failed to remember the correct response (i.e., the sum of the nonblack cells for each column). We take care of this bias in Experiment 2 by counterbalancing the sources that are associated with the middle and outer keys.

Experiment 2

The previous experiment demonstrates a method and model for separately measuring specific- and partial-source information. One domain to which we think this methodology can be applied usefully is in examining the role of attention during retrieval. Much research has shown that although old–new recognition performance is unaffected by dividing attention during retrieval, source identification is disrupted by this manipulation (e.g., Baddeley, Lewis, Eldridge, & Thomson, 1984; Craik, Govoni, Navich-Benjamin, & Anderson, 1996; Dodson & Johnson, 1996; Gruppuso et al., 1997; Jacoby, 1991; Mulligan & Hirshman, 1997). Researchers have interpreted this pattern of data as reflecting differences between the two kinds of memorial information (i.e., familiarity and source information) that drive performance on old–new recognition and source identification tests. The activation of familiarity is faster and relatively more automatic than the process of recollecting source information (e.g., Atkinson & Juola, 1974; Hintzman & Curran, 1994; Jacoby, 1991; Johnson, Kounios, & Reeder, 1994; Mandler, 1980). Presumably, dividing attention at retrieval has little effect on old–new recognition performance because this decision is based primarily on the contributions of overall familiarity. Judgments about the source of an item, however, are susceptible to the harmful effects of dividing attention because they rely on the recollection of source information.

It is not clear what the precise role of attention is during retrieval of source information. Some have speculated that attention is needed to establish and maintain a retrieval agenda or mode that sets the goal of remembering (e.g., Baddeley, 1982; Craik et al., 1996; Johnson, 1992; Moscovitch, 1994; Tulving, 1983). Once retrieval is initiated, the process of activating memorial information is relatively automatic (Baddeley et al., 1984; Moscovitch, 1994). To the extent that attention is needed to establish and maintain a retrieval agenda, dividing attention may disrupt the recovery of all source information, regardless of the specificity of the memorial information (i.e., memory for voice and gender).

Experiment 2 investigated the role of attention during retrieval of voice and gender information. The stimuli and procedure were very similar to those used in Experiment 1. Participants initially heard words spoken by two different women and two different men and then completed a surprise source test under conditions of either full or divided attention.

Method

Participants. The participants were 60 volunteers who were students at the University of California, Berkeley.

Materials. The stimuli were similar to those used in Experiment 1, except participants heard words spoken only by two women and two men. The target materials consisted of 75 nouns that were divided into five sets of 15 words that were matched for length (5 letters) and frequency (M = 75; Kučera & Francis, 1967). The sets of words were rotated in the experimental design so that each set was spoken by each of the four speakers (i.e., Anthony, Eric, Julie, and Lucy) and also served as new items on the test. Five counterbalancing conditions were created to rotate the sets of words through the five different sources. The study list contained 70 words, with the words in the first five and last five positions serving as buffer items. Of the remaining 60 target words, 15 words were spoken by each of the four voices. The test list consisted of 75 words (60 old words randomly mixed with 15 new words). In addition to the 75 words, there were another 10 words in the beginning of the list that were practice items and were not scored.

To control for a slight bias in Experiment 1 to choose the sources associated with the two middle keys of the keyboard, we rotated the sources associated with the middle keys of the keyboard with those sources associated with the outer keys of the keyboard. Specifically, for half the participants the keys \( a, j, j, \) and \( j \) were labeled with the first letter of the names \( \text{Anthony, Eric, Julie, Lucy} \), respectively, whereas for the remaining participants these keys were labeled with the first letter of the names \( \text{Eric, Anthony, Lucy, and Julie} \).

In the auditory distractor task in the divided-attention condition was similar to one used by Haist, Joyce, and Kutas (1995). Before each word on the source-memory test, participants heard a sc-
quence of five single-digit numbers (e.g., 5, 3, 2, 9, 0) and were instructed to remember the numbers by repeating them aloud. Then, participants were presented with a test word and made a source judgment. After making this judgment, participants heard one of the numbers from the sequence of five numbers (e.g., 2). Participants were instructed to remember the number that came before the heard number in the initial sequence (i.e., 3). After a response, participants were presented with a new sequence of five numbers (e.g., 8, 1, 0, 5, 7). The sequences of five numbers were formed by randomly choosing five numbers with the following two constraints: Each number in a sequence was unique and each sequence of five numbers was unique. In every other respect the materials were identical to those used in Experiment 1.

Procedure. Participants were tested individually and were randomly assigned to either the full-attention or divided-attention test conditions. The procedure for the study phase was similar to that used in Experiment 1. At the beginning of the experiment, participants were told that we were interested in the processes that are involved in perceiving and imagining voices. No mention was made of a later memory test.

After the study phase, all participants were given a surprise source-memory test. They were informed that the test would contain both new words and old words spoken by the four voices. The procedure for the full-attention condition was identical to that used in Experiment 1. In the divided-attention condition, participants were informed that they would do two tasks simultaneously. For one task, participants were informed that they would complete a source-memory test and were given the same instructions that the participants in the full-attention condition were given. Participants were informed that the second task was a short-term auditory memory test. They were given the following instructions:

Before seeing each test word you will hear five numbers (from 0 to 9). Please remember these numbers by continually repeating them aloud. After answering the source question for the test word, you will hear a number and you must remember the number that came before this number.

Participants were given practice on the task during the initial five test trials.

Results

The average accuracy on the auditory distractor task was 91%, with no individual performing worse than 75% correct. Moreover, the results of and conclusions from the source-memory data remain unchanged when these data are conditioned on correct performance on the distractor task. Therefore, we present and analyze the overall, unconditionalized results. Appendix E presents the response frequency matrices for the full- and divided-attention conditions. Table 5 displays the corresponding response percentages for the two conditions. These sections address the basic question, What is the effect of dividing attention during the act of retrieval on old–new recognition performance and specific-and partial-source identification performance?

Participant-based analyses. As seen in Table 6, manipulating the presence of attention at test had little effect on the recognition of previously studied items. Regardless of source accuracy, the probability of recognizing as “old” previously heard words was the same in the divided-attention condition (.84) and in the full-attention condition (.81), t(58) = .61. The false alarm rate to new items was nonsignificantly higher in the divided attention (.22) than in the full attention (.14) conditions, t(58) = 1.68. Because of the possibility that dividing attention is affecting old–new recognition by increasing the overall bias to call studied and new items “old,” we analyzed both the A’ discrimination and the corresponding B’b bias scores (Donaldson, 1992). In the analyses of both scores, there was no effect of attention either on A’ (.90 in the full-attention condition vs. .88 in the divided-attention condition, F[1, 58] < 1) or on B’b (.06 in the full-attention condition vs. -.17 in the divided-attention condition, F[1, 52] = 1.67).

Unlike old–new recognition performance, dividing attention during retrieval did affect source identification performance. Participants were less likely to remember specific source (voice) information in the divided-attention condition (.46) than in the full-attention condition (.56), t(58) = 2.52, p < .01. In contrast, there was no effect of dividing attention on memory for partial source (gender) information (.31 in both the divided- and full-attention conditions), t(58) = .03.

Model-based analyses. We used six parameter submodels of our partial-source model to analyze the data from both the full- and divided-attention conditions. See Appendix F for the parameter values from the unconstrained 19-parameter models for both conditions. The models were nonparametric in the divided attention (.22) than in the full attention (.14) conditions, t(58) = 1.68. Because of the possibility that dividing attention is affecting old–new recognition by increasing the overall bias to call studied and new items “old,” we analyzed both the A’ discrimination and the corresponding B’b bias scores (Donaldson, 1992). In the analyses of both scores, there was no effect of attention either on A’ (.90 in the full-attention condition vs. .88 in the divided-attention condition, F[1, 58] < 1) or on B’b (.06 in the full-attention condition vs. -.17 in the divided-attention condition, F[1, 52] = 1.67).

Unlike old–new recognition performance, dividing attention during retrieval did affect source identification performance. Participants were less likely to remember specific source (voice) information in the divided-attention condition (.46) than in the full-attention condition (.56), t(58) = 2.52, p < .01. In contrast, there was no effect of dividing attention on memory for partial source (gender) information (.31 in both the divided- and full-attention conditions), t(58) = .03.

Model-based analyses. We used six parameter submodels of our partial-source model to analyze the data from both the full- and divided-attention conditions. See Appendix F for the parameter values from the unconstrained 19-parameter models for both conditions. The models were nonparametric in the divided attention (.22) than in the full attention (.14) conditions, t(58) = 1.68. Because of the possibility that dividing attention is affecting old–new recognition by increasing the overall bias to call studied and new items “old,” we analyzed both the A’ discrimination and the corresponding B’b bias scores (Donaldson, 1992). In the analyses of both scores, there was no effect of attention either on A’ (.90 in the full-attention condition vs. .88 in the divided-attention condition, F[1, 58] < 1) or on B’b (.06 in the full-attention condition vs. -.17 in the divided-attention condition, F[1, 52] = 1.67).

Unlike old–new recognition performance, dividing attention during retrieval did affect source identification performance. Participants were less likely to remember specific source (voice) information in the divided-attention condition (.46) than in the full-attention condition (.56), t(58) = 2.52, p < .01. In contrast, there was no effect of dividing attention on memory for partial source (gender) information (.31 in both the divided- and full-attention conditions), t(58) = .03.

Model-based analyses. We used six parameter submodels of our partial-source model to analyze the data from both the full- and divided-attention conditions. See Appendix F for the parameter values from the unconstrained 19-parameter models for both conditions. The models were
constructed by setting the parameters $D, P, d, a,$ and $g$ to be equal for each voice (plus the $b$ parameter). Both submodels fit the data: For the full-attention data set, $G^2(14) = 12.21$, $p = .59$; for the divided-attention data set, $G^2(14) = 20.75$, $p = .11$.

As seen in Table 7, the results are very straightforward. The division of attention at retrieval had no effect on the old-new detectability parameter, $D$, $G^2(1) = .15$, nor on the partial-source (gender) identification parameter, $P$, $G^2(1) = .01$. However, voice identification, indexed by $d$, was much lower in the divided- than in the full-attention condition, (.19 vs. .35, respectively), $G^2(1) = 24.22$, $p < .001$. Also, the false alarm rate, $b$, was higher in the divided- than in the full-attention condition (.22 vs. .14), $G^2(1) = 8.80$, $p < .01$. Finally, although there was no difference in the values of the $a$ guessing parameters for the different sources, $G^2(1) = .68$, the $g$ guessing parameters were significantly different in the two attention conditions, $G^2(1) = 10.51$, $p < .01$.

Discussion

As in Experiment 1, when attention is full at test, participants remembered varying degrees of specific information about past events. When participants could not remember the correct voice, they remembered information that identified the correct gender of the voice. Dividing attention during retrieval had interesting and selective effects on memory performance. As in previous studies, dividing attention did not affect old-new recognition performance (e.g., Baddeley et al., 1984; Craik et al., 1996). In terms of source memory, however, divided attention during retrieval selectively disrupted access to memory for voice information and had no effect on the recovery of memory for gender information. In other words, the recollection of different kinds of source memories requires different amounts of attention, with more specific voice information needing more attention than less specific gender information.

Recently, both Gruppuso et al. (1997) and Mulligan and Hirshman (1997) found that dividing attention during retrieval impairs the retrieval of source information, its use, or both of these processes. Gruppuso et al. examined the effects of list similarity on the contributions of recollection and familiarity to the process-dissociation procedure’s inclusion and exclusion recognition memory tasks. In their easy discrimination condition, List 1 and List 2 items were judged in two different ways (i.e., judging the monetary value of List 1 items and judging the frequency of occurrence of List 2 items). Gruppuso et al. found that dividing attention during retrieval decreased the value of recollection and increased the value of familiarity. They suggested that dividing attention in this condition impairs the use of diagnostic source information in such a way that this information no longer contributes to the estimate of recollection and that instead it contributes to the estimate of familiarity. Alternatively, we believe that their results are consistent with our results that dividing attention disrupts the retrieval of specific source information but has less impact on memory for partial source information. That is, in terms of Gruppuso et al.'s task, dividing attention disrupts the retrieval of information about the specific judgment that was performed on the item (i.e., whether the item was judged at study in terms of monetary value or frequency), and thus reduces participants' ability to distinguish between List 1 and List 2 items. However, dividing attention does not disrupt the recovery of partial source information about having performed some kind of judgment on the item that contributes to estimates of familiarity in this paradigm.

Finally, an alternative account of our results is that dividing attention at retrieval affects memory of all source information in the following manner. Memory for voice information degrades into memory for gender information; similarly, memory for gender information degrades into nondiagnostic source information. This account requires the parametric assumption that dividing attention degrades memory for voice and gender information by exactly the same amounts in order to explain why the partial source parameter remains unchanged across the full- and divided-attention conditions. That is, the amount of voice information that degrades into gender information completely offsets the loss of gender information that has degraded into nondiagnostic source information. We are skeptical of this alternative account and are inclined to believe the first explanation, that only memory for voice information is affected by our divided attention task, for the following reasons: It is more parsimonious and does not require the parametric assumption of this latter account; second, the first account is consistent with existing data.

General Discussion

We introduce a method and multinomial model for measuring the separate contributions of specific- and partial-source information to performance. In both experiments participants initially heard two men (i.e., Anthony and Eric) and two women (i.e., Julie and Lucy) say words and were later asked to remember the source of test words (i.e., who said the word earlier). We used sources that vary in similarity to each other in order to measure the separate contributions of specific- and partial-source information to performance. For example, specific information about the particular quality of Anthony’s voice is necessary to discriminate the Anthony words from the Eric words, whereas only general information about the gender of the speaker is sufficient for distinguishing the Anthony source from the female sources. With this method we show that dividing

<table>
<thead>
<tr>
<th>Retrieval condition</th>
<th>Parameter</th>
<th>$G^2(14)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full attention</td>
<td>$D=.78$</td>
<td>$.41$</td>
</tr>
<tr>
<td>Divided attention</td>
<td>$D=.79$</td>
<td>$.40$</td>
</tr>
</tbody>
</table>

*p < .01; significantly different; $G^2(1) = 8.80$. **p < .000001; significantly different; $G^2(1) = 24.22$.
attention during retrieval impairs memory for voice information but does not affect memory for gender information.

A similar pattern of impaired memory for precise source information and preserved memory for partial-source information has occurred in studies investigating developmental differences in memory for source information (e.g., Ferguson et al., 1992; Johnson et al., 1995; Lindsay, Johnson, & Kwon, 1991). In studies by Ferguson et al. (1992) and Johnson et al. (1995), younger and older adults heard words spoken by either two similar sources (two women) or two dissimilar sources (female and male). During the test, participants indicated the source of previously heard and new words (e.g., spoken by the woman, spoken by the man, or new). These researchers found that older adults were worse than younger adults at identifying the speaker of a previously heard word when the words were spoken by two women. However, older adults were just as able as younger adults to remember the speaker of words spoken by a woman and a man. These results suggest that a symptom of aging is a disproportionate loss of or failure to recover information about relatively specific source information as compared with partial-source information (see Johnson et al., 1995).

Our results for divided attention in combination with these developmental studies suggest that we may need to reconsider the proposed role of attention as setting and maintaining a retrieval agenda (e.g., Craik et al., 1996). If establishing and maintaining a retrieval mode uses attention, then when attention is divided during retrieval one might expect both the voice and gender identification parameters to decrease. That is, participants should be impaired at making all source judgments. Instead, as Johnson (1997) suggested, dividing attention may interfere with the activation of source information. More specific memories about voice information require more attention for recovery than do less specific memories about gender information. These results suggest that attention may support a "recollective focusing mechanism." Dividing attention during retrieval may blur this focusing mechanism, making it more difficult to recover relatively precise information than less precise information.

It is important to comment on two aspects of using multinomial models to analyze data from source tests. First, in both experiments the model-based analyses paralleled the participant-based analyses of the identification-of-origin and partial-source scores. Because of this correspondence between the two analyses, why present both analyses? We present the participant-based analyses to follow the traditional way of analyzing these data. However, as Batchelder and Riefer (1990) have explained well, there are potentially many problems with the participant-based analyses. One problem is that multiple processes, such as response bias or memory for the item, can influence the identification-of-origin and partial-source scores. This can make these scores misleading. The multinomial model, in contrast, provides a method of separating and directly measuring both the influence of response bias and memory for the item (for further discussion of this point, see Batchelder & Riefer, 1990; Riefer et al., 1994).

Second, there appears to be a discrepancy between our belief that memory for source varies along a continuum of specificity, whereas the multinomial model assumes that source memory is best characterized as a threshold-like phenomenon. Essentially, our multinomial model assumes that memory for source is described as being in one of the following states: (a) remember the specific voice, (b) remember the correct gender of the voice and guess the particular voice, or (c) do not remember any source information and guess. We view the model as an approximation to reality and as currently the best option available to analyze source data. In addition, there are some important reasons to use the multinomial model. First, in recent work, we have compared the results from both a multinomial analysis and a signal-detection analysis of data from a source-monitoring task (Slotnick, Klein, Dodson, & Shimamura, 1998). The data suggest that the multinomial and signal-detection analyses will closely correspond to each other except at extremely high and low rates of source misattributions and false alarms for new items. Second, the multinomial model has a nice safeguard for signaling when its analysis does not fit reality. That is, we agree with Batchelder, Riefer, and Hu (1994) that the multinomial model should provide a poor fit of the data when it does not approximate the situation well.

Finally, our finding in Experiment 1 that old-new recognition performance was lower when participants heard words spoken by similar voices (e.g., four different men) than when participants heard words spoken by dissimilar voices (e.g., two men and two women) conflicts with recent studies that have found no effect of source similarity on old-new recognition performance (e.g., Bayen et al., 1996; Ferguson et al., 1992; Lindsay et al., 1991). However, other data are consistent with our finding and indicate that similarity affects old-new recognition performance. For instance, both Nelson, Brooks, and Wheeler (1975) and Runquist (1978) found lower old-new recognition performance when participants studied physically similar words (e.g., fish, flush, etc.) than physically dissimilar words. Similarly, Schmidt (1985) found that increasing the conceptual similarity of words decreased recognition performance. He found better recognition of a set of target words (e.g., four-footed animals) when these words were mixed with conceptually different words (e.g., chemical elements) than when these words were mixed with words from the same conceptual category (e.g., list is totally composed of four-footed animals). Moreover, another reason for lower old-new recognition in the high similarity condition is the particular strategy that participants adopt when completing the five-alternative, forced-choice source test. That is, participants may use a strategy of responding "new" to studied items in order to avoid guessing a source when they cannot remember any source-identifying information about the item. This strategy would occur more often in the high-similarity condition, thus lowering the recognition rate, because of the greater difficulty in identifying the source in this condition. Clearly, some work remains to explain the effect of similarity on recognition performance.

In conclusion, the richness of the experience of recollection is due to the fact that memories vary in their precision (e.g., Johnson et al., 1993). People's memories are more similar to John Dean's memory, who remembered the events
involving Nixon and Watergate with different degrees of detail and specificity, than they are to S’s unvarying perfect memory (Luria, 1968; Neisser, 1981).

References


### Appendix A

**Response Frequencies for the Different-Gender- and Same-Gender-Voices Conditions in Experiment 1**

<table>
<thead>
<tr>
<th>True source</th>
<th>Response</th>
<th>“Anthony”</th>
<th>“Eric”</th>
<th>“Julie”</th>
<th>“Lucy”</th>
<th>“New”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different-gender-voices condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anthony</td>
<td>151</td>
<td>80</td>
<td>19</td>
<td>21</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Eric</td>
<td>49</td>
<td>171</td>
<td>31</td>
<td>15</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Julie</td>
<td>19</td>
<td>24</td>
<td>175</td>
<td>54</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Lucy</td>
<td>29</td>
<td>31</td>
<td>67</td>
<td>153</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>New</td>
<td>13</td>
<td>9</td>
<td>8</td>
<td>13</td>
<td>272</td>
<td></td>
</tr>
<tr>
<td>Same-gender-voices condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anthony</td>
<td>94</td>
<td>71</td>
<td>43</td>
<td>36</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Eric</td>
<td>44</td>
<td>83</td>
<td>57</td>
<td>45</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>Mike</td>
<td>30</td>
<td>54</td>
<td>127</td>
<td>30</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Tom</td>
<td>43</td>
<td>54</td>
<td>42</td>
<td>94</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>New</td>
<td>17</td>
<td>18</td>
<td>10</td>
<td>15</td>
<td>255</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Correct responses are in boldface type.

### Appendix B

**Formulas**

The following are formulas for computing the old–new recognition, exact- and partial-source identification scores for the Anthony and Eric items.

#### Old–new recognition score

\[
1 - \frac{P(\text{"N"}|A) + P(\text{"N"}|E)}{2}.
\]  

(B1)

#### Specific-source identification score:

\[
\frac{(P(\text{"A"}|A) + P(\text{"E"}|E))}{((1 - P(\text{"N"}|A)) + (1 - P(\text{"N"}|E))}.
\]  

(B2)

#### Partial-source identification score:

\[
\frac{(P(\text{"E"}|A) + P(\text{"A"}|E)) - (5 \cdot (P(\text{"J"}|A) + P(\text{"L"}|A) + P(\text{"J"}|E) + P(\text{"L"}|E)))}{((1 - P(\text{"A"}|A) - P(\text{"N"}|A)) + (1 - P(\text{"E"}|E) - P(\text{"N"}|E))}.
\]  

(B3)
### Appendix C

#### Parameter Estimates for Experiment 1 From the Unconstrained 19- and 15-Parameter Models

<table>
<thead>
<tr>
<th>Source</th>
<th>Parameter</th>
<th>Anthony</th>
<th>Eric</th>
<th>Julie</th>
<th>Lucy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different-gender-voices condition (19-parameter model)</td>
<td>D</td>
<td>.84</td>
<td>.82</td>
<td>.84</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>.59</td>
<td>.39</td>
<td>.55</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>.31</td>
<td>.44</td>
<td>.36</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>.24</td>
<td>.28</td>
<td>.29</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>g</td>
<td>.30</td>
<td>.21</td>
<td>.19</td>
<td>.30</td>
</tr>
<tr>
<td>Same-gender-voices condition (15-parameter model)</td>
<td>D</td>
<td>.72</td>
<td>.66</td>
<td>.71</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>.24</td>
<td>.02</td>
<td>.41</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>.21</td>
<td>.36</td>
<td>.25</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>g</td>
<td>.29</td>
<td>.30</td>
<td>.17</td>
<td>.25</td>
</tr>
</tbody>
</table>

Note. D = probability of detecting studied items as old; P = probability of identifying the correct gender of studied items; d = probability of identifying the correct speaker of studied items; a = probability of guessing that a detected item is from a particular source; g = probability of guessing that an undetected item is from a particular source; and b = probability of guessing an item is old.

### Appendix D

#### Further Thoughts on the Partial-Source Model

On the surface, our "full" four-source, two-gender model, as depicted in Figure 2, has 19 parameters (1 b, 4 Ds, 4 ds, 4 Ps, 3 as and 3 gs) for 20 independent degrees of freedom. Hence, this model appears to be fully identified and has 1 degree of freedom for testing the fit of the model. However, it is always foolish to regard models that posit latent or unobservable states that "lie behind" the observed data as behaving in the same way as models that are based completely on observed data. To illustrate what we mean, consider the submodel of our "full" model that assumes that same-gender sources are remembered in the same way but different-gender sources might be remembered in different ways (i.e., within-gender indistinguishability). We operationalize this submodel as: $D_{\text{Anthony}} = D_{\text{Eric}}$, $D_{\text{Julie}} = D_{\text{Lucy}}$, $d_{\text{Anthony}} = d_{\text{Eric}}$, $d_{\text{Julie}} = d_{\text{Lucy}}$, $P_{\text{Anthony}} = P_{\text{Eric}}$, $P_{\text{Julie}} = P_{\text{Lucy}}$, $a_{\text{Anthony}} = a_{\text{Eric}}$, $a_{\text{Julie}} = a_{\text{Lucy}} = .5 = a_{\text{Anthony}}$, $g_{\text{Anthony}} = g_{\text{Eric}}$, $g_{\text{Julie}} = g_{\text{Lucy}} = .5 = g_{\text{Anthony}}$. In this submodel there are 9 free parameters. Or are there? It is easy (but somewhat tedious) to see that this submodel produces the following pattern of predicted cell values for the 5 × 5 table of response probabilities illustrated in Table 1 and the bottom of Table 2.

An examination of the cells of Table A1 reveals that there are only 8 free parameters in it: $t$, $u$, $x$, $r$, $s$, $x^*$, $y$, and $y^*$ (because the rows sum to 1, the parameters $v$ and $w$ are not free). Hence, the 9-parameter model in the "psychological space" governed by Figure 2 only produces an 8-parameter model in the "data space" described by Table 1. A more detailed examination of the parameters of the "within-gender indistinguishability" model reveals that the problem of identifiability resides solely within the triple $(P_1, P_3, a_1)$. The other six parameters are completely identified.

This fact is disconcerting because it is the $P$ parameters that interest us regarding partial source memory. Their usefulness depends on what we can assume about the $a$ parameters. In our experiment, assuming the $a$ parameters were all approximately .25 did not conflict with the data very much—see the parameter estimates in Table 4. Hence, we can use this to identify $P_1$ and $P_3$. On the other hand, this lack of identifiability also suggests that the "full" model actually fits the data better than it looks like it does. This argument goes as follows. Problems of identifiability do not go away when more parameters are considered, they go away only when fewer parameters are considered (i.e., when restrictions are placed on the nonidentified parameters). Hence, the identifiability problem we have discovered in the "within-gender indistinguishability" model must remain in the full, 19-parameter model. The only problem to solve is to see how many of the 19 parameters are involved in this lack of identifiability. We have not done this carefully, and think that it is a complicated task. However, the end result is that the 19-parameter model does not really have 19 parameters when looked at in "data space" (i.e., Table 1). It must have somewhat fewer parameters—maybe 2 or 3 fewer. Thus, the degrees of freedom for lack of model fit actually must be larger than the nominal value of 1 that we have used to assess the lack of model fit, and hence the fit is probably better than it appears to be when $G^2$ is referred to the chi-square distribution on 1 degree of freedom.

(Appendix continues)
Table D1
Pattern of Response Probabilities Predicted by the “Within-Gender Indistinguishability” Model

<table>
<thead>
<tr>
<th>True source</th>
<th>“Anthony”</th>
<th>“Eric”</th>
<th>“Julie”</th>
<th>“Lucy”</th>
<th>“New”</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthony</td>
<td>r</td>
<td>u</td>
<td>v</td>
<td>v</td>
<td>x</td>
<td>1</td>
</tr>
<tr>
<td>Eric</td>
<td>u</td>
<td>t</td>
<td>v</td>
<td>v</td>
<td>x</td>
<td>1</td>
</tr>
<tr>
<td>Julie</td>
<td>w</td>
<td>w</td>
<td>r</td>
<td>s</td>
<td>x*</td>
<td>1</td>
</tr>
<tr>
<td>Lucy</td>
<td>w</td>
<td>w</td>
<td>s</td>
<td>r</td>
<td>x*</td>
<td>1</td>
</tr>
<tr>
<td>New</td>
<td>y</td>
<td>y</td>
<td>y*</td>
<td>y*</td>
<td>z</td>
<td>1</td>
</tr>
</tbody>
</table>

Appendix E
Response Frequencies for the Full-Attention and Divided-Attention Conditions in Experiment 2

<table>
<thead>
<tr>
<th>True source</th>
<th>“Anthony”</th>
<th>“Eric”</th>
<th>“Julie”</th>
<th>“Lucy”</th>
<th>“New”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full attention at test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anthony</td>
<td>225</td>
<td>70</td>
<td>41</td>
<td>34</td>
<td>80</td>
</tr>
<tr>
<td>Eric</td>
<td>90</td>
<td>196</td>
<td>47</td>
<td>30</td>
<td>87</td>
</tr>
<tr>
<td>Julie</td>
<td>33</td>
<td>40</td>
<td>205</td>
<td>85</td>
<td>87</td>
</tr>
<tr>
<td>Lucy</td>
<td>38</td>
<td>39</td>
<td>93</td>
<td>199</td>
<td>81</td>
</tr>
<tr>
<td>New</td>
<td>20</td>
<td>23</td>
<td>14</td>
<td>7</td>
<td>386</td>
</tr>
</tbody>
</table>

| Divided attention at test |
| Anthony     | 225       | 70     | 41      | 34     | 80    |
| Eric        | 90        | 196    | 47      | 30     | 87    |
| Julie       | 33        | 40     | 205     | 85     | 87    |
| Lucy        | 38        | 39     | 93      | 199    | 81    |
| New         | 20        | 23     | 14      | 7      | 386   |

Note. Correct responses are in boldface type.

Appendix F
Parameter Estimates for Experiment 2 From the Unconstrained 19-Parameter Models

<table>
<thead>
<tr>
<th>Source</th>
<th>Parameter</th>
<th>Anthony</th>
<th>Eric</th>
<th>Julie</th>
<th>Lucy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full attention at test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>.79</td>
<td>.77</td>
<td>.77</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>.33</td>
<td>.47</td>
<td>.47</td>
<td>.35</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>.45</td>
<td>.28</td>
<td>.26</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>.22</td>
<td>.24</td>
<td>.31</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>.31</td>
<td>.36</td>
<td>.22</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Divided attention at test |
| Anthony     | 225       | 70     | 41    | 34    | 80   |
| Eric        | 90        | 196    | 47    | 30    | 87   |
| Julie       | 33        | 40     | 205   | 85    | 87   |
| Lucy        | 38        | 39     | 93    | 199   | 81   |
| New         | 20        | 23     | 14    | 7     | 386  |

Note. D = probability of detecting studied items as old; P = probability of identifying the correct gender of studied items; d = probability of identifying the correct speaker of studied items; a = probability of guessing that a detected item is from a particular source; g = probability of guessing that an undetected item is from a particular source; and b = probability of guessing an item is old.

Received May 1, 1996
Revision received November 10, 1997
Accepted November 18, 1997