Recollection is a continuous process: Evidence from plurality memory receiver operating characteristics

Scott D. Slotnick1, Brittany M. Jeye1, and Chad S. Dodson2

1Department of Psychology, Boston College, Chestnut Hill, MA, USA
2Department of Psychology, University of Virginia, Charlottesville, VA, USA

(Received 10 April 2014; accepted 25 September 2014)

Is recollection a continuous/graded process or a threshold/all-or-none process? Receiver operating characteristic (ROC) analysis can answer this question as the continuous model and the threshold model predict curved and linear recollection ROCs, respectively. As memory for plurality, an item’s previous singular or plural form, is assumed to rely on recollection, the nature of recollection can be investigated by evaluating plurality memory ROCs. The present study consisted of four experiments. During encoding, words (singular or plural) or objects (single/singular or duplicate/plural) were presented. During retrieval, old items with the same plurality or different plurality were presented. For each item, participants made a confidence rating ranging from “very sure old”, which was correct for same plurality items, to “very sure new”, which was correct for different plurality items. Each plurality memory ROC was the proportion of same versus different plurality items classified as “old” (i.e., hits versus false alarms). Chi-squared analysis revealed that all of the plurality memory ROCs were adequately fit by the continuous unequal variance model, whereas none of the ROCs were adequately fit by the two-high threshold model. These plurality memory ROC results indicate recollection is a continuous process, which complements previous source memory and associative memory ROC findings.

Keywords: Dual-process; Plural; Receiver operating characteristic; Relational memory; Signal detection.

Memories can be based on either specific recollection or non-specific familiarity. Familiarity is widely believed to be a continuous process in which memory for previously presented information is graded, ranging in strength from very weak to very strong. By contrast, recollection has long been thought to be an all-or-none threshold process (Mandler, 1980; Yonelinas, 2002; Yonelinas and Parks, 2007). However, recent evidence indicates that recollection, like familiarity, is a continuous process (Slotnick, 2013; Wixted, 2007; Wixted & Mickes, 2010). As such, the nature of recollection is currently a heated source of debate.

One way to isolate the process of recollection is to investigate relational memory. Relational memory refers to retrieval of information related to an item such as its previous context (i.e., source memory), its previous paired associate (i.e., associative memory) or whether it was previously singular or plural (i.e., plurality memory). Receiver operating characteristic (ROC) analysis has been used to distinguish between the continuous model of recollection and the threshold model of recollection as these models predict curved and linear relational memory ROCs, respectively.
In the present study, we investigated the nature of recollection by evaluating the shape of the plurality memory ROC. During encoding, participants were shown singular or plural words (Experiment 1) or single/singular or duplicate/plural objects (Experiment 2; Figure 1, top). During retrieval, words or objects were presented with the same plurality or different plurality, and participants made a confidence rating ranging from “very sure old” (i.e., same plurality) to “very sure new” (i.e., different plurality; Figure 1, bottom).

The two formal models of recollection are referred to as the unequal variance model and the two-high threshold model (Green & Swets, 1966; Macmillan & Creelman, 2005). The continuous unequal variance model of recollection dictates that same plurality and different plurality items have Gaussian distributions of memory strength in decision space and can have unequal variance (Figure 2, top left; memory strength is the distance between the distribution means). Each confidence rating depends on an item’s plurality memory strength and criteria placement ($c_1$–$c_6$). That is, memory strength greater than $c_6$ would produce a “very sure old” response, memory strength between $c_6$ and $c_5$ would produce a “moderately sure old” response, memory strength between $c_5$ and $c_4$ would produce a “less sure old” response, and so on. To generate the ROC (Figure 2, top centre), hit rate is plotted against false alarm rate for each criterion/confidence rating (e.g., for the leftmost ROC point, hit rate is the probability of a “very sure old” response for a same plurality item, and false alarm rate is the probability of a “very sure old” response for a

**Figure 1.** In Experiment 1 (left), singular or plural words were presented during encoding. During retrieval, old words were presented with the same or different plurality and participants classified each item as either “old” (i.e., same plurality) or “new” (different plurality). Experiment 2 (right) used the identical protocol except that single/singular or duplicate/plural objects were presented.
different plurality item). The continuous model ROC is always curved, and the specific curvature of the ROC is dictated by the shapes of the underlying Gaussian distributions. In addition, the continuous model z-transformed ROC (zROC; Figure 2, top right) is always linear. The two-high threshold model of recollection dictates that there are two thresholds in a decision space beyond which only one distribution exists and between which the distributions are completely overlapping (Figure 2, bottom left). Recollection occurs if memory strength is beyond one of the two thresholds but otherwise fails, which is why this model has been referred to as the all-or-none model (e.g., Hilford, Glanzer, Kim, & DeCarlo, 2002; Slotnick & Dodson, 2005; Wixted & Stretch, 2004). The threshold model ROC is always linear (Figure 2, bottom centre). In addition, the threshold model zROC always has a positive curvature (Figure 2, bottom right).

Two methods have been employed to evaluate relational memory ROC/zROC curvature. Linearity analysis determines whether or not the ROC/zROC has quadratic curvature (i.e., if the quadratic coefficient, c, in \( a + bx + cx^2 \) is significant). However, linearity analysis is an inconclusive method of evaluating ROC/zROC shape. The presence or absence of a significant quadratic component in this analysis allows for the conclusion that the ROC/zROC does or does not have a quadratic curvature. Critically, it is not the case that a non-significant quadratic component means that the ROC is a straight line, as predicted by the threshold model. Instead, it is possible that even though the quadratic component is non-significant, the ROC has a non-quadratic curvature, as predicted by the continuous unequal variance model. For instance, a quadratic curvature might not account for an unequal variance ROC that is too bowed, which can occur when memory strength is very high (see Slotnick, 2010). Overall, then, linearity analysis is a problematic method of model assessment.

By comparison, goodness-of-fit analysis determines whether an ROC has the specific type of curvature dictated by the continuous unequal

Figure 2. Decision space, predicted ROC and predicted zROC for the continuous unequal variance model (top) and the two-high threshold model (bottom).
variance model or whether an ROC is linear as dictated by the two-high threshold model. For example, chi-squared analysis measures the deviation between the predicted ROC, which reflects the underlying distributions in a decision space, and the empirical ROC, with a lower chi-square value indicating a better fit. Therefore, goodness-of-fit analysis provides a better method of model assessment.

Rotello, Macmillan, and Van Tassel (2000) used a word plurality paradigm that was identical to the present Experiment 1 (Figure 1, left) and conducted linearity analysis. The plurality memory ROC in one experiment did not have significant quadratic curvature, and the plurality memory ROC in another experiment had a relatively small (but significant) positive quadratic curvature, which the investigators took as a support for the threshold model of recollection. Heathcote, Raymond, and Dunn (2006) also used a word plurality paradigm but with a greater number of shorter lists (five lists with 32 encoding items as compared to three lists with 48 encoding items) and conducted goodness-of-fit analysis. In two experiments, the plurality memory ROCs were adequately fit by the continuous unequal variance model and were not adequately fit by the two-high threshold model.

Given that Rotello et al. (2000) did not conduct a goodness-of-fit analysis, their results are inconclusive. However, Heathcote et al. (2006) employed a different paradigm than Rotello et al. (2000). As such, it is possible that the seemingly linear plurality ROCs reported by Rotello et al. (2000) might be obtained under the experimental conditions they employed. The aim of the present investigation was to use the same experimental procedures as Rotello et al. (2000) and conduct a goodness-of-fit (chi-squared) analysis to determine whether plurality memory ROCs support the continuous unequal variance model of recollection or the two-high threshold model of recollection. This is the first time that a chi-squared analysis has been conducted using the plurality paradigm of Rotello et al. (2000).

**METHODS**

Participants were recruited from the Boston College undergraduate community. Each participant received 1 credit or $10 for a 1-hour session. The protocol was approved by the Boston College Institutional Review Board, and informed consent was obtained from each participant. Eighteen participants completed each of the four experiments, and each experiment had a different set of participants.

The experiments were conducted using E-Prime (Psychology Software Tools, Inc., Sharpsburg, PA). Stimuli consisted of line drawings selected from a pool of 520 objects (International Picture Naming Project at the UCSD Center for Research in Language) or the word labels/names of those objects. Selected objects had at least a 96% valid response rate, at least a 90% name agreement and a name with one or two syllables.

Each participant completed a one-quarter length practice run and three encoding-retrieval runs. During each encoding phase in Experiments 1a and 1b (Figure 1, top left), 48 singular or plural words were centrally presented in random order. Two additional buffer words were presented at the beginning and at the end of each study list. Each word was presented for 3 seconds and spanned 2° of visual angle in height × 3–12° of visual angle in width. Participants were instructed to remember each word and its plurality. During each retrieval phase (Figure 1, bottom left), the 48 words from the study phase, half with the same plurality and half with a different plurality, and 24 new words were presented in random order. For each word, participants made an old-new confidence rating where 1 = “very sure old”, 2 = “moderately sure old”, 3 = “less sure old”, 4 = “don’t know”, 5 = “less sure new”, 6 = “moderately sure new” and 7 = “very sure new”. Participants were instructed that an “old” response referred to old words with the same plurality and a “new” response referred to new words or old words with a different plurality (e.g., if the word “computers” was presented at a study and “computer” was presented at retrieval, the correct response should have been “new”). They were also informed that accurate confidence judgements were of primary importance, but response speed was also of importance. The retrieval phase was self-paced. An equal number of each item type was presented in each encoding and retrieval phase, and no more than three items of a given type were presented sequentially. Lists (encoding–retrieval format: singular–singular, plural–plural, singular–plural, plural–singular, new–singular and new–plural) were counterbalanced across participants using a Latin Square design. A subsequent remember-know-new or remember-know-guess-new response was made in Experiments 1a and 1b, respectively, which
was the only protocol difference between these experiments; we collapsed over these responses to be consistent with previous plurality memory ROC studies. Eight participants, five from Experiment 1a and three from Experiment 1b, were excluded from the analysis because they incorrectly classified a majority of different plurality items as “old” (i.e., their different plurality memory performance was less than chance). It should be noted that the present paradigm included new items to replicate the paradigm employed by Rotello et al. (2000); however, only old items were included in the ROC analysis because our aim was to evaluate the nature of recollection.

Experiments 2a and 2b used the same experimental protocol previously specified, except that singular or plural objects were used as stimuli. During each encoding phase (Figure 1, top right), objects were presented to the left and/or to the right of a red fixation cross. Each object was contained within a bounding box 8° of visual angle along each side with the nearest edge 3° of visual angle from fixation. Participants were instructed to maintain central fixation and remember each object and its plurality. They were also encouraged to remember the entire display, the object(s) and the fixation cross, rather than use non-visual strategies. During each retrieval phase (Figure 1, bottom right), old and new objects were presented and participants made old-new confidence ratings as described earlier. One participant was excluded from the Experiment 2a analysis because they made almost no high confidence responses to same plurality items (i.e., their proportion of “very sure old” responses was less than three standard deviations from the mean), and two participants were excluded from Experiment 2b because they classified a majority of different plurality items as “old” (i.e., performance for these items was less than chance).

For each experiment, the plurality memory ROC was generated by plotting same plurality item hit rate versus different plurality item false alarm rate. Specifically, for successive confidence ratings, hit rate was the cumulative probability of classifying same plurality items as “old”, and false alarm rate was the cumulative probability of classifying different plurality items as “old”. The unequal variance and two-high threshold models were fit to each plurality memory ROC by adjusting model parameters using maximum likelihood estimation. Log-likelihood chi-square was used to assess the adequacy of each model, with lower chi-square values reflecting a better fit ($p > 0.05$ indicated an adequate fit). To make inferences across participants or experiments, independent chi-square values and the corresponding degrees of freedom were summed (Rice, 2007). Even though its results are inconclusive, to be consistent with the literature, we conducted linearity analyses to assess whether each ROC had significant quadratic ($cx^2$) curvature.

**TABLE 1**

<table>
<thead>
<tr>
<th>Response matrices</th>
<th>← “Old”→</th>
<th>“New”→</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4</td>
<td>5 6 7</td>
</tr>
<tr>
<td><strong>Experiment 1a</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same plurality</td>
<td>623 64 46</td>
<td>15 34 71 83</td>
</tr>
<tr>
<td>Different plurality</td>
<td>78 40 46</td>
<td>27 55 105 585</td>
</tr>
<tr>
<td><strong>Experiment 1b</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same plurality</td>
<td>727 38 51</td>
<td>36 49 59 120</td>
</tr>
<tr>
<td>Different plurality</td>
<td>99 40 38</td>
<td>46 67 76 714</td>
</tr>
<tr>
<td><strong>Experiment 2a</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same plurality</td>
<td>820 114 64</td>
<td>30 51 61 84</td>
</tr>
<tr>
<td>Different plurality</td>
<td>68 60 44</td>
<td>35 78 179 760</td>
</tr>
<tr>
<td><strong>Experiment 2b</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same plurality</td>
<td>774 67 52</td>
<td>23 47 69 120</td>
</tr>
<tr>
<td>Different plurality</td>
<td>79 41 32</td>
<td>27 44 131 798</td>
</tr>
</tbody>
</table>

**RESULTS**

For Experiments 1a and 1b, the word plurality memory ROCs, zROCs and best-fit unequal variance and two-high threshold models are shown in Figure 3 (response matrices for all the experiments are shown in Table 1). Plurality memory strength was relatively high, which is illustrated by the ROCs being far from the chance line (Experiment 1a, unequal variance $d' = 2.15$, two-high threshold $R_1 = 0.57$, $R_2 = 0.64$; Experiment 1b, unequal variance $d' = 2.17$, two-high threshold $R_1 = 0.58$, $R_2 = 0.58$). Chi-squared analysis revealed that both plurality memory ROCs were adequately fit by the unequal variance model (Experiment 1a, $\chi^2(4) = 8.44$; Experiment 1b, $\chi^2(4) = 4.49$). By contrast, neither of the plurality memory ROCs was adequately fit by the two-high threshold model (Experiment 1a, $\chi^2(4) = 23.39$, $p < .001$; Experiment 1b, $\chi^2(4) = 12.97$, $p < .05$). Linearity analysis produced complementary results for the most part, as both ROCs were

---

1The present old-new recognition memory $d'$ ranged from 1.86 to 2.63, which was similar to Rotello et al. (2000).
curved (Experiment 1a, quadratic coefficient $c = -1.43$, $F(1, 3) = 25.55, p < .05$; Experiment 1b, $c = -1.04$, $F(1, 3) = 11.14, p < .05$) and the Experiment 1a zROC was linear ($c = 0.22$, $F(1, 3) = 3.71$). The Experiment 1b zROC had a slight but significant positive curvature ($c = 0.36$, $F(1, 3) = 14.79, p < .05$), which is addressed in the Discussion.

For Experiments 2a and 2b, the object plurality memory ROCs, zROCs and best-fit unequal variance and two-high threshold models are shown in Figure 4. Plurality memory strength was somewhat higher than the previous experiments (Experiment 2a, unequal variance $d' = 2.16$, two-high threshold $R_1 = 0.62, R_2 = 0.70$; Experiment 2b; unequal variance $d' = 2.35$, two-high threshold $R_1 = 0.59, R_2 = 0.68$). Like the previous experiments, chi-squared analysis revealed that the plurality memory ROCs were adequately fit by the unequal variance model (Experiment 2a, $\chi^2(4) = 5.42$; Experiment 2b, $\chi^2(4) = 7.86$) and were not adequately fit by the two-high threshold model (Experiment 2a, $\chi^2(4) = 63.65, p < .001$; Experiment 2b, $\chi^2(4) = 34.79, p < .001$). Linearity analysis produced complementary results, as both ROCs were curved (Experiment 2a, quadratic coefficient $c = -3.04$, $F(1, 3) = 49.42, p < .01$; Experiment 2b, $c = -3.36$, $F(1, 3) = 147.11, p < .01$) and both zROCs were linear (Experiment 2a, $c = -0.087$, $F(1, 3) = 6.64$; Experiment 2b, $c = -0.0048$, $F(1, 3) < 1$).

To ensure the previous group ROC analysis results were not due to averaging effects, an
individual participant ROC analysis was conducted, and the identical pattern of results was observed. Across all four experiments, the individual participant ROCs were adequately fit by the unequal variance model ($\chi^2(201) = 141.82$) and were not adequately fit by the two-high threshold model ($\chi^2(201) = 347.36, p < .001$). Furthermore, the unequal variance model provided a better fit than the two-high threshold model in the large majority (51/59) of participants.

DISCUSSION

In the present study, the continuous unequal variance model provided an adequate fit to all of the plurality memory ROCs. By contrast, the two-high threshold model of recollection did not provide an adequate fit to any of the plurality memory ROCs. Of direct relevance to the aim of the present investigation, we employed the same experimental procedures as Rotello et al. (2000), conducted a goodness-of-fit (chi-squared) analysis and found that plurality memory ROCs supported the continuous unequal variance model of recollection. The present results replicate previous memory ROC findings based on
chi-squared analysis using a word plurality paradigm (Heathcote et al., 2006) and, for the first time, extend these findings using an object plurality paradigm.

Rotello et al. (2000) conducted linearity analysis and found that the plurality memory ROC in one experiment had relatively small but significant positive quadratic curvature, and the plurality memory ROC in another experiment did not have significant quadratic curvature. The fact that the quadratic component was significant in one of their experiments is inconsistent with the two-high threshold model, which predicts a linear plurality memory ROC. Although the non-significant quadratic component in their other experiment is seemingly at odds with the present findings, this result is inconclusive as the unequal variance model does not predict quadratic curvature. To illustrate this point empirically, we fit the linear plus quadratic model (i.e., \(a + bx + cx^2\)), which corresponds to linearity analysis, to the present plurality memory ROCs and evaluated goodness-of-fit. Across all four experiments, this model did not adequately fit the plurality memory ROCs (\(\chi^2(16) = 52.83; p < .001\)). As the unequal variance model adequately fit the plurality memory ROCs and the linear plus quadratic model did not adequately fit the plurality memory ROCs, quadratic ROC curvature should not be taken as a proxy for the curvature predicted by the unequal variance model.

Linearity analysis showed that a majority of zROCs in the current study were linear, which supports the unequal variance model. However, the Experiment 1b zROC had significant positive quadratic curvature, which could be taken as a support for the two-high threshold model. Previous work has indicated that inclusion of non-diagnostic information (e.g., “new” responses) in relational memory analysis can flatten the ROC and produce positive quadratic curvature in the zROC (Mickes, Johnson, & Wixted, 2010; Slotnick & Dodson, 2005). To assess whether that was the case here, the Experiment 1b zROC was generated from “remember” responses alone. Figure 5 shows the corresponding conditional plurality memory zROC. Linearity analysis showed that this conditional zROC had negative quadratic curvature that did not significantly deviate from linearity (quadratic coefficient \(c = -3.56, F(1, 3) = 7.41\)). Thus, the positive zROC curvature in the Experiment 1b zROC (Figure 3, bottom right) can be attributed to distortion due to inclusion of non-diagnostic information. We also generated the conditional zROCs (generated from “remember” responses) for the other three experiments, and they all had negative quadratic curvature as well (Experiment 1a, \(c = -1.41, F(1, 3) = 11.62, p < .05\); Experiment 2a, \(c = -0.54, F(1, 3) = 18.47, p < .05\); Experiment 2b, \(c = -1.02, F(1, 3) < 1\)). Such negative zROC curvature is in direct opposition to the positive quadratic curvature predicted by the two-high threshold model, but is consistent with a modified unequal variance model that accounts for relational memory misattribution (Dodson et al., 2007; Slotnick, 2010). It is also notable that although the Experiment 1b zROC had significant positive quadratic curvature, the two-high threshold model did not provide an adequate fit. Similar to the conclusion in the preceding paragraph, this shows that positive quadratic zROC curvature should not be taken as a proxy for the curvature predicted by the two-high threshold model.

The present findings are in line with previous relational memory ROC findings. Source memory ROCs have been evaluated using linearity analysis and have been consistently found to have quadratic curvature (Dodson, Bawa, & Slotnick, 2007; Glanzer, Hilford, & Kim, 2004; Hilford et al., 2002; Qin et al., 2001; Slotnick & Dodson, 2005; Slotnick, Klein, Dodson, & Shimamura, 2000; but see Yonelinas, 1999). However, as underscored previously, the ROC curvature
predicted by the continuous unequal variance model is not quadratic; therefore, linearity analysis results are inconclusive. Many source memory ROC studies have employed chi-squared analysis (Dodson et al., 2007; Slotnick & Dodson, 2005; Slotnick et al., 2000, 2010). Some of these studies maximised the process of recollection by analysing source responses that were “remembered” (Slotnick, 2010) or were based on high confidence item responses (Slotnick et al., 2000; Slotnick & Dodson, 2005). Across these studies, a large majority of the source memory ROCs were adequately fit by the continuous unequal variance model (or a modified continuous unequal variance model that accounted for source memory misattribution), but the source memory ROCs were never adequately fit by the two-high threshold model (the same pattern of results was obtained by Hautus, Macmillan, & Rotello, 2008, and Dubé, Payne, Sekuler, & Rotello, 2013). An associative memory ROC study that employed chi-squared analysis (Kelley & Wixted, 2001) found that all of the associative memory ROCs in the higher memory strength condition, which can be assumed to reflect a greater degree of recollection, were adequately fit by the unequal variance model, and none of the associative memory ROCs were adequately fit by the two-high threshold model (a similar pattern of findings was reported by Mickes et al., 2010). Thus, there is convergent relational memory evidence in support of the continuous model of recollection and against the threshold model of recollection.

Although the two-high threshold model of recollection has been widely assumed to predict linear ROCs (e.g., Dodson et al., 2007; Glanzer et al., 2004; Heathcote et al., 2006; Hilford et al., 2002; Qin et al., 2001; Rotello et al., 2000; Slotnick, 2010; Slotnick & Dodson, 2005; Slotnick et al., 2000; Wixted, 2007; Wixted & Mickes, 2010; Yonelinas, 1999, 2002; Yonelinas & Parks, 2007), it is also possible for this model to predict curved ROCs when generated from confidence ratings (Bröder & Schütz, 2009; Malmberg, 2002; Schütz & Bröder, 2011). Critically, all threshold models of recollection dictate that there are thresholds in the decision space above and below which only one distribution exists (Figure 2, bottom left). To assess whether there was evidence for thresholds in the current study, we fit a four-parameter two-high threshold unequal variance model to the ROCs (see Kapucu, Macmillan, & Rotello, 2010). Note that this threshold model predicts a curved ROC. Of importance, the two-high threshold model is nested within the two-high threshold unequal variance model, which means the models can be directly compared. If the two-high threshold unequal variance model produces a significantly better fit than the unequal variance model, this would provide evidence for a threshold recollection process (Slotnick & Thakral, 2013). Across all four experiments, chi-squared analysis revealed that the two-high threshold unequal variance model did not provide a better fit than the unequal variance model ($\chi^2(8) = 14.37$). Thus, there is no evidence for threshold recollection in the present study.

The present findings indicate that recollection, like familiarity, is a continuous process. Meta-analysis results have indicated that recollection and familiarity represent a single process and correspond to strong memory and weak memory, respectively (Dunn, 2004, 2008; Slotnick, 2013). For instance, in one meta-analysis of 37 remember-know studies with 230 conditions, a single-process model accounted for the empirical findings (Dunn, 2008). By contrast, Wixted and colleagues have proposed a dual-process model in which both recollection and familiarity are separate but continuous processes (Ingram, Mickes, & Wixted, 2012; Wixted, 2007; Wixted & Mickes, 2010). In a recent study, remember judgements with intermediate item recognition confidence were associated with higher source memory accuracy than familiarity/know judgements with high item recognition confidence, which provides support for the dual-process model of memory (Ingram et al., 2012). Further empirical study is needed to distinguish between the continuous single-process model of memory and the continuous dual-process model of memory.

The present plurality memory ROC findings unequivocally support the continuous model of recollection and complement previous source memory and associative memory ROC results. Thus, although the threshold model of recollection might be intuitively appealing, the evidence indicates that recollection is a continuous process.

REFERENCES

Cengage Learning.


