Some Memories are Odder than Others:

Judgments of Episodic Oddity Violate Known Decision Rules

Akira R. O'Connor, University of St Andrews

Emily N. Guhl, University of Chicago

Justin C. Cox, Washington University in St Louis

Ian G. Dobbins, Washington University in St Louis

Corresponding author: Akira O'Connor

Address: School of Psychology, University of St Andrews, St Mary's College, South Street, St Andrews, Fife, KY16 9JP, Scotland, UK

E-mail: aoconnor@st-andrews.ac.uk

Telephone: +44(0)1334 46 2077

Fax: +44(0)1334 46 3042

Abstract

Current decision models of recognition memory are based almost entirely on one paradigm, single item old/new judgments accompanied by confidence ratings. This task results in receiver operating characteristics (ROCs) that are well fit by both signal-detection and dual-process models. Here we examine an entirely new recognition task, the judgment of episodic oddity, whereby participants select the mnemonically odd members of triplets (e.g., a new item hidden among two studied items). Using the only two known signal-detection rules of oddity judgment derived from the sensory perception literature, the unequal variance signal-detection model predicted that an old item among two new items would be easier to discover than a new item among two old items. In contrast, four separate empirical studies demonstrated the reverse pattern: triplets with two old items were the easiest to resolve. This finding was anticipated by the dual-process approach as the presence of two old items affords the greatest opportunity for recollection. Furthermore, a bootstrap-fed Monte Carlo procedure using two independent datasets demonstrated that the dual-process parameters typically observed during single item recognition correctly predict the current oddity findings, whereas unequal variance signal-detection parameters do not. Episodic oddity judgments represent a case where dual- and single-process predictions qualitatively diverge and the findings demonstrate that novelty is "odder" than familiarity.

Keywords: Episodic Memory, Recognition, Cognitive Models

Some Memories are Odder than Others: Judgments of Episodic Oddity Violate Known Decision Rules

The number and nature of processes contributing to episodic recognition memory and confidence remains heavily debated in both behavioural and neuroscience literatures (e.g., Diana, Reder, Arndt, & Park, 2006; Donaldson, 1996; Dunn, 2004; Vilberg & Rugg, 2008; Wais, 2008; Wixted & Stretch, 2004). Given this, it is somewhat surprising that for the most part, research has focused on a single task, namely, the judgment of items presented in isolation as either studied or novel, usually supplemented with ratings of confidence (e.g., Bayley, Wixted, Hopkins, & Squire, 2008; Jang, Wixted, & Huber, 2009; Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000). As an illustration of this heavy task reliance, we conducted an informal search of prior work published in the Journal of Memory and Language using the key word “recognition memory”. The search returned 109 articles, 66 of which were actually focused on episodic recognition. Of these, 58 (88%) employed the single item old/new paradigm. As we argue below, this heavy reliance on this one task has potential drawbacks. Before doing so however, we first briefly describe the two dominant decision models of recognition memory.

One of the most successful decision models of single item recognition is the unequal variance signal-detection model–a unidimensional model which provides a straightforward decision-rule (see Figure 1 panel a. for a graphical representation of the unequal variance model). During single item recognition, observers are assumed to evaluate a continuous strength of evidence (or ‘familiarity’) variable evoked by each item relative to an internal old-new criterion: if signal strength exceeds the old/new criterion, the item is judged old; if it falls below the criterion, the item is judged new. The distribution of evidence values across the test is characterized by two normal distributions, one for old items and one for new items, separated by a distance, *d′*, that corresponds to the observer’s sensitivity to the category distinction. Considerable research examining the cumulative relationship between the confidence and accuracy of reports (viz., the receiver operating characteristic, ROC) indicates that in order for the normal distribution model to hold, the variance of old item evidence must be assumed greater than that of the new item evidence (Egan, 1975; Heathcote, 2003; Ratcliff, Sheu, & Gronlund, 1992).

Figure 1 about here

In contrast to the unidimensional signal-detection approach to recognition judgments, dual-process models assume that studied materials are capable of evoking both a relatively continuous sense of prior occurrence (i.e. familiarity) and recollections of specific contextual details associated with the prior encounter of the memory probe (i.e. episodic recollection). The specifics of the models vary, but all assume that it is inappropriate to characterize recognition evidence as solely unidimensional (e.g., Atkinson & Juola, 1973; Hintzman & Curran, 1994; Jacoby, 1991; Mandler, 1980; Yonelinas, 1994). Furthermore, most dual-process models assume that in comparison to familiarity, recollection affords greater behavioural control, leads to more confident endorsement, and requires deliberate retrieval attempt (Atkinson & Juola, 1973; Jacoby, 1991; Mandler, 1980).

Compared to the dual-process model, the unequal variance signal-detection model appears to have a slight but reliable advantage in fitting confidence-based ROCs during single item recognition (Heathcote, 2003; Jang, et al., 2009; Smith & Duncan, 2004; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996). However, the small magnitude of this advantage is arguably underappreciated. For example, Figure 2 shows the aggregate ROC from 26 participants in a single item recognition task fit by both the dual-process model of Yonelinas (1994) and the unequal variance signal-detection model (see General Discussion for details of the procedure). The ROC is formed by taking the most confident “old” response proportions to old and new items as the initial point on the function. By convention the y-axis specifies the response rates to old items and the x-axis to new items. Following this, the next most confident “old” response rates are added to the initial response rates, forming the second point on the ROC. This cumulative process is repeated until the data are exhausted and results in a function linking cumulative accuracy to confidence (Macmillan & Creelman, 2005).

For the aggregate ROC in Figure 2, both models fit the data extremely well, accounting for more than 99% of the variance. When applied to individuals, there is a slight advantage for the unequal variance signal-detection model (99.90% vs. 99.86% variance accounted for) although it is not reliable in this data set (15 of 26 subjects; $Z = .59$, $p > .55$, sign test). The recognition confidence ROC is typically convex and asymmetrical about the negative diagonal. As noted above, the unequal variance model accounts for this asymmetry by assuming that the old item evidence is more variable than the new (Figure 1 panel a.) In contrast, the dual-process model assumes that the most confident hit rates reflect a mix of recollection and familiarity, with all other responses reflecting the contribution of an equal variance signal-detection familiarity process. Critically, because the single item recognition ROC has a form that is often easily accommodated by both models it is important to look for other tasks in which the two models may make more divergent predictions. Furthermore, examining other decision tasks may help elucidate which model more naturally generalizes across tasks and provides a more general understanding of how observers actually decide about their own recognition evidence in various situations. The task we consider here is one that is historically used in perception research, often in cases where the basis of the perceptual distinction subjects are asked to make would be difficult to verbalize, rendering a single item procedure less than ideal (e.g., Brandt & Arnold, 1977; Helm & Trolle, 1946).

Indeed, the ease with which subjects appear to be able grasp this task's requirements is perhaps best illustrated by its frequent use on children's educational programs such as Sesame Street, where amusingly it even has its own song: "One of these things is not like the others."

Figure 2 about here

Judgments of Oddity

The oddity task requires that the observer identify the 'odd-man-out' of a triplet. As noted above, this is a judgment task that is often understandable even when the exact basis for discrimination is quite difficult to articulate (Amerine, Pangbourn, & Roessler, 1965; Peryam, 1958; Frijters, Kooistra, & Vereijken, 1980; Macmillan & Creelman, 2005; Versfeld, Dai, & Green, 1996), As an example, Macmillan and Creelman (2005) discuss the question of whether novices are able to discriminate between Burgundy and Claret. For novices, one assumes that the verbal descriptions of the bouquets would not be particularly helpful since they do not presumably have an explicit characterization of "Burgundyness". However, they are capable of attempting to judge which of three glasses of wine tastes different from the others and may be successful even if they cannot articulate the basis of the difference they are using for the discrimination. Similarly, one could question whether observes can discriminate organic from traditionally grown fruit, 80% dark chocolate from 85% dark chocolate, etcetera. Here we consider the relative oddness of recognition memories. Namely, we consider how subjects identify either experienced novelty or experienced familiarity as contextually odd. Interestingly, one of the most oft cited anecdotes in support of dual-process models is linked to the detection of a mnemonically odd item. Specifically, Mandler (1980) describes a seemingly ubiquitous experience that suggests the potential need for dual retrieval processes. During the anecdote an observer notices a face on a bus that strikes him as surprisingly familiar but is initially unable to specifically note why the face seems so familiar. This is followed by a period of deliberate memory search wherein he or she tries to resolve this ambiguous sense of familiarity, in this example, ultimately retrieving the appropriate contextual detail indicating that the individual is in fact the butcher from the supermarket. Critically, it is the familiarity of the passenger that is "odd" in the context of the other unfamiliar passengers, and it is the use of recollection that successfully resolves the rapidly perceived familiarity oddness.

Returning to the use of the oddity task, within the context of basic recognition memory its use is straightforward. Subjects are presented with triplets in which the odd item is either studied or new to the experiment. Critically, because the query "Which is odd?" does not indicate whether the isolate is old or new, the observer cannot reduce the task to a maximum or minimum evidence selection rule; sometimes the odd item will be studied whereas other times it will be novel.

Under the signal-detection model, there are currently only two possible decision rules assumed for oddity: the triangular rule (Ennis, Mullen, & Frijters, 1988; Frijters, 1979; Peryam, 1958); and the independent-observation rule (Macmillan & Creelman, 2005; Versfeld, et al., 1996). Which rule observers should select depends upon how the triplets are constructed and whether observers are sensitive to the locations of the putative underlying evidence distributions. The triangular rule is simplest, and is most appropriate for designs that use a 'roving' standard. During roving designs there are no stationary evidence distributions across the trials. For example, an auditory frequency discrimination oddity task may use triplets from many different portions of the frequency continuum, although on any given trial those items would have the same hypothetical relative distance between the isolate and non-isolates. As noted by Versfeld, Dai, and Green (1996), under these conditions the triangular rule is statistically optimal. During this rule, the observer simply compares the relative evidence distances among the members of the triplet (see Figure 1 panel b.) The two items closest on the evidence dimension are assumed similar and the remaining item is classified as odd. The triangular rule is optimal when there is a large rove as there is no other basis of information upon which to make the judgment. Because the items can come from any portion of the evidence axis, the absolute location on this axis of any particular item is wholly uninformative, and only the relative distances among the items matters. The algebraic equation describing the triangular rule is:

[Equation 1]

$$Z_i = |x_i - \bar{x}|$$

where Z_i is the absolute value of the difference between each member of the triplet and the mean of the triplet, $\bar{x}$. The member with the largest Z value is the item selected as odd. This is equivalent to the psychological characterization of the rule above, namely, that observers find the two most similar items (in terms of evidence values) and then select the remaining item as the isolate.

The triangular rule is not statistically optimal if the stimuli are drawn from two constant locations (with added noise) on the evidence axis. Under these conditions, if the observer also explicitly knows the locations of the two evidence distributions, he or she can use this additional information to help solve the oddity problem by using a rule termed “independent-observation” (Macmillan & Creelman, 2005; Versfeld, et al., 1996). This rule can be thought of as an augmentation of the triangular rule where the observer uses both relative evidence among the items, and the position of the evidence along the evidence axis to render a judgment. Using this rule, the observer examines the evidence for all items (x_1, x_2, x_3) and the selects the item with the greatest absolute distance (Z) from the mean evidence of the current triplet ($\bar{x}$), adjusted by a constant, namely the central tendency of the two evidence distributions (μ).

[Equation 2]

$$Z_i = \left| x_i - \frac{3\bar{x} - \mu}{2} \right|.$$

The link between the independent-observation rule and the mental operations of the observer is not particularly transparent from the form of Equation 2. However, the rule reduces to the following steps. First the observer orders the stimuli along the evidence axis. Second, he or she classifies the center item along this axis using a standard single item recognition judgment and the criterion location μ in Equation 2, which corresponds to the optimal criterion midway between the old and new evidence distributions for such a judgment. If the item falls above this criterion then the minimum of the triplet is designated odd, if the intermediate item falls below this location then the maximum of the triplet is designated as odd (see Figure 1 panel c.) The "independent" nomenclature refers to the fact that the criterion, μ, is independent of the relative distances between the items.

If the two classes of items under discrimination are chosen from stationary distributions, then the independent-observation rule is superior to the triangular rule because the criterion, μ, is optimal for classifying any given item as old or new (Versfeld, et al., 1996). In the case of typical recognition memory experiments, the independent-observation rule would appear to be superior because the evidence distributions are assumed to be stationary. However, the ability of observers to use such a rule depends critically upon their awareness of the locations of the distributions and their appreciation of the superiority of this more complex rule. Since there is considerable doubt about the ability of observers to directly appreciate or deduce the positions and shapes of hypothetical evidence distributions during recognition (e.g., Benjamin & Bawa, 2004; Criss, 2009; Healy & Jones, 1975) we examine both rules here.

The Importance of the Unequal Variance Assumption

Although it is not widely appreciated, the unequal variance assumption of the standard signal-detection approach to recognition bears critically on judgments of mnemonic oddity. This is because the model is a statement regarding the perceived similarity of old and new item memoranda during recognition. Put simply, the model assumes that the evidence values for new items are more similar to one another than the evidence values for old items (Figure 1 panel a.) Because successful judgments of oddity are heavily dependent on the similarity of the non-odd items, the model predicts an asymmetry in performance depending upon whether the odd item is old or new. Looking at Figure 1 panel a. it is intuitively clear that under the situation in which the new item is odd (new item isolate trial) performance is more likely to be disrupted by the high variability of old items than under a situation in which the old item is odd. In this latter case, the lower variability of new item evidence is less likely to obscure the oddness of the single old item of a triplet. This intuitive prediction is verified below in the simulation section, which demonstrates that for the range of d' and old item variance values typically assumed in the literature, it should be quite robust. The mechanics of the decision rules are almost wholly unaffected by the unequal variance assumption. In the case of the triangular rule, Equation 1 holds unaltered; the observers again find the two most similar items and then select the remaining item as odd. In the case of the independent-observation rule, μ in the equation must be replaced with a criterion on the evidence axis that is optimal for a single item recognition judgment. Thus to bring the rule into line with the unequal variance signal-detection approach typically used in the literature (e.g., Glanzer, Hilford & Maloney, 2009; Ratcliff, Sheu & Gronlund 1992; Squire, Wixted & Clark, 2007; Stretch & Wixted, 1998) , the μ term in Equation 1 is replaced by an optimal single criterion (μ_{opt}) for the specific unequal variance model being simulated using the following equation (with d' and σ^2 representing the mean and variance of the old item distribution with the new item distribution mean and variance set at 0 and 1 respectively)[1]:

[Equation 3]

$$\mu_{opt} = \frac{\sqrt{d'^2\,\sigma^2 + 2\sigma^4 \ln(\sigma) - 2\sigma^2 \ln(\sigma)} - d'}{\sigma^2 - 1}.$$

Figure 3 about here

Figure 3 shows that this approach indeed appears to yield the maximum proportion of overall correct responding during the independent-observation rule when variances are unequal. The figure illustrates the proportion of overall correct responding, estimated through Monte Carlo simulation, for observers with an old item standard deviation of 1.25 and various criterion positions relative to μ_{opt} of Equation 3. As the criterion shifts from this position, overall performance declines.

[1] This equation provides the optimal criterion placement on the *x*-axis at which old and new item distributions intersect, above which returned evidence is more indicative of an old item than a new item, and below which returned evidence is more indicative of a new item than an old item. For formal derivation see the appendix of Stretch & Wixted (1998). We have not ascertained that this solution is analytically optimal for oddity judgments by maximum likelihood estimation, but we assume that this is the case since it would be optimal under single item recognition (see Figure 3).

It should also be noted that there is an additional intersection (μ_{alt}) of old and new item distributions at the lower end of the new item distribution located at:

[Equation 4]

$$\mu_{alt} = \frac{-\sqrt{d'^2\,\sigma^2 + 2\sigma^4 \ln(\sigma) - 2\sigma^2 \ln(\sigma)} - d'}{\sigma^2 - 1}.$$

Below this intersection, driven by its greater variance, the probability density function for the old item distribution once again exceeds the probability density function for the new item distribution. Whilst a comprehensively optimal decision rule would incorporate this lower criterion, the proportions of old and new item evidence distributions falling below this intersection are extremely small (4.02×10^{-10} and 3.21×10^{-10} respectively for a d' of 1.5 and a variance ratio of .80). Thus consistent with prior approaches we disregard this second location and the rule we present reflects the maximal performance achievable if the observer applies a *single* evidence criterion at the upper intersection in order to parse the old/new continuum.

In summary, the literature contrasting the unequal variance signal-detection and dual-process models has relied almost entirely on a single recognition paradigm, namely single item old/new recognition judgment accompanied by confidence reports. Although the fits of the models systematically favor the unequal variance signal-detection model, both signal-detection and dual-process models typically account for the vast majority of data and make qualitatively similar predictions. However, when the only two decision rules known for oddity judgments are considered in light of the unequal variance assumption, an arguably surprising prediction is made; as long as optimal decision rules from the sensory-perceptual domain hold within the memory domain, it should be more difficult to detect an odd new item than an odd old item. As we discuss more fully later, this is not a natural prediction for dual-process model approaches because they assume that individuals have more information at hand for old items (familiarity and recollection) versus new items (familiarity information only). Given this, there is more information to work with given a triplet with one new item and two old items (new item isolate trial) compared to a triplet with only one old item and two new items (old item isolate trial). If, as asserted by the Mandler (1980) anecdote, recollection is useful for disambiguating the familiarity of items then it is reasonable to assume that subjects will perform better for resolving triplets with two versus one old item present. This prediction is opposite that of the unequal variance signal-detection model and is explored more formally in the General Discussion.

While the decision modeling considerations above suggest that the oddity task may inform the single- versus dual-process memory debate, it is also important to emphasize that this research question bears critically upon the comparability of decision models for recognition memory and perception judgments. The original application of the theory of signal-detection to recognition memory hinged on the assumption that recognition memory, like perceptual discrimination, could be treated as a unidimensional statistical decision problem (e.g., Parks, 1966; Banks, 1970) and the oddity task provides us with another opportunity to critically re-evaluate this assumed comparability outside of single item judgments. Additionally, unlike single item recognition testing, recognition judgment outside the laboratory is often conducted against a background of simultaneous novelty or familiarity. For example, searching for a new acquaintance in the midst of an unfamiliar crowd is arguably a mnemonic oddity task. Similarly, although great pains are taken to warn eyewitnesses that perpetrators may be absent from lineups, it is nonetheless the case that the observers may treat these tasks as ones of relative mnemonic oddity judgment (for discussion see Wells et al. 1998). That is, they may assume that the array has an old item isolate (the perpetrator) present and search for the member whose familiarity isolates him or her most from the other members of the array. Thus understanding whether observers find new or old item isolates easier to resolve and evaluating which if any current decision models anticipate these observed differences (and generalizes among tasks), has importance outside of the single- versus dual-process model debate.

Before turning to simulations to solidify the predictions of the single- and dual-process theories outlined above, we first summarize the findings of four separate experiments contrasting the accuracy of oddity judgments for new versus old item isolate triplets. To preview the data, under oddity participants reliably find it easier to detect a new item hidden among two old items compared to the reverse. Basic Monte Carlo simulations then show this finding to be at odds with the predictions of the unequal variance signal-detection model using known optimal decision rules. To address this, in the General Discussion we develop entirely new decision rules to verify that the dual-process model anticipates the data and to attempt to salvage the unequal variance signal-detection model.

Empirical Findings

General Methods

Overview. We conducted four computer-presented experiments focusing on oddity recognition judgments. In all experiments, a three-alternative forced-choice (3-AFC) control task was also administered. In the 3-AFC task, participants were cued to identify the old item or the new item according to triplet construction (e.g., "Which is old?" or "Which is new?" as opposed to "Which is odd?" under oddity). The 3-AFC task provides a useful point of comparison due to the identical triplet construction and analogous requirement of participants to identify one item that satisfies the criteria indicated in the cue.

Before describing the procedures for each experiment we briefly describe the rational for the conditions that were considered across the experiments. Experiment 1 contrasted forced-choice recognition and oddity recognition judgments for identically constructed triplets. As demonstrated below, only oddity judgment was sensitive to the construction of the triplets, demonstrating an advantage for new item isolate trials versus old item isolate trials. Because performance was fairly low during the oddity judgments of Experiment 1, Experiment 2 incorporated a levels of processing (LOP) manipulation in order to improve general performance on the task and to determine whether the new item isolate advantage remained at higher levels of performance (Craik & Lockhart, 1972). The experiment again demonstrated that triplet construction did not affect forced-choice decisions but did influence oddity judgments, with the new item isolate advantage present at fairly high levels of performance. Although forced-choice remained insensitive to the construction of the triplets, we wondered if this factor would affect outcomes if observers were instead asked to rate each member of a triplet independently for simple item recognition (without confidence). Thus Experiment 3 introduced this independent-classification task using the triplets, while again examining forced-choice and oddity in order to replicate the prior findings. Neither the independent classification or forced-choice tasks were sensitive to triplet construction, but again, the oddity judgment was, confirming that the effect seems fairly specific to this task.

Finally, because all three experiments intermixed trials of different types, Experiment 4 re-examined forced-choice and oddity, but in a blocked fashion, in order to confirm that the new item isolate advantage during oddity was not somehow an artifact of trial mixing.

Stimuli. For each participant, a different set of words was randomly sampled from source lists comprising 1,607 common nouns in Experiment 1 and 1,216 common nouns in Experiments 2, 3 and 4. Randomization ensured the linguistic characteristics were matched across new and old item isolate oddity tasks.

Participants. Twenty-two (18 women), thirty-one (20 women), sixteen (10 women) and twenty-six (18 women) participants were tested. Data from an additional three participants in Experiment 4 were discarded due to overall performance being at chance. Participants in Experiments 1, 2 and 4 were compensated with course credit and participants in Experiment 3 received payment. Informed consent was obtained in accordance with the Institutional Review Board of Washington University. The behavioral data in Experiment 3 were collected during a functional magnetic resonance imaging study with participants situated in a scanner: the imaging data are not presented here.

Procedure: Experiment 1. After the presentation of on-screen instructions and a practice phase in which participants were familiarized with all the judgments to be made, there were four self-paced study-test cycles, each consisting of 120 words at study and 80 triplets at test. During encoding participants counted the number of syllables in each serially presented item. In each self-paced encoding trial, single words were presented above the cue "Syllables? 1/2/3/4 or more" and participants responded by pressing the appropriate key on the keyboard numberpad (1 through 4). Immediately following each study list, a test comprising randomly mixed trials was administered. The recognition trials were 20 oddity trials and 20 3-AFC trials (each with 10 new item isolate trials and 10 old item isolate trials). In all self-paced test trials, triplets were presented below a cue and were composed of either two old words (previously studied) and one new word (not previously studied), or two new words and one old word. Words were presented in a single column in three rows numbered 1 through 3; the isolate was equally likely to occupy each of the positions. Participants responded by pressing the number on the keyboard numberpad corresponding to the item they believed satisfied the criteria indicated by the cue. Oddity trials were cued by the prompt “old/new: ODD?” indicating the participant should select the item whose recognition status was odd. 3-AFC trials were cued by the prompt “old/new: NEW?” or “old/new: OLD?” indicating that participants should select either the new item or the old item. Across the entire experiment, there were a total of 40 trials of each type (oddity new item isolate, oddity old item isolate, 3-AFC new item isolate, 3-AFC old item isolate). Equivalent oddity and 3-AFC judgments were made using semantic criteria (living/non-living judgments) in other intermixed trials, but these do not bear centrally on the current hypotheses about episodic memory and are not discussed.

Procedure: Experiment 2. After the presentation of on-screen instructions and a practice phase in which participants were familiarized with all the judgments to be made, there were four self-paced study-test cycles, each consisting of 96 words at study and 64 triplets at test. In each study phase half the study items were encoded using a shallow LOP syllable counting task cued by the prompt “Syllables? 1/2/3/4 or more” and the remainder were encoded using a deep LOP pleasantness rating task cued by the prompt “Pleasantness? 1/2/3/4” (it was explained to participants in the instructions that a rating of 1 corresponded to the least pleasant rating and a rating of 4 corresponded to the most pleasant rating). Deep and shallow LOP encoding tasks were randomly intermixed within the same study phase. All other aspects of the study procedure were identical to those of Experiment 1. In each test run, there were 16 oddity judgments for triplets using shallowly encoded materials, 16 oddity judgments for triplets using deeply encoded materials, 16 forced 3-AFC trials for triplets using shallowly encoded materials and 16 3-AFC trials for triplets using deeply encoded materials (each with 8 new item isolate trials and 8 old item isolate trials). Encoding types were never mixed within a single triplet. Across the entire experiment this yielded a total of 32 triplets for each of the four recognition conditions under both oddity and 3-AFC (deeply encoded old items [new and old item isolate triplets]; shallowly encoded old items [new and old item isolate triplets]). Trials were cued by the same prompts as in Experiment 1. For oddity trials, following the indication of which member of the triplet they believed to be odd, participants also justified their selection by indicating whether they believed the selected item was old or new. These justifications are not presented or discussed here.

Procedure: Experiment 3. After the presentation of on-screen instructions and a practice phase in which participants were familiarized with all the judgments to be made, there were two study-test cycles, each consisting of 72 words at study and 48 triplets at test. During self-paced, unscanned encoding trials, participants made pleasantness ratings cued by the prompt “Is Pleasant?”, responding with MRI-safe button-box button-presses corresponding to either “yes” or “no”. Immediately following each study list, a scanned test comprising randomly mixed trials was administered. In each test run, there were 16 oddity trials, 16 3-AFC trials and 16 independent classification trials (each with 8 new item isolate trials and 8 old item isolate trials). Oddity and 3-AFC trials were presented in the same manner as previous experiments and participants responded with button-presses corresponding to the numbered item they believed satisfied the criteria indicated in the cue. Oddity trials were cued by the prompt “Which is Different?” and 3-AFC trials were cued by the prompt “Which is New?” or “Which is Old?” Independent classification trials were presented and responded to in a different manner. In each triplet, the three item numbers were replaced by an arrow that moved progressively down the list of three items indicating to which item participants were required to make “Old/New?”-cued responses. Responses were made using button-presses corresponding to either “new” or “old”. In each trial, participants had 7 seconds to render a response (or all three responses in the independent classification condition). If responses were not rendered completely by the end of the response period, the response was coded as incorrect and the next trial was initiated. Across the entire experiment, this procedure yielded a total of 16 trials of each isolate type (new and old item isolate triplets) for oddity, 3-AFC and independent classification. Passive fixation trials, in which participants were instructed to “relax”, were intermixed throughout the recognition judgments.

Procedure: Experiment 4. After the presentation of on-screen instructions and a practice phase in which participants were familiarized with all the judgments to be made, there were four study-test cycles, each consisting of 72 words at study and 48 triplets at test. There were two cycles of intermixed test trials and two cycles of blocked test trials. During intermixed cycles, as in the previous experiments, participants made oddity judgments intermixed amongst 3-AFC judgments in a randomized order within the same cycle. During blocked cycles, participants made only oddity judgments (one cycle) or 3-AFC judgments (one cycle). The order of the cycles was counterbalanced among participants. For all cycles, during encoding participants made self-paced pleasantness ratings in a procedure identical to Experiments 1 and 2. Immediately following each study list, a self-paced test was administered. In intermixed cycles, test runs comprised 24 oddity trials and 24 3-AFC trials (each with 12 new item isolate trials and 12 old item isolate trials). Across the entire experiment this yielded a total of 24 trials of each isolate type (new and old item isolate triplets) for oddity and 3-AFC during intermixed cycles. In one blocked cycle the test comprised 48 oddity trials and in the other the test comprised 48 3-AFC trials, again yielding 24 trials of each isolate type for both oddity and 3-AFC. In all cycles, oddity trials were cued by the prompt “which is ODD?” and 3-AFC trials were cued by the prompt “which is NEW?”

or “which is OLD?” Participants made responses using the keyboard numberpad in a manner identical to Experiments 1 and 2.

Results

Mean (and standard deviation) isolate identification accuracies for oddity, 3-AFC and independent classification tasks are shown in Table 1.

Table 1 about here

Oddity. Figure 4 shows six separate comparisons of new item isolate and old item isolate oddity performance across all experiments. In Experiment 1, $t(21) = 4.02$, $p < .001$, $d = .852$, and Experiment 3, $t(15) = 3.157$, $p = .006$, $d = .753$, performance was significantly better for new item isolates than old item isolates. Results from Experiment 2 were entered into a 2 x 2 (LOP x isolate type) repeated measures ANOVA revealing a main effect of LOP, $F(1,30) = 63.06$, $p < .001$, $\eta^2 = .678$, indicative of an overall accuracy advantage for triplets containing deeply processed items. There was no main effect of isolate type across LOPs, $F(1,30) = 1.87$, $p = .182$, $\eta^2 = .059$, but crucially there was a significant interaction, $F(1,30) = 23.31$, $p < .001$, $\eta^2 = .437$. Post hoc pair-wise contrasts demonstrated performance that was significantly better for new item isolates than old item isolates for triplets containing deeply encoded items, $t(30) = 3.89$, $p < .001$, $d = .666$, but not for triplets containing shallowly encoded items, $t(30) = -1.17$, $p = .249$, $d = -.254$. Results from Experiment 4 were entered into a 2 x 2 (presentation condition [blocked or intermixed] x isolate type) repeated measures ANOVA revealing no main effect of presentation condition, $F(1,25) = 1.07$, $p = .310$, $\eta^2 = .041$, suggesting that accuracy was similar for oddity regardless of whether these judgments were in isolation (blocked) or mixed amongst 3-AFC judgments (intermixed). There was a main effect of isolate type, $F(1,25) = 23.80$, $p < .001$, $\eta^2 = .488$, with performance better for new item isolate identification than old item isolate identification collapsed across blocked and intermixed trials, and there was no significant interaction, $F < 1$. Planned contrasts confirmed the new item isolate identification advantage in both blocked, $t(25) = 3.55$, $p = .002$, $d = .492$ and intermixed cycles, $t(25) = 3.83$, $p < .001$, $d = .478$.

Figure 4 about here

3-AFC and Independent Classification. Under 3-AFC, in all experiments there were no significant differences in performance across new and old item isolate identification trials: Experiment 1, $t(21) = 1.39$, $p = .178$, $d = -.255$; Experiment 2 shallow encoding, $t(30) = -1.29$, $p = .206$, $d = -.163$ and deep encoding, $t(30) = -1.01$, $p = .322$, $d = -.125$; Experiment 3, $t(15) = 0.85$, $p = .410$, $d = .101$; and Experiment 4 blocked cycles, $t(25) = -0.72$, $p = .480$, $d = -.109$ and intermixed cycles, $t(25) = -0.97$, $p = .340$, $d = -.088$.

The only task not considered above was the independent classification task in Experiment 3. As with the forced-choice comparisons, the construction of the triplets did not significantly affect accuracy on the task, which was scored as the proportion of trials in which all three separate judgments were correct, $t(15) = 1.87$, $p = .080$, $d = .274$.

Results Summary. The only recognition judgment that was reliably affected by the construction of the triplets was the judgment of mnemonic oddity. In five of the six comparisons there was a significant advantage for the new item isolate triplets over the old item isolate triplets. Thus the effect is quite reliable. In contrast, forced-choice judgments were insensitive to this stimulus difference in all six cases, and the independent classification task of Experiment 3 was also not sensitive to the construction of the triplets. Thus there is an empirical regularity in recognition data demonstrating that observers find it easier to identify a new item hidden among two old items compared to an old item hidden among two new items. Analogously, one could characterize this pattern as demonstrating that pairs of old items are more easily perceived as mnemonically similar to one another than pairs of new items. Below we formally simulate the dual-process and unequal variance models discussed in the introduction and demonstrate that the dual-process account more easily anticipates this finding based on the parameters that one typically observes during standard single item recognition procedures.

Before turning to the simulation results we comment on the one oddity comparison that failed to demonstrate a new isolate advantage, namely, the triplets constructed with shallowly encoded old items in Experiment 2. As is clear from Experiment 1, the effect does occur when such triplets are judged in isolation, thus it appears that the trial mixing of these types of triplets, with triplets whose old items have been encoded deeply, is what results in the elimination of the effect. Here we can only speculate on why this occurs, but it presumably reflects the fact that the retrieval of deep or vivid encoding information somehow serves to overshadow episodic information potentially available for the shallow triplets. We suspect this effect is linked to phenomena noted in the reality monitoring literature of Johnson, Raye, Foley & Foley (1981), often referred to as the "it-had-to-be-you" effect. During the task, observers attempt to discriminate whether test probes were previously self-generated or instead provided by an external agent. Critically, the test also contains new items. The general finding is that when observers incorrectly endorse new items as coming from the study episode, they also tend to ascribe them to the external source. This pattern is thought to reflect that fact that observers are monitoring their memories in search of evidence particularly diagnostic of prior self generation (e.g., evidence of cognitive operations). While some portion of new materials may seem spuriously familiar, they will generally lack this type of information, leading observers to conclude they originated from the external source.

In the current Experiment 2, we suspect that observers are similarly looking for vivid evidence indicative of having performed the deep processing task on the items (did I rate this for pleasantness?) Such a retrieval orientation arguably renders it less likely that they will recover or consciously apprehend the episodic information linked to the shallowly encoded materials. Since, as we more fully detail below, the dual-process model assumes the new item isolate advantage is a recollective retrieval phenomenon, this deep retrieval orientation would necessarily reduce the phenomenon for study materials whose prior processing poorly matches the sought after episodic content (for similar ideas see Jacoby, Shimizu, Daniels, & Rhodes, 2005). Although consistent with prior frameworks, this account clearly should be replicated in the context of oddity judgment, where support for it would strengthen the assumption that the effect is heavily linked to recollection. Nonetheless, it is important to note that the null effect for the shallowly processed materials in Experiment 2 does not weaken the general finding that the new item isolate advantage was found in four experiments and five separate contrasts, and thus is a robust phenomenon.

Turning to the simulation results, predictions of the oddity decision rules were generated by simple Monte Carlo procedures in which random samples of triplets (500,000) were drawn from old and new item distributions to generate old or new item isolate triplets (e.g., two draws from the old item distribution and one from the new item distribution for new item isolate triplets). For each simulation, the odd item was selected according to the relevant decision rule and the proportion of correct selections recorded. Oddity results from Experiments 1 to 4 are shown alongside simulation results for the unequal variance signal-detection model across a range of selected key parameter values in Figure 5. In the empirical data, the results are clear. Participants find the detection of new isolates easier than the detection of old isolates. That is, their performance is superior for the condition with two old items and a new item, compared to the condition with one old item and two new items. In contrast, the shaded symbols of the figure demonstrate the expectation of the only two known signal-detection rules assumed for oddity judgments. For both the simple triangular rule, in which the evidence values are directly compared, and the more complex independent classification rule with an optimal criterion, all data points lie clearly below the diagonal indicating that the model generally predicts that performance on the old item isolate trials is expected to be higher than that on the new item isolate trials. Thus the empirical data are categorically at odds with the only two decision rules known for oddity judgment under signal-detection theory; rules which are assumed valid in the domain of sensory perception (Frijters, 1979; though see also O'Mahoney, 1995)

As Figure 5 demonstrates, the mis-prediction of the signal-detection rules spans the range of *d'* values and old to new item variance differences that one would typically expect during single item recognition paradigms if the unequal variance signal-detection model were correct. Again, this is noteworthy because these are the only known signal-detection rules for such judgments and typically assumed correct for perceptual judgments. Furthermore, the independent-observation rule is in fact the optimal rule for recognition memory given the assumption of stationary evidence distributions (although it is unclear how observers would determine this). Nonetheless, both rules categorically fail in the way expected given simple visual consideration of the unequal variance model in Figure 1. Thus even without consideration of the dual-process predictions formalized below, the data clearly show that the optimal decision rule for oddity, based on a likelihood ratio criterion, fails to predict the current empirical findings. Given that several current approaches assume that observers use an optimal likelihood criterion during recognition (e.g., Glanzer, Hilford, & Maloney, 2009) this finding is newsworthy.

Figure 5 about here

In the General Discussion we consider ways in which the independent-observation signal-detection decision rules might be brought more in line with the data. We then propose a second viable decision rule that is consistent with dual-process models and move on to suggest ways in which the two competing decision rules might be further tested.

General Discussion

With respect to the unequal variance signal-detection model, the current data present two possibilities. First, the only two known decision rules for oddity may be incorrect for episodic recognition judgments, indicating that rules that work for perceptual judgments do not transfer to recognition memory. Second, the unequal variance signal-detection model itself may be fundamentally flawed. If the first possibility is true then a viable signal-detection rule can be found. If the second possibility is true, then the signal-detection model should be abandoned. Of course, a challenge to the signal-detection approach would benefit by demonstrating a plausible decision rule derived from an alternate model. Below we present a new decision rule appropriate to each of the circumstances outlined above: first, an alternative signal-detection-based rule; and second, a dual-process-based rule. Finally, we evaluate the two new rules using empirical data from single item recognition via a bootstrap-fed Monte Carlo simulation method to see which model actually generalizes from independent single item recognition data to the current oddity findings.

Before contrasting the novel signal-detection and dual-process rules presented below, it is important to note that they both share a commonality. Both rules start from the premise that observers pay unique attention to the intermediate item of the triplet. In the case of the original independent-observation rule this was already clear given the single item classification of the intermediate item via an optimal criterion. As we detail below, the dual-process rule also assumes that the intermediate item receives additional processing. In this case however, it is assumed that observers attempt to recollect context for this item. Importantly, the reason that both models assume additional processing for the intermediate item of the triplet is because it is this item that is decidedly ambiguous under oddity instructions. Under oddity instructions, observers explicitly know that every triplet contains at least one old item and one new item. Given this, the maximum and minimum items, with respect to the single evidence dimension under signal-detection or the familiarity process under dual-process, are most likely to have arisen from the old and new evidence distributions regardless of model. In short, on the vast majority of trials, the identities of these two items are not ambiguous, however, this ease of attribution does not hold for the intermediate evidence item whose likely origin is unclear. In the original triangular and independent-observation rules the inherent ambiguity of the intermediate item is resolved either one of two ways; by either determining whether it is most similar to the maximum or minimum (triangular rule), or by evaluating it with respect to an optimal absolute criterion (μ_{opt}; independent-observation rule). Given the categorical failure of these two rules, we examine two different ways to further process the intermediate item that might yield the correct pattern of performance for new and old item isolate triplets.

Alternative Viable Decision Rules

Flexible Criterion Independent-Observation Rule. The standard independent-observation rule uses an optimal criterion (μ_{opt}) to facilitate performance and improve upon the triangular rule. This fixed, optimal criterion placement maximizes accuracy for stationary distributions (Figure 3) but one can instead ask, what would happen if subjects systematically chose a non-optimal location for the criterion position. That is, what if all subjects tended to be liberal or conservative in the criterion that is applied to the intermediate item during the independent-observation rule. Unlike the two parameter model previously discussed, this flexible criterion model uses three key parameters: the distance between old and new distributions (*d'*); the variance of the old item distribution (σ^2); and crucially, a non-optimal criterion chosen for the classification of the intermediate item (*c*). As shown in Figure 6, assuming a generally liberal criterion (rectangles), as opposed to conservative criterion (crosses), aligns the decision rule with the experimental findings in that it now predicts a new item isolate advantage. Thus, if one takes the independent-observation rule, adds an unequal variance assumption, and then assumes that observers interrogate the intermediate using an old/new discrimination with a generally liberal criterion, then the signal-detection model is able to accommodate the current data. However, this approach faces at least two criticisms. First, the model is clearly ad hoc. As emphasized above, under signal-detection this is not the optimal approach to the task and the need to assume a generally liberal criterion directly contradicts with the assumption that in general observers use an optimal likelihood ratio criterion for single item recognition judgment (e.g., Hirshman, 1995; Shepard, 1967). Indeed, in the empirical data illustrated in Figure 2 and used for the first bootstrap Monte Carlo simulation, the 26 participants' mean criterion estimates did not differ from the mean optimal criterion (*M*s = 0.916 vs. 0.999; optimal criterion determined using Equation 3), $t(25) = -1.327$, $p = .196$, $d = -.219$. Thus not only would the current participants need to use a liberal criterion that reduces overall oddity performance (Figure 3), but this criterion deviates from the neutral criterion observed under the signal-detection model for single item recognition. Second, this version of the independent-observation model is arguably over-parameterized in that it can accommodate every possible new and old item isolate performance relationship during the oddity task. That is, as Figure 6 shows, it is quite capable of producing an old item isolate advantage even though one never actually occurs in the data. Far from representing admirable characteristic, such flexibility is the hallmark of an unfalsifiable model. If there were a model that, with an equal or lesser number of parameters, could not produce data in the lower half of Figure 6, it would be preferable on these grounds alone. As we show below, the dual-process model makes this more restricted and hence riskier prediction.

Figure 6 about here

Dual-Process Rule. Before outlining a viable dual-process decision rule, it is important to clarify the assumed nature of 'recollection' in order to establish how it might contribute to successful decision-making during the oddity task. Within dual-process theories, there are several postulated characteristics that distinguish recollection from familiarity. Here we focus on the assumption that compared with familiarity, recollection is a deliberate retrieval process that is resource intensive and is therefore employed in a strategic fashion (as opposed to relatively automatic employment of familiarity; e.g., Dehn & Engelkamp, 1997; Gruppuso, Lindsay, & Kelley, 1997; Jacoby, 1991; Yonelinas & Jacoby, 1996). This assumption is key to Mandler's (1980) 'butcher on the bus' anecdote and was also formalized in Atkinson & Juola's (1974) dual-process model, in which recollection search is initiated only if the initial rapid interrogation of familiarity returns an inconclusive outcome.

In the proposed dual-process decision rule for oddity, we draw on this assumption that recollection is resource-intensive and therefore strategically employed when it can decisively resolve a clear ambiguity in familiarity. As noted above, during oddity there is only one item whose familiarity is routinely ambiguous, namely, the item of the three with the intermediate familiarity value on the evidence dimension (Figure 1 panel c.) Whereas the independent-observation signal-detection decision rule attempts to resolve the ambiguity of this item by applying an old/new unequal variance judgment, the dual-process decision rule instead attempts to resolve the ambiguity of this intermediate item through a recollection attempt. If this attempt is successful, then the observer can conclude that the intermediate and maximum item are in fact old and hence will select the minimum item as odd. However, if recollection fails, then the item indicated by the triangular rule is selected. Thus the dual-process decision rule is as follows: participants order the stimuli along the equal variance signal-detection familiarity axis and evaluate the median item for recollection; if recollection succeeds then the minimum item is taken as odd; in the absence of recollection for the intermediate item, the triangular method is applied to the unidimensional familiarity values. In short, the dual-process rule is a simple extension of the triangular rule, in which the intermediate item is evaluated for recollection. If successful then the outcome of this retrieval attempt dictates the odd item. If it fails, then subjects fall back on the item that appears odd based on the triangular rule. Dual-process decision rule simulation outcomes are shown in Figure 7 for several values of recollection and familiarity.

Figure 7 about here

The dual-process rule is able to accommodate the observed oddity findings: it predicts a prominent new item isolate advantage. This occurs because participants can potentially recover recollective context for the intermediate familiarity items-an important source of additional evidence that can override the triangular rule on trials in which its use alone would otherwise lead to an error[2]. It is important to note that these predictions are not strictly

[2] The triangular rule utilizes the distances from the maximum and minimum familiarity items to the intermediate item. It can lead to errors even when the ordering of items is correct (a new item presents with the lowest familiarity and an old item presents with the highest familiarity) if the two like items are further apart than the odd and intermediate items. The additional context recovered when using the dual-process rule can sometimes overcome these distance-related errors, but only for new item isolate trials (where the intermediate item affords the potential for context recovery). For example, consider a potential recollection rate of 30% with a moderate *d'* of 1.5. Under these conditions, approximately 8% of trials that would have been incorrect using the triangular rule alone are saved by successful recollection with the intermediate item using the dual-process rule. It should also be noted that the dual-process rule can very rarely lead to errors in new item isolate trials that would not result from using the triangular rule alone. In just under 1% of trials, the new item isolate will spuriously present with maximum familiarity (displacing both of the old items), and recollection will succeed for the

dependent upon recollection search being applied only to the intermediate strength item. We have assumed this for simplicity and because it is consistent with the idea of observers limiting recollection attempts to materials whose familiarity content is ambiguous and hence conserving resources. Nonetheless, the predictions hold even if observers interrogate both the intermediate and the maximum familiarity items. This is because, in all possible outcomes, the recollection status of the most familiar item is incidental to the rendered judgment. For example, if the intermediate item evokes recollection, but the maximum item does not, this would still lead to the classification of the minimum item as the odd item: failures of recollection are not diagnostic of novelty and the maximum item has already returned high levels of familiarity. Conversely, if the observer recovers recollection for the maximum familiarity item but not for the intermediate item, the observer identifies the maximum as old via recollection-this would already have been assumed as it had returned the maximum familiarity. Thus the observer would still have to resort to the triangular rule in order to determine the status of the intermediate item. If recollection occurred for both items, the observer again gains no benefit from this outcome on the maximum familiarity item because he or she would have classified the minimum as odd based solely on the successful recollection for the intermediate item. Thus in general, the decision rule we present here, in which only the ambiguous intermediate item is interrogated for recollection, is optimal in terms of effort expended but identical patterns of behavior would also be obtained from a more comprehensive recollection search in which both the intermediate and most familiar items of the triplet were interrogated for recollection. Finally, we do not assume recollection is attempted for the minimum familiarity valued item because it conceptually makes little sense for observers to interrogate items perceived as novel for recollection of the study

---

intermediate (old) item, leading to the incorrect conclusion that the minimum item is odd. However, because of the relative infrequency of recollection induced losses compared to gains (an 8% gain vs. a 1% loss), use of the dual-process rule results in increased accuracy overall for new item isolate trials compared to when only the triangular rule is used (58% correct vs. 51% correct).

context. Critically, not only does the dual-process model correctly predict the new item isolate advantage using the typical parameters shown in Figure 7, it is notable for what it does not or cannot predict. Unlike the flexible criterion independent-observation rule, the dual-process model cannot produce an old item isolate advantage. If recollection completely fails (i.e., R = 0) then the model predicts that new and old item isolate performance is equivalent. Otherwise, as recollection increases, a new item isolate advantage is correctly predicted. Thus the model cannot produce patterns that do not seem to occur in actual empirical data.

Which Approach is Superior?

The two newly derived rules accommodate the clear new item isolate advantage in the empirical data via two fundamentally different mechanisms. In the case of the signal-detection approach, the asymmetry arises strictly because observers are systematically liberal in their old/new classifications of the intermediate item. Thus the effect is not a retrieval phenomenon but a systematic decision bias that just happens to favor the new item isolate triplets. In contrast, the dual-process approach assumes that observers attempt to retrieve additional recollective information for the intermediate item in order to resolve its ambiguous familiarity. Thus the explanation is one of retrieval not decision bias. Below we discuss both approaches before presenting two bootstrap Monte Carlo simulations that address the question "which of the two approaches is most likely to generalize to the current empirical data given the parameters actually observed in standard single item recognition?"

The newly derived signal-detection rule for oddity developed above may strike some as a suitable alternative given the errant predictions of the traditional triangular and independent-observation rules. For example, the fact that the intermediate item is evaluated with a single item recognition judgment process means a similar recognition process is applied during the task as would be applied during a test of single item recognition (albeit one that is augmented by an ordering process, an inferential selection rule, and a liberal criterion placement). However, the revised decision rule is potentially open to criticism on a number of fronts.

The first criticism is that summaries of the prior literature do not suggest that observers are systematically liberal during standard single-item recognition paradigms. We demonstrate this empirically below, however the aggregate data of the prior literature also suggests that observers, if anything, are typically somewhat conservative with respect to recognition judgments across a broad range of conditions. For example, Ratcliff, McKoon and Tindall's (1994) participants demonstrated conservative criteria for strongly encoded recognition items and also weakly encoded items during standard recognition procedures. Similarly, Snodgrass and Corwin (1988) demonstrated that elderly control participants typically held conservative criteria with only neurologically impaired participants sometimes demonstrating liberal criteria.

The second criticism, as noted earlier, is that the model is overly flexible and in fact capable of producing an old item isolate advantage that does not occur in the empirical findings. The third criticism focuses on the number of parameters that are necessary in order to simulate the empirical effect. Whereas the viable signal-detection model utilizes three key parameters (d', σ^2, and c), the dual-process decision rule utilizes only two (d' and R). Once again, this affords the signal-detection rule flexibility not afforded to the dual-process rule, which nevertheless makes the correct prediction despite the smaller number of parameters.

Because arguments of parsimony and excessive flexibility can be subjective, we statistically evaluate which model more naturally predicts the current data given what is already known about single item recognition performance. More specifically, we take the parameters obtained from fitting two sets of independent single item recognition data and use these parameters to generate a distribution of expected oddity outcomes using the bootstrap principle in combination with Monte Carlo simulation. We then evaluate whether the current findings are statistically unlikely given these outcome distributions. The use of independent datasets to obtain typical parameters from the models is necessitated by the fact that single item recognition data with confidence judgments were not collected in the current study, and so fits of ROCs cannot be estimated for these particular participants. Nonetheless, it is important to note that in terms of testing model generalizability, the use of independent data is in fact preferable because the starting parameters obtained from such data cannot in any way be contaminated by the specific procedures used in the current multi-item experiments. Despite this complete independence, a truly generalizable model should nonetheless be able to easily generate the correct predictions.

Bootstrap Monte Carlo A: As noted in the Introduction, Figure 2 shows an aggregate ROC from a standard single item recognition test with 26 participants. These data are wholly independent of Experiments 1-4 and were used to generate the parameters for Bootstrap Monte Carlo A. Following a syllable counting task performed on 180 items, each participant immediately performed old/new recognition classifications followed by a three point confidence rating (3 = high, 2 = medium, 1 = low) for 120 test items (60 old, 60 new). We fit the ROCs of each individual using the dual-process model of Yonelinas (2004) and the unequal variance signal-detection model, using nonlinear regression, a least squares criterion, and the Solver add-in in Microsoft Excel (for a detailed explanation see Harris, 1998). Table 2 shows the mean parameter estimates obtained for each model and the average quality of fit. Critically, for each model this provides a set of parameters sampled from a theoretical population of parameters that, for single item recognition, are ideal. We then applied these parameters to the current decision rules using a bootstrap principle. That is, a random sample of 20 parameter fits (with replacement) was repeatedly drawn from each full set of fitted parameters. This is referred to as a parametric bootstrap and the purpose of the procedure is to introduce variability across the sampled sets of parameters (in this case samples of 20) that should closely approximate how one might expect the parameters to vary if we had conducted a large number of identical experiments using 20 participants, repeatedly drawing from the student population (Efron & Tibshirani, 1993). Following each re-sampling, the mean old and new item isolate performance estimates for the sample were then generated using the Monte Carlo simulation applied to each case within the sample. Forty trials of each isolate type were drawn for each simulated participant. For example, if Subject #1 had a *d′* of 1.25, an old item standard deviation of 1.25, and a criterion location of .50, then these three values would be used in the Monte Carlo simulation to generate the percentages for the new item and old item isolate trials using the independent-observation signal-detection rule applied to the 80 draws. This procedure is repeated for all 20 re-sampled subjects' parameter sets and then the mean new and old isolate performance across the 20 cases is plotted for this particular bootstrap re-sample. Following this, the next bootstrap resample is drawn and the process repeated again to generate the next pair of mean percentages for the old versus new item isolate tasks. This was repeated for 1000 experiment resamples.

Bootstrap Monte Carlo B: The same bootstrap procedure was carried out on a second independent dataset with 40 participants to generate Bootstrap Monte Carlo B. To generate this independent dataset, participants performed a syllable counting task on 60 items, and then immediately performed old/new recognition classifications followed by a three point confidence rating (3 = high, 2 = medium, 1 = low) for each of 120 test items (60 old, 60 new). Table 2 shows the mean parameter estimates and quality of fit for each model.

Table 2 about here

This bootstrap-fed Monte Carlo process yields an expected distribution of experiment outcomes (1000 experiments each with 20 participants) under each model. If the empirical oddity findings are considerably more unlikely under one rule than the other, then the former is rejected in favor of the latter, which by simulation demonstrates greater generalization across single item recognition and the current oddity recognition paradigm. This approach favors the model that is capable of generating accurate predictions of the current empirical data pattern based solely on the bootstrapped population of parameters typically observed during single item recognition performance. Thus the analysis is a formal test of model generalizability. Critically, the predictions are not in any way statistically dependent upon the empirical findings in the current experiments because the parameters are obtained through the fitting of independent data gathered during single item recognition.

Figure 8 about here

As shown in Figure 8, the bootstrap-fed Monte Carlo simulations from both sets of parameters favor the dual-process model. In 98.8% (Bootstrap Monte Carlo A) and 100% (Bootstrap Monte Carlo B) of the bootstrapped samples, the dual-process model yielded a new item isolate advantage consistent with the pattern in the empirical data. This means that the parameters that are typically encountered when fitting standard recognition data following syllable encoding, lead directly to the prediction of a new-isolate item advantage in the current oddity task. In contrast, the unequal variance signal-detection model yielded this outcome in only 18.4% (Bootstrap Monte Carlo A) and 6.2% (Bootstrap Monte Carlo B) of the bootstrapped samples-predictions that are markedly different to that of the dual-process model, and far less consistent with the overall pattern of results observed in Experiments 1-4. Put simply, the current empirical outcomes, in which four separate experiments demonstrated a new item isolate advantage, are anticipated well by the recollection (*R*) and familiarity (d') parameters one typically obtains in fits of the dual-process model to single item recognition data. The same cannot be said for the unequal variance model and its parameters (d', σ^2, and c), which in turn means that it requires parameters of a direction and magnitude that would be unusual in order to capture the current phenomenon. Thus the simulations favor the interpretation of the current new item isolate advantage in terms of episodic retrieval, not an unusually liberal response bias.

Although the data and simulations favor the dual-process model and decision rule further testing is warranted given this new and interesting task. Given that recollection is often argued to decline more rapidly than familiarity with aging (e.g., Grady & Craik, 2000; Jacoby, 1999; Jennings & Jacoby, 1997; Titov & Knight, 1997) the model also predicts that the new item isolate advantage should consequently decline with age. Additionally, because recollection is effortful and presumably mediated by resource-limited prefrontal cortex mechanisms (e.g., Dobbins, Foley, Schacter, & Wagner, 2002; Henson, Shallice, & Dolan, 1999; Rugg, Fletcher, Chua, & Dolan, 1999; Simons, Owen, Fletcher, & Burgess, 2005), one might also predict that the new item isolate advantage may be disrupted by the addition of a demanding secondary task during oddity judgment or the speeding of oddity decisions (e.g., Yonelinas & Jacoby, 1996). These testable hypotheses follow directly from the rationale on which the decision rule is founded–a failure to support them would therefore pose problems to the dual-process decision rule we have proposed to explain the oddity findings.

Conclusion

The current data demonstrate a novel, robust, and important empirical phenomenon; a decided accuracy advantage for new versus old item isolates in judgments of mnemonic oddity. This effect is not predicted or accommodated by existing oddity task decision rules and represents the only case that we are aware of in which decision rules with support in the perception literature, are nonetheless demonstrably invalid in the case of recognition memory judgment. To address this, we developed two novel decision rules capable of accommodating the new item isolate advantage in oddity. Although these competing decision rules both required additional processing of the intermediate familiarity item of each triplet, they differed in their theoretical foundations (signal-detection vs. dual-process) and how they addressed the inherent ambiguity of the intermediate strength item of the triplets. The flexible criterion independent-observation rule assumes the phenomenon reflects a systematic decision bias on the part of observers, whilst the dual-process rule assumes the effect results from context recollection. Given its consistency with findings supporting the use of recollection to disambiguate trace strength when classifying memory probes (Hintzman, 1988; Huppert & Piercy, 1978; Mandler, 1980), we favor the retrieval-driven dual-process decision approach. Additionally, the bootstrap-fed Monte Carlo simulations demonstrated that the observed empirical findings are entirely consistent with the generated predictions of the dual-process model based on the fitting of simple, single item recognition data. Thus the model clearly generalizes across two fairly different recognition tasks. In contrast, the simulation establishes that endorsement of the unequal variance signal-detection model would require one to assume that oddity judgments represent a special case of recognition task, whereby for entirely unknown reasons, observers deviate systematically from how they typically place the old/new criterion during standard single item recognition and do so in a decidedly non-optimal manner. It would also require one to favor the model that is arguably over-parameterized as it is also capable of accommodating findings that do not actually occur in the empirical data (viz., an old item isolate advantage).

Returning to the actual empirical finding, it is perhaps important to note the uniqueness of the current effect. Traditional dissociations in recognition memory research often revolve around manipulations of the featural properties of the test stimuli (e.g., the word frequency effect), the processes engaged at encoding (e.g., levels of processing manipulations), or tasks engaged during recognition (e.g., item vs. associative recognition).

In contrast, the current finding illustrates a robust difference that arises solely from the memorial context. That is, the only thing separating new and old item isolate triplets during oddity judgments is the background of memory evidence against which the isolate occurs. Critically however, all four experiments demonstrated that triplet construction had absolutely no effect on forced-choice recognition performance. Thus, overall, the empirical findings suggest that there is something very special about the way decision processes interact with triplet construction during oddity judgments.

[Acknowledgement] This work was supported by National Institutes of Health Grant MH07398

References

Amerine, M. A., Pangbourn, R. M., & Roessler, E. B. (1965). *Principles of Sensory Evaluation of Food*. New York: Academic Press.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and Performance IV*. New York: Academic Press.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74,* 81-99. doi: 10.1037/h0029531

Bayley, P. J., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2008). Yes/No Recognition, Forced-choice Recognition, and the Human Hippocampus. *Journal of Cognitive Neuroscience*, *20*(3), 505-512. doi: 10.1162/jocn.2008.20038

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language*, *51*(2), 159-172. doi: 10.1016/j.jml.2004.04.001

Brandt, F. I., & Arnold, R. G. (1977). Sensory tests used in food product development. *Food Product Development*, *11*(8), 56.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671-684. doi: 10.1016/S0022-5371(72)80001-X

Criss, A. H. (2009). The distribution of subjective memory strength: List strength and response bias. *Cognitive Psychology*, *59*(4), 297-319. doi: 10.1016/j.cogpsych.2009.07.003

Dehn, D. M., & Engelkamp, J. (1997). Process Dissociation Procedure: Double Dissociations Following Divided Attention and Speeded Responding. *The Quarterly Journal of Experimental Psychology A, 50*(2), 318-336. doi: 10.1080/027249897392116

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual process account. *Psychonomic Bulletin & Review, 13*, 1-21.

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive Control during Episodic Retrieval: Multiple Prefrontal Processes Subserve Source Memory. *Neuron, 35*(5), 989-996.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*(4), 523-533.

Dunn, J. C. (2004). Remember-Know: A Matter of Confidence. *Psychological Review, 111*(2), 524-542. doi: 10.1037/0033-295X.111.2.524

Efron, B., & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Egan, J. P. (1975). *Signal detection theory and ROC-analysis*. New York: Academic press.

Ennis, D. M., Mullen, K., & Frijters, J. E. R. (1988). Variants of the method of triads: Unidimensional Thurstonian models. *British Journal of Mathematical and Statistical Psychology, 41*, 25-36.

Frijters, J. E. R. (1979). The paradox of discriminatory nondiscriminators resolved. *Chemical Senses and Flavour, 4*(4), 355-358. doi: 10.1093/chemse/4.4.355

Frijters, J. E. R., Kooistra, A., & Vereijken, P. F. G. (1980). Tables of d' for the triangular method and the 3-AFC signal detection procedure. *Perception & psychophysics., 27*(2), 176-178.

Glanzer, M., Hilford, A., & Maloney, L. T. (2009). Likelihood ratio decisions in memory: Three implied regularities. *Psychonomic Bulletin & Review, 16*, 431-455. doi: 10.3758/PBR.16.3.431

Grady, C. L., & Craik, F. I. M. (2000). Changes in memory processing with age. *Current Opinion in Neurobiology, 10*(2), 224-231. doi: 10.1016/S0959-4388(00)00073-8

Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The Process-Dissociation Procedure and Similarity: Defining and Estimating Recollection and Familiarity in Recognition Memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 23*(2), 250-278. doi: 10.1037//0278-7393.23.2.259

Harris, D. C. (1998). Nonlinear Least-Squares Curve Fitting with Microsoft Excel Solver. *Journal of Chemical Education, 75*(1), 119-121. doi: 10.1021/ed075p119

Healy, A. F., & Jones, C. (1975). Can subjects maintain a constant criterion in a memory task? *Memory & Cognition, 3*(3), 233-238.

Helm, E., & Trolle, B. (1946). Selection of a taste panel. *Wallerstem Laboratory Communications, 9,* 181-191.

Heathcote, A. (2003). Item Recognition Memory and the Receiver Operating Characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1210-1230. doi: 10.1037/0278-7393.29.6.1210

Henson, R. N. A., Shallice, T., & Dolan, R. J. (1999). Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. *Brain, 122*(7), 1367-1381. doi: 10.1093/brain/122.7.1367

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528-551. doi: 10.1037/0033-295X.95.4.528

Hintzman, D. L., & Curran, T. (1994). Retrieval Dynamics of Recognition and Frequency Judgments: Evidence for Separate Processes of Familiarity and Recall. *Journal of Memory and Language*, *33*(1), 1-18. doi: DOI: 10.1006/jmla.1994.1001

Hirshman, E. (1995). Decision Processes in Recognition Memory: Criterion Shifts and the List-Strength Paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 302-313. doi: 10.1037//0278-7393.21.2.302

Huppert, F. A., & Piercy, M. (1978). The role of trace strength in recency and frequency judgements by amnesic and control subjects. *Quarterly Journal of Experimental Psychology, 30*(2), 347-354. doi: 10.1080/14640747808400681

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513-541. doi: 10.1016/0749-596X(91)90025-F

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. . *Journal of Experimental Psychology: Learning, Memory & Cognition, 25*, 3-22. doi: 10.1037/0278-7393.25.1.3

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*(5), 852-857.

Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology General, 138*(2), 291–306. doi: 10.1037/a0015525

Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology & Aging, 12*, 352-361. doi: 10.1037/0882-7974.12.2.352

Johnson, M K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology, 94*(1), 37-64. doi: 10.2307/1422342

Khoe, W., Kroll, N. E. A., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes-no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia, 38*(10), 1333-1341. doi: 10.1016/s0028-3932(00)00055-5

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (Second ed.). New York: Lawrence Erlbaum.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252-271. doi: 10.1037/0033-295X.87.3.252

O’Mahoney, M. (1995). Who told you the triangle test was simple? *Food Quality and Preference,* 6, 227-238. doi: 10.1016/0950-3293(95)00022-4

Parks, T. E. (1966). Signal detectability theory of recognition memory performance. *Psychological Review, 73*, 44-58. doi:10.1037/h0022662

Peryam, D. R. (1958). Sensory difference tests. *Food Technology*, *12*, 231-236.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 763-785. doi: 10.1037/0278-7393.20.4.763

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing Global Memory Models Using ROC Curves. *Psychological Review*, *99*(3), 518-535. doi: 10.1037/0033-295X.99.3.518

Rugg, M. D., Fletcher, P. C., Chua, P. M. L., & Dolan, R. J. (1999). The Role of the Prefrontal Cortex in Recognition Memory and Memory for Source: An fMRI Study. *NeuroImage*, *10*(5), 520-529. doi: 10.1006/nimg.1999.0488

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Leaning and Verbal Behavior, 6,* 156-163. doi: 10.1016/S0022-5371(67)80067-7

Simons, J. S., Owen, A. M., Fletcher, P. C., & Burgess, P. W. (2005). Anterior prefrontal cortex and the recollection of contextual information. *Neuropsychologia, 43*(12), 1774-1783. doi: 10.1016/j.neuropsychologia.2005.02.004

Smith, D. G., & Duncan, M. J. J. (2004). Testing Theories of Recognition Memory by Predicting Performance Across Paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 615-625. doi: 10.1037/0278-7393.30.3.615

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of Measuring Recognition Memory: Applications to Dementia and Amnesia. *Journal of Experimental Psychology: General, 117*(1), 34-50. doi: 10.1037/0096-3445.117.1.34

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: a new perspective. *Nature Reviews Neuroscience, 8,* 872-883. doi: 10.1038/nrn2154

Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1397-1410. doi: 10.1037/0278-7393.24.6.1397

Titov, N., & Knight, R. G. (1997). Adult age differences in controlled and automatic memory processing. *Psychology & Aging, 12*, 565-573. doi: 10.1037//0882-7974.12.4.565

Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Neuropsychologia, 46*(7), 1787-1799. doi: 10.1016/j.neuropsychologia.2008.01.004

Versfeld, N. J., Dai, H., & Green, D. M. (1996). The optimum decision rules for the oddity task. *Perception & Psychophysics, 58*(1), 10-21.

Wais, P. E. (2008). fMRI signals associated with memory strength in the medial temporal lobes: A meta-analysis. *Neuropsychologia, 46*(14), 3185-3196. doi: 10.1016/j.neuropsychologia.2008.08.025

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads. *Law and Human Behavior, 22*(6), 603-647. doi: 10.1023/A:1025750605807

Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of Remember/Know judgments. *Psychonomic Bulletin & Review, 11*, 616-641.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory & Cognition, 20*, 1341-1354. doi: 10.1037/0278-7393.20.6.1341

Yonelinas, A. P., Dobbins, I. G., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-Detection, Threshold, and Dual-Process Models of Recognition Memory: ROCs and Conscious Recollection. *Consciousness and Cognition, 5*(4), 418-441. doi: 10.1006/ccog.1996.0026

Yonelinas, A. P., & Jacoby, L. L. (1996). Noncriterial Recollection: Familiarity as Automatic, Irrelevant Recollection. *Consciousness and Cognition, 5*, 131-141. doi: 10.1006/ccog.1996.0008

Tables

Table 1

New and old item isolate identification accuracy

	Oddity		3-AFC		Independent Classification	
Isolate	New	Old	New	Old	New	Old
Experiment 1	.46 (.16)	.33 (.15)	.62 (.17)	.58 (.15)	-	-
Experiment 2: Shallow	.54 (.15)	.58 (.18)	.71 (.16)	.74 (.13)	-	-
Experiment 2: Deep	.78 (.18)	.66 (.19)	.88 (.11)	.89 (.10)	-	-
Experiment 3	.70 (.20)	.53 (.25)	.85 (.16)	.83 (.20)	.66 (.20)	.60 (.24)
Experiment 4: Blocked	.69 (.23)	.57 (.27)	.80 (.28)	.83 (.17)	-	-
Experiment 4: Intermixed	.66 (.22)	.55 (.25)	.77 (.26)	.79 (.23)	-	-

Note. Mean isolate identification accuracy for new and old item isolate triplets under conditions of oddity, three-alternative forced-choice (3-AFC) and independent classification. Standard deviations are shown in parentheses. Under oddity and 3-AFC, accuracy reflects the proportion of trials in which participants selected the isolate as the odd item (oddity) or the new or old item (3-AFC). Under independent classification, accuracy reflects the proportion of trials in which all items were classified correctly.

Table 2

Single item recognition parameter estimates.

Bootstrap	Model	d'	c	σ^2	R	% variance explained
A	Signal-detection	1.77 (0.81)	0.92 (0.40)	2.30 (1.80)	-	99.90 (0.15)
	Dual-process	1.11 (0.58)	0.88 (0.37)	1 (-)	.29 (.20)	99.86 (0.17)
B	Signal-detection	2.35 (1.00)	1.15 (0.41)	2.29 (1.62)	-	99.95 (0.05)
	Dual-process	1.50 (0.57)	1.13 (0.40)	1 (-)	.35 (.22)	99.92 (0.10)

Note. Mean parameter estimates and % variance explained under signal-detection and dual-process models for datasets used in Bootstraps A and B. Standard deviations are shown in parentheses. d' (old item distribution mean) and c (criterion) are estimated for both models though c is not utilized within the dual-process oddity decision rule. σ^2 (old item distribution variance) is fixed at 1 under the dual-process model and R (recollection rate) is estimated only for the dual-process model.

Figures

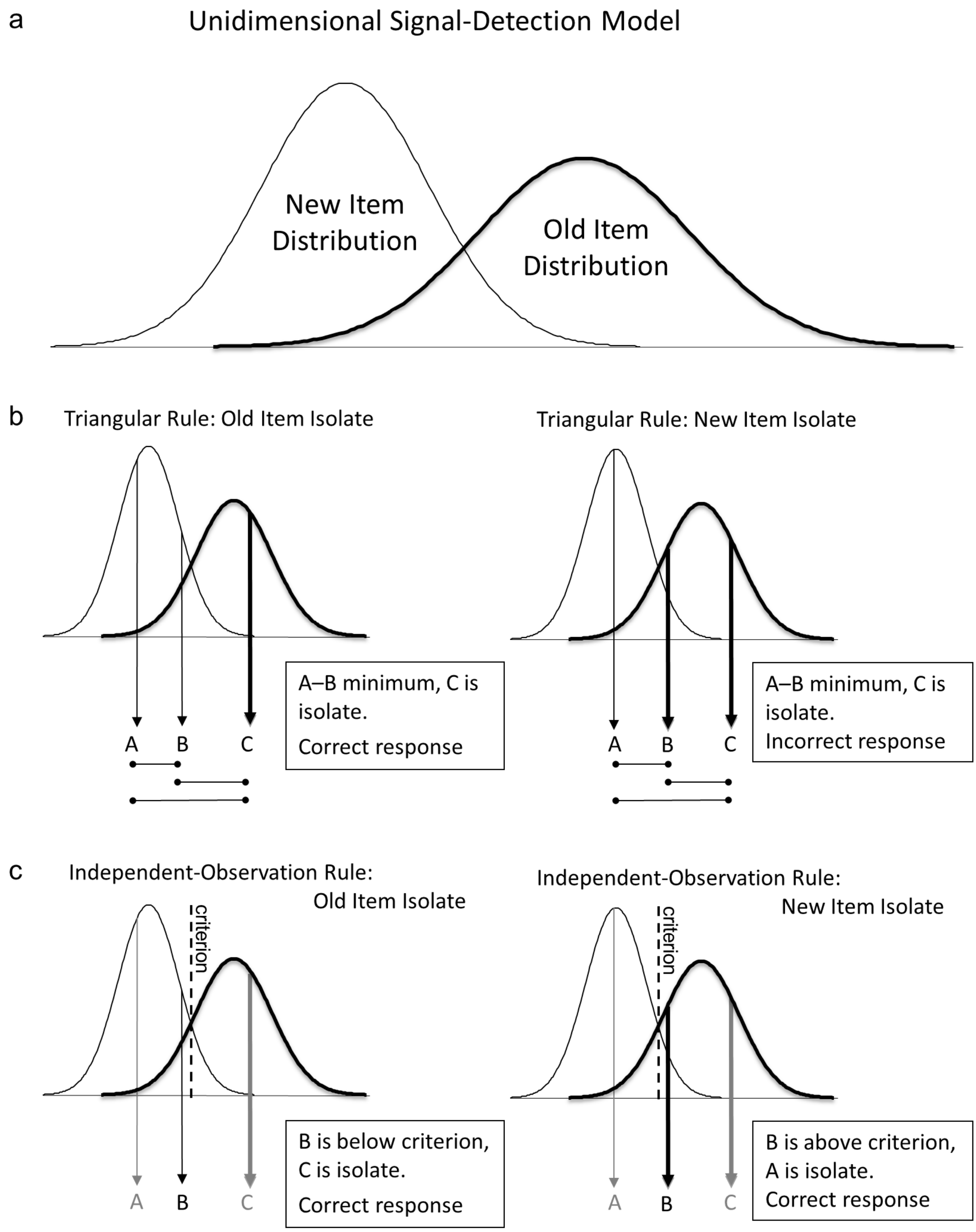


Figure 1. (Panel a.) The unidimensional signal-detection model with old item distribution variance greater than new item distribution variance. Evidence in favor of an item being old increases the further an item is positioned to the right. (Panel b.) The triangular decision rule

applied to old and new item isolate oddity trials. The triangular rule is the optimal signal-detection decision rule for the oddity task when a rove is in place. (Panel c.) The independent-observation decision rule applied to old and new item isolate trials. The independent-observation rule is the optimal signal-detection decision rule for the oddity task when no rove is in place.

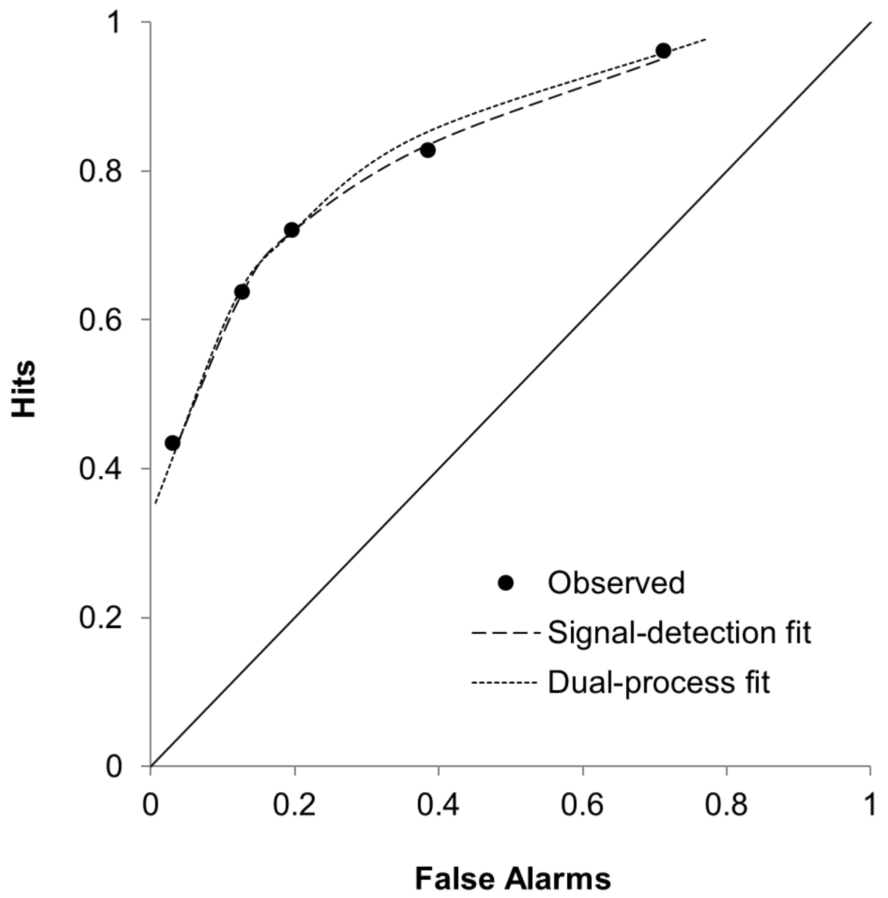


Figure 2. Aggregate ROC fits from 26 participants in a single item recognition task.

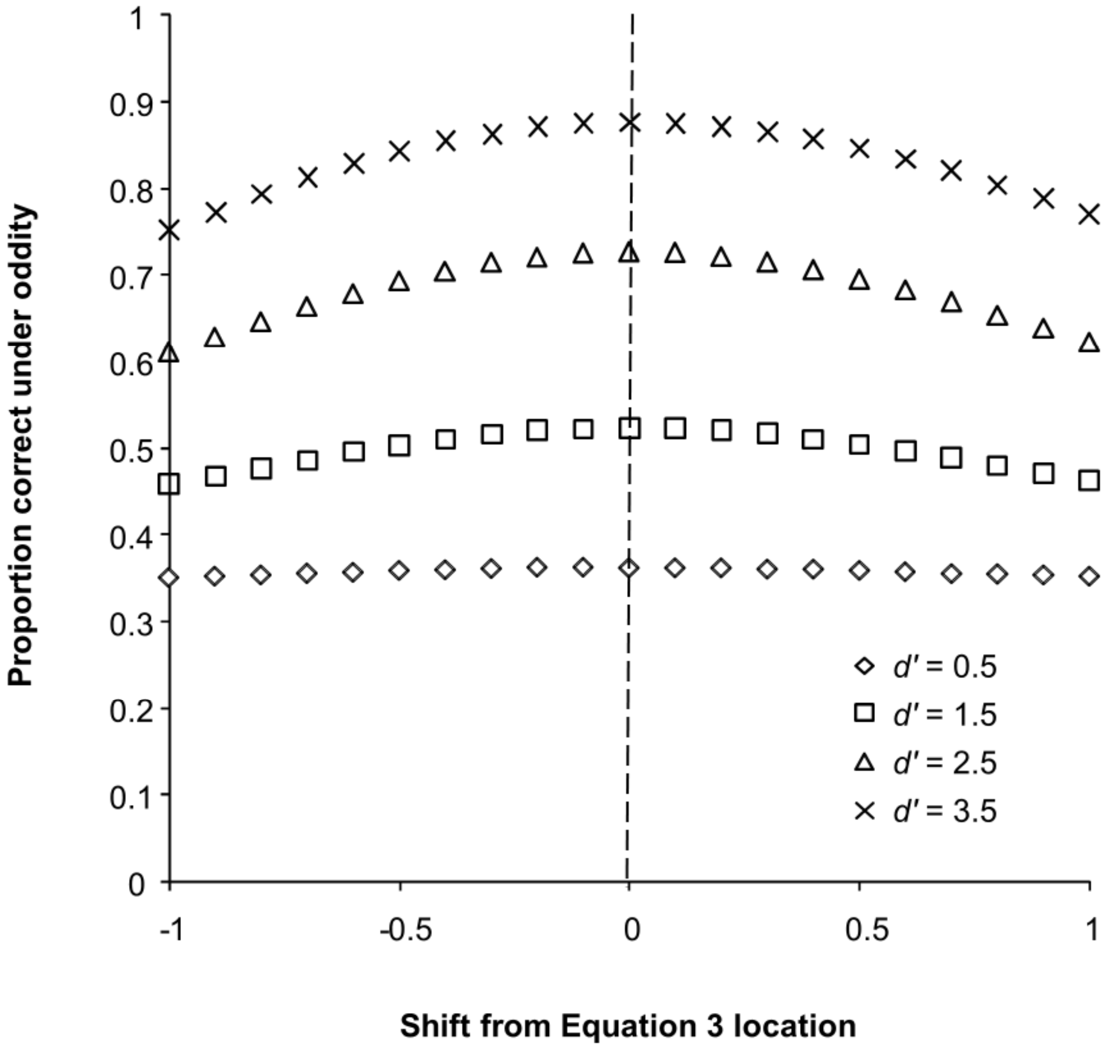


Figure 3. Simulated proportion of responses correct using the independent-observation decision rule. In all simulations, *d'* and criterion location (represented on *x*-axis as deviation from μ_{opt} calculated using Equation 3) are varied but old item distribution variance is kept constant at 1.25 (new item distribution variance is 1). As the criterion shifts from μ_{opt}, overall performance declines.

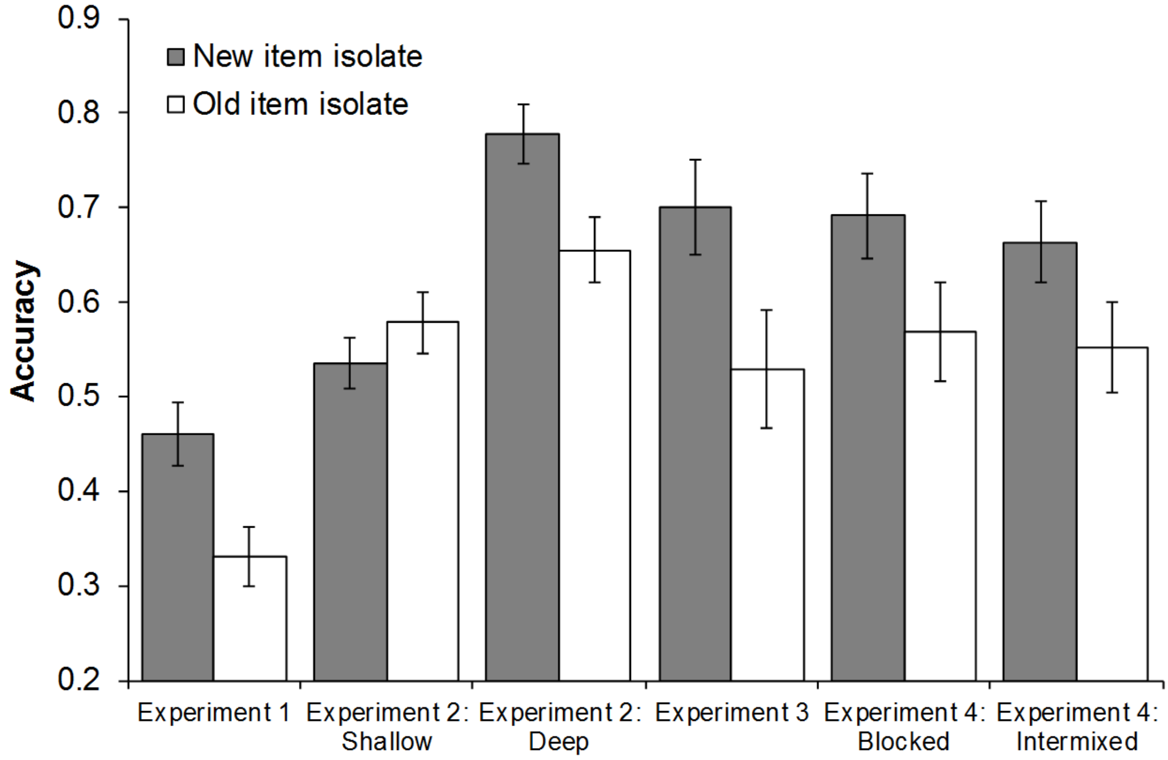
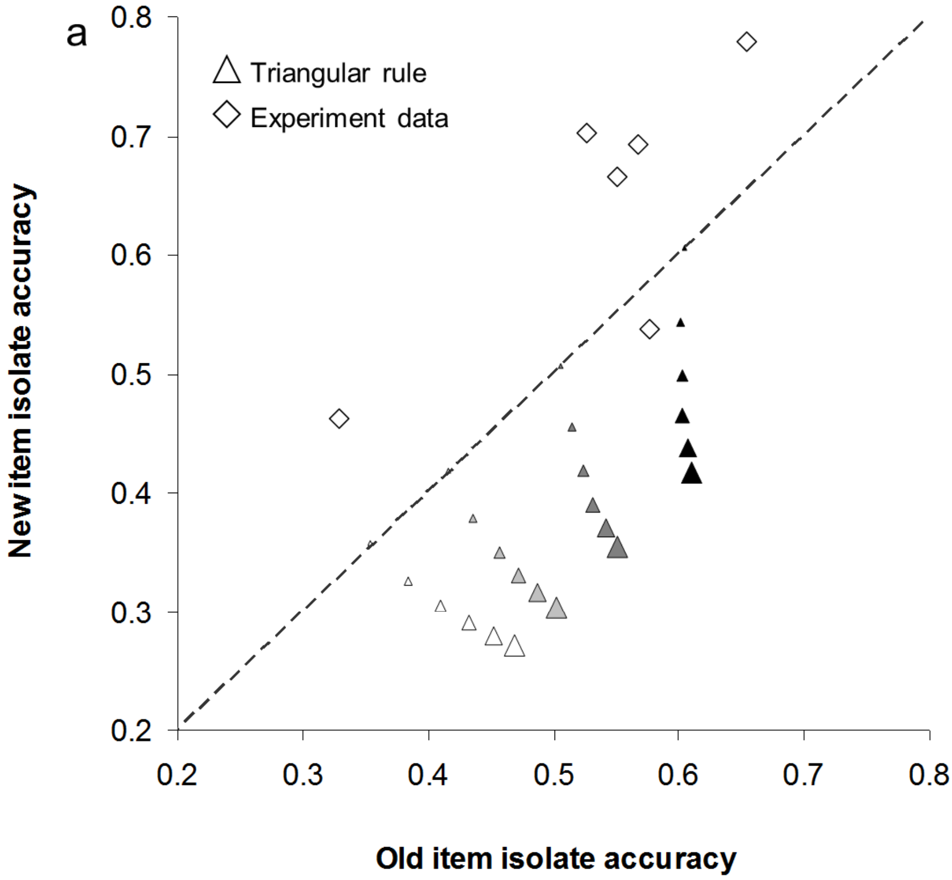


Figure 4. Mean proportion of responses correct in Experiments 1-4 for the oddity task. Error bars represent standard errors of the means.


a
Triangular rule
Experiment data
New item isolate accuracy
Old item isolate accuracy
0.2
0.3
0.4
0.5
0.6
0.7
0.8

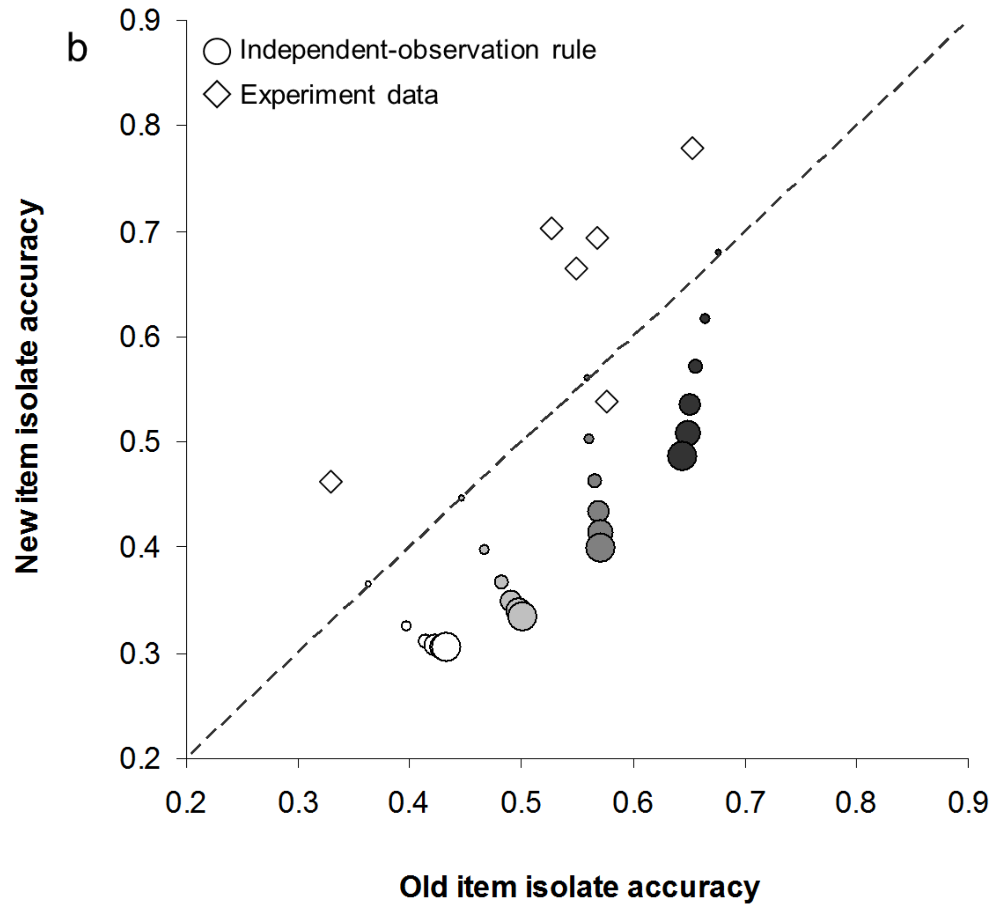

b
Independent-observation rule
Experiment data
New item isolate accuracy
Old item isolate accuracy
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Figure 5. Simulated old and new item isolate oddity task accuracies for the triangular (Panel a.) and independent-observation (Panel b.) decision rules. The dashed line represents the line along which new- and old item isolate accuracies are equal. Simulations with varying *d'*s (from 0.5 to 2.0 in steps of 0.5, represented by points shaded from white to black) and old item distribution variances (from 1.0 to 3.0 in steps of 0.4, represented by points with an increasing area; the new item distribution variance is fixed at 1.0) are shown.

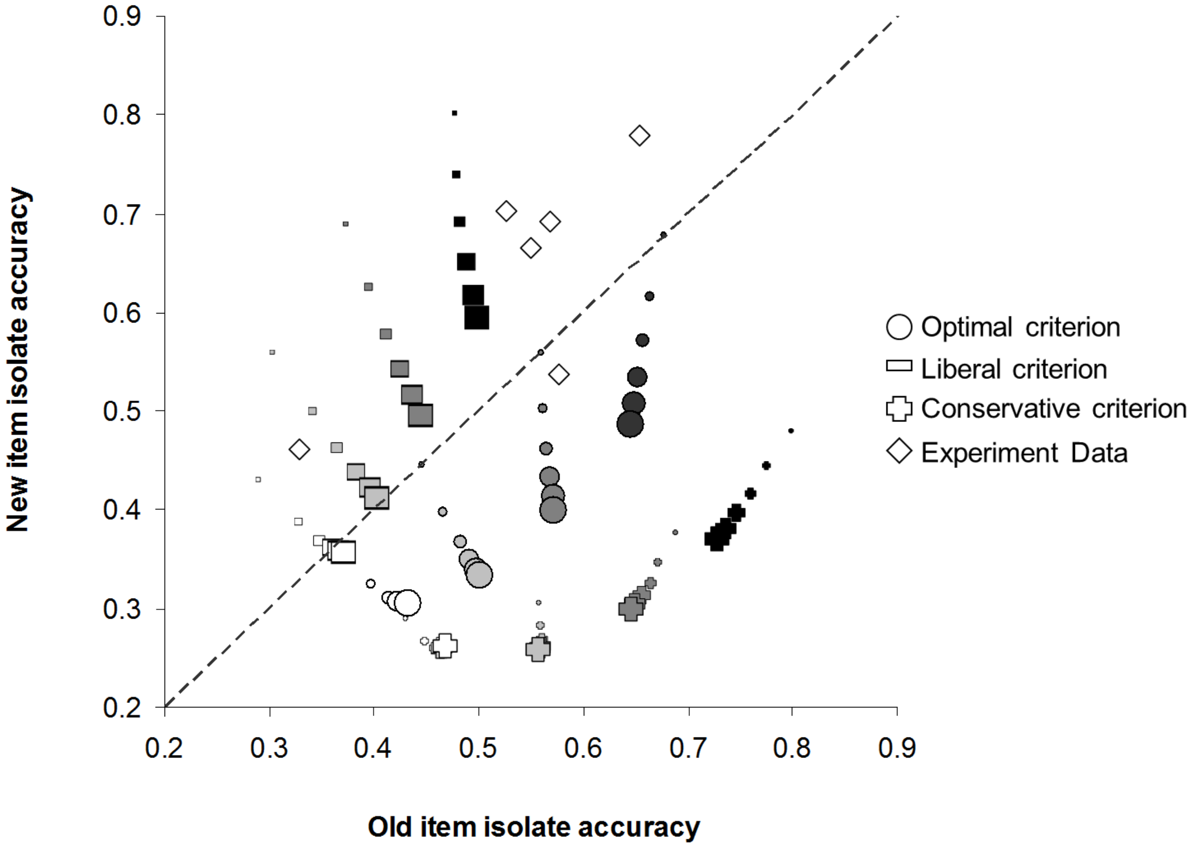


Figure 6. Simulated old and new item isolate oddity task accuracies for the independent-observation decision rule with optimal (circles) and flexible (rectangles and crosses) criterion placements. Simulations with varying *d*'s (from 0.5 to 2.0 in steps of 0.5, represented by points shaded from white to black) and old item distribution variances (from 1.0 to 3.0 in steps of 0.4, represented by points with an increasing area; the new item distribution variance is fixed at 1.0) are shown. The dashed line represents the line along which new and old item isolate accuracies are equal. The conservative criterion placement was 0.5 above optimal and the liberal criterion placement was 0.5 below optimal.

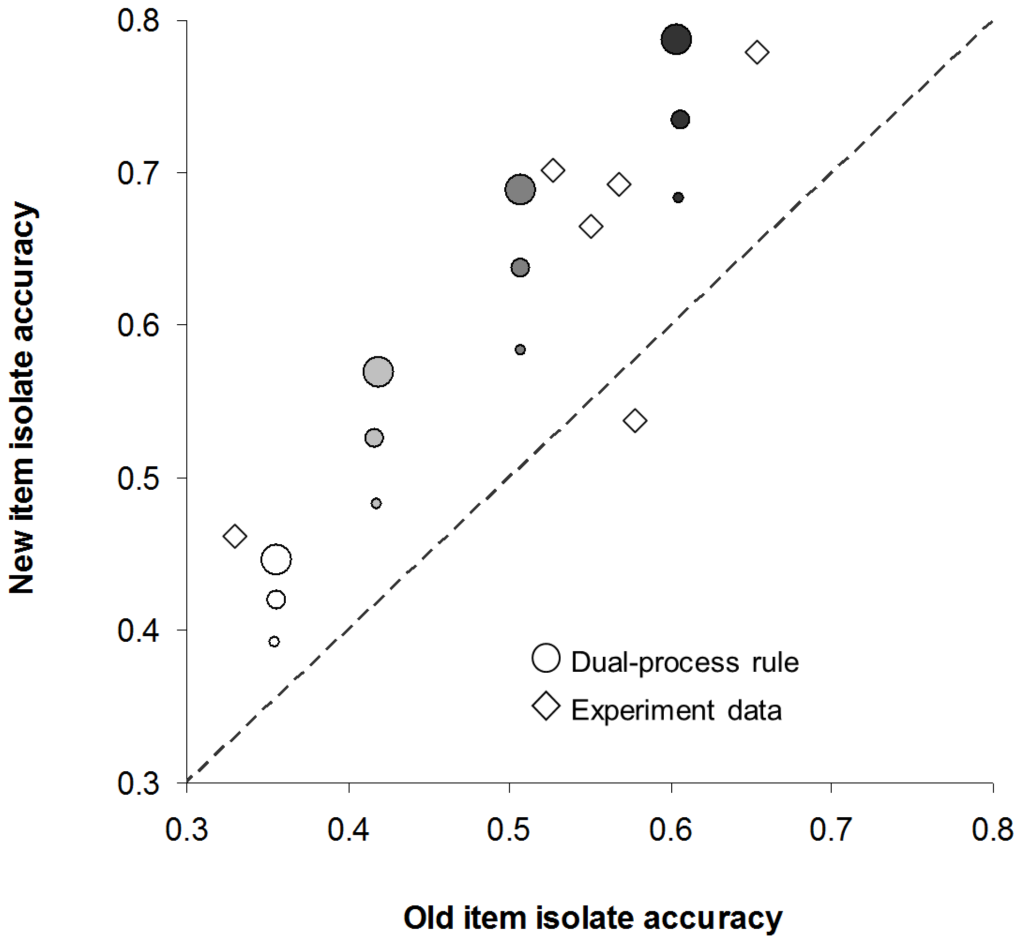


Figure 7. Simulated old and new item isolate oddity task accuracies for the dual-process decision rule. The dashed line represents the line along which new- and old item isolate accuracies are equal. Simulations with varying *d'*s (from 0.5 to 2.0 in steps of 0.5, represented by circles shaded from white to black) and recollection thresholds (*R*; from .3 to .7 in steps of 0.2, represented by circles with an increasing radius; the new and familiarity item distribution variances are fixed at 1.0) are shown.

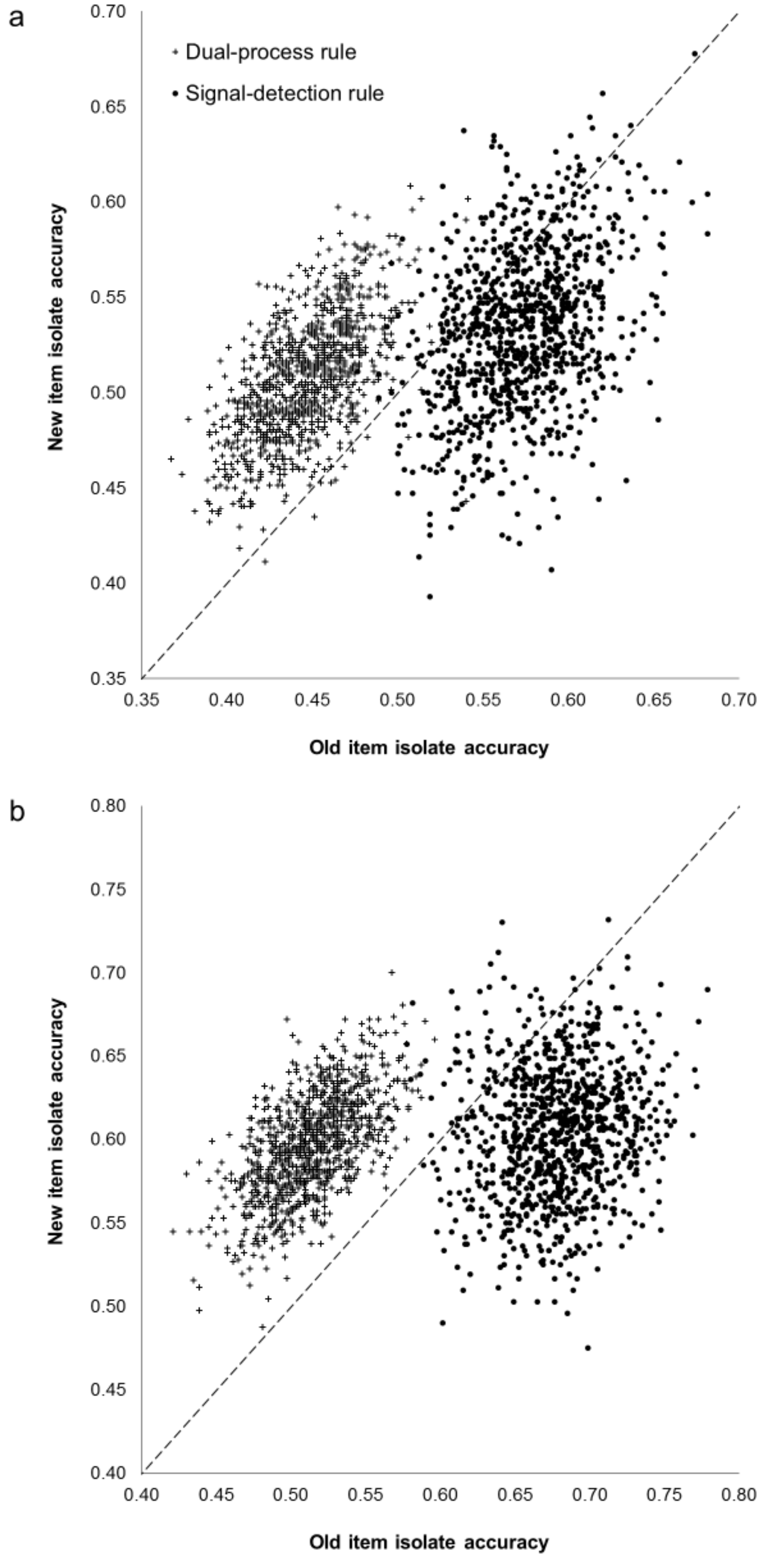


Figure 8. Bootstrap simulation outcomes for Bootstrap Monte Carlo A (panel a.) and Bootstrap Monte Carlo B (panel b.) for dual-process and independent-observation (with flexible criterion) decision rules. 1000 experiments each with 20 participants were simulated

using parameters estimated from single item recognition data. Each point represents the mean outcome from one simulated experiment.