Distinguishing between the success and precision of recollection

Iain M. Harlow & Andrew P. Yonelinas

Routledge
Taylor & Francis Group

Distinguishing between the success and precision of recollection

Iain M. Harlow and Andrew P. Yonelinas

Department of Psychology, University of California, Davis, Davis, CA, USA

Recollection reflects the retrieval of complex qualitative information about prior events. Recently, Harlow and Donaldson developed a method for separating the probability of recollection success from the precision of the mnemonic information retrieved. In the current study, we ask if these properties are separable on the basis of subjective reports—are participants aware of these two aspects of recollection and can they reliably report on them? Participants studied words paired with a location on a circle outline, and at test recalled the location for a given word as accurately as possible. Additionally, participants provided separate subjective ratings of recollection confidence and recollection precision. The results indicated that participants either recollected the target location with considerable (but variable) precision or retrieved no accurate location information at all. Importantly, recollection confidence reliably predicted whether locations were recollected, while precision ratings instead reflected the precision of the locations retrieved. The results demonstrate the experimental separability of recollection success and precision, and highlight the importance of disentangling these two different aspects of recollection when examining episodic memory.

Keywords: Recollection; Precision; Episodic memory.

Episodic recollection—the conscious retrieval of details from personal experience—plays a crucial role in people's lives, underpinning our identity and relationship with the world. Recollection provides rich and detailed information about the past, and can be distinguished from other cognitive processes that support recognition, such as familiarity (Yonelinas, 2002). Recollection can be measured using source memory tests, in which one estimates the ability to determine where or when an item was studied (Jacoby, 1991); remember/know methods, in which one examines the proportion of items that are accompanied by subjective experiences of conscious recollection (Gardiner, 1988; Tulving, 1985); and response confidence methods, in which one estimates the frequency of recollection on the basis of receiver operating characteristics (ROCs; Yonelinas, 1994). Such studies indicate that recollection is particularly vulnerable to distraction (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996), brain injury (Vann et al., 2009) and age-related cognitive decline (Light, Prull, La Voie, & Healy, 2000)—vulnerabilities which make recollection a particularly important focus of memory research.

All of these approaches, however, focus on measuring the frequency of recollection success and largely overlook the quality of the information recollected. Yet when recollection succeeds, it does not only indicate that an item was previously encountered: It also provides rich qualitative information about a prior event, and that information can be quite precise in some cases and very imprecise in others (e.g., "I parked in a space 40 feet north of the entrance" versus "I parked my car in the parking structure"). Traditional approaches to measuring recollection tend to interrogate memory in a binary way (e.g., "Was the item in the study list?" or "Was the item on the left or right side of the screen?"), and thus tell us relatively little about the precision of information retrieved.

Recently, Harlow and Donaldson (2013) developed a "positional response accuracy" paradigm that allows the probability of recollection success to be separated from the precision of information recollected. In that study, participants memorised unique word/location pairs, in which each location was a random position on a circle. At retrieval, participants were presented with each studied word and asked to recall, as precisely as possible, the location it was paired with (see Figure 1). For some trials, participants were very accurate at recollecting the study location, whereas for others, recollection of the location failed entirely and participants performed at chance. Examination of the response error distributions allows estimates to be drawn of both the proportion of items recollected and the mean precision of location information

retrieved on recollected trials (see Harlow & Donaldson, 2013; Figure 3).

The separation of overall recollection into two theoretical components—success and precision—was found to provide an excellent account of the data. A similar approach has also proven useful in studies of visual working memory (Bays, Catalao, & Husain, 2009; Zhang & Luck, 2009), though in that case to separate precision from working memory capacity. Along with statistical validity, however, an important test of a model is its theoretical validity. Do the mathematical estimates of recollection success and precision accurately capture these concepts? The current study extends this previous work by testing whether these two aspects of recollection are qualitatively separable in terms of subjective experience. Specifically, we required participants to judge how confident they were that they recollected each location, and to separately rate the precision of the information retrieved. If information about recollection success and precision is available to subjective experience, then participants should be able to report on these two aspects of recollection separately. Furthermore, if the estimates of success and precision derived using response error distributions are accurate, then participant judgements of success and precision should selectively track changes in the relevant parameter. Alternatively, if there is only one dimension to recollection—a more parsimonious account—then reports of precision and success should be directly coupled and index a single underlying dimension. For example, it may be the case that some recollections are simply weak, being both imprecise and difficult to recollect, whereas others are
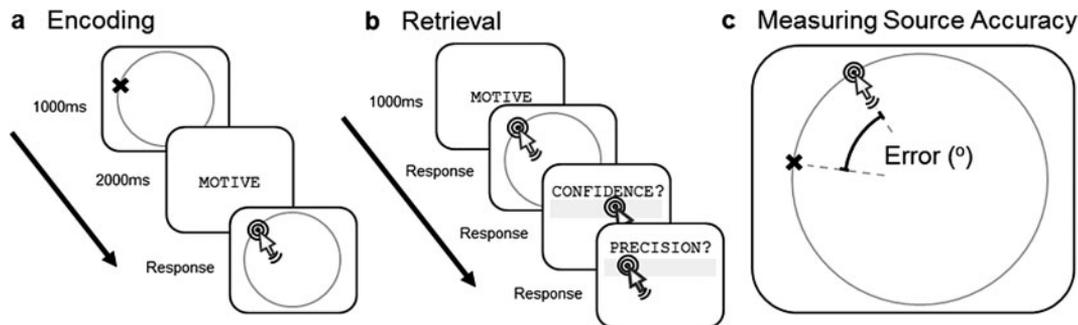
Figure 1. Source memory task. (a) At encoding, participants memorised unique word/location pairs, indicating the location after each trial to confirm attention and provide a baseline measure of response error. (b) At retrieval, participants indicated the recollected location for each studied word and made two separate ratings about their memory using a mouse: How confident they were that they had recollected the study location related to the word, and how precise their memory for this location was. (c) Source accuracy was measured by calculating the radial error between the correct and recollected locations. Participants studied 18 blocks, 12 trials per block.

stronger and have more precision. In that case, recollection confidence and precision should be functionally equivalent.

Subjective reports provide a means of testing the theoretical validity of the success/precision distinction, analogous to how the mapping between subjective remember/know judgements and ROC-derived estimates of recollection and familiarity corroborates the qualitative distinction between those processes (Yonelinas, Aly, Wang, & Koen, 2010). Subjective reports of precision would, however, provide additional advantages. By analysing response errors, precision can be estimated across groups of trials. Yet if participants could provide accurate trial-by-trial assessments of their memory precision—beyond simply reporting the presence or absence of recollection—this would allow the relationship between precision and other experimental variables (such as activity in a particular brain region) to be investigated.

The distinction between success and precision is also important because it has implications for existing studies. If the two aspects are not separable, the conclusions of most previous studies of recollection would not be challenged by the more detailed characterisation of recollection emerging from Harlow and Donaldson's (2013) study. If, however, recollection is underpinned by two separable elements or components, many prior studies become inherently ambiguous. For example, a manipulation or disease-affecting recollection might influence the probability of recollection, the precision of recollection or both. Two apparently similar deficits in recollection could reflect functionally and neurally distinct problems that would be overlooked by a unidimensional measure of recollection.

To determine whether success and precision are distinct and useful aspects of recollection, we used the positional response accuracy paradigm from Harlow and Donaldson (2013), but introduced two subjective response confidence decisions on each trial. Now, after indicating the location they believed was associated with each cue word, participants made a "recollection confidence" judgement indicating their certainty that they had recollected the location, followed by a "precision rating" judgement indicating the relative precision of their memory for the location. We then examined the correspondence between these subjective ratings of recollection success and precision with the objective measures derived from response errors. First, we expected that the success/precision distinction should be reflected in subjective experience: Participants should be able to judge the likelihood of successful recollection, and also (for trials on which recollection occurred) the relative precision of their response. Second, subjective ratings of success and precision should track the objective estimates of each: Recollection confidence should carry information about whether a trial was recollected, while precision ratings should instead distinguish between recollected trials of lower or higher precision.

METHODS

Thirty-one undergraduate students at the University of California, Davis participated in the 90-minute study and received course credits. To ensure that participants engaged consistently with the task, only participants making proactive responses (moving the mouse from its starting point in the centre of the scale) on more than 95% of their ratings were included for analysis ($N = 27$). These participants (16 female, mean age 19.7, range 18–31) all had normal or corrected-to-normal vision.

Each participant completed 18 blocks of 12 trials, for a total of 216 trials. The study phase of each block comprised 12 word/location pairs (Figure 1a). At the start of each trial, participants were presented with a black cross on a grey circle outline (600 ms), followed by a word in the centre of the screen (1500 ms). Participants verified attention by indicating the (now hidden) location using the mouse. Responses within a strict 20 pixels (around 6°, 75% of trials) of the target advanced participants to the next trial, otherwise the location was re-presented (600 ms) and the verification task repeated. Note that while 25% of trials fell outside this strict criterion, modelling showed fewer than 2% of the responses were information-free guesses (see Harlow & Donaldson, 2013 or Results section below for details about the Cauchy-plus-guessing model used). The median error was 5°.

At test (Figure 1b), participants were cued with each of the 12 studied words in a random order (1500 ms), and then selected a location on the circle outline, as close as possible to the original location paired with that word (self-paced). The location response on each test trial was compared to the target location, producing a
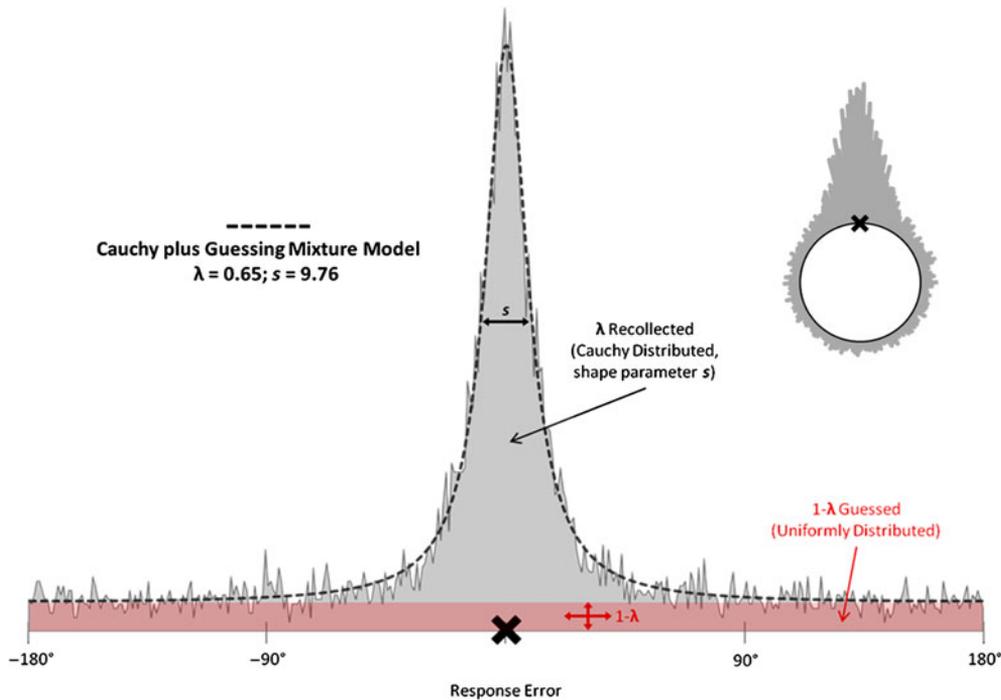
Figure 2. Distribution of location responses relative to the correct location (x). The response distribution (grey histogram) is fit to a mixture model of recollection, in which a given trial is either recollected (λ) and lands near the target (with Cauchy distributed error) or is not recollected (1–λ) and lands a random distance from the target (with Uniform distributed error). The Cauchy distribution is described by the shape parameter *s*, which reflects the precision of recollected trials. [To view this figure in colour, please visit the online version of this Journal.]

radial error between 0° and 180°, rounded up to the nearest whole degree ([Figure 1c](#)).

This was followed by two separate subjective ratings, both made by clicking along a near-continuous (600-pixel) horizontal scale. First, participants judged how confident they were that they had recollected the location associated with the target word (regardless of how precise they might be), prompted by the question: "How confident are you that you remembered the right location?" Next, participants rated how precisely they recalled the location: "How precise is your memory for the location?" Participants were encouraged to use the scale consistently throughout the experiment and to make use of the full scale width across trials, but to avoid the very edges. To minimise averaging artefacts due to differences in scale use across participants or rating types, and because the value of each rating is only informative relative to other trials, confidence ratings were converted to z-scores before analysis: Each subjective rating is expressed in standard deviations above or below the mean of that judgement type, for that participant.

Statistical tests on log-normally distributed variables such as the Cauchy distribution scale parameter *s* and standard deviation ratios were performed after log-transforming the raw statistics, to preserve normality. Likewise, we report geometric means for these statistics. Confidence intervals for correlations were calculated using Fisher's z-transformation.

RESULTS

Objective measures of location memory

[Figure 2](#) shows the distribution of radial errors for all 5832 trials (216 trials per participant). The figure demonstrates that a proportion of responses clustered tightly around the target location, indicating that participants often were able to report accurate location information, whereas the remainder were evenly distributed between 0 and 180° from the target, indicating that participants were otherwise guessing about the locations.

To quantify the results, the data were fit to a mixture model with two free parameters. A mixing parameter, λ, denotes the proportion of trials on which recollection succeeds.[1] Errors on these recollected trials follow a wrapped Cauchy distribution[2] with shape parameter s, denoting the spread of responses around the target such that higher values of s indicate a greater mean error (lower mean precision). The remaining $1 - \lambda$ non-recollected trials are guesses, randomly located relative to the target, resulting in a uniform distribution of errors. As with previous data obtained using this paradigm (Harlow & Donaldson, 2013), responses comprised a mixture of guesses and highly accurate recollected trials. G-tests showed that the Cauchy-plus-guessing model provided a good fit to both the aggregate and individual data [aggregate: $\chi^2(177) = 205.34$, $p = .071$; individual: $\chi^2(4779) = 4862.05$, $p = .197$]. In contrast, a continuous Cauchy-only model was strongly rejected [aggregate: $\chi^2(178) = 1171.88$, $p < .001$; individual: $\chi^2(4806) = 5502.15$, $p < .001$]; thus, a single parameter model was not able to account for the results.

The best-fitting parameters for the aggregate data ($\lambda = .65$; $s = 9.76$) were very close to the mean estimates obtained when the model was fit to participants individually (mean $\lambda = .65 \pm .04$; mean $s = 9.77 \pm 0.48$). On average, therefore, participants recollected 65% (λ) of the locations, and the median recollected trial was $9.77°$ (s) from the target. Thus, in line with previous work (Harlow & Donaldson, 2013), the results indicate that the accuracy of objective location memory can be well characterised as reflecting recollection success and recollection precision.
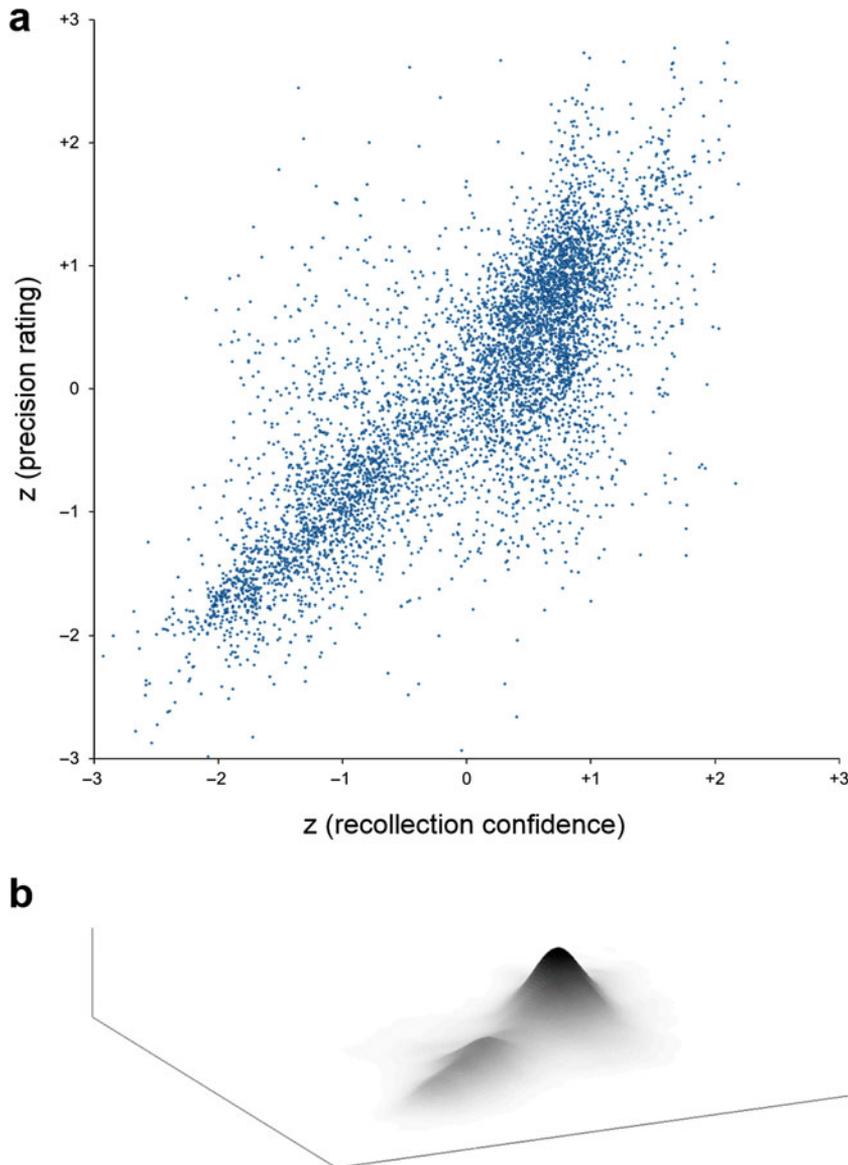
Subjective ratings of recollection success and precision

Figure 3a shows the subjective ratings of recollection success plotted against ratings of precision. Each point represents a single trial and its position in space reflects how confident the participant was that they had recollected the correct location ("recollection confidence", x-axis) and how precise they believed the recollected information to be ("precision rating", y-axis). Figure 3b shows a smoothed three dimension rendering of the same data; both figures reveal a strongly bimodal pattern of confidence ratings. Each individual's data were also clearly bimodal. Likelihood ratio tests performed separately for each participant showed that, in every case, a mixture of two bivariate Gaussian distributions (11 parameters[3]) drastically improved the fit compared to a single bivariate Gaussian distribution (mean $\chi^2(10) = 222.0$; all ps $< .001$). These bimodal distributions are consistent with both the analysis of the error data, showing a mix of accurate responses and guesses, and previous evidence that recollection is a thresholded process which either provides strong evidence about past experience, or fails completely (Harlow & Donaldson 2013; Parks & Yonelinas, 2009; Yonelinas & Parks, 2007). Under this interpretation, the bimodal distribution occurs because two different types of responses—guesses and recollected trials—are present in the data. Nevertheless, a bimodal distribution in confidence responses alone could potentially arise if participants were simply reluctant to use the middle of the response scale. It is

---

[1] We operationalized recollection in the current study as the ability to retrieve the arbitrary location associated with the studied word, which is based on the generally accepted notion that recollection reflects the retrieval of qualitative information associated with a prior episode.

[2] A Cauchy distribution (of which the t-distribution with one degree of freedom is a familiar example) differs from the Normal distribution by having a higher, narrower peak and heavier tails, as well as a more rapid transition between the two. The Cauchy distribution provides a significantly better fit than the Normal to both these data, and those from Harlow and Donaldson (2013).

[3] A single bivariate Gaussian distribution is described by 5 parameters: the mean and standard deviation in both x and y dimensions, plus the correlation between x and y. Two bivariate Gaussian distributions are therefore described by 11 parameters; the original 5 parameters for each distribution, plus a mixing parameter λ. Since the data were z-transformed before analysis, in practice the means (0,0) and standard deviations (1,1) are fixed for the single bivariate Gaussian distribution, which can therefore be completely described by the correlation parameter. The fixed overall means and standard deviations will also restrict the bivariate mixture more than the 11 parameters would suggest, so model comparisons used here should be considered conservative, i.e., they will tend to favour the simpler model.
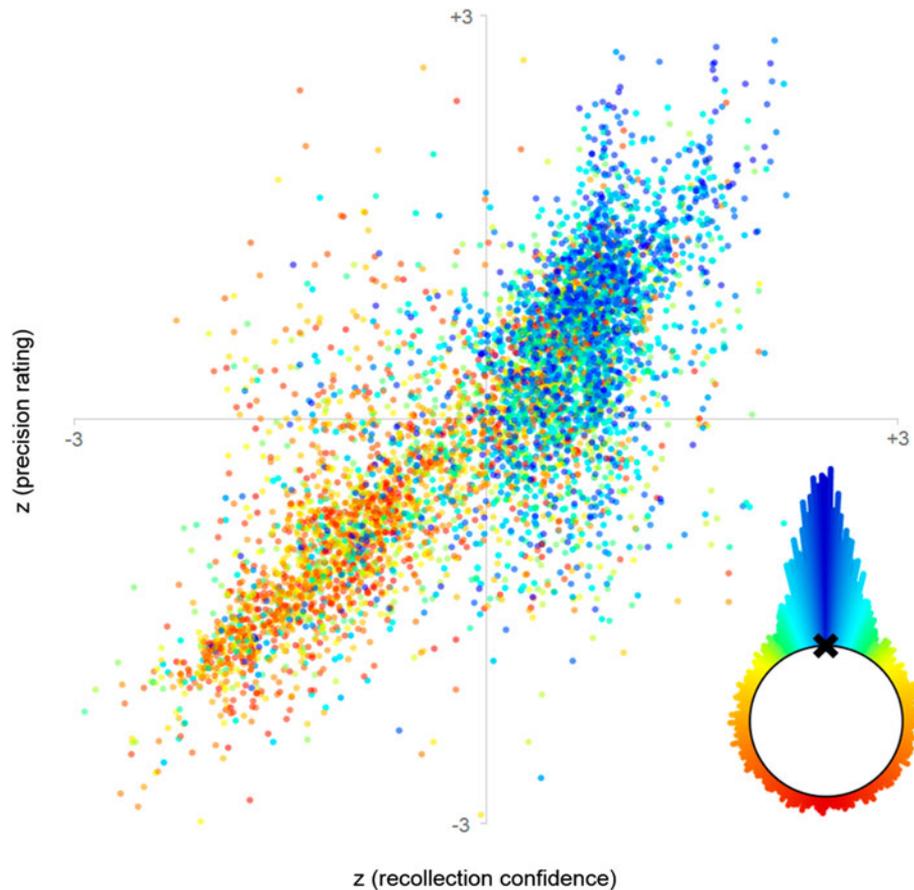
Figure 3. Distribution of subjective ratings of recollection confidence and precision. (a) Each trial is plotted according to the normalised recollection confidence (*x*-axis) and precision rating (*y*-axis) it received. The two responses are highly correlated for low-confidence trials, which comprise mainly guesses and therefore little or no precision information. Conversely, trials that are more likely to be recollected (top–right) are discriminated more (show greater variance) in terms of their precision ratings, which also become more independent of recollection confidence. (b) A smoothed density plot of the same data more clearly reveals the bimodal distribution of subjective responses.

therefore important to test whether this bimodal distribution is similarly reflected in response accuracy.

The relationship between objective and subjective measures of recollection

Figure 4 displays objective (radial error) data in conjunction with the subjective (recollection confidence and precision rating) measures of memory. Markers are positioned according to the two subjective judgements, as in Figure 3, and the colour of each marker reflects the radial error on that trial, i.e., the distance in degrees between the original studied location and that supplied by the participant at test. Specifically, the hue ranges from blue to red, and is linearly dependent on the trial's rank in terms of error.

Figure 4. Trial error as a function of recollection confidence and precision rating. Each trial is plotted in space according to its (subjective) recollection confidence and precision rating, and is coloured according to its (objective) accuracy. Colour ranges from entirely blue (error < 1°) to entirely red (180° > error > 179°), and hue changes linearly with the number of trials, meaning that as many trials are "blue" as are "red". The mapping between subjective ratings and objective accuracy becomes clearly visible: The guess and recollected distributions calculated from the subjective data (position) correspond strikingly with the probability of recollection implied by the objective error (colour). Furthermore, while subjective ratings for guessed trials show no relationship with error, precision rating does appear to be related to the error associated with recollection: Within the high-confidence distribution, the precision rating (*y*-value) of a trial tracks the shade of blue. [To view this figure in colour, please visit the online version of this Journal.]

Figure 4 demonstrates a clear and specific correspondence between the objective and subjective measures of recollection. First, items subjectively rated as being recollected (right side of figure) were associated with higher objective location accuracy (more blue) than those that were considered less likely to be recollected (left side of figure). Furthermore, this shift from non-recollected to recollected is not a linear transition; as seen both in Figure 4, where the proportion of precise trials (blue points) suddenly increases on the right side of the figure and more clearly in Figure 5a, where the average error rates for non-recollected items (left side of figure) are very high, but then

quickly transition to being very low (right side of figure) for the items that are reported as recollected. Thus, the bimodal pattern of guessed and recollected trials present in the confidence data is directly paralleled in objective measures of memory accuracy.

In addition, for recollected items, increases in precision ratings reflected increases in objective precision: On the right side of Figure 4 (but not the left), trials become darker blue moving towards the top of the figure. This relationship is laid out concretely in Figure 5b, which shows that the unique variance from the precision rating (i.e., after partialling out recollection confidence) linearly tracks radial error on recollected

trials—those given high recollection confidence by participants—but not on trials which were likely to be guesses.[4]

To summarise, recollection confidence appears to dichotomise trials into those that have highly accurate location information and those that have essentially no location information. In contrast, precision ratings are directly related to objective measures of location precision for the items that are recollected, but they are not related to objective precision for non-recollected items.

The results indicate that participants were able to separate recollection success from recollection precision. Nonetheless, Figures 3 and 4 show that there was an overall correlation between these two types of ratings. These ratings are positively correlated for four main reasons. First, the ratings were made immediately following each other by participants, so non-mnemonic influences (hand position, mood, etc.) were likely shared. Second, guesses are rated low for both judgements. Third, these low-confidence guesses show a strong correlation between probability and precision confidence (mean $r = .70$, SD = .25). Since participants did not feel they had retrieved the correct location at all on these trials, presumably meaningful information about precision was infrequent and so participants tended not to separate the two different properties. Thus, this might be viewed as a "baseline" correlation, for trials—guesses—on which there are no two meaningfully

separate dimensions of recollection. Finally, recollected trials (the higher confidence distribution) showed a correlation between precision confidence and recollection confidence (mean $r = .48$, SD = .25), but importantly, this correlation was significantly reduced compared to guess trials [$t(26) = 3.11$; $p = .005$], indicating greater independence between the two ratings. Furthermore, the ratio of standard deviations between precision rating and recollection confidence increased significantly for recollected trials compared to guesses [Guesses: mean $\alpha^2_{\text{prec}}/\alpha^2_{\text{rconf}} = 1.14$; recollected: mean $\alpha^2_{\text{prec}}/\alpha^2_{\text{rconf}} = 1.95$; $t(26) = 3.27$, $p = .003$]. That is to say, for recollected trials (relative to guesses) comparatively more of the variance in subjective judgements was driven by the precision rating. When participants felt they were guessing, their precision rating did not stray far from their probability rating, but when they felt they recollected any information about the location, they assigned the ratings more independently and differentiated between trials comparatively more on the basis of perceived precision.

To further verify that confidence and precision ratings captured genuine differences in performance, we examined model parameter estimates for items that were expected to be recollected. To maximise the number of recollected trials in the sample and minimise guesses, for which precision ratings are uninformative, we selected 65% (i.e., the estimated recollection rate) of trials: Those with the greatest likelihood of belonging to a participant's higher confidence bivariate Gaussian distribution in confidence space. Fitting the errors on these 3890 trials to the Cauchy-plus-guessing model confirmed that this set comprised mainly recollected trials ($\lambda = .84$). Next, a median split on the precision rating for these trials produced high- and low-precision groups comprising 1895 trials each, and the mean and standard deviation of recollection confidence were matched across the two groups by systematically removing as few trials as necessary (i.e., in descending order of confidence from the higher confidence group, and ascending order of confidence from the lower confidence group, along with a small number of outlying low-confidence trials from the high-confidence group to equalise standard deviation). This selection—on the basis of subjective ratings—produced groups of 1375 and 1349 trials, respectively, closely matched in

---

[4]Figure 5b contains some points that may have been outliers, particularly at the extreme left and right sides of the functions. Removing these points would not fundamentally change the conclusions because the same overall effects are observed for the middle points of those functions. Nonetheless, we examined this issue further by conducting a regression analysis across all the trials (i.e. without binning the data). This analysis suggested that a small linear trend ($R = 0.05$) could not be ruled out (marginal $p = .072$) for the lower confidence trials; by comparison the relationship within confidently recollected trials was relatively strong ($R = -0.30$) and highly significant ($p < .001$, more than 10 orders of magnitude smaller than the p for the equivalent relationship in guess trials). Of course, since the two groups are sorted on the basis of subjective ratings (i.e., imperfectly), even a relationship present only in recollected trials may be expected to manifest to a smaller degree in those assigned lower confidence ratings. Thus, while the results cannot conclude definitively the null result that there is absolutely no relationship between precision ratings and objective precision for low confidence ('guess') trials, they do demonstrate that any such relationship is at most trivial compared to the strong relationship observed for recollected trials.
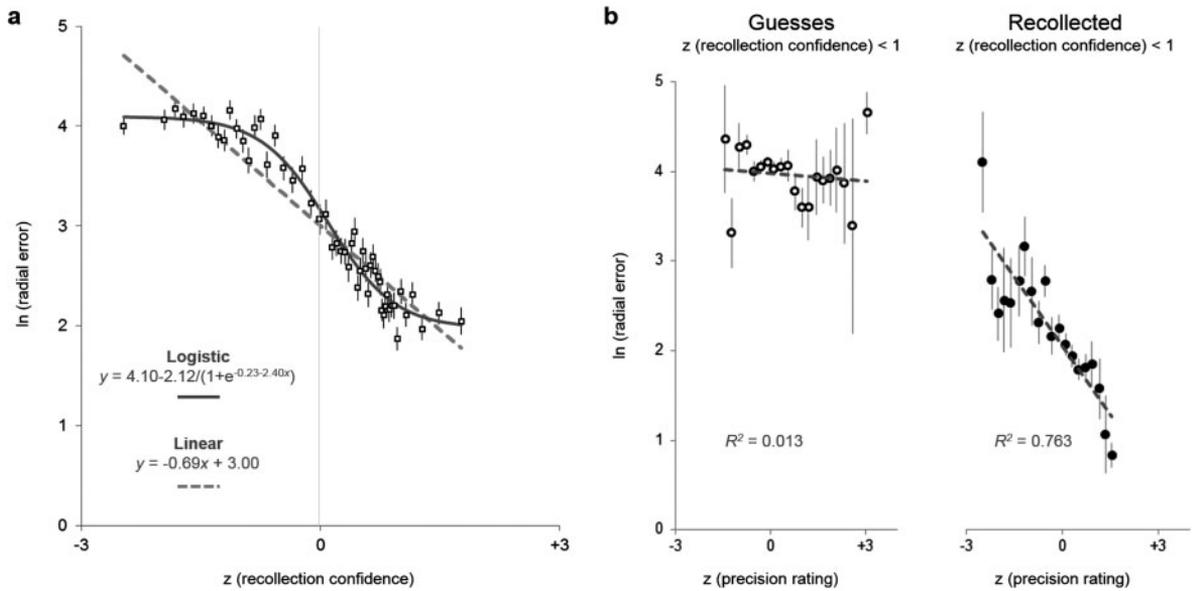
Figure 5. The relationship between accuracy and subjective ratings. (a) Mean radial error decreases sigmoidally (BIC = 193) not linearly (BIC = 378), as confidence in recollection increases. Recollection confidence divides trials into two distinct groups, guesses and recollected trials, but provides little information about the relative accuracy of trials within each group. (b) The relationship between precision ratings and log-transformed radial error, controlling for recollection confidence, depends on the type of trial. Guess trials (those with recollection confidence z-score <–1) show little relationship between precision rating and error because precision is ill-defined when memory is absent. In contrast, when trials were likely to have been recollected (recollection confidence z-score > 1) precision ratings closely track the relative accuracy of a trial. Trials were binned into 20 equally sized precision-rating ranges, each point shows the mean log-transformed error of all trials within that range.

recollection confidence [mean .62 vs. 0.62; SD = .31 vs. .31; $t(2722) = .13$, $p = .895$] but differed in precision ratings [non-overlapping distributions; mean .97 vs. .02; SD = .35 vs. .43; $t(2722) = 64.05$, $p < .001$]. Crucially, objective response errors followed the same pattern: According to the Cauchy-plus-guessing model, the high precision-rating group comprised trials that were more precise ($s = 7.78$ vs. 10.66) but no more frequently recollected ($\lambda = .91$ vs. .88). Hierarchical likelihood ratio tests confirmed that fixing s across both groups impaired the fit [$\chi^2(1) = 21.79$, $p < .001$] but fixing λ did not [$\chi^2(1) = 1.39$, $p = .238$].

A similar selection process produced groups of 1124 and 1405 trials, respectively, with differing recollection confidence [non-overlapping distributions; mean .90 vs. .30; SD = .23 vs. .33; $t(2527) = 52.66$, $p < .001$], but matched precision ratings [mean .46 vs. .46; SD = .42 vs. .42; $t(2527) = .16$, $p = .869$]. Consistent with participants' perception of their memory, response errors for these groups showed corresponding differences in recollection rate [$\lambda = .96$ vs. .78; $\chi^2(1) = 45.10$, $p < .001$] but not precision [$s = 9.34$ vs. 9.99; $\chi^2(1) = .81$, $p = .367$]. Together, these analyses demonstrate that the information underlying participants' precision

ratings selectively reflected the relative precision of recollected trials, while probability ratings conversely predicted recollection success.

Finally, we investigated in greater detail the information carried by each subjective judgement. Earlier, Figure 5b, we assessed the strength of relationship between precision rating and objective error for different types of trials (low confidence guesses and high confidence recollected trials). This elucidates the type of information captured by the precision rating: It distinguishes between recollected trials of varying accuracy, but carries little information about the error on guess trials. Here, we expand this technique by examining partial correlations across subsets of trials to more continuously examine the nature of each subjective judgement. In Figure 6 we plot partial correlations between each subjective judgement and (the log of) the objective response error for overlapping sets of 1250 trials with increasing overall confidence. Using this technique, it is possible to infer in greater detail the types of trials distinguished by recollection confidence and precision ratings, and therefore the information these subjective judgements are based on. A peak (or dip) in correlation magnitude indicates that
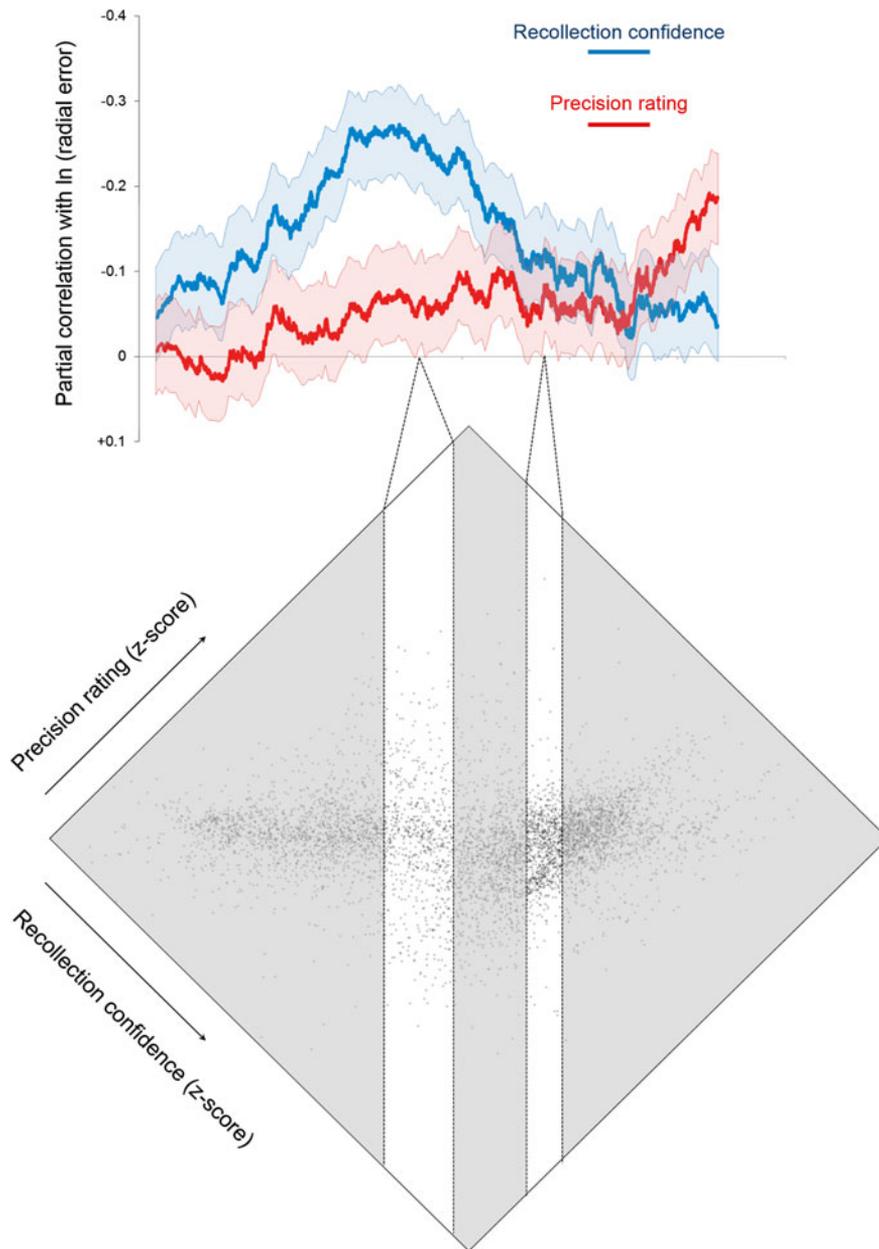
Figure 6. Relative accuracy information of recollection confidence and precision ratings. The partial correlation between each subjective rating (controlling for the other) and the log-transformed radial error on each trial can be used as a measure of how much information each judgement carries about trial accuracy, with higher magnitude negative correlations indicating a stronger relationship between the subjective rating and the objective error. We computed this correlation across overlapping sets of 1250 trials with increasing aggregate confidence (recollection confidence + precision rating). Correlations are plotted together with 95% confidence intervals at the mean aggregate confidence for the given set, and the distribution of recollected trials is included and scaled for reference. Two sample correlations demonstrate schematically how individual correlation values are drawn from sets of trials on the distribution. Recollection confidence and precision ratings provide qualitatively distinct information, consistent with the interpretation that they respectively distinguish trials based on likelihood of recollection, and precision. Recollection confidence is a strong predictor of objective error at moderate confidence, i.e., when there is a mix of guessed and recollected trials to be distinguished. As confidence increases, the proportion of guesses in the sample approaches zero and recollection confidence no longer discriminates between trials of differing error. In contrast, at this same point, precision ratings actually become *more* informative, since they instead reflect fine-grained differences in the precision of *recollected* trials. [To view this figure in colour, please visit the online version of this Journal.]

the subjective judgement concerned carries more (or less) information about the accuracy of trials within the corresponding confidence range. For example, the peak in correlation magnitude for recollection confidence around the centre of the plot shows that these ratings correlate strongest with error when overall confidence is near the mean—i.e., when roughly equal numbers of guesses and recollected trials exist to discriminate between. Conversely, recollection confidence correlates weakly with error when confidence is very low or very high. In those cases, the proportion of guesses approaches zero or one, so recollection confidence does not distinguish between trials within the sample. This pattern—like Figure 5a—suggests that recollection confidence is selectively based upon a binary assessment of whether a trial is recollected or guessed (and thus distinguishes between a guess and a successful recollection), but does not carry additional information about precision (and so does not distinguish between two recollected trials, or between two guesses).

Precision ratings display a very different pattern, reflecting their basis on qualitatively different aspects of the memory. Since this judgement discriminates between recollected trials of differing precision (but not between guesses and recollected trials), as the proportion of recollected trials in the sample increases, the precision rating becomes correspondingly more informative. This visualisation makes clear how the unique variance in the two judgements capture qualitatively separate dimensions of memory: The probability of recollection success and the precision of recalled information.

DISCUSSION

Here we examined memory on a fine-grained source task, allowing recollection success and precision to be quantitatively and qualitatively distinguished. In line with previous work (Harlow & Donaldson, 2013), we found that recollection cannot be described as reflecting a single dimension, but rather it reflects two properties that are functionally distinct. That is, performance on a word-location source test was fit well by a model that included a measure of recollection success and another measure of how precise that recollected location information was. In contrast, a single-dimensional model of recollection failed to account for the data. The current results further

indicated that these two properties of recollection are available to subjective experience and that participants can accurately judge the relative precision of their responses on a trial-by-trial basis. That is, subjective reports of recollection confidence (after partialling out precision ratings) tracked whether participants were able to objectively recollect location information compared to cases in which recollection failed. In contrast, subjective reports of recollection precision (after partialling out recollection confidence) directly tracked the objective measures of word-location precision for the recollected items. Although ratings of recollection success and precision were correlated with one another, they each account for unique variance in memory performance and they are functionally and phenomenologically distinct.

At the broadest level the results tells us that the overall confidence or strength of a recollection can vary at least partly independently of how precise that recollected information is. That is, the results show that two items can be associated with identical levels of recollection confidence yet one can be highly precise whereas the other is very imprecise. More strongly, while the raw ratings were correlated with each other as noted in detail above, once mutual information was partialled out the unique information present in each rating tracked only the aspect of recollection associated with it (e.g., the unique information in precision ratings predicted the precision, but not the success rate, of the associated response).

Most importantly, these two aspect of recollections can vary independently, and it is not possible to characterise recollection in the current study as falling on a single dimension of strength. In particular, two ratings based on differential response biases to high-accuracy and low-accuracy trials could not explain the results. The analyses showed that the precision ratings and the confidence ratings captured mutually exclusive variance in the memory judgements, leading to a double dissociation between recollection confidence and precision rating in terms of their relationships with recollection success and precision. When one type of rating was held constant (and the other allowed to vary), analysis of the objective error rates in each group showed that the associated aspect of recollection was also constant (and the other aspect varied). If both ratings were measuring the same underlying construct, this qualitative dissociation should not be present. Furthermore, when we examined the

unique variance in location accuracy accounted for by the confidence and precision ratings (Figure 6), we found the two measures produced very different profiles across overall confidence, which would again not be expected if they measured the same underlying memory construct.

Why is it important that recollection precision can be intuited? Most approaches to measuring recollection consider only whether it has succeeded. This is appropriate for many recognition memory tasks, such as item recognition, where successful recollection normally provides sufficient information to identify a stimulus as previously encountered. Undoubtedly, however, in many other tasks—as in real life—the quality of information recollected has greater impact. In a source task such as the location retrieval task used here, the precision of memory is crucial to the accuracy of a response. Outside the laboratory, a witness in a criminal case may recollect an event (and therefore be certain they witnessed it) but the detail, precision and accuracy of what they recollect may nevertheless be of crucial importance to the outcome of the case. For example, remembering seeing the suspect "fifty yards from the bank between 2 pm and 2.30 pm on the day of the crime" is very different from remembering seeing the suspect "somewhere in town on the day of the crime". Even if the witness's memory is equally strong or confident in these two cases, the precision of the memory is of the upmost importance.

The present study therefore fits into a broader trend within the memory literature which holds that qualitative measures of memory are as important an area of study as quantitative measures such as hit and false alarm rates (Koriat, Goldsmith, & Pansky, 2000). This view is compatible with the probabilistic model of recollection described above, but places additional emphasis on the memory experience and the quality of the information recollected, for example, to explain the experience of false recollection (Scimeca, McDonough, & Gallo, 2011). The angular location task used here makes it possible to investigate these important qualitative aspects of memory in finely grained quantitative detail.

Understanding what determines the precision of recollected material is also important for practical reasons. Preserving the ability to recollect in the face of ageing, trauma or disease is personally and societally beneficial, but a decrease in the quality of recollected information

also poses difficulties. Similarly, techniques for improving learning may be incomplete if these only focus on successfully triggering recollection, and not the quality of the information memorised and retrieved. In all these cases, assessing recollection uni-dimensionally (e.g., by using a single confidence scale, or a remember/know judgement) would conceal important changes or differences in the quality of information retrieved.

Recollection can vary meaningfully in precision, and this can be measured using a source task similar to that described above (Harlow & Donaldson, 2013). This approach yields good estimates of mean precision across responses, suitable for comparing the effect of two different conditions; measuring the average error rate alone, however, provides little trial-specific information. This is one reason why it is important that participants are able to identify the precision of their responses on a trial-by-trial basis. By collecting a participant's estimate of their relative precision on every trial, relationships between precision and other experimental variables can be measured. This allows researchers to investigate how precision of memory is affected not only by a pre-defined group-level manipulation, but other factors measured during the study, such as activity in particular brain regions or strategies at encoding.

This ability to provide specific and fine-grained subjective reports is important for a second reason: It suggests that participant introspection—commonly acquired in the form of confidence ratings, remember/know judgements or judgements of learning—remains a rich potential source of psychological data. Collecting unidimensional confidence ratings have revealed a number of interesting findings about the nature and neurobiology of cognitive processes that would have been invisible using only accuracy data (e.g., Aly, Ranganath, & Yonelinas, 2013; Onyper, Zhang, & Howard, 2010). More specific ratings, such as memory precision, have great potential to uncover further interesting findings.

The finding that memory confidence and precision can be measured and separated on the basis of subjective reports may have important implications for studies of working memory, as well as other psychological phenomena. As mentioned in the introduction, a method similar to the current one has been used in studies of visual working memory in which participants are presented with a small set of coloured squares, and then after a very brief delay are asked to recall

identify the colour of one of the studied items using a continuous colour wheel (Zhang & Luck, 2009). The approach has been used to measure how precisely the participants can reproduce the colour, and, separately, their working memory capacity. There is an extensive literature indicating that working memory and long term memory are functionally and neuroanatomical separable, and the theoretical interpretations of the parameters in these studies are quite different to our study of episodic recollection (e.g., the capacity of short-term visual memory versus the probability of recollecting information already stored in long-term memory). Nonetheless, they may share some common underlying processes (for further discussion of this possibility see Yonelinas, 2013), and the rigorous and quantitative approaches we introduce here to investigate subjective judgements are certainly applicable more generally to studies of psychological phenomena, including working memory.

These results also suggest that the exact question asked of participants may be crucial to extract specific factors underlying the perception of the strength of memory. In Harlow and Donaldson's (2013) study, participants were simply asked to rate their confidence in each response without explicit reference to either precision, or the probability of recollection engagement. This question, analogous to that asked in most memory studies, requires participants to summarise the separable factors of precision and recollection engagement into a single variable. As a result, participants apparently relied heavily on their sense of recollection engagement: Confidence correlated strongly with recollection rate across participants, but not with precision. The method used by participants to reduce memory to a single rating may vary across individuals and tasks, potentially obscuring differences. Asking precise, concrete and theoretically driven questions about the subjective experiences related to memory retrieval should yield more informative data, advancing our understanding of memory storage and retrieval. Furthermore, the information that participants based each judgement on can be carefully measured after a study, for example by using the approach detailed in Figure 6. Helpfully, this allows the consistency of instructions or interpretations to be measured across participants or studies, as well as providing a clear and detailed empirical basis for making inferences about the psychological processes being measured.

The current model is consistent with a number of previous models of recognition memory that assume that recollection can fail entirely (e.g., DeCarlo, 2003; Mickes, Johnson, & Wixted, 2010; Onyper et al., 2010; Yonelinas, 1994); though note that Harlow and Donaldson (2013) explicitly rules out the theory proposed by some of these models that non-recollected trials might be explained only by encoding or attention failure). The new model however, builds on earlier versions that measured recollection with a single parameter, by assuming that this single parameter can be unpacked further to reflect two different aspects of recollection (i.e., successful engagement of the recollection process itself, and the precision of the information retrieved from long-term memory). The current experiment shows that there are conditions in which a more complex view of recollection is necessary, but it is worth noting that in some circumstances, a simplified model of recollection may yet be appropriate to apply. For example, the benefits of assessing recollection precision in an item recognition task might be outweighed by over-fitting, since recollection precision should have little effect on the confidence that an item was previously encountered. When interpreting such results, however, it remains important to consider that successful recollection still varies in precision (even when not captured by the task or rating designed by the experimenter), and that this precision may be salient to participants.

Disentangling the success and precision of recollection is also important for understanding how memory declines with healthy ageing, disease or brain injury. For example, although it is clear that normal ageing is associated with decreases in recollection (e.g., Howard, Bessette-Symons, Zhang, & Hoyer, 2006; Light et al., 2000), future studies will need to be conducted to determine whether those deficits reflect a decreased likelihood of successful recollection or a decrease in the precision of that recollected information, distinct effects which could imply different neural causes.

Episodic recollection is a fundamentally important component of cognition, giving us access to the world outside the present moment. Recollection can vary both in terms of its probability of success, and the precision, richness and fidelity of the episodic information retrieved. These properties are not only separable on the basis of objective data, but also lead to distinct, salient and reportable memory experience on

individual trials. A greater understanding of recollection and its preservation can be achieved by separating these two aspects of recollection.

REFERENCES

Aly, M., Ranganath, C. R., & Yonelinas, A. P. (2013). Detecting changes in scenes: The hippocampus is critical for strength-based perception. *Neuron*, *78*, 1127–1137. doi:10.1016/j.neuron.2013.04.018

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 1–11. doi:10.1167/9.10.7

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159–180. doi:10.1037/0096-3445.125.2.159

DeCarlo, L. T. (2003). An application of signal detection theory with finite mixture distributions to source discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 767–778. doi:10.1037/0278-7393.29.5.767

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313. doi:10.3758/BF03197041

Harlow, I. M., & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin & Review*, *20*, 318–325. doi:10.3758/s13423-012-0340-9

Howard, M. W., Bessette-Symons, B., Zhang, Y., & Hoyer, W. J. (2006). Aging selectively impairs recollection in recognition memory for pictures: Evidence from modeling and receiver operating characteristic curves. *Psychology and Aging*, *21*(1), 96–106. doi:10.1037/0882-7974.21.1.96

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

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, *51*, 481–537.

Light, L. L., Prull, M. W., La Voie, D. J., & Healy, M. R. (2000). Dual process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 238–300). Oxford: Oxford University Press.

Mickes, L., Johnson, E. M., & Wixted, J. T. (2010). Continuous recollection vs. unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 843–863. doi:10.1037/a0019755

Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General*, *139*, 341–364. doi:10.1037/a0018926

Parks, C. M., & Yonelinas, A. P. (2009). Evidence for a memory threshold in second-choice recognition memory responses. *Proceedings of the National Academy of Sciences of the USA*, *106*, 11515–11519. doi:10.1073/pnas.0905505106

Scimeca, J. M., McDonough, I. M., & Gallo, D. A. (2011). Quality trumps quantity at reducing memory errors: Implications for retrieval monitoring and mirror effects. *Journal of Memory and Language*, *65*(4), 363–377.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, *26*(1), 1–12. doi:10.1037/h0080017

Vann, S. D., Tsivilis, D., Denby, C. E., Quamme, J. R., Yonelinas, A. P., Aggleton, J. P., … Mayes, A. R. (2009). Impaired recollection but spared familiarity in patients with extended hippocampal system damage revealed by 3 convergent methods. *Proceedings of the National Academy of Sciences of the USA*, *106*, 5442–5447. doi:10.1073/pnas.0812097106

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

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517. doi:10.1006/jmla.2002.2864

Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behavioral Brain Research*, *254*, 34–44.

Yonelinas, A. P., Aly, M., Wang, W.-C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*, 1178–1194.

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, *133*, 800–832. doi:10.1037/0033-2909.133.5.800

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*, 423–428. doi:10.1111/j.1467-9280.2009.02322.x