Psychon Bull Rev (2013) 20:318–325
DOI 10.3758/s13423-012-0340-9

BRIEF REPORT

Source accuracy data reveal the thresholded nature of human episodic memory

Iain M. Harlow · David I. Donaldson

Published online: 29 November 2012

Abstract Episodic recollection supports conscious retrieval of past events. It is unknown why recollected memories are often vivid, but at other times we struggle to remember. Such experiences might reflect a recollection threshold: Either the threshold is exceeded and information is retrieved, or recollection fails completely. Alternatively, retrieval failure could reflect weak memory: Recollection could behave as a continuous signal, always yielding some variable degree of information. Here we reconcile these views, using a novel source memory task that measures retrieval accuracy directly. We show that recollection is thresholded, such that retrieval sometimes simply fails. Our technique clarifies a fundamental property of memory and allows responses to be accurately measured, without recourse to subjective introspection. These findings raise new questions about how successful retrieval is determined and why it declines with age and disease.

Keywords Human memory · Episodic memory · Recollection · Familiarity · Signal detection theory

Electronic supplementary material The online version of this article (doi:10.3758/s13423-012-0340-9) contains supplementary material, which is available to authorized users.

I. M. Harlow
Neuroinformatics Doctoral Training Centre,
University of Edinburgh, Edinburgh, Scotland

D. I. Donaldson
Department of Psychology, School of Natural Sciences,
University of Stirling, Stirling, Scotland

I. M. Harlow (✉)
Department of Psychology, University of California, Davis,
One Shields Avenue,
Davis, CA 95616, USA
e-mail: imharlow@ucdavis.edu

Does recollection sometimes fail completely? This question has remained at the forefront of memory research for half a century. The common experience of forgetting where you left your keys fits well with threshold theories (Atkinson & Juola, 1974; Yonelinas, 1994) that state that retrieval can indeed fail. Alternatively, however, recollection may be a continuous signal (Green & Swets, 1966; Mickes, Wais, & Wixted, 2009), and retrieval "failure" may reflect weak or inaccurate, but not absent, recollection. Today, the fundamental nature of recollection remains as fiercely disputed as ever (Wixted, Mickes, & Squire, 2010; Yonelinas, Aly, Wang, & Koen, 2010).

Characterizing recollection correctly is necessary for it to be accurately dissociated from other memory processes. In particular, one highly influential dual-process theory distinguishes thresholded recollection from continuous familiarity (Yonelinas, 2002). This important functional dissociation underpins the widely used separation of recollection and familiarity using confidence ratings, yet it remains valid only if recollection really is thresholded (Wixted, 2007). If, instead, recollection were found to be continuous, many existing conclusions within memory research, from the specific decline of recollection in aging (Howard, Bessette-Symons, Zhang, & Hoyer, 2006) to the mapping between cognitive processes and neurobiological structures (Eichenbaum, Fortin, Sauvage, Robitsek, & Farovik, 2010; Peters, Thoma, Koch, Schwarz, & Daum, 2009), would need to be reinterpreted.

To date, the most widely cited evidence for a retrieval threshold comes from receiver-operating characteristic (ROC) curves (Yonelinas & Parks, 2007). Theoretically, the shape of the ROC provides information about the processes supporting memory. In practice, however, ROCs have been used to argue both for and against a recollection threshold (Wixted, 2007). One reason may be that ROCs, like old/new judgments, reflect subjective confidence rather than memory directly. The validity of this approach—especially the assumption that confidence and memory strength are directly and consistently related—is questionable (Bröder & Schütz, 2009; Malmberg, 2002; but see also Dube & Rotello, 2012). In short, unambiguous conclusions about the nature of recollection are difficult to draw from confidence ratings alone.

Here, we introduce an alternative measure of recollection: positional response accuracy. To assess accuracy, we used a source task that provides fine-grained assessments of retrieval errors (Fig. 1). At study, words were presented visually, each paired with a unique source location that participants would reproduce (Fig. 1a). At test, the studied words were re-presented, and participants recollected the associated source location as accurately as possible (Fig. 1b). Thus, instead of relying on (subjective) confidence to infer memory strength on each trial, we used (objective) error: the angle between the recollected and true source locations (Fig. 1c). To do this, we defined strong memories as those with high fidelity: On average, they should be recalled with smaller errors than weak memories. Note, of course, that the observed (circular) error distribution is not itself a (linear) evidence distribution: For example, a very small error on a trial would not require that the memory strength be high, since even a pure guess could result in an accurate response, through chance alone. We therefore recovered the strength or precision of recollection by examining the distribution of errors across a number of trials (see the supplementary online materials for more details).

Continuous and threshold models of recollection make conflicting predictions about the distribution of response errors that we should expect (see Fig. 2a below). Specifically, if a recollection threshold operates, recollection must fail on a subset of trials, and these trials would necessarily be distributed randomly relative to the target locations. Successfully recollected trials would cluster closely around the target location, and the final distribution would be a mix of these two types of responses. Alternatively, if recollection is not subject to a threshold, a single distribution of responses around the target should be observed, since trials with moderate accuracy would be possible.

Although response error data are inherently informative about the underlying basis of memory, it is important to note that a mixture of guesses and accurate responses need not necessarily imply a threshold at retrieval. Such a mixture could instead be introduced at study, for example by attending to or encoding only some subset of the to-be-remembered items (DeCarlo, 2003; Mickes, Johnson, & Wixted, 2010). Here, therefore, we directly tested the plausibility of a retrieval account by varying the delay between encoding and retrieval. If encoding failure were the real reason for the appearance of a thresholded pattern, and retrieval is actually continuous, the proportions of trials on which recollection failed should be the same at both short and long study–test delays; that is, the proportion should be exactly the proportion of trials for which encoding did not succeed. As we report below, by examining response errors in the context of a novel source memory task, we were able to demonstrate that recollection really is associated with a threshold.

Method

Seventeen University of Stirling students (11 female, six male; mean age 19.1 years, range 17–24) gave informed consent (approved by the University of Stirling Psychology Ethics Board) and received course credit or £5/hour for participation. The study blocks comprised nine word/location pairs (Fig. 1a). The participants pressed a button to begin each trial and were presented with a black cross located on a gray circle outline (600 ms), followed by a word (1,500 ms). Participants verified their attention by indicating the (now hidden) location using the mouse. Responses within 20 pixels (around 6º; 75 % of trials) from the target advanced participants to the next trial; otherwise, the location was re-presented (250 ms) and the verification task repeated. The first study response was analyzed for each trial; any additional responses for repeated presentations were excluded before the final analysis.

At test (Fig. 1b), participants were cued with each studied word (1,500 ms) and indicated the recalled location, followed by a confidence judgment made by clicking along a near-continuous (600-pixel) scale. There was no response

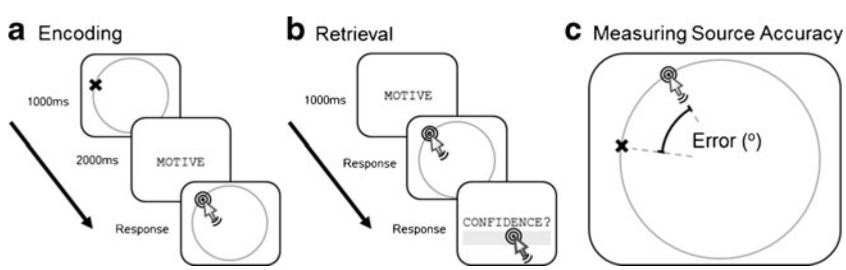

Fig. 1 Source memory task. **a** At encoding, the participants memorized unique word/location pairs, indicating the location after each trial so as to confirm attention and provide a baseline measure of response error. **b** At retrieval, participants indicated the recollected location for each studied word and rated their confidence, using a mouse. **c** Source accuracy was measured by calculating the arc length (in degrees) between the correct and recollected locations. Participants studied 36 blocks, with nine trials per block, in two intermixed conditions: 18 of the blocks were tested 10 s after encoding (short delay), and 18 were tested two blocks later (long delay)

time limit, and participants could change their selections freely, confirming their final choice with a buttonpress. The participants were tested after a short delay (10 s after the study phase, or approximately 2 min after studying the location) or a long delay (two blocks, or approximately 7 min, later), completing 162 trials in each condition. Participants did not know when they would be tested until after completing each study phase, making the encoding phases identical for both conditions.

Test trials separated from the corresponding study trial by fewer than three other trials (i.e., when the study trial occurred very near the end of the study block, and the test trial near the start of the following test block) were removed before analysis, to ensure that all trials relied on episodic memory. Source ROCs were constructed by defining correct responses as those falling within 90º of the target location—that is, on the same half of the circle, analogous to selecting the correct one of two choices in a standard source task. The errors and ROC curves were fit separately for each participant using maximum likelihood estimation, to a mixture of λ recollected trials and $(1 - \lambda)$ guesses. Details about the error distribution modeling, ROC construction, and choice of stimuli are included in the supplementary materials.

Results

Figure 2 shows the predicted and observed mean distributions of response errors across all participants, relative to the target source locations. To control for encoding, participants were required to indicate each location during the study phase (Fig. 1a). These study data (Fig. 2b) indicate that participants were very accurate at reproducing the target location within a few seconds of seeing it: 93 % of the trials were accurate within 10º. At study, errors to the attended stimuli were best described by a wrapped[1] normal distribution, consistent with findings from working memory (Zhang & Luck, 2009). By contrast, at test (Fig. 2c and d), errors to successfully recollected trials followed a wrapped Cauchy distribution, which contains greater proportions of very accurate and very inaccurate trials than does the normal distribution, and fewer trials of moderate accuracy. The qualitative difference between the distributions seen at study and test is consistent with the widely held view that working and episodic memory reflect the operation of different underlying memory systems (Atkinson & Shiffrin, 1968); our source retrieval task relies on episodic memory.

We next investigated whether responses made at test exhibited a thresholded or continuous pattern, by fitting each participant's error distributions to a thresholded model of recollection, with two free parameters. A threshold parameter, λ, denoted the proportion of trials on which recollection succeeded. Errors on recollected trials followed a Cauchy distribution with the shape parameter s, the spread of responses around the target, such that higher values of s indicate a greater mean error. The remaining $1 - \lambda$ nonrecollected trials were guesses, randomly distributed around the circle relative to the target, resulting in a uniform distribution of errors (Fig. 2a).

When $\lambda = 1$, all responses are based on some, variable, recollection: No threshold is present, rendering the model continuous (Fig. 2a). Alternatively, if $\lambda < 1$, recollection fails for a subset of responses: Therefore, a threshold exists. The presence of a threshold was assessed using a likelihood ratio test. Allowing λ to vary below 1 significantly improved the likelihood of the observed data, as compared to fixing $\lambda = 1$ [mean $\lambda_{\text{short}} = .698$, $\chi^2(17) = 376.68$, $p < .001$; mean $\lambda_{\text{long}} = .593$, $\chi^2(17) = 292.61$, $p < .001$].

We also tested goodness of fit using the G statistic, confirming that the thresholded model fit the aggregate data well [$\chi^2(354) = 328.46$, $p = .831$] but that the continuous model did not [$\chi^2(354) = 1{,}116.78$, $p < .001$]. The adjusted fit residuals for each model (Student's t statistic; insets in Fig. 2c and d) make clear the reason: The thresholded model fits consistently well across the entire range of errors, but the continuous model systematically underestimates the numbers of highly accurate and highly inaccurate responses. In essence, the continuous model fits poorly because the data are more thresholded than a continuous account would predict: Recollection is highly accurate most of the time, but fails completely on over a quarter of trials.

Although the data are consistent with recollection being thresholded, recollection "failure" may occur simply because contextual source information has not been successfully encoded (DeCarlo, 2003). Under the encoding failure model, some items are completely unattended at encoding, and hence no usable information about their encounter can be later retrieved. If this account is correct, some proportions of both the long- and short-block items would not be encoded, resulting in guessing at later retrieval for these items. Critically, because the short and long encoding blocks were not predictable, these proportions must be equivalent across the two conditions. Separate to encoding failure, it is also possible that variable accuracy for recollected trials might reflect imprecise encoding at study, not variable retrieval. This possibility could be tested using the same logic as for the encoding failure account, since it implies that the precision parameter would be determined at encoding.

To eliminate encoding failure or imprecise encoding as confounds, we compared responses that had identical encodings but different retrieval delays. Specifically, we examined how the threshold and precision parameters

[1] In circular statistics, distributions are *wrapped*, meaning that the tails continue indefinitely around the circle. Thus, at angle x, the probability density is given by the sum of probabilities for x, $x \pm 2\pi$, $x \pm 4\pi$, etc.

responded to an increase in retrieval delay. If either the threshold or the variability that we observed in the data was simply a product of encoding, λ or s, respectively, should be matched across short and long delays. In fact, after a longer delay, responses were significantly less precise [s_{short} = 9.68, s_{long} = 13.93; $t(17)$ = 2.83, p = .012] and less frequently correct [λ_{short} = .69, λ_{long} = .59; $t(17)$ = 4.02, p < .001], while fixing either s [$\chi^2(17)$ = 72.33, p < .001] or λ [$\chi^2(17)$ = 38.71, p = .002] across conditions significantly reduced the likelihood of the data. This violates the encoding failure account, since the probability of retrieval was sensitive to the postencoding manipulations; the change in the recollection rates, λ, across study–test delays rules out encoding failure as a sufficient explanation for the threshold that we observed. Instead, the pattern was consistent with a retrieval threshold model in which forgetting increases the proportion of items falling below a psychological threshold as retention intervals increase. Similarly, the change in precision, s, over delays was incompatible with an all-or-none model of recollection. Recollection was graded but subject to a threshold, independent of encoding.

We also ruled out one other, subtle way in which a continuous model might predict a thresholded pattern. Perhaps “subthreshold” trials were subject to source confusion (i.e., they reflected retrieval of another location from the study

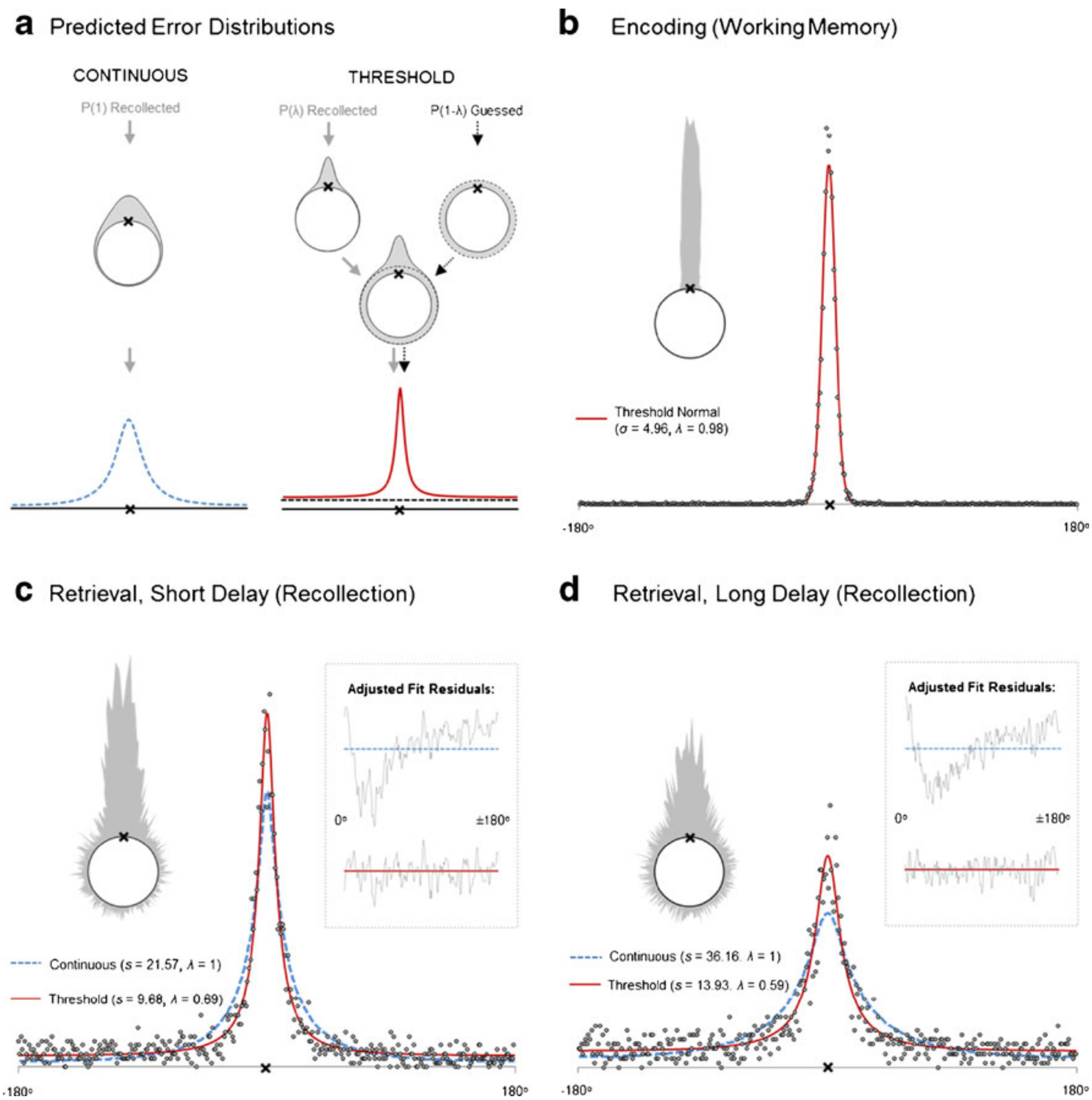

Fig. 2 Predicted and observed error distributions. **a** Continuous recollection predicts a continuous distribution, strictly decreasing away from the target (x). Thresholded recollection predicts a mixture of successful recollection (λ) and subthreshold guesses that uniformly raise the distribution above zero. **b** At encoding, responses were highly accurate, with 98 % of the trials attended. At retrieval, both immediate (**c**) and delayed (**d**) responses clustered around the target but exhibited significant subthreshold guessing (recollection failure). Increasing study–test delays caused recollection to be less precise, but also to fail more frequently. The adjusted fit residuals show that recollection was thresholded (*solid red*), not continuous (*dashed blue*)

block). We tested whether this was the case by examining the trials on which the correct location had not been recollected: that is, responses more than 90º away from the target. Did they fall closer to other locations from that study block than would be expected by chance? Errors were redefined as the distance between each response and the closest studied location from the same block. These were then fit to the same model of recollection described above, with the exception that the guess rate was replaced with the distribution of errors expected by chance (enumerated exactly for each study block). Likelihood ratio tests confirmed that allowing s and λ to vary did not improve the fit, as compared to the guess distribution alone [$\chi^2(2) = 1.04$, $p = .594$]; that is, the distribution was best fit by pure guessing. This result is more clearly visualized in Fig. 3, which shows that incorrect trials were no closer to other studied locations than would be expected by chance, and that a continuous model cannot, therefore, account for the threshold.

It is important to recognize that the threshold model that we found here was not all or none: Errors could be introduced through both memory failure and imprecision. How important were each of these effects? We tested this by calculating the proportion of each participant's mean errors accounted for by memory failure, using each participant's estimate of λ and assuming a mean error of 90º on guessed trials. We removed nonmnemonic response (e.g., motor) errors, operationalized as the mean error observed during the study task, which was 5.4º. The remaining errors were assumed to reflect a loss of precision for graded recollection. For trials tested after a short delay, the mean error was 42.1º, with 27.9º of this number resulting from subthreshold guessing, and only 8.8º through memory-related loss of precision.

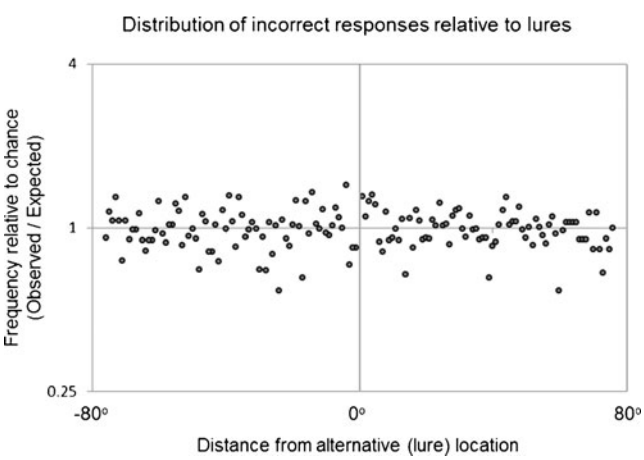

Fig. 3 Normalized distribution of guess responses, relative to alternative locations from the same study set. Each clearly incorrect response (>90º error) was compared to each of the other eight locations from the same study set. If participants incorrectly recalled a different studied location (instead of guessing), these trials should cluster around the remaining studied locations, leading to a peak around 0º. No such peak is apparent: The inaccurate responses fall no closer to alternative studied locations than would be expected by chance

Thus, the proportion of trials on which recollection occurred was more critical to overall performance than was the precision with which the locations were recalled [$t(16) = 5.09$, $p < .001$]. Similarly, for trials tested after a longer delay, the mean error was 53.0º, with 36.6º of this resulting from subthreshold guessing, and only 11.0º through memory-related loss of precision [$t(16) = 6.32$, $p < .001$]. Viewed another way, λ explains virtually all of the variance in overall performance (mean errors): These two variables are strongly correlated after both short and long delays [$r(15) = -.97$, $p < .001$, and $r(15) = -.84$, $p < .001$, respectively], whereas memory precision was not correlated with overall performance [$r(15) = -.00$, $p = .997$, and $r(15) = -.33$, $p = .196$]. Performance on even this task, which was highly sensitive to the precision of recollection, was primarily determined by the rate—and not the strength—of recollection.

Finally, to demonstrate the practical effects of using confidence data to examine memory, we also constructed symmetric source ROC curves and tested whether a retrieval threshold was still visible. These results are presented in the supplementary online materials; the shapes of these ROCs fit a mixture of high strength recollection and information-free guessing, and are therefore consistent with a graded, thresholded model of recollection (and not with a continuous model). Significantly, however, the error data revealed a significant decrease in recollection rates as test delay increased, ruling out pure encoding failure and confirming that recollection was thresholded. By contrast, when using the confidence data, this crucial difference was only visible in ROCs with a far higher resolution than are generally used (in this case, 20 points: $\lambda_{short} = .82$, $\lambda_{long} = .67$ [$t(17) = 2.40$, $p = .029$]). When confidence ratings were binned to form six- or ten-point source ROCs, the resolutions employed in the overwhelming majority of previous ROC studies (e.g., Howard et al. 2006; Slotnick, 2010; Yonelinas, 1994), there was insufficient power to detect a change in recollection rates (Fig. 4). Of course, nothing is particularly special about the number 20: The resolution required to detect a threshold for a given paradigm will depend on many factors, including the type of judgment being made. However, if a resolution on the order of 6–10 points is even sometimes insufficiently sensitive, as our results suggest, that may explain why previous attempts to characterize recollection using confidence data have provided conflicting evidence for a threshold. In short, direct comparison suggests that memory is characterized better using objective measures of memory strength, such as accuracy, than by subjective measures, such as confidence.

Discussion

The ability to recollect individual events is fundamental to episodic memory, yet the nature of recollection has long

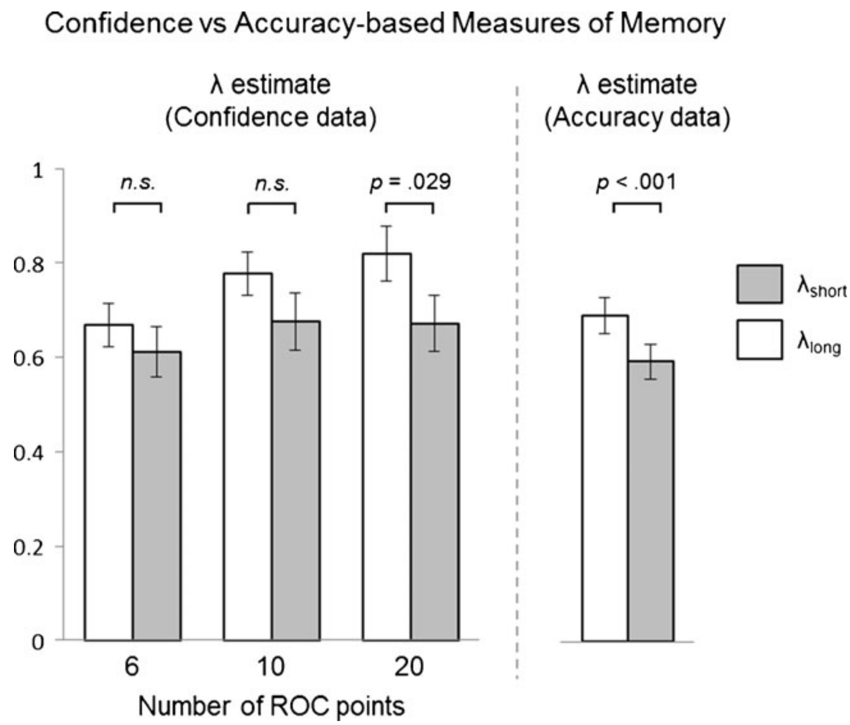

Fig. 4 Estimates of the critical threshold parameter λ. Confidence provides accurate parameter estimates when the correct (variable and thresholded) model is used, but it is underpowered for characterizing recollection itself. A reduction in λ with study–test delay is evidence of a recollection threshold, but this difference is noisy in the confidence data and is statistically invisible in ROC curves with commonly used resolutions (6–10 points). In contrast, the source accuracy task eliminates the need to model the relationship between memory strength and confidence for each participant, resulting in more stable parameter estimates. Using accuracy, but not confidence, recollection can be clearly characterized as thresholded

been disputed. Here, by using a direct test of source accuracy, we were able to clearly characterize recollection: It exhibits a threshold and sometimes fails, even when the information being sought was successfully encoded. In addition, although recollection is not a continuous process, it is graded (i.e., variable) when it succeeds. Quantifying the accuracy of the information retrieved from memory, as we have done here, provides a useful and less noisy alternative to the common approach adopted when studying memory, which is to examine metacognitive judgments of previous occurrence made by participants. It is important to realize that any such judgments, including confidence ratings, necessarily introduce noise, since they depend on nonmnemonic factors such as mood, task instructions, previous responses, and the maintenance of decision criteria. This certainly does not mean that confidence ratings cannot tell us anything about memory—indeed, when analyzed at sufficiently high resolution, the confidence data corroborate the conclusions from the accuracy data—but the effects of metacognitive noise must nonetheless be carefully factored in. This is especially important in studies that aim to characterize recollection, because such noise alone can produce curvilinear ROCs even in the absence of graded memory strength, and can make the underlying threshold more difficult to see as a result.

The threshold that we observed in source memory has also been reported in associative recognition (Parks & Yonelinas, 2009); significantly, Parks and Yonelinas's study also eschewed confidence ratings in favor of a second-choice procedure. The results presented here allow us to go farther than previous studies, however. Crucially, the present findings demonstrate that the threshold cannot be accounted for by attention or encoding effects (e.g., DeCarlo, 2003; Mickes et al., 2010), and instead they support the competing view that recollection is a threshold process (e.g., Malmberg, 2008; Parks, Murray, Elfman, & Yonelinas, 2011; Sherman, Arri, Hasselmo, Stern, & Howard, 2003). To be absolutely clear, it is possible that some trials were not encoded (although participants did attend to over 98 % of the study trials), but this cannot be the only explanation for the threshold in our data. The attention-driven mixture signal detection model suggested by DeCarlo can mathematically fit the data for a single, fixed delay, but it is incompatible with our finding that test delay increased the proportion of guessed trials. Thus, the present study suggests that the threshold is, instead, a real property of episodic recollection.

Accurately describing how recollection operates is of considerable practical importance because the interpretations of many behavioral and neurophysiological memory studies ultimately rely on how memory processes are modeled. The fact that recollection has a threshold allows it to be objectively measured and separated from other processes, supporting an extensive body of research that continues to illuminate the function (Elfman, Parks, & Yonelinas, 2008; Greve, van Rossum, & Donaldson, 2007), neurobiology (Eichenbaum et al., 2010; Peters et al., 2009), and decline (Howard et al., 2006) of episodic memory. Although we have ruled out a continuous account of recollection, it is important to highlight that our findings are in agreement with others showing that when recollection does occur, it yields information of varying quality (Mickes et al., 2009;

Slotnick, 2010). Drawing process estimates from models that treat recollection as being thresholded but not variable, as we and others have previously done (Eichenbaum et al., 2010; Harlow, MacKenzie, & Donaldson, 2010; Yonelinas, 1994), will in many cases underestimate the contribution of recollection. To be accurate, future studies must address the variable nature of recollection explicitly by characterizing it as a graded, thresholded process.

The present findings rule out encoding as a sufficient explanation either for the threshold or for imprecision in recollection. Why, then, are memories less frequently recollected after a longer delay? One possibility is that interference during the retention interval reduces the distinctiveness of the stored information, impairing the ability of the network to successfully pattern complete an episodic memory from a given cue. By this view, the precision and probability of recollection should move in parallel: Precision should be a proxy measure for the integrity of the stored information, which in turn determines how accessible the information is at the time of retrieval. Alternatively, the success of recollection might depend on the cue–target binding, but not on the precision of the target itself. In this case, the precision and probability of recollection could theoretically move in opposite directions: Manipulations that specifically interfere with the cue–target relationship should reduce the frequency of recollection, while manipulations that interfere only with the target should separately reduce the precision.

Another important question about the recollection threshold is exactly why it arises. Do memory traces "burn out" in a probabilistic way as new memories are stored? Or, are the computational mechanisms underlying recollection-based retrieval intrinsically thresholded? One way in which these questions can be addressed is by bridging the predictions of neural network models with the empirical data. The characterization of recollection proposed here is in striking agreement with the bimodal distribution of recollection strength observed in the complementary-learning-systems model of hippocampus and medial temporal cortex (Elfman et al., 2008; Norman & O'Reilly, 2003). This model suggests that the thresholded nature of recollection might reflect particular properties of the hippocampal network (Norman & O'Reilly, 2003) or the distinctiveness of representations within the hippocampus (Elfman et al., 2008). Intriguingly, however, computational modeling has also demonstrated that the presence of a threshold may depend on the way in which a retrieval network is interrogated, such that the same memory representation can give rise to either a continuous or a thresholded signal (Greve, Donaldson, & van Rossum, 2010). According to the latter view, the mechanism underlying recollection is itself inherently thresholded, not the representations upon which it operates. More broadly, retrieval processing may operate within the context of support processes that can themselves influence whether recollection succeeds (Rugg & Wilding, 2000). For example, electrophysiological evidence has shown that the adoption of an appropriate retrieval orientation is predictive of recollection success (Herron & Rugg, 2003). One important implication of this view is that even when recollection fails, the memory being sought may nonetheless be retrievable using different cues, or by an alternative process such as familiarity.

If the mechanisms underlying recollection are thresholded, this may have wider cognitive implications. Recollection or a similar set of processes may retrieve not only episodic, but also semantic, information: Does the failure to bring a particular word to mind reflect very different cognitive processing than does the failure to recollect an episode? More widely, other cognitive processes may be subject to a threshold—for example, perception (Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009) and working memory (Zhang & Luck, 2009). A threshold might arise as a general consequence of information processing in the brain, which is simply most noticeable when memory fails.

Accurately modeling recollection is important for practical reasons, as we have noted above, but it is also important for a theoretical understanding of how memory declines with age or disease. This is because the deficits shown to specifically affect recollection (Howard et al., 2006) could reflect either the reduced probability or precision of retrieval, and these two deficits could reflect distinct patterns of damage. Our results suggest that the rate of recollection—that is, the thresholded aspect—is the most critical aspect of retrieval. Further research will be required, however, to determine whether this is true in general or is task-specific, and in particular whether age-related memory deficits are caused by disruption of the memory trace (reducing precision) or of the mechanism retrieving it (reducing recollection rate).

Author note I.M.H. is now at the Department of Psychology, University of California, Davis, and was supported by the EPSRC/MRC-funded Doctoral Training Centre in Neuroinformatics at the University of Edinburgh. D.I.D. is a member of the SINAPSE collaboration (www.sinapse.ac.uk), a pooling initiative funded by the Scottish Funding Council and the Chief Scientific Office of the Scottish Executive.

References

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 1. Learning, memory, and thinking* (pp. 243–293). San Francisco: Freeman.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed systems and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory. Vol. 2* (pp. 89–195). New York: Academic Press.

Bröder, A., & Schütz, J. (2009). Recognition ROCs are curvilinear—or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 587–606. doi:10.1037/a0015279

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.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain, 132,* 2531–2540.

Dube, C., & Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 130–151. doi:10.1037/a0024957

Eichenbaum, H., Fortin, N., Sauvage, M., Robitsek, R. J., & Farovik, A. (2010). An animal model of amnesia that uses Receiver Operating Characteristics (ROC) analysis to distinguish recollection from familiarity deficits in recognition memory. *Neuropsychologia, 48,* 2281–2289. doi:10.1016/j.neuropsychologia.2009.09.015

Elfman, K. W., Parks, C. M., & Yonelinas, A. P. (2008). Testing a neurocomputational model of recollection, familiarity, and source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 752–768. doi:10.1037/0278-7393.34.4.752

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Greve, A., Donaldson, D. I., & van Rossum, M. C. (2010). A single-trace dual-process model of episodic memory: A novel computational account of familiarity and recollection. *Hippocampus, 20,* 235–251. doi:10.1002/hipo.20606

Greve, A., van Rossum, M. C., & Donaldson, D. I. (2007). Investigating the functional interaction between semantic and episodic memory: Convergent behavioral and electrophysiological evidence for the role of familiarity. *NeuroImage, 34,* 801–814. doi:10.1016/j.neuroimage.2006.07.043

Harlow, I. M., MacKenzie, G., & Donaldson, D. I. (2010). Familiarity for associations? A test of the domain dichotomy theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1381–1388.

Herron, J. E., & Rugg, M. D. (2003). Strategic influences on recollection in the exclusion task: Electrophysiological evidence. *Psychonomic Bulletin and Review, 10,* 703–710.

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,* 96–106. doi:10.1037/0882-7974.21.1.96

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 380–387.

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57,* 335–384.

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

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science, 20,* 509–515. doi:10.1111/j.1467-9280.2009.02324.x

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110,* 611–646. doi:10.1037/0033-295X.110.4.611

Parks, C. M., Murray, L. J., Elfman, K., & Yonelinas, A. P. (2011). Variations in recollection: The effects of complexity on source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 861–873.

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, 106,* 11515–11519.

Peters, J., Thoma, P., Koch, B., Schwarz, M., & Daum, I. (2009). Impairment of verbal recollection following ischemic damage to the right anterior hippocampus. *Cortex, 45,* 592–601.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences, 4,* 108–115. doi:10.1016/S1364-6613(00)01445-5

Sherman, S. J., Arri, A., Hasselmo, M. E., Stern, C. E., & Howard, M. W. (2003). Scopalomine impairs human recognition memory: Data and modeling. *Behavioural Neuroscience, 114,* 526–539.

Slotnick, S. D. (2010). Remember source memory ROCs indicate recollection is a continuous process. *Memory, 18,* 27–39.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152–176. doi:10.1037/0033-295X.114.1.152

Wixted, J. T., Mickes, L., & Squire, L. R. (2010). Measuring recollection and familiarity in the medial temporal lobe. *Hippocampus, 20,* 1195–1205.

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., 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