



Predicting and “predicting” high confidence misses

Henry L. Roediger III^{*}, Ian Dobbins

Washington University in St. Louis, USA

Goshen-Gottstein et al. (2022; hereafter GGLM) argue that criterion placement is “trivial” and they intended to illuminate “the triviality of high confidence judgments.” In their earlier comment, they referred to confidence ratings as “epiphenomenal” (Levi et al., 2022; hereafter LMGG). All that matters to them is the memory signal as reflected in d' . Confidence in memory decisions, reflected by placement of criteria, doesn't matter according to their point of view. That idea seems odd, because confidence is often critical within signal detection theory (SDT) to sweep out the receiver operating characteristic (ROC). The only way insight into the structure of memory evidence can be achieved is through asking subjects to make decisions about that evidence, such as confidence or source judgments. Still, that is what they argue; only d' matters. Criterion placement, and hence confidence, is trivial.

Of course, from the point of view of the rememberer, nothing could be further from the truth. The rememberer – all of us – has the experience of remembering something, and confidence comes along for the ride, part and parcel of one's experience of remembering (Wixted and Stretch, 2000). Witt et al. (2015, 2017) have made a similar argument about perceiving, in particular in reference to visual illusions; it can be impossible to separate sensitivity from criterion in experiencing a perceptual illusion, just as with a memory illusion (Wixted and Stretch, 2000).

Contrary to LMGG's argument, criteria also matter greatly in applied settings. Consider the case of Watson vs. Commonwealth of Virginia in 2019 (Wixted and Roediger, 2022):

The case stemmed from the robbery of Joseph Jackson and Paul Abbey and the murder of Abbey by three men in a parking lot. Jackson's identification of Watson as one of the assailants was a key issue at trial. Jackson initially identified Watson in a photo lineup, but indicated that he was “not sure.” At the preliminary hearing, however, Jackson expressed certainty that Watson was one of the assailants (p. 3).

To some theorists, this shift in confidence from quite low to quite high might be conceptualized as a criterion shift (perhaps owing to the

suggestive nature of the courtroom test setting). However, to Watson, this is anything but trivial since the placement and flexibility of confidence criteria would be pivotal in the determination of guilt or innocence.

However, to GGLM, the matter is trivial and all that matters is the witness's memory signal or d' (which of course cannot be known in legal cases). Of course, the shift in this case could be due to an external influence, but it could also be due to the witness's continuing to think about the case and shifting the criterion from internally generated “evidence” due to repeated retrieval. Dozens of legal cases like this one exist, and they have led to grievous errors in the legal system, ones that have sentenced innocent people to prison or to death. Shifting criteria, and hence confidence, do matter in the legal system and in many other contexts outside the lab; they are not trivial in this or in many other situations.

In GGLM's (2021) response to our criticism of their position, they say that high confidence misses will occur in virtually all experiments involving a decision component. SDT is said to “predict” the outcome, and they refer to their examples in vision as ones of “everyday blindness,” apparently thinking that is a ridiculous idea. The examples they use involve detecting “briefly presented lights just above threshold,” but a better analogy to our experiments would be presenting bright and dim lights, not a faintly appearing stimulus, with high confidence misses occurring in these discriminations. Moreover, GGLM's examples suggest that they believe the prevalence of HCMs is wholly unmoored from considerations about the performance levels of the subjects, the characteristics of stimuli, or the characteristics of the design (e.g., payoff manipulations, biasing instructions, target prevalence, etc.). This makes it very hard to pin down effects they view as surprising versus unsurprising.

In the data presented in Roediger and Tekin (2020a) paper, we had assumed that the presented items were encoded. In Roediger and Tekin (2020b) experiments, we went to even greater lengths to ensure the words were well encoded. In one experiment, we had subjects repeat words aloud to ensure accurate encoding. In the other experiment, we had subjects perform a semantic encoding task on the words to ensure

^{*} Corresponding author.

E-mail address: roediger@wustl.edu (H.L. Roediger).

enhanced memorability during encoding (Craig and Tulving, 1975). Yet, despite these definite encodings of the items, subjects rejected having seen 20% of the items they missed, with high confidence, on a test shortly after list presentation. This outcome seems more like the perceptual analogue of confidently claiming a bright light was in fact dim, despite squinting in response to its brightness. GGLM argue that such a finding is “trivial,” yet we think it has the potential to be interesting and to tell us something fundamental about rapid forgetting.

As noted above, GGLM appeal to several perceptual phenomena where high confidence misses involving weak signals occur. They say SDT “predicted” these phenomena. Then they later provide data that preceded their prediction as confirming it. Dobbins (2022) and Roediger and Tekin (2022) explained that these predictions are actually post-dictions. Prediction means forecasting or, according to the *Pocket Oxford English Dictionary* (2001), to predict means to “state that (an event) will happen in the future.” LMGG and GGLM use “predict” in the unusual sense of “to explain after the fact by appeal to a model or theory.” We have put predict in quotes when we use this meaning of *ex post facto* reasoning. Of course, there is nothing wrong with saying how a model fits data after the data have been collected, but such reasoning is not predicting the data. The perceptual phenomena they discuss existed before GGLM came along and noted that SDT could account for them, after the fact. And we have no argument there. As we stated in Roediger and Tekin (2022), SDT is a beautiful framework for conceiving of decision processes in making judgments about perceiving, remembering, making medical decisions and many other topics (Wixted, 2020). It is an abstract model and not a model of psychological or neural processes that give rise to the representation of evidence or the execution of decisions regarding the evidence. Indeed, the fact that GGLM use exactly the same model to describe visual detection experiments and recognition memory experiments means that SDT is not a substantive psychological model, because surely the psychological and neural processes differ markedly in the two cases.

GGLM belittle the concept of “everyday blindness,” as they called it. We don’t know why. After all, perceptual psychologists have discovered exactly such phenomena and they have aroused intense interest. Consider phenomena such as inattentive amnesia (Wolfe, 1999), inattentive blindness (Castel et al., 2012), and change blindness (Pashler, 1988; Simons and Rensink, 2005), among others. These phenomena are accounted for by lapses in, or the limitations of, attention, where objects in plain sight are not attended and, largely, not remembered. We would like to think that what we are calling everyday amnesia belongs in the same class – the failure to recognize (with high confidence) an event that occurred 5–10 minutes previously. The everyday amnesia hypothesis is different than these attentional accounts, because Roediger and Tekin (2020a) proposed that even events that were perceived and attended can be rapidly forgotten such that only 5 min later subjects fail to recognize them with high confidence. Still, perceptual phenomena like everyday blindness are of intense interest today, not ideas to be mocked.

In Levi et al. (2022), the claim was made that “The absence of such errors [high confidence misses or false alarms] would undeniably provide a refutation of SDT ...” (p. 7, MS). In GGLM, the authors wish “to revise” that opinion, because Dobbins (2022) and Roediger and Tekin (2022) showed that it was wrong. We applaud their belated acknowledgment of this fact, because SDT can easily accommodate the finding of no, a few, or an abundance of high confidence misses. In fact, as we argued previously, SDT can account for almost any set of hit and false alarm rates observed in real data. That is, in accounting for a finding in recognition memory, an SDT model can be drawn after the fact to account for virtually any outcome. LMGG (2021) have SDT “predicting” about 20% of high confidence misses, but of course SDT can “predict” any level of high confident misses from 0% to 50%–100%. If SDT can “predict” any recognition outcome, then it really predicts nothing. Rather, the SDT model that LMGG use is more a conceptualization than an explanation of memory or decision making.

Let us assume that the basic SDT conceptualization of confidence espoused by GGLM is correct, namely, that it reflects the strategic placement of numerous ordered confidence criteria along a single, continuous dimension of normally distributed old and new item recognition evidence. In the absence of any further stipulations, such a model is fairly weak in terms of genuine predictions. Indeed, at this level of minimal theoretical constraint, the model doesn’t even predict that confidence will increase across individuals with increasing discrimination accuracy or will be on average higher in experiments that promote better versus worse episodic encoding. Instead, the model only predicts that when performance is above chance, individuals’ correct reports will be more confident on average than their incorrect reports (provided subjects make minimal use of all the possible confidence options). However, if we also assume that old item evidence can be more variable than new item evidence, then even this prediction cannot be sustained, because it is possible that increased old item variance will ‘push’ the left tail of the old item distribution sufficiently leftward to produce a greater relative proportion of misses falling below the high confidence “new” criterion compared to correct rejections. In their commentary GGLM discuss various possible positions of confidence criteria as more versus less surprising, but the basis of this surprise is a mystery because it is not a property of the unequal variance (UEV) SDT model they present. To be clear, we would be quite surprised and intrigued by a design that produced misses with higher average confidence than correct rejections, or a design in which the mean confidence of the observers was wholly unrelated to their discrimination abilities. However, since it is trivial to position confidence criteria and unequal variance normal evidence distributions to produce such outcomes, our surprise would arise from a broader theoretical basis of understanding than afforded by the basic UEV SDT model discussed by GGLM.

To make this excessive flexibility more concrete, consider Fig. 1 of Roediger and Tekin (2020a), re-presented below. It shows accuracy for old responses in the highest confidence bin (4) of greater than 90%, and accuracy in the highest confidence 4 bin of New responses of about 80%, with these differences being clearly reliable. Can SDT produce this? Of course. For example, one could assume that the greater assumed variance of the old item distribution ‘pushes’ a considerable portion of the right tail of the old item evidence distribution to fall above the highest ‘old’ confidence criterion; such an effect would not occur for the less variable new items that lead to high confidence false alarms. Since this variance effect will be prominent for the right tails of the old and new evidence distributions, but not the left tails, one might claim to predict the greater observed accuracy shown in the figure. Although, this seems intuitively satisfying, this intuition would be illusory since it would be possible to equally easily explain data with exactly the opposite pattern; that is higher accuracy for the most confident new responses compared to the most confident old responses. To do so would merely require adjustment of the highest old and new confidence criteria. To be sure, there are ways of placing additional constraints on the UEV SDT model to reduce this extreme flexibility and lack of definitive prediction (e.g., Selker, van den Bergh, Criss & Wagenmakers, 2019), but this is not done by GGLM.

In addition, there is serious reason to believe that subjects may not in fact produce memorial confidence decisions by strategically placing decision criteria along a single continuous evidence dimension. Dobbins (2022) mentioned papers by Kellen and Klauer (2015) and Province and Rouder (2012) as providing data strongly challenging this idea in humans. Moreover, this idea of strategically parsing a continuous decision space to optimize explicit goals, on an experiment-by-experiment basis, seems a bit strained in animals that produce data patterns similar to humans, such as pigeons and rats (e.g., Fortin et al., 2004). This is probably why, when researchers consider animal memory, they often propose models without strategic decision criteria. For example, Wixted and Gaitan (2002) proposed a model that completely eschewed strategic decision criteria, as did Kepecs et al. (2008). This abandonment of strategic confidence criteria also fits recent work on distractor free

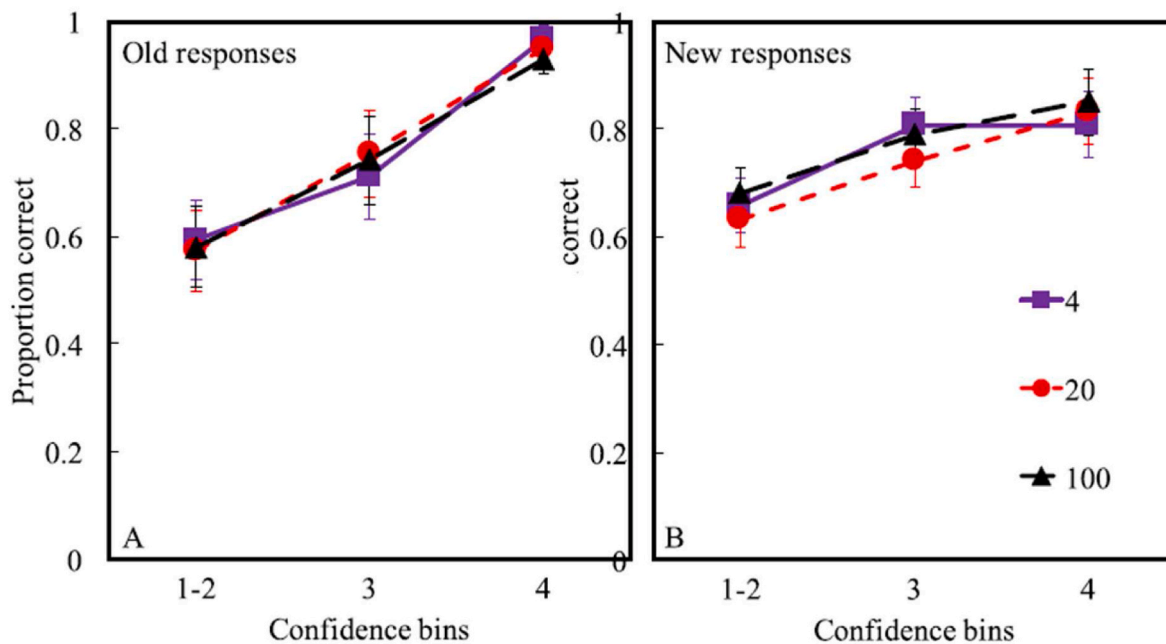


Fig. 1. Comparison of scale types for old responses (A), and new responses (B). The 20- and 100-point scales are divided into four bins. 1–2 combines lowest two confidence bins, and 4 indicates highest confidence. For both types of responses, there was a positive relation between confidence and accuracy; however, this relation was more pronounced for old responses. Error bars indicate 95% confidence intervals. Data from Tekin and Roediger (2017, Experiment 2).

recognition following an earlier paradigm developed by Wallace (1980). For example, Cox and Dobbins (2011) showed that hit rates and confidence ratings on a recognition test without any distractors or lures were highly similar to a test with lures (see too Koop et al., 2015). It seems odd to assume that subjects strategically locate confidence criteria yet do not alter these in response to a test that contains absolutely no lures. Finally, Miller and Kantner (2019) have argued that “not all people are cut out for strategic criterion shifting,” because they consistently find a groups of people who do not shift criteria under experimental conditions that would seem to call for strong shifts (Layher et al., 2020). These findings demonstrate that confidence behavior during recognition has many aspects that simply are not anticipated given the basic UEV SDT model espoused by GGLM.

GGLM end their comment by discussing remember/know judgments (Tulving, 1985) and their proper account. We had not introduced remember/know judgments in any of our work, so we did not see the relevance of the discussion. Apparently, GGLM intended this discussion, to paraphrase the argument, as “here’s another silly idea that signal detection models showed could be modelled as a single process rather than two.” Yes, much debate has occurred on this score, but once again we do not think the solution is as simple as they make out. Much work by Mandler (1980), Jacoby (1991), Yonelinas (2002), and Wixted and Mickes (2010), among others, has found evidence for two components in recognition memory reflecting recollection (akin to remembering) and fluency (akin to knowing, by one view).

Much hinges on the idea that confidence can in some sense explain the remember/know distinction: remembering occurs when confidence is high, knowing occurs when it is low. Confidence is assumed to be the fundamental quality underlying judgements of remembering and knowing. Of course, Tulving (1985) noted, from data in his second experiment, that “a positive correlation between confidence and remembering” existed. This fact did not trouble him in the least. In all likelihood, the experience of remembering is what gives rise to confidence. That is, if a person can remember the time and place that an event occurred, as well as details surrounding the event, one can be more confident that it occurred. In short, remembering may explain confidence rather than the other way around. Confidence in judgments of knowing is often thought to reflect the quality familiarity or fluency of

processing.

To return to our main point, we continue to believe that the mystery of highly confident misses shortly after study of material is worthy of further attention. Waving the wand of signal detection theory and shifting criteria over it does not make the phenomenon better understood. Rather, we are only provided with an illusion of explanation.

Authors’ note

We thank Chris Zerr for his comments and help with the manuscript.

References

- Castel, A.D., Vendetti, M., Holyoak, K.J., 2012. Fire drill: inattention blindness and amnesia for the location of fire extinguishers. *Attention, Perception, & Psychophysics* 74 (7), 1391–1396. <https://doi.org/10.3758/s13414-012-0355-3>.
- Cox, J.C., Dobbins, I.G., 2011. The striking similarities between standard, distractor-free, and target-free recognition. *Mem. Cognit.* 39 (6), 925–940. <https://doi.org/10.3758/s13421-011-0090-3>.
- Craik, F.I., Tulving, E., 1975. Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* 104 (3), 268–294. <https://doi.org/10.1037/0096-3445.104.3.268>.
- Dobbins, I.G., 2022. Hindsight and the Theories of Signal Detection: A Comment on Levi, Mickes, and Goshen-Gottstein. *Neuropsychologia*.
- Fortin, N.J., Wright, S.P., Eichenbaum, H., 2004. Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature* 431, 188–191. <https://doi.org/10.1038/nature02853>.
- Goshen-Gottstein, Y., Levi, A., Mickes, L., 2022. Signal-detection Theory Separates the Chaff of Bias from the Wheat of Memory: Illuminating the Triviality of High-Confidence Judgments. *Neuropsychologia*.
- Jacoby, L.L., 1991. A process dissociation framework: separating automatic from intentional uses of memory. *J. Mem. Lang.* 30 (5), 513–541. [https://doi.org/10.1016/0749-596X\(91\)90025-F](https://doi.org/10.1016/0749-596X(91)90025-F).
- Kellen, D., Klauer, K.C., 2015. Signal detection and threshold modeling of confidence-rating ROCs: a critical test with minimal assumptions. *Psychol. Rev.* 122 (3), 542–557. <https://doi.org/10.1037/a0039251>.
- Kepecs, A., Uchida, N., Zariwala, H.A., Mainen, Z.F., 2008. Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455, 227–231. <https://doi.org/10.1038/nature07200>.
- Koop, G.J., Criss, A.H., Malmberg, K.J., 2015. The role of mnemonic processes in pure-target and pure-foil recognition memory. *Psychonomic Bull. Rev.* 22 (2), 509–516. <https://doi.org/10.3758/s13423-014-0703-5>.
- Layher, E., Dixit, A., Miller, M.B., 2020. Who gives a criterion shift? A uniquely individualistic cognitive trait. *J. Exp. Psychol. Learn. Mem. Cognit.* 46 (11), 2075–2105. <https://doi.org/10.1037/xlm0000951>.

- Levi, A., Mickes, L., Goshen-Gottstein, Y., 2022. The new hypothesis of everyday amnesia: an effect of criterion placement, not memory. *Neuropsychologia*.
- Mandler, G., 1980. Recognizing: the judgment of previous occurrence. *Psychol. Rev.* 87 (3), 252–271. <https://doi.org/10.1037/0033-295X.87.3.252>.
- Miller, M.B., Kantner, J., 2019. Not all people are cut out for strategic criterion shifting. *Curr. Dir. Psychol. Sci.* 29 (1), 9–15. <https://doi.org/10.1177/0963721419872747>.
- Pashler, H., 1988. Familiarity and visual change detection. *Percept. Psychophys.* 44 (4), 369–378. <https://doi.org/10.3758/BF03210419>.
- Pocket Oxford English Dictionary, 2001. Oxford University Press, Oxford.
- Province, J.M., Roudier, J.N., 2012. Evidence for discrete-state processing in recognition memory. *Proc. Natl. Acad. Sci. Unit. States Am.* 109 (36), 14357–14362. <https://doi.org/10.1073/pnas.1103880109>.
- Roediger, H.L., Tekin, E., 2020a. Recognition memory: Tulving's contributions and some new findings. *Neuropsychologia* 139, 107350. <https://doi.org/10.1016/j.neuropsychologia.2020.107350>.
- Roediger, H.L., Tekin, E., 2020b. Everyday amnesia: High confidence misses in recognition memory. In: Paper presented at the meeting of the Psychonomic Society.
- Roediger, H.L., Tekin, E., 2022. Can signal detection theory explain everyday amnesia (high confident misses)? *Neuropsychologia*.
- Selker, R., van den Bergh, D., Criss, A.H., Wagenmakers, E.J., 2019. Parsimonious estimation of signal detection models from confidence ratings. *Behav. Res. Methods* 51 (5), 1953–1967. <https://doi.org/10.3758/s13428-019-01231-3>.
- Simons, D.J., Rensink, R.A., 2005. Change blindness: past, present, and future. *Trends Cognit. Sci.* 9 (1), 16–20. <https://doi.org/10.1016/j.tics.2004.11.006>.
- Tulving, E., 1985. Memory and consciousness. *Canadian Psychology/Psychologie canadienne* 26 (1), 1–12. <https://doi.org/10.1037/h0080017>.
- Wallace, W.P., 1980. On the use of distractors for testing recognition memory. *Psychol. Bull.* 88 (3), 696–704. <https://doi.org/10.1037/0033-2909.88.3.696>.
- Witt, J.K., Taylor, J.E.T., Sugovic, M., Wixted, J.T., 2015. Signal detection measures cannot distinguish perceptual biases from response biases. *Perception* 44 (3), 289–300. <https://doi.org/10.1068/p7908>.
- Witt, J.K., Tenhundfeld, N.L., Bielak, A.A., 2017. Dissociating perception from judgment in the action-specific effect of blocking ease on perceived speed. *Atten. Percept. Psychophys.* 79 (1), 283–297. <https://doi.org/10.3758/s13414-016-1222-4>.
- Wixted, J.T., 2020. The forgotten history of signal detection theory. *J. Exp. Psychol. Learn. Mem. Cognit.* 46 (2), 201–233. <https://doi.org/10.1037/xlm0000732>.
- Wixted, J.T., Gaitan, S.C., 2002. Cognitive theories as reinforcement history surrogates: the case of likelihood ratio models of human recognition memory. *Anim. Learn. Behav.* 30 (4), 289–305. <https://doi.org/10.3758/BF03195955>.
- Wixted, J.T., Mickes, L., 2010. A continuous dual-process model of remember/know judgments. *Psychol. Rev.* 117, 1025–1054. <https://doi.org/10.1037/a0020874>.
- Wixted, J.T., Roediger, H.L., 2022. Signal detection theory and eyewitness identification. In: Logie, R.H., Wen, Z., Gathercole, S., Cowan, N., Engle, R. (Eds.), *Memory in Science for Society*. Oxford University Press.
- Wixted, J.T., Stretch, V., 2000. The case against a criterion-shift account of false memory. *Psychol. Rev.* 107 (2), 368–376. <https://doi.org/10.1037/0033-295X.107.2.368>.
- Wolfe, J.M., 1999. Inattention blindness. In: Coltheart, V. (Ed.), *Fleeting Memories: Cognition of Brief Visual Stimuli*. The MIT Press, pp. 71–94.
- Yonelinas, A.P., 2002. The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46 (3), 441–517. <https://doi.org/10.1006/jmla.2002.2864>.