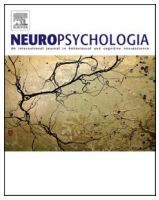
Neuropsychologia 166 (2022) 108116

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Signal-detection theory separates the chaff of bias from the wheat of memory: Illuminating the triviality of high-confidence judgments

Yonatan Goshen-Gottstein [a,*], Adva Levi [a], Laura Mickes [b]

[a] Tel-Aviv University, Ramat Aviv, Israel
[b] University of Bristol, United Kingdom

ARTICLE INFO

Keywords
Recognition
Signal-detection theory
Sensitivity
Bias
Criterion
High-confidence misses
Prediction
Process model

It's an idea that piques one's curiosity: high-confidence misses (HCMs) in recognition memory may represent 'everyday amnesia' (EA; Roediger and Tekin, 2020; R&T1). In our commentary (Levi et al., 2021; LMGG), we argue otherwise. We posit that HCMs are derivations of signal-detection theory (SDT) that reflect response bias. As such, their existence is unsurprising and investigating them does not seem constructive to understanding memory processes.

Four criticisms stand out in rebuttals to our commentary authored by Dobbins (2021) and Roediger & Tekin (2021; R&T2). Criticism 1 (C1): The predictions we attribute to SDT are not predictions but are explanations made in hindsight. Criticism 2 (C2): SDT would not be refuted by finding no HCMs. Criticism 3 (C3): SDT does not provide a process explanation of HCMs; it merely provides a re-description of the data and as such, adds little to our understanding of HCMs. Criticism 4 (C4): History teaches us that had criticisms like ours been afforded serious consideration in the past, important contributions to memory research would have been stifled.

C1. To address the question of hindsight, we articulate below several novel a-priori SDT-based predictions regarding the emergence of HCMS in domains of inquiry other than memory (e.g., vision, audition, reading). By virtue of their novelty—none of the predictions can be attributed to hindsight. Regrettably, none of the findings will justify a claim about any domain of discovery other than decision bias.

According to SDT, two cumulative conditions must be met for HCMs to emerge in any domain of psychological research: Participants must make decisions using confidence judgments and an objective standard must exist to categorize task performance as incorrect. Leave it to nature that a gush of HCMs will ensue.

Formulating novel predictions that will unquestionably be materialized in empirical data may sound like a good idea. But this is so only if the predictions are interesting. If asked to place our novel predictions of HCMs along a surprise-signal distribution ranging from high-surprise to low-surprise (e.g., boring, R&T2; expected, Wilson and Wixted, 2018; trivial), we would undoubtable place them on the low-surprise end of the distribution. Henceforth, we use the term 'trivial' to describe our predictions, as shorthand to represent the idea of a low-surprise signal. Let the predictions begin.

We predict that in a visual-detection task to briefly presented lights just above threshold, with lights present on half the trials and absent on the other half, if participants judge the appearance of light on a 6-point confidence scale ('1'- very confident the light was not presented, '6'- very confident the light was presented), a number of those responses will be HCMs (i.e., the light was in fact presented but was judged '1'). These HCMs represent a condition of 'everyday blindness.' We make

; EA, everyday amnesia; HCMs, high-confidence misses; HCCRs, high-confidence correct rejections; SD, standard deviation; SDT, signal-detection theory; UVSD model, unequal variance signal-detection model.

* Corresponding author. School of Psychological Sciences, Tel-Aviv University, Ramat Aviv, 69978, Israel.
E-mail address: goshengott@gmail.com (Y. Goshen-Gottstein).

https://doi.org/10.1016/j.neuropsychologia.2021.108116
Received 8 May 2021; Received in revised form 6 December 2021; Accepted 7 December 2021
Available online 11 December 2021
0028-3932/

corresponding predictions for tone-, olfactory-, tactile- and taste-detection tasks, hypothesizing that HCMs in these tasks will be observed, representing everyday deafness, everyday anosmia, everyday anaphia, and everyday ageusia, respectively. Likewise, we predict the existence of everyday neglect, everyday dyslexia, everyday dyscalculia. Disturbing is the thought that if HCMs would be subject to neuroscientific investigation (in memory, vision or any other domain of research), neural signatures in both time and space would likely be discovered (e.g., Curran, 2004; Selimbeyoglu, Kesin- Ergen and Demiralp, 2012).

We could then introduce a "HCM syndrome" that combines all of the predictions, including HCMs in memory. The syndrome comprises three components: 1) the very existence of HCMs, sometimes made with high prevalence; 2) individual differences, with some participants showing few HCMs and others exhibiting many; 3) HCMs will sometimes decrease with an increase in overall memory performance (measured by, say, d').

We have formulated many trivial, apriori predictions regarding HCMs in domains other than memory. All predictions were derived from SDT and none were the product of hindsight. We invite readers to ponder the status of HCMs in memory described by R&T1. Are they different in any way from our predicted HCMs in, say, vision and audition? Do they not omit the identical weak surprise signal? Are they not just as trivial and, as such, not a product of hindsight?

The formal SDT model (or specifically the UVSD model) is not required to appreciate why this is the case.[1] The technical trees that spawn SDT are merely mathematical formalizations of psychological principles rooted in nature regarding decision-making. For it is the nature of any dynamic, noisy system that errors will emerge whenever decisions are made using different thresholds (including, but not limited to, thresholds pertaining to confidence). Such systems surround us everywhere, ranging from anomaly-detection of micro-processing chips at Intel to the psychological world of decision-making, where recognition is but one instance. For decisions to be made, a decision boundary must be set. And such boundaries entail errors.

That decision boundaries must mediate judgments, such as those of confidence, drove us to list six concerns regarding T&R1's mnemonic interpretation (i.e., EA) of confidence ratings (see the penultimate paragraph in LMGG). These concerns should be addressed by EA proponents. All concerns are based on the idea that all judgments of confidence (in SDT and in any noisy system in nature), and in particular HCMs, are mediated by the setting of decision boundaries. R&T2 did not address these concerns. Is the only competitor in the market of ideas regarding confidence rating that they are a function of criteria placement, not memory? If so, we are left with no option but to subscribe to the lone product on the theoretical shelf that views HCMs as trivial and thus insignificant to the enterprise of understanding vision, audition, and, yes, memory.

Finally, and importantly, we do not deem HCMs to be trivial in principle; our point is simply that no data have been presented thus far to signal any interest in them. Still, we can imagine a pattern of results that would boost the interest signal of HCMs. Specifically, SDT predicts that more high-confidence correct rejections (HCCRs) should be observed than HCMs (see Fig. 1). Both response types are items judged 'new' and both are conceptualized by SDT as representing items (lures and targets) with strength lower than the lowest criterion (c1). Because HCCRs are responses for items from the lure distribution, which is assumed to have a lower mean than that of the target distribution, more items ought to fall to the left of the relevant criterion in the lure distribution (HCCRs) than in the left of the target distribution (HCMs). SDT could not be said to predict the finding if the reverse were found. A strong surprise-signal would emerge, suggesting that further research is warranted. Note that it was the complementary pattern—fewer hits than false alarms—that highlighted the surprising nature of false alarms in DRM lists (see our discussion below of Roediger and McDermott, 1995; see Fig. 1). At present, though, nothing in the data described by R&T1 should increase our epistemological certainly that such data will, in fact, be found.

C2. We acknowledge and appreciate the claims made by R&T2 and Dobbins that SDT is a flexible model and would not be refuted by an absence of HCMs. We thus wish to modify our arguments in LMGG and claim that within the parameter-space of reasonable (realistic) SDT parameters, HCMs will always be found. The prediction is thus phrased in probabilistic, not deterministic, terms. Note that probabilistic predictions are abundant in physics and chemistry, e.g., the field of stochastic chemistry, and are true even for the Second Law of Thermodynamics, Batalhao et al. (2015).

Critical to our argument is the idea that a scenario of zero HCMs is so off the charts, requiring an extreme decision criterion so far out of left field—both literally and figuratively—that though not impossible, it is highly unlikely. To the point, when modelling the R&T1 data, the parameter value for c1, the left-most criterion, was −1.045, roughly 1.5 SDs the left of the mean of the distribution from which misses are sampled—the target distribution (mean = 1.08, SD = 1.29) .[2] Setting of this criterion at 1.5 SDs below the mean seems like a judicious strategy—well within the reasonable parameter space. Importantly, this setting seems to provide a sound balance between the need for few HCMs while still enabling high-confidence correct rejections[3] (also see our reply to C3, below).

In summary, the setting of an extreme left decision criterion is so unlikely as to sustain positive trivial predictions for the presence of HCMs in any domain where the two conditions (confidence judgments, responses can be objectively incorrect) are satisfied. Such predictions are true for vision, audition, and, well, memory.

C3. HCMs reflect the operation of a domain-general system that makes boundary decisions under noisy conditions. As such, they are the product of the decision process, not memory. If HCMs are not memory-based, then there is no need for a process account. We agree with R&T2 that SDT does not provide a mnemonic process explanation for HCMs. We disagree that such an explanation is warranted for HCMs.

Decision-making scientists, less so memory scientists, may well be interested in questions of criteria placement, including why some participants are liberal and others conservative in their criteria settings. R&T2 asked why, according to SDT, some participants place their most liberal criterion in a way that they miss, with high confidence, items that they recently studied. Answers have been proposed to this question, not under the rubric of memory, but under that of "decision goals that mediate criteria placement" (for an overview, see Macmillian and Creelman, 2005, pp. 42–44). One answer, for example, is that the goals of criterion placement is to provide a balance between HCMs that constitute errors, and high-confidence correct-rejections that constitute correct judgments (see Fig. 1).

The paths of memory scientists and decision-making scientists do sometimes cross. For example, Stretch and Wixted (1998) demonstrated

[1] As Dobbins correctly pointed out, there is a family of SDT models, not just the single UVSD model we described. Common to all, is the conceptualization that when boundaries must be set in noisy dynamic environments, errors will ensue. Our adoption of the UVSD model was due to its relative simplicity and because our reading of the literature suggests that it best accounts for existing recognition data (e.g., Pazzaglia et al., 2013).

[2] We estimated the parameters for R&T's data (the HCM data in Table 1 of R&T1; Experiment 1 of Tekin and Roediger, 2017) using MLE of the UVSD model (see LMGG). Assuming a mean and SD of 0 and 1 for the lure distribution, the parameter values obtained were a mean of 1.08 and SD of 1.29, and for c1-c5 criteria, were −1.045, −0.393, 0.385, 0.962, 1.477, respectively.

[3] Note that even if the criterion had been set as far left as an extreme 2 SDs below the mean—a value considered so extreme as to represent a significant result in a two-tailed test—2.28% of the observations would still be HCMs, yielding almost 250 HCMs in some of the data sets analysed by R&T1.

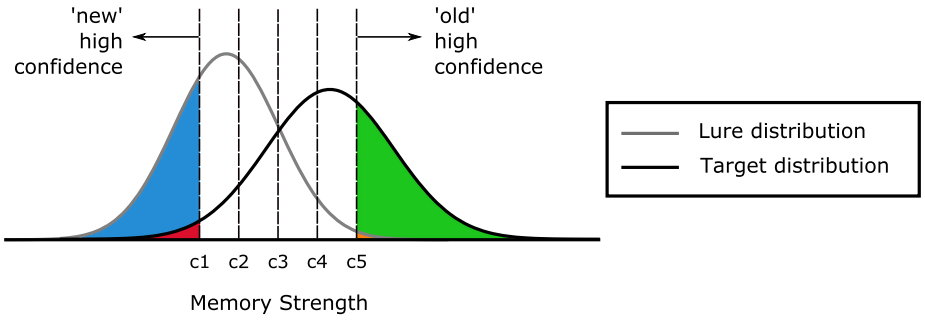


Fig. 1. An illustration of the unequal variance signaldetection (UVSD) model. The lure distribution, in grey, and the target distribution, in black, represent the unstudied ("new") and studied ("old") items, respectively. Five criteria are spread across the memory-strength axis and represent different confidence judgments. Studied items given the highest confidence that the item is "new" (c1) result in high-confidence misses (HCMs). HCMs are the items that fall under the target distribution to the left of the c1 criterion and their probability is represented in red. Unstudied items given the highest confidence that the item is "new" (c1) results in high-confidence correct rejections (HCCRs). HCCRs are the items that fall under the lure distribution to the left of the c1 criterion and their probability is represented in blue. Because the target distribution is assumed to have a higher mean than the lure distribution, this model predicts a higher rate of HCCRs than HCMs. If the reverse finding was observed, a strong 'surprise' signal would emerge. In parallel, SDT predicts a higher rate of high-confidence hits (in green) than high confidence false alarms (in orange), the reverse of which was observed by Roediger and McDermott (1995), yielding a strong surprise-signal. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

how decision criteria fan out on the memory-strength axis as memory performance decreases (as measured by d'). Such an association between memory-strength and criterion placement yields questions regarding the very foundations of memory (for another example, see Selmeczy and Dobbins, 2014). As memory scientists, we would embrace an association between HCMs and genuine mnemonic processes. At present, we cannot envision such an association.

C4. If taken seriously, would the SDT arguments advanced by LMGG have stifled classic advances in research, memory or otherwise? Our answer is a resounding "No." R&T2 listed a series of classic mnemonic effects they suggest would have been trivialized had SDT formulations been proposed upon their inception. By extension, our trivialization of HCMs should be dismissed.

To refute this argument, we counter that important effects listed by R&T2 would not have been dismissed based on SDT analysis. As we describe below, from the onset, the surprise signal in these effects would have been sufficiently strong as to invite further research rather than dismiss it. We contend that any observation that would have been interpreted by SDT to affect sensitivity, not bias, would prompt researchers to pursue what mediated the effect, including its cognitive and neural underpinnings.

Fortunately, SDT separates the chaff (bias) from the wheat (memory) of interesting effects. Dense bilateral medial lobe amnesia was presented by R&T2 as a condition in which there is little or no overlap between distributions. In SDT terms, for these patients, studying an item does not boost its memory representation. We cannot think of a more compelling finding. Thus, SDT would increase understanding of the mystifying experience afforded by amnesia, not trivialize it (like it does with HCMs). In fact, much research has been directed towards depicting the most precise SDT model to describe amnesia (e.g., Didi, Pereman & Goshen-Gottstein, 2016) and to uncover its neurological underpinnings. All interpretations of this phenomenon are of memory sensitivity, not bias. SDT's surprise-signal for amnesia as reflecting memory is very strong, guiding researchers to pursue it further. The same is true for hyperthymesia. According to the SDT interpretation (Fig. 1B of R&T2), for some individuals, the boost to the target distribution is substantial. SDT signals both memory-related questions as important.

Perhaps most interesting are high-confidence false memories created by DRM lists (Roediger and McDermott, 1995). At first glance, such false memories resemble HCMs by representing an inevitable error of the placement of a decision boundary—false alarms. In their comment, R&T2 argue that history has taught us that arguments such as ours (LMGG) were raised against false alarms in DRM lists (Miller and Wolford, 1999), trivializing such false alarms to represent bias. Such arguments were eventually rejected (Roediger and McDermott, 1999; Wickens and Hirshman, 2000; Wixted and Stretch, 2000).

Whether that debate was fruitful is a matter of opinion. Roediger and McDermott (1995; R&M) documented a data pattern that provided a strong surprise signal. In their Experiment 2, a higher false-alarm rate (0.72) than hit rate (0.65) was found. Importantly, SDT interprets both hits and false alarms as 'old' responses made to items (both targets and lures, respectively) with strength higher than criterion. Additionally, according to SDT, the target distribution has a higher mean than the lure distribution. The conjunction of these two propositions reveals that SDT predicts that for 'old' responses, the rates of false alarms will inevitably be lower, not higher, than those of the hits (see Fig. 1 caption). That the reverse was found provided a strong surprise-signal, steering SDT theorists away from their (typically well-justified) knee-jerk SDT interpretation. We posit that R&M implicitly applied a SDT analysis to their data, which is why they called their data "remarkable"—a concept that is shorthand for "emitting a strong surprise signal." We propose that it would be difficult to explain why these data were remarkable without resorting to insights afforded by SDT.[4]

We thus disagree with R&T2's recount of the role that SDT played in high-confidence false alarms in the DRM task, according to which "the SDT debate was a sideshow." Quite the contrary. Only because of SDT was the data pattern found by Roediger and McDermott experienced as 'remarkable.'

We end by presenting a counterexample to the idea that criticisms such as ours would have stifled classic research. When first introduced by Tulving (1985), it was proposed that remember-know (RK) judgments reflect the operation of two qualitative mnemonic processes. We suggest that had SDT interpretations been taken more seriously (e.g., Donaldson, 1996), dual-process assumptions would have been tested more vigorously and as a result, not been supported by the data. Today—35 years after the introduction of the RK task and thousands of articles later—it is debateable whether RK responses truly reflect two processes or are better interpreted by a unidimensional memory process. The evidence may weigh in favor of the latter interpretation (e.g., Brezis et al., 2017; Dunn, 2004; Smith et al., 2011; but see Wixted and Mickes, 2010). Our take home message from the history of RK research is that

[4] The Miller and Wolford (1999) SDT interpretation of high confidence false alarms is qualitatively different from our SDT trivialization of HCMs (LMGG). Miller and Wolford argued that SDT could perhaps accommodate the DRM data by making a novel, memory-based (!) explanation. They proposed that participants respond 'old' to unstudied items not because they falsely remember studying them but because they realized that these items were closely related to the semantic theme defined by the list items. Thus, a memory mechanism (i.e., similarity to semantic theme) was proposed that differed from the memory mechanism proposed by R&M (the creation of false memories). Miller and Wolford's analysis ultimately led to a criterion-shift account, grounded in memory. Not so for our interpretation of HCMs, that trivialized them to be an epiphenomena of the decision process.

ruling out trivial interpretations, like the one we argue HCMs to be, should be the first order of business. All the more so if to begin with, the surprise-signal is ever so weak.

Conclusion: Psychology should advance as a cumulative science, by attempting to build on, not eschew, previous theoretical triumphs; in this case—SDT. Otherwise, our science is in danger of what Ernst Rutherford, 1908 Nobel Prize laureate in Physics, famously labelled as stamp-collecting, including collectables such as everyday blindness, everyday deafness and everyday amnesia.

Acknowledgements

This work was supported by the Israel Science Foundation [Grant number 1163–20].

References

Batalhao, T.B., Souza, A.M., Sarthour, R.S., Oliveira, I.S., Paternostro, M., Lutz, E., Serra, R.M., 2015. Irreversibility and the arrow of time in a quenched quantum system. Phys. Rev. Lett. 115, 190601.

Brezis, N., Bronfman, Z.Z., Yovel, G., Goshen-Gottstein, Y., 2017. The electrophysiological signature of remember-know is confounded with memory strength and cannot be interpreted as evidence for dual-process theory of recognition. J. Cognit. Neurosci. 29, 322–336.

Curran, T., 2004. Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. Neuropsychologia 42, 1088–1106.

Didi-Barnea, C., Peremen, Z., Goshen-Gottstein, Y., 2016. The unitary zROC slope in amnesics does not reflect the absence of recollection: critical simulations in healthy participants of the zROC slope. Neuropsychologia 90, 94–109.

Dobbins, Ian, 2021. Hindsight and the theories of signal detection: commentary on Levi, Mickes and Goshen-Gottstein (2022). Neuropsychologi. https://doi.org/10.1016/j.neuropsychologia.2021.108121.

Donaldson, W., 1996. The role of decision processes in remembering and knowing. Mem. Cognit. 24, 523–533.

Dunn, J.C., 2004. Remember–know: a matter of confidence. Psychol. Rev. 111, 524–542.

Levi, A., Mickes, L., Goshen-Gottstein, Y., 2021. The New Hypothesis of Everyday Amnesia: an Effect of Criterion Placement, Not Memory. Neuropsychologia.

Miller, M.B., Wolford, G.L., 1999. Theoretical commentary: the role of criterion shift in false memory. Psychol. Rev. 106 (2), 398–405.

Pazzaglia, A.M., Dube, C., Rotello, C.M., 2013. A critical comparison of discrete-state and continuous models of recognition memory: implications for recognition and beyond. Psychol. Bull. 139, 1173–1203.

Roediger, H.L., McDermott, K.B., 1995. Creating false memories: remembering words not presented in lists. J. Exp. Psychol. Learn. Mem. Cognit. 21 (4), 803–814.

Roediger III, H.L., McDermott, K.B., 1999. False alarms and false memories. Psychol. Rev. 106, 406–410.

Roediger, H.L., Tekin, E., 2020. Recognition memory: Tulving's contributions and some new findings. Neuropsychologia 139, 107350.

Roediger, H.L., Tekin, E., 2021. Can signal detection theory explain everyday amnesia (high confident misses)? Neuropsychologia.

Selmeczy, D., Dobbins, I.G., 2014. Relating the content and confidence of recognition judgments. J. Exp. Psychol. Learn. Mem. Cognit. 40 (1), 66–85.

Selimbeyoglu, A., Kesin-Ergen, Y., Demiralp, T., 2012. What if we are not sure? Electroencephalographic correlates of subjective confidence level about a decision. Clin. Neurophysiol. 123, 1158–1167.

Smith, C.N., Wixted, J.T., Squire, L.R., 2011. The hippocampus supports both recollection and familiarity when memories are strong. J. Neurosci. 31, 15693–15702.

Stretch, V., Wixted, J.T., 1998. Decision rules for recognition memory confidence judgments. J. Exp. Psychol. Learn. Mem. Cognit. 24, 1397–1410.

Tulving, E., 1985. Memory and consciousness. Can. Psychol. 26, 1–12.

Wickens, T.D., Hirshman, E., 2000. False Memories and Statistical Design Theory: Comment on Miller and Wolford (1999) and Roediger and McDermott (1999), vol. 107. Psychological Review, pp. 377–383, 2.

Wixted, J.T., Mickes, L., 2010. A continuous dual-process model of remember/know judgments. Psychol. Rev. 117, 1025–1054.

Wilson, B.M., Wixted, J.T., 2018. The prior odds of testing a true effect in cognitive and social Psychology. Advances in Methods and Practices in Psychological Science 1, 186–197.

Wixted, J.T., Stretch, V., 2000. The case against a criterion-shift account of false memory. Psychol. Rev. 107 (2), 368–376.