CRITICAL INVESTIGATIONS OF THE Z-ROC CURVE: EVIDENCES FOR A SINGLE-PROCESS MODEL OF RECOGNITION MEMORY

Ziv Peremen, Chen Didi-Barnea, Yonatan Goshen-Gottstein
Department of Psychology, Tel-Aviv University, Israel
zperemen@gmail.com, chen.barnea@gmail.com, goshen@post.tau.ac.il

Abstract

A dominant approach to recognition memory is dual-process theory, comprising both signal-detection (familiarity) and high-threshold (recollection) mechanisms. The finding that z-ROC slopes are invariably smaller than unity—establishing larger variance for the hypothetical target distribution than that of the lure distribution—is taken as support for the involvement at retrieval of recollection, in addition to familiarity (thereby increasing the variance of the target distribution). Critically, in a seminal study, Yonelinas et al. (2001) obtained confidence ratings (from which z-ROC slopes could be calculated) and, within each level of confidence, a subjective Remember-Know (R-K) judgment was made. In an attempt to simulate recognition unaffected by recollection, 'Remember' (R) trials were then selectively excluded. Consistent with dual-process theory, the z-ROC slope increased to unity, demonstrating the involvement of only a single process—Familiarity. Here we argue that Yonelinas et al. inadvertently biased participants to respond R only when they were highly confident that the item had been studied and that, in addition, the authors did not consider the possible existence of Recollection-based false alarms. When running the experiment with unbiased instructions, R responses spread across all confidence levels. We next excluded both R and K responses for both FAs and hits. Similar to Yonelinas, we too found an increase in the slope of the z-ROC curve. Our analysis suggests a strength-based interpretation to the data which include not only recollection-based hits, but also recollection-based FAs.

A prominent theory of recognition memory is the dual-process theory. This theory describes recognition memory as based on two processes: Recollection and Familiarity. Recollection involves slower, conscious remembering of contextual detail about a prior learning episode, whereas Familiarity involves a fast-acting, relatively automatic assessment that an item was presented, without having available to consciousness any further information about the learning episode. Dual-process theory suggests that recognition judgments are based on both a high-threshold process (e.g., when qualitative information about the study event exists) and a signal-detection process (e.g., the assessment of familiarity). Recollection is an all-or-none threshold process. In contrast, Familiarity adheres to the assumptions of signal-detection theory including the existence of two equal-variance normal distributions, one representing unstudied, lure items and the other, studied, target items, as well as the use of a subjective decision criterion (Yonelinas, 1994). Much empirical data has supported dual-process theory (Diana, Reder, Arndtt, Park, & Arndt, 2006). Still, several assumptions remain to be proven. One such intrinsic assumption is that because recollection entails remembering of specific episodic details from the encoding episode, therefore, when such details are remembered, they, by definition, stem from an event that occurred. Therefore, there are no recollection-based false alarms.

The current research focuses on one of the strong pieces of evidence for the dual-process model. While many pieces of evidence for the dual-process model have been challenged (Wixted, 2007), one particular finding still stands out in the crowd. It has been found that the z-ROC slope, which is representative of the ratio between the 'old' and 'new' distributions' variance is smaller than 1, indicating a higher variance in the 'old' distribution. Yonelinas (1994) suggested that the unequal variance stems from the involvement of high-threshold process of recollection, predicting that in its absence, the variance between the 'old'

and ‘new’ item distributions will be equal. Yonelinas demonstrated this effect using both amnesic patients (1994) and healthy participants (2001).

To support this idea, Yonelinas combined two common recognition procedures within a single design. After presenting participants with study words, they answered a two-stage test for each presented item. The first test was a confidence test (using six-point scale ranging from 1- sure ‘new’ to 6- sure old) and the second, a Remember-Know judgment task. The first test, based on confidence ratings is often used to obtain Receiver Operating Characteristics (ROC) in the z-space (z-ROC). A confidence level analysis is a parsimonious alternative of calculating different subjects' criterion within a single subject. The 6-levels confidence scale can be used to reflect five different data sets representing five hypothetical criterions; responses provided with high confidence (i.e., 6), can be viewed as representing responses of hypothetical participants with a very strict response bias, who only judge an item to be ‘old’ only if they are extremely confident the item had appeared. Similarly, responses 5 and 6 can be grouped together and viewed as representing hypothetical participants with slightly more lenient response bias. At the other extreme is the grouping of responses 2-6, viewed as simulating the most lenient response bias of all. Thus, from a six-point confidence scale, the responses of five hypothetical participants, each with a different response bias, can be extracted. For each subject’s five possible criterions, hits and false alarms can be tabulated as z-scores and plotted with each criterion’s hits as a function of its false alarms score. Linear regression on these hits and false-alarms intersections will provide the z-ROC function parameters including the z-ROC slope.

Following the confidence rating of each item, a Remember-Know (R-K) judgment was made (Tulving, 1985). In the R-K paradigm, participants were asked to indicate whether they were able to “remember” the prior occurrence of the item or whether they simply “knew” on some other basis that it was old. Studies that use the R-K paradigm frequently do so under the assumption that R and K responses reflect different forms of memory retrieval. According to Yonelinas’ interpretation R and K responses closely reflect the processes of recollection and familiarity, respectively. In support of dual-process theory, Yonelinas found that when calculating the z-ROC with both the R and K trials, the z-ROC slope was ~0.8, whereas when the correct R responses excluded from the z-ROC calculations, the z-ROC slope increased to unity (Yonelinas, 2001). These results clearly support the prediction that once the recollection-based threshold responses were removed, the familiarity-based responses reflect indeed classic equal-variance signal-detection distributions.

However, a closer look at their experiment indicates that almost all R responses were given in the highest confidence level and that there were virtually no recollection-based false alarms (only 2% of the R trials). These findings support Yonelinas’ view of recollection as a high-threshold process leading to high-confidence correct responses. Unlike knowing (K responses), which can sometimes lead to errors, remembering (R responses) indicates retrieving of accurate episodic details leading to correct judgment. Critically, reviewing the instructions given to their participants, reveals that in addition to the standard R-K instructions, participants were encourage to choose R only if “they could, if asked, tell the experimenter what they recollected about that study event”. With this kind of instructions, there is no wonder that participants used R responses only when they were absolutely sure that the word had appeared before, confounding recollection with the strength of memory trail. Not surprisingly, memory trace strength is at the basis of an alternative theory to dual-process model. Single-process theory suggests that recognition is based on a single process of evaluating the strength of the memory signal. Presumably, studying an item temporarily increases the item's strength, which is based on a combination of item's familiarity and recollection (Wixted, 2007). According to this theory, the unequal variance between target

and lure distributions is due to difference in the amount of strength that is added during the study phase to different target items.

To consider the single-process theory interpretation, one should refer to the question whether R responses (reflecting recollection) indeed represent an error-free conscious process, which is qualitatively different from familiarity. Strong indication to the contrary can be found in the high proportion of R responses received for lab-induced false memories generated by the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995). Additional real life support can be demonstrated by the lively and rich recollection memories of people claiming to have unrealistic experiences (e.g., aliens' abduction). Here we argue that R-based false alarms do exist. We therefore reasoned that the remarkable increase of the z-ROC to 1 would remain if we excluded both R hits and R false alarms and if we used instructions that assert that the recognition confidence test and the R-K test are unrelated. Furthermore, we examined whether the increase of z-ROC slope is unique to the exclusion of R responses or whether, as predicted by the single-process theory, any reduction in the target items variance, which can be obtained via the exclusion of the strongest confidence sublevels from the calculation, also leads to an increase in the z-ROC slope.

Methods

Sixteen students (2 males, aged 24±2.3 years old) participated in the experiment for a small fee (30 NIS) or credit in an undergraduate psychology course at the University of Tel Aviv.

The experiment included three blocks. Each block included a study phase and a test phase. During each study phase, 80 single words were presented randomly for 2 seconds each (for a total of 240 words across the experiment). Displays were generated by an Intel Pentium 4 computer attached to a 15-inch CRT monitor, using 1024 × 768 resolution graphics mode. Viewing distance was set at 70 cm. The stimuli were presented in 20-point Arial font. Participants were briefed to repeat vocally the displayed words.

Following each study phase, participants completed a recognition memory test. 80 studied words and 80 ‘new’ words were presented one at a time in random order on the screen (for a total of 480 words throughout the experiment). Participants made recognition judgments on the computer keyboard using a 6-point confidence scale. Participants were instructed to respond ‘6’ if they were sure the item was presented in the study list, ‘5’ if they were less sure, and ‘4’ if very unsure. They were instructed to respond ‘1’ if they were sure the word was not presented earlier, ‘2’ if they were less sure and ‘3’ if they were very unsure.

Within each level of ‘old’ item judgments (4-6), a standard Remember-Know (R-K) judgment was presented. Participants made R-K judgments on the computer keyboard using a left/right arrow keys. Participants were instructed to respond "remember" (R) if they could recollect specific details regarding the item's occurrence during the study and to respond "know" (K) if they recognized the item was in the study list on the basis that it was familiar, but it did not evoke specific recollection. The instructions included a comment that the responses to the tests need not be related to their rating of confidence.

Four-hundred and eighty words with similar frequency were used to create six lists of 80 words each. All words in the lists were between three and four letters in length. Both the lists order and the study-test lists were counterbalanced between participants, such that each list served in each condition for an equal number of times. Participants were tested individually, and each session lasted approximately 45 min.

Results and Discussion

One participant was dropped from the analysis hereby since its z-ROC slope was more than two standard deviations lower than average slope.

Remember responses

As predicted, R responses were spread out throughout the range of confidence levels.

R correct responses (i.e., hits) reflected 25%, 37% and 85% of the correct responses at the 4, 5, and 6 levels of confidences, respectively. Similarly, R incorrect responses (i.e. false alarms) reached 25%, 30% and 80% of the total incorrect responses with 4, 5, and 6 levels of confidences, respectively. Total percentage of false alarms out of total responses increased as the confidence level decreased, so it encompassed 7% for the 6^{th} confidence level, 24% for the 5^{th} confidence level and 45% for the 4^{th} level.

z-ROC Analysis

The hits and false alarms in each of the confidence levels for each of the participants were summed and plotted as z-ROC curves. A standard linear regression was conducted to determine the slope of each of the z-ROC plots. The average slope value of all participants was 0.73 with SDT of 0.07 which was significantly less than 1 ($t=-14.7$, $p<0.0001$).

Remember-Know and the z-ROC Analysis

When we excluded the total Remember responses (hits and false alarms), which reflected 21% of the trials, the z-ROC slope of the accumulative data has increased to 0.89 with SDT of 0.15, which was significantly less than 1 ($t_{(15)}= 2.98$, $p=0.007$).

In order to assess the impact of strength we compared exclusion of the three strongest sublevels defined as 6-Remember, 6-Know, and 5-Remember. Exclusion of these sublevels resulted in z-ROC slope of 0.98 with SDT of 0.12, which was not significantly different than 1 ($t_{(15)}= 0.59$, p=ns.).

Additionally, more challenging examination reducing 6-Remember, 6-Know, and 5-Know produce z-ROC slope of 0.99 with SDT of 0.13, which was not significantly different than 1 ($t=0.37$, p=ns.).

Conclusions

In this study we demonstrated that the exclusion of sublevels with high confidence from the z-ROC calculation leads to an increase in the z-ROC slope to 1, which is similar to the results reached by Yonelinas when correct R sublevels were excluded. These results, together with the spreading of R responses across the confidence levels, bring out the possibility that the subjective R-K judgment, if administered at a second stage following confidence rating, might reflect a second-order strength test. Therefore, participants seem to report the strength-of-evidence within the chosen confidence level instead of indicating a qualitatively different mental processing. These findings are best interpreted as supporting a single-process theory.

References

Diana, R., Reder, L. M. L. M., Arndtt, J., Park, H., & Arndt, J. (2006). Models of recognition: A review of arguments in favor of a dual- process account. *Psychological Bulletin Review*, *13*(1), 1-21. doi:10.1016/j.bbi.2008.05.010

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803. American Psychological Association.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, *26*(1), 1-12. doi:10.1037/h0080017

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological review*, *114*(1), 152-76. doi:10.1037/0033-295X.114.1.152

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*(6), 1341-54. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7983467

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, *130*(3), 361-379. doi:10.1037//0096-3445.130.3.361