Commentary

ROCs in rats? Response to Wixted and Squire

Howard Eichenbaum,[1,3] Magdalena M. Sauvage,[1] Norbert J. Fortin,[1] and Andy P. Yonelinas[2]

[1]*Center for Memory and Brain, Boston University, Boston, Massachusetts 02215, USA;* [2]*Center for Neuroscience, University of California, Davis, California 95616, USA*

The article, "Constructing receiver operating characteristics (ROCs) with experimental animals: Cautionary Notes" offered by Wixted and Squire (2008, this issue) attempts to dismiss our observation of linear ROC functions in rats performing recognition tasks in two recent reports (Fortin et al. 2004; Sauvage et al. 2008). The motive for their effort is that the model of recognition memory they choose to adopt (the single-process model of recognition memory known as the Unequal Variance Signal Detection Model) cannot account for a linear ROC, so they are convinced that there must be a violation of key assumptions in signal detection analysis methodology that artificially causes the functions to be linear. Of note, our approach has been different; we examined our data using both the single- and dual-process models and found that the dual-process model provided a better fit in both studies.

Wixted and Squire begin with the assertion that linear ROC functions are almost never observed in humans. However, their analysis focuses on the linearity of the probability ROCs, which rests only on the failure to find statistically significant curvilinearity. More compelling are analyses that calculate performance in z-scores, which reveal a statistically significant U-shape curve in z-space that provides direct evidence for ROC linearity. On those grounds, our ROCs are in perfect agreement with the vast majority of the findings from associative recognition tests (59 conditions from 17 different studies) (Parks and Yonelinas 2007), wherein a large majority showed the U-shaped zROCs as seen in our studies. U-shaped zROC functions were observed across a variety of different associative recognition tasks, including tests of word pairs, memory for location, list membership, and more.

Also, Wixted and Squire emphasize that our use of differential reward payoffs to manipulate response biases is unique and suspect. In contrast to this claim, the manipulation of payoff ratios is considered a valid way to obtain a range of response criteria to determine ROCs, and this method has been used since the beginning of signal detection analyses. The use of payoff ratios was initially used in signal detection analyses on visual detection performance (Tanner and Swets 1954) and produced curvilinear ROCs. In their comprehensive review of the literature, Macmillan and Creelman (1991) conclude: "Payoffs and verbal instructions, although more time consuming than ratings provide one important advantage: The data points are statistically independent, as they are not in a rating-experiment ROC. If the aim of an experiment is to evaluate theoretical assumptions (e.g., that the ROC is regular), then a separate-session procedure may be indicated. For most practical applications, however, the rating task is recommended because of its efficiency." Therefore, while confidence ratings are more convenient for studies on humans, the use of payoff ratios is considered valid and does not necessarily produce linear ROCs in humans. Indeed, in our particular application of payoff ratios, we observed curvilinear ROCs under some memory demands and linear ROCs under others, using identical payoff ratio conditions for all conditions (see below). Therefore, while we appreciate Wixted and Squire's suggestion of a different method that has been used to manipulate biases, the literature provides many precedents for our use of payoff ratios as equally valid.

We will not further argue these general issues here, except to highlight that Wixted and Squire's interpretation of the literature on the shape of ROC functions and on the use of payoff ratios is far from consensual. Rather, we will focus on their three main criticisms: that a valid ROC requires equal accuracy across bias levels, a "differential outcomes effect" explains the observations of linear ROCs, and other aspects of our protocol might force the ROC curve to be linear.

Equal accuracy

Wixted and Squire expressed concern that ROC data are valid "only if accuracy (measured, for example, as d′) remains constant across all biasing conditions." However, d′ cannot be calculated for each bias condition separately without a theory-driven model. Notably, Wixted and Squire only consider the d′ measure from the view of their single-process model, and the requirement applies only within that model, i.e., there is a key circularity in their logic. Nevertheless, to address Wixted and Squire's concern, when we calculated d′ according to their model, we found that the variations in accuracy across levels were not significantly different between our ROCs and theirs (for each ROC, we calculated the overall d′, the d′ for each of the five biases, and five "variation scores," each representing the difference between the d′ of one bias and the overall d′). We used unpaired *t*-tests to compare the variation scores of each of our ROCs in both the Fortin et al. (2004) and the Sauvage et al. (2008) studies to those of their 1-h control data reported in Wais et al. ([2006], their Fig. 3, top, left; all *P*-values >0.05). So, our data are as valid as those reported by Wixted and Squire, according to their own measure of accuracy.

An alternative approach to evaluate accuracy involves the most commonly used accuracy score—percent correct $[P(\text{hit}) + (1 - P(\text{FA}))]/2 \times 100$—which can be readily calculated for each bias. Notably, ROC functions in both humans and rats are typically asymmetrical, characterized by an upward shift on the left side of the function, so percent correct necessarily varies across levels, and this is the case both for curvilinear and linear ROCs and both for humans and rats. For example, in a recent study by Wixted and coworkers (Wais et al. 2006), accuracy across levels for the asymmetrical ROC observed for young human subjects at 1-h delay varied from 69% to 85%, a 16-point range. By this standard, all of our data, including those derived from linear as well as curvilinear ROCs, involved smaller ranges of scores (8%–13%) and, therefore, are at least as valid as their own. Measuring accuracy this way, equal accuracy is obviously not required for construction or interpretation of ROC functions. Importantly, Wixted and Squire can hardly complain that percent correct is an invalid measure of accuracy, because they relied

[3]Corresponding author.
E-mail hbe@bu.edu; fax (617) 358-3296.
Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.1133808.
Freely available online through the *Learning & Memory* open access option.

on this accuracy measure in describing the differential outcomes effect, discussed next.

The differential outcomes effect

The differential outcomes effect is NOT relevant to our protocol. In every study, including all three that are cited by Wixted and Squire (Trapold 1970; Santi and Roberts 1985; DeMarse and Urcuioli 1993), the differential outcomes (rewards) occur following a correct response to a stimulus or stimulus pairing, and learning of a set of stimulus–response–reward combinations is measured by the increase in performance accuracy (always measured as percent correct) over the course of many repetitions of the same combinations. In other words, this effect involves learning to associate specific stimuli and responses with different rewards, and the effect emerges only after hundreds of repetitions of these specific stimulus–response–reward combinations. This is not at all similar to our experimental design, in which each stimulus is presented only once and rewarded equally prior to a single memory test. It is important to understand the details of the behavioral protocol in each case of the differential outcomes effect and of our studies. In Trapold (1970), rats hear one of two sounds, and then must press the appropriate one of two bars to be rewarded—differentially or not differentially—for each Stimulus–Response combination. They learn faster, after a few hundred repetitions of the S–R–reward combinations, under differential rewards. In Santi and Roberts (1985), pigeons are presented with a red or green sample key, and then must peck the matching color key in a test to obtain different or the same rewards for each color. They are trained with the differential stimulus–match–reward contingencies for hundreds of trials. Subsequently, their accuracy over hundreds of additional repetitions of those stimulus–match–reward combinations is superior with different rewards. In DeMarse and Urcuioli (1993), pigeons learn a set of two-stimulus sequences associated with differential or the same reward outcomes. For each sequence, they first peck at one of two first visual patterns, then must choose the assigned second visual pattern to obtain reward, wherein different stimulus pairings are rewarded differently or the same. They learn these $stimulus_1$–$stimulus_2$–response–reward combinations more rapidly with differential rewards, albeit only after hundreds of repetitions on all of the combinations.

In contrast, in our protocol each sample stimulus is presented just once and responding to that stimulus always produces the *identical* reward. Then, recognition memory is assessed by a *single* go or no-go response to each new and old stimulus, respectively. Only AFTER the recognition response is made can the subject obtain differential rewards whose values vary across bias levels. The effect of differential reward for the response to a particular test stimulus is not subsequently measured, so there is no opportunity for a differential outcomes effect. Wixted and Squire are mistaken in their application of this effect, and this erroneous claim is most egregious in their interpretation of the DeMarse and Urcuioli (1993) study, where a superficial reading can be deceptive. They conclude that, "... even when the rewards are not predictable during sample presentation" the differential outcomes effect is observed. What Wixted and Squire do not mention is that the "sample" in the DeMarse and Urcuioli (1993) study is just the first element of each stimulus pair, the response to the entire correct pairing is associated consistently with a differential reward, and the benefit of differential rewards emerges only after hundreds of repetitions of these specific stimulus–response–reward combinations.

Furthermore, it turns out that the predictions of the differential outcomes effect are *opposite* to the pattern of findings in our experiments. As mentioned earlier, accuracy as measured by d′ does not vary significantly across the biases in our studies, providing strong support for the validity of our approach in producing ROC curve. However, when accuracy is measured using percent correct, as in all of the differential outcome effect studies mentioned above, it is clear that animals actually perform less accurately with the most differential amounts of rewards (bias level 1, right side of curve, see Fig 1) compared with the same amount (bias 5, left side) for go and no-go responses. This pattern corresponds to the typical asymmetry in both linear and curvilinear ROC functions and is opposite to the pattern seen under the differential rewards effect. Clearly, the differential outcomes effect could not work and is not at work in our studies.

Other unusual task parameters that might affect accuracy

Does any conceivable aspect of our experiments, such as reward contingencies, differences in the total amount of reward that can be obtained or in response effort across bias levels, or other variables unique to our behavioral protocol, drive our ROC functions to be linear? No. Wixted and Squire neglected to mention that while using the identical reward and effort contingencies, we observed either a linear or a curvilinear ROC in Controls by simply varying the memory demands. Furthermore, these data provide a compelling test of what factors in our protocol determine whether the ROC is curvilinear or linear. Combining the data across our two studies, the key comparisons involve the data from Control rats tested on item recognition with 30 and 75 min memory delays in the Fortin et al. (2004) study and the Control rats tested on associative recognition in the Sauvage et al. (2008) study. Other than the differences in delays (30 or 75 min) and in the stimuli (items or pairs) just mentioned, we used the identical protocol, including all of the reward contingencies and differential effort, under all conditions, in both experiments. In the item recognition task, Control rats had a curvilinear ROC (Fig. 1A). In contrast, in the associative recognition study, Control rats had a linear ROC (Fig. 1B). Because the reward contingencies, as well as all other aspects of the protocol, were the same for both studies, we can conclude that variations in the reward contingencies did not cause the ROC function to change from curvilinear to linear. Instead, a curvilinear ROC resulted when the stimuli were single items, whereas a linear ROC resulted when the stimuli were item-medium pairs, consistent with predictions of the dual-process model. Another useful comparison involves the performance of the same Control rats in the Fortin et al. (2004) study tested at 30 min versus 75 min memory delays using identical protocols, including the reward contingencies. Whereas a curvilinear ROC was observed with a 30-min delay (Fig. 1A), a linear ROC was obtained with a 75-min delay (Fig. 1C). It was the delay and not any variations in the reward contingencies or other task variables that determined the shape of the ROC function. These comparisons provide compelling evidence that the ROCs can be linear or curvilinear, depending on the memory demands, under identical reward contingencies and other task parameters. Importantly, our observations of linear ROCs in some of these conditions are similar to observations in humans where, contrary to Wixted and Squire' claims, linear ROCs have been observed in associative recognition (see above) and with elongated memory delays (Hockley 1992; Yonelinas and Levy 2002).

In conclusion, Wixted and Squire make a flawed attempt to dismiss our findings because they strongly challenge the single-process theory of recognition memory. Linear ROC functions are common in particular types of recognition performance and differential reinforcement is a legitimate procedure for directly manipulating response biases. We encourage readers to devote the significant effort required to carefully examine the literature on the differential outcomes effect, because they will find that the

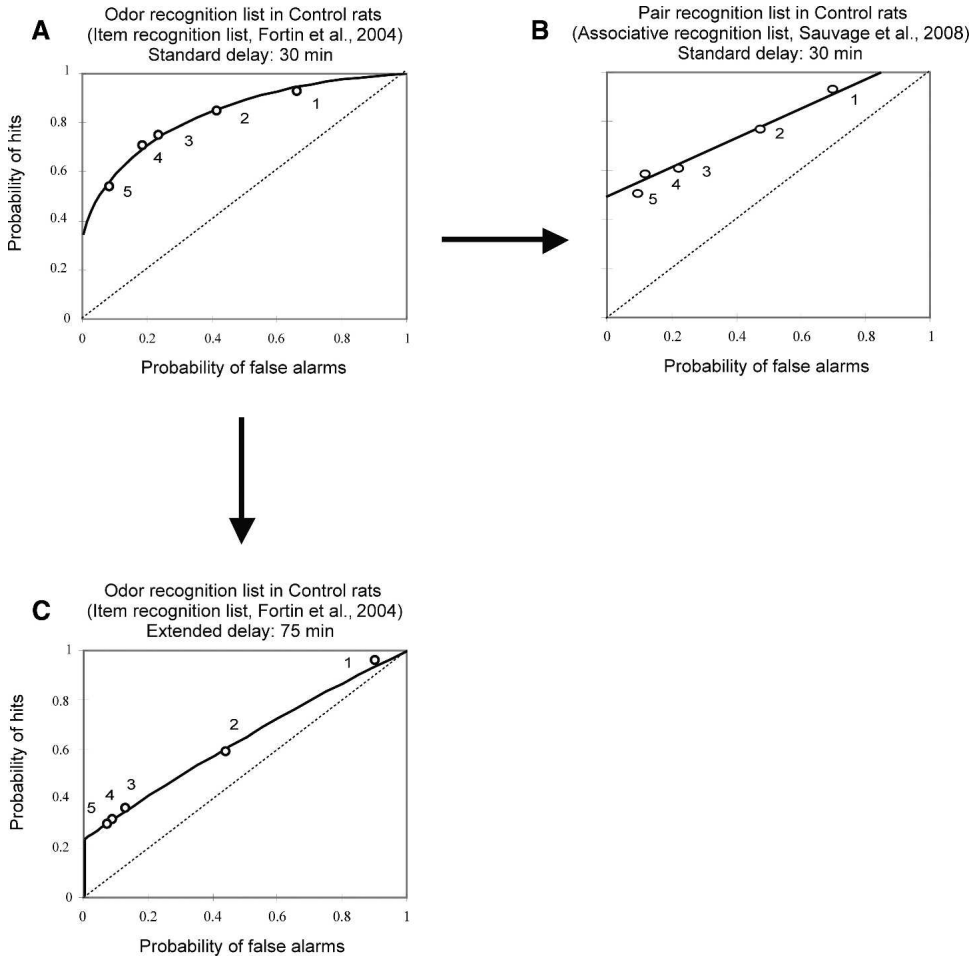

Figure 1. Two manipulations of memory demands that make a curvilinear ROC become more linear. (*A*) Curvilinear ROC observed in item recognition. (*B*) Linear ROC observed for associative recognition. (*C*) Linear ROC observed in item recognition with increased memory delay. Note that the reward/effort contingencies and all other aspects of the behavioral protocol are identical across these conditions. Numbers in ROC functions refer to bias levels (see text).

effect is not relevant to our experiments and it is misleading to claim that it is. Nor do differences in maximum rewards or effort across bias levels, or any other aspect of our protocol, explain why the ROC function of intact rats is linear under some memory demands and curvilinear under others, under the identical reward and effort contingencies. Instead, different memory demands consistent with dual-process theory determine whether the ROC function is linear or curvilinear.

References

DeMarse, T.B. and Urcuioli, P.J. 1993. Enhancement of matching acquisition by differential comparison–outcome associations. *J. Exp. Psychol. Learn. Mem. Cogn.* **19:** 317–326.

Fortin, N.J., Wright, S.P., and Eichenbaum, H. 2004. Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature* **431:** 188–191.

Hockley, W.E. 1992. Item versus associative information: Further comparison of forgetting rates. *J. Exp. Psychol. Learn. Mem. Cogn.* **18:** 1321–1330.

Macmillan, N.A. and Creelman, C.D. 1991. *Detection theory: A users guide*. Cambridge University Press, New York.

Parks, C.M. and Yonelinas, A.P. 2007. Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychol. Rev.* **114:** 188–202.

Santi, A. and Roberts, W.A. 1985. Reinforcement expectancy and trial spacing effects in delayed matching-to-sample by pigeons. *Anim. Learn. Behav.* **13:** 274–284.

Sauvage, M.M., Fortin, N.J., Owens, C.B., Yonelinas, A.P., and Eichenbaum, H. 2008. Recognition memory: Opposite effects of hippocampal damage on recollection and familiarity. *Nat. Neurosci.* **11:** 16–18.

Tanner, W.P. and Swets, J.A. 1954. A decision-making theory of visual detection. *Psychol. Rev.* **61:** 401–409.

Trapold, M.A. 1970. Are expectancies based upon different positive reinforcing events discriminably different? *Learn. Motiv.* **1:** 129–140.

Wais, P.E., Wixted, J.T., Hopkins, R.O., and Squire, L.R. 2006. The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron* **49:** 459–466.

Wixted, J.T. and Squire, L.R. 2008. Constructing receiver operating characteristics (ROCs) with experimental animals: Cautionary notes. *Learn. Mem.* **15:** (this issue).

Yonelinas, A.P. and Levy, B.J. 2002. Dissociating familiarity from recollection in human recognition memory: Different rates of forgetting over short retention intervals. *Psychon. Bull. Rev.* **9:** 575–582.

Received July 2, 2008; accepted in revised form July 2, 2008.