

# LEARNING & MEMORY

## Constructing receiver operating characteristics (ROCs) with experimental animals: Cautionary notes

John T. Wixted and Larry R. Squire

*Learn. Mem.* 2008 15: 687-690

Access the most recent version at doi:[10.1101/lm.1077708](https://doi.org/10.1101/lm.1077708)

---

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

---

---

To subscribe to *Learning & Memory* go to:  
<http://learnmem.cshlp.org/subscriptions/>

---

## Commentary

# Constructing receiver operating characteristics (ROCs) with experimental animals: Cautionary notes

John T. Wixted<sup>1</sup> and Larry R. Squire<sup>2-4</sup>

<sup>1</sup>Department of Psychology, University of California, San Diego, California 92161, USA; <sup>2</sup>Veterans Affairs Medical Center, San Diego, California 92161, USA; <sup>3</sup>Departments of Psychiatry, Neurosciences, and Psychology, University of California, San Diego, California 92161, USA

A recent article reported findings from an associative recognition memory procedure in which the receiver operating characteristic (ROC) obtained for rats with hippocampal lesions was compared with that of controls (Sauvage et al. 2008). An ROC is a plot of the hit rate (percent correct choices of old items) vs. the false alarm rate (percent incorrect choices of new items) across several biasing conditions. The shape of the ROC for control rats was linear, but it was curvilinear for lesioned rats. These results were interpreted in terms of a particular psychological model (the high-threshold/signal-detection model) (Yonelinas 1994) to mean that rats, like humans, recognize stimuli on the basis of two memory processes (recollection and familiarity) and that hippocampal lesions impair one process (recollection) while enhancing the other (familiarity) (Fig. 1).

The procedure used by Sauvage et al. (2008) was similar to an odor-recognition procedure used earlier by the same laboratory (Fortin et al. 2004) to study Old/New recognition, and both studies are noteworthy in that the control rats produced linear ROCs following a long retention interval (30–75 min). Aside from these results, between 1958 and 2007, no other linear ROCs for Old/New recognition and very few linear ROCs for associative recognition have been reported.

The extreme rarity of linear ROCs in the long history of recognition memory research raises the question as to why the odor-recognition procedure yielded that pattern. Since ROC analysis of recognition memory was introduced by Egan (1958), dozens of Old/New recognition ROCs have been collected using a wide variety of stimuli (words, pictures, faces, etc.) and a wide variety of subjects (young, old, amnesic, etc.) and species (rats, pigeons, humans) who were tested under a wide variety of conditions (different retention intervals, different durations of study, different list lengths, etc.). Until Fortin et al. (2004) reported their findings, no other linear Old/New ROC had ever been reported.

Fortin et al. (2004) observed a linear Old/New ROC in their control rats when a long retention interval was used, and they suggested that this result occurred because familiarity faded more rapidly than recollection as the delay between study and test increased. However, every other relevant study that has investigated the effect of retention interval on the shape of the ROC has found that a long retention interval yields a curvilinear function (Donaldson and Murdock Jr. 1968; Gehring et al. 1976; Wixted 1993; Francis and Irwin 1998; Wais et al. 2006; Tunney and Bezzina 2007). In most of these studies, the ROC was symmetrically curvilinear after a long retention interval, which was the pattern that Fortin et al. (2004) observed for their lesioned rats after a short retention interval. Indeed, as a general rule, weak memory conditions are associated with a symmetrical curvilinear ROC regardless of how memory strength is manipulated (Glanzer et al. 1999). It is therefore puzzling that Fortin et al. (2004) instead

obtained a linear ROC when they weakened memory in their control rats using a standard retention interval manipulation.

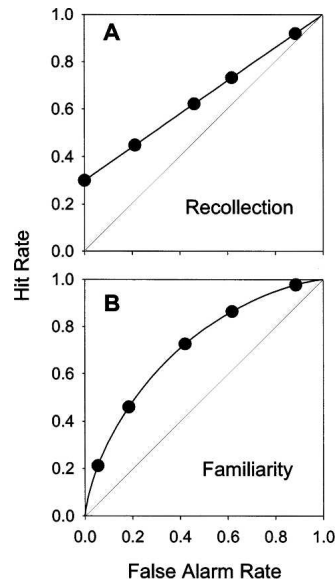
Sauvage et al. (2008) used an associative recognition version of the odor-recognition task, in which rats were presented with a list of odor pairs and were later tested with intact pairs and recombined pairs. Again, the ROC for control rats was linear. A few early reports in the human literature suggested that ROCs based on associative recognition and related tasks might be linear (Yonelinas 1997, 1999; Rotello et al. 2000). However, a later review of 13 associative recognition studies found them to be much better fit by a curvilinear function than by a linear function, and six new associative recognition ROCs (three from younger subjects and three from older subjects) were all found to be curvilinear (Healy et al. 2005). Moreover, the two laboratories that initially reported linear associative recognition ROCs have, since that time, reported 15 new associative recognition ROCs between them, and all have been curvilinear (Verde and Rotello 2004; Quamme et al. 2007). Thus, although the linear associative recognition ROC obtained by Sauvage et al. (2008) is somewhat less exceptional than the linear Old/New ROC obtained by Fortin et al. (2004), it is still an unusual result.

Why did the odor-recognition procedure in rats yield linear ROCs when other conceptually identical procedures almost never do? In the odor-recognition memory procedure, rats were presented with a list of odors and then were given test odors one at a time. Each odor was mixed with sand in a cup. For the test odors, the animal could dig in the test cup to indicate a New decision (and obtain a reward if the choice was correct) or go to a different cup at the back of the cage to indicate an Old decision (and obtain a reward if that decision was correct). To produce the data needed to plot an ROC with this procedure, reward magnitude and reward accessibility were differentially manipulated so as to create different amounts of bias toward either an old or a new decision. Thus, for example, in one biasing condition, correctly reporting an item to be Old yielded three pieces of food in a shallow cup, whereas correctly reporting an item to be New yielded only a quarter of one piece presented at the bottom of a particularly deep cup (thereby biasing the rat to choose Old on both Old and New test trials). This biasing manipulation is novel, and it produces valid ROC data only if accuracy (measured, for example, as  $d'$ ) remains constant across all biasing conditions. Whether this biasing procedure in fact maintains constant accuracy across conditions is not known and has not been tested. If accuracy were instead to vary across biasing conditions, then interpreting the shape of the ROC becomes complicated. That is, a linear ROC would not be indicative of recollection even if the high-threshold/signal-detection model that was used to interpret that result were valid (which is itself a debatable issue) (see Heathcote 2003; Slotnick and Dodson 2005; Wixted 2007). Note that whether accuracy is constant or not across biasing conditions cannot be determined from the data collected at each bias condition without assuming the validity of a particular theory.

<sup>4</sup>Corresponding author.

E-mail [lsquire@ucsd.edu](mailto:lsquire@ucsd.edu); fax (858) 552-7457.

Article is online at <http://www.learnmem.org/cgi/doi/10.1101/lm.1077708>.



**Figure 1.** Predicted shapes of the ROC if responding were based exclusively on recollection (A) or exclusively on familiarity (B) according to the high-threshold/signal-detection model (Yonelinas 1994). The recollection-based ROC is linear, whereas the familiarity-based ROC is symmetrically curvilinear. In contrast, the traditional signal-detection model (Egan 1958) is incompatible with linear ROCs, and the extreme rarity of linear ROCs partly explains why the signal-detection model has long been the dominant model of decision making in recognition memory tasks. The traditional signal-detection model is fully compatible with the idea that two processes (recollection and familiarity) contribute to the recognition decision. However, if the traditional model is correct, then quantitative estimates of recollection and familiarity that are derived from the high-threshold/signal-detection model are not valid. According to the traditional model, separate estimates of recollection and familiarity cannot be derived from an ROC curve. Instead, the ROC yields estimates of strength and variance of the memory signal (with the estimated strength value being a joint function of recollection and familiarity).

The difficulty is that one cannot determine whether performance is being influenced only by memory (modulated by bias) or whether performance is being modulated by a change in motivation, attention, or some other non-memory factor. Independent assessments are needed to determine whether the method used to manipulated bias might also affect accuracy (see below).

The odor-recognition procedure used differential rewards to manipulate bias. The use of differential rewards has long been known to affect accuracy as well as bias under a variety of conditions. One relevant phenomenon is known as the differential outcomes effect (Trapold 1970). In a typical study involving the delayed matching-to-sample task, each sample stimulus is consistently associated with a unique reward outcome (e.g., a correctly remembered red stimulus reliably yields food, whereas a correctly remembered green stimulus reliably yields sucrose). Compared with the nondifferential case (i.e., when correct choices are equally likely to be rewarded with food or sucrose), performance is reliably enhanced. Santi and Roberts (1985) found that this phenomenon can occur even when the differential outcomes consist of unequal probabilities of reward (e.g., a correctly remembered red stimulus yields food with probability 1.0, whereas a correctly remembered green stimulus yields food with probability 0.2). Thus, whenever differential outcomes are used to create an ROC, as they were in the odor-recognition procedure, the possibility that accuracy as well as bias has been manipulated should be considered.

In most studies using differential reward outcomes, the re-

ward characteristics were predictable during the presentation of the sample (e.g., when the green sample appeared, the animal could predict that a correct response at the end of the trial would yield sucrose). In contrast, in the odor-recognition procedure, the differential rewards that were in effect for a particular biasing condition were not apparent to the rat until after the list of odors had been presented. Thus, the traditional differential outcomes effect cannot account for the data obtained in the odor recognition procedure. Nevertheless, DeMarse and Urcuioli (1993) showed that even when the rewards are not predictable during sample presentation, differential outcomes associated with the two response options could facilitate performance on a delayed matching to sample task.

Still other studies show that manipulating the accessibility of reward (e.g., by making the animal work harder to obtain the reward) also affects accuracy (Wilkie and Spetch 1978). In the odor-recognition procedure, the two most liberal biasing conditions linked the New response to deeper cups, which made it more difficult to access the food reward, thereby biasing the animal to make an Old response by choosing not to dig and instead to approach a shallower cup at the back of the cage. However, in addition to affecting bias, this manipulation may have affected accuracy. For example, the presence of a deep cup might reduce the willingness of the rat to carefully sample the test odor and instead to simply search for a more accessible reward in the shallower cup. In the ROC data for the control rats, the two conditions that used deep cups do appear to be associated with less accurate performance in both Fortin et al. (2004) and Sauvage et al. (2008). That is, for those two conditions, performance was somewhat closer to the diagonal line on the ROC (which corresponds to chance accuracy) than was performance in the other three biasing conditions.

Finally, manipulating the overall amount of food reward can affect accuracy as well. For example, using a delayed matching-to-sample procedure, Brown and White (2005) varied the amount of food reward for correct responses within a session. A signal was presented following each sample, which indicated whether the upcoming reward for a correct response would be large or small. When the overall reward amount was large, delayed matching accuracy was enhanced compared with when it was small (which may reflect altered motivation during the test). In the odor-recognition procedure, the amount of food earned for correct responses was considerably higher in some biasing conditions than it was in other biasing conditions.

It is important to emphasize that the odor-recognition procedure differs in many ways from the procedures used to study the differential outcomes effect and the additional effects discussed above. For example, in most of the studies discussed above, a delayed matching-to-sample procedure was used in which only a few stimuli were used repeatedly across many trials and many sessions. In the odor-recognition procedure, by contrast, memory was tested for a novel list of 10 odors in each session (Fortin et al. 2004; Sauvage et al. 2008). Thus, our point is not that the past literature directly indicates what aspect of the odor-recognition procedure might have influenced accuracy. Instead, our point is that a number of variables that were not held constant across biasing conditions in the odor-recognition task have reliably influenced accuracy in previous studies of animal memory. As such, it does not seem safe to assume that no such effect occurred in this novel procedure, particularly since it yielded a very rare result (linear ROCs). It is also worth noting that similar concerns would also apply to humans if bias were manipulated to generate ROC data. That is, it would be important to demonstrate that accuracy is constant across biasing conditions.

In both Fortin et al. (2004) and Sauvage et al. (2008), the control rats produced linear ROCs, but the rats with hippocampal lesions did not. When interpreted in terms of the high-threshold/signal-detection model (Yonelinas 1994), this pattern is consistent with the notion that hippocampal lesions selectively impaired recollection. However, before drawing strong theoretical conclusions about the neuroanatomy of recollection and familiarity on the basis of a procedure that regularly yields what is otherwise an extremely rare result (namely, a linear ROC), the possibility should be investigated that variables other than recollection and familiarity influenced performance. This point was recently underscored by an ROC analysis of a patient with bilateral hippocampal lesions (Bird et al. 2008). When memory for pictures was tested, this patient (Patient Jon) produced the first linear Old/New ROC known to us in the extensive human literature on ROC analysis. According to the model used by Fortin et al. (2004) and Sauvage et al. (2008) to interpret their ROC data, Patient Jon's ROC indicates purely recollection-based responding. That is, the results would suggest that hippocampal lesions selectively impair familiarity while preserving recollection (exactly the opposite of the conclusion reached by Fortin and colleagues and Sauvage and colleagues). An alternative and more likely possibility is that the extremely rare linear Old/New ROC was an artifact. Indeed, Bird et al. (2008) suggest that Patient Jon's linear ROC probably reflects some idiosyncratic strategy for coping with weak memory. In a similar vein, we suggest that the rare linear ROC produced by the odor-recognition procedure may reflect an artifact as well. More specifically, it may reflect the influence of uncontrolled variables, which are known to affect accuracy on recognition memory procedures, and that might have exerted similar effects in the odor-recognition task.

Instead of taking the results reported by Fortin et al. (2004) and Sauvage et al. (2008) as evidence that the hippocampus selectively supports recollection, the novel biasing procedure that they used to produce their unusual ROC data should be further scrutinized. At a minimum, ROC data should be generated by an alternate method that does not involve differential difficulty, and differential amounts of reward. For example, one could use identical reward outcomes for correct Old and New decisions, but manipulate bias by varying signal-presentation probability across test conditions (McCarthy and Davison 1979). In this procedure, each biasing condition involves a disproportionate percentage of New and Old stimuli. To induce a liberal bias (i.e., a disproportionate tendency to choose Old), 80% of the test trials might involve Old stimuli, whereas only 20% would involve New stimuli. To induce a conservative bias, the opposite would be true. If this procedure produced the same results, it would suggest that the linear ROCs observed on the odor-recognition task may reflect memory processes (though it would still not explain why rats yield this result when humans almost never do).

Another approach to this issue would be to test directly the effects on accuracy of differential reward magnitudes and different cup sizes, all within the context of the odor-recognition procedure that has already been used (Fortin et al. 2004; Sauvage et al. 2008). For example, in one condition, the food cups used to reward both correct and incorrect responses should both be deep, and in another condition, they should both be shallow. If accuracy were higher in the latter condition, this result would indicate that the size of the food cups should be held constant across biasing conditions when trying to construct ROCs. In addition, overall food amounts for correct and incorrect responses could be varied across conditions. In one condition, both rewards should be small, and in another condition, both rewards should be large. Again, if accuracy were higher when the reward

amounts were higher, then this result would indicate that the overall reward for correct and incorrect responses should be held constant across biasing conditions when trying to construct ROCs. If none of these variables affected accuracy on the odor-recognition task, then this outcome would suggest that the linear ROCs reported by Fortin et al. (2004) and Sauvage et al. (2008) may indeed reflect recognition memory processes (even though humans almost never exhibit the same result). Until such issues are carefully investigated, the idea that the ROCs produced by the odor-recognition procedure imply that hippocampal lesions selectively impair recollection should be viewed with caution.

## References

- Bird, C.M., Vargha-Khadema, F., and Burgess, N. 2008. Impaired memory for scenes but not faces in developmental hippocampal amnesia: A case study. *Neuropsychologia* **46**: 1050–1059.
- Brown, G.S. and White, K.G. 2005. On the effects of signaling reinforcer probability and magnitude in delayed matching to sample. *J. Exp. Anal. Behav.* **83**: 119–128.
- Egan, J.P. 1958. *Recognition memory and the operating characteristic* (Tech. Note AFCRC-TN-58-51). Hearing and Communication Laboratory, Indiana University, Bloomington, IN.
- DeMarse, T.B. and Urciuoli, P.J. 1993. Enhancement of matching acquisition by differential comparison-outcome associations. *J. Exp. Psychol. Learn. Mem. Cogn.* **19**: 317–326.
- Donaldson, W. and Murdock Jr., B.B. 1968. Criterion change in continuous recognition memory. *J. Exp. Psychol.* **76**: 325–330.
- 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.
- Francis, M.A. and Irwin, R.J. 1998. Stability of memory for colour in context. *Memory* **6**: 609–621.
- Gehring, R.E., Togli, M.P., and Kimble, G.A. 1976. Recognition memory for words and pictures at short and long retention intervals. *Mem. Cogn.* **4**: 256–260.
- Glanzer, M., Kim, K., Hilford, A., and Adams, J.K. 1999. Slope of the receiver-operating characteristic in recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **25**: 500–513.
- Healy, M.R., Light, L.L., and Chung, C. 2005. Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *J. Exp. Psychol. Learn. Mem. Cogn.* **3**: 768–788.
- Heathcote, A. 2003. Item recognition memory and the ROC. *J. Exp. Psychol. Learn. Mem. Cogn.* **29**: 1210–1230.
- McCarthy, D. and Davison, M. 1979. Signal probability, reinforcement, and signal detection. *J. Exp. Anal. Behav.* **32**: 373–386.
- Quamme, J.R., Yonelinas, A.P., and Norman, K.A. 2007. Effect of unitization on associative recognition in amnesia. *Hippocampus* **17**: 192–200.
- Rotello, C.M., Macmillan, N.A., and Van Tassel, G. 2000. Recall-to-reject in recognition: Evidence from ROC curves. *J. Mem. Lang.* **43**: 67–88.
- 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.
- Slotnick, S.D. and Dodson, C.S. 2005. Support for a continuous (single process) model of recognition memory and source memory. *Mem. Cognit.* **33**: 151–170.
- Trapold, M.A. 1970. Are expectancies based upon different positive reinforcing events discriminably different? *Learn. Motiv.* **1**: 129–140.
- Tunney, R.J. and Bezzina, G. 2007. Effects of retention intervals on receiver operating characteristics in artificial grammar learning. *Acta Psychol. (Amst.)* **125**: 37–50.
- Verde, M.F. and Rotello, C.M. 2004. Strong memories obscure weak memories in associative recognition. *Psychon. Bull. Rev.* **11**: 1062–1066.
- Wais, P., 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**: 459–458.
- Wilkie, D.M. and Spetch, M.L. 1978. The effect of sample and

- comparison ratio schedules on delayed matching to sample in the pigeon. *Anim. Learn. Behav.* **6**: 273–278.
- Wixted, J.T. 1993. A signal detection analysis of memory for nonoccurrence in pigeons. *J. Exp. Psychol. Anim. Behav. Proc.* **19**: 400–411.
- Wixted, J.T. 2007. Dual-process theory and signal-detection theory of recognition memory. *Psychol. Rev.* **114**: 152–176.
- Yonelinas, A.P. 1994. Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *J. Exp. Psychol. Learn. Mem. Cogn.* **20**: 1341–1354.
- Yonelinas, A.P. 1997. Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Mem. Cogn.* **25**: 747–763.
- Yonelinas, A.P. 1999. The contribution of recollection and familiarity to recognition and source memory: An analysis of receiver operating characteristics and a formal model. *J. Exp. Psychol. Learn. Mem. Cogn.* **25**: 1415–1434.

Received May 21, 2008; accepted in revised form July 2, 2008.