Journal of Mathematical Psychology 55 (2011) 251–266

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Evaluating models of recognition memory using first- and second-choice responses

David Kellen *, Karl Christoph Klauer

Albert-Ludwigs-Universität Freiburg, Germany

ARTICLE INFO

Article history:
Received 6 August 2010
Received in revised form
8 November 2010
Available online 5 January 2011

Keywords:
Recognition memory
Second-choice responses
Signal detection theory
Discrete-state models
Hybrid models

ABSTRACT

Swets, Tanner Jr., and Birdsall (1961) proposed a 4-alternative forced-choice task with two choices (4AFC-2R) for distinguishing between the Equal-Variance Signal Detection model and the One-High Threshold model. This task was recently implemented in the field of recognition memory (Parks & Yonelinas, 2009), a field in which several candidate models have been proposed. One advantage of the 4AFC-2R task is that it permits parameter estimation and goodness of fit testing, something which so far was only possible through the use of Receiver Operating Characteristic (ROC) functions for the more complex candidate models. The present article provides a thorough characterization and comparison of the main recognition memory models in the context of this task. Results are illustrated by a reanalysis of Parks and Yonelinas' original data, revealing a preference for hybrid approaches to recognition memory, more specifically for the dual-process model (Yonelinas, 1997), whereas pure signal detection models performed poorly. The present analysis provides an assessment of the merits and limitations of this task, highlighting future research applications.

The ability to recognize previously acquired information is one of the aspects of human memory that has been extensively studied by means of mathematical models (for a review, see Malmberg, 2008). In the past decades several models have been proposed and discarded, each with distinct assumptions and predictions: Some approaches assume that mnesic information is adequately represented as an all-or-none process (e.g., Batchelder & Riefer, 1990; Klauer & Kellen, 2010), while others postulate a continuous representation of mnesic evidence (e.g., Wixted, 2007) or a combination of both possibilities (e.g., DeCarlo, 2002; Yonelinas, 1997). The major models, represented in Fig. 1, can thus be divided into three classes: discrete-state, continuous, and hybrid models.

The Signal Detection model (SDT; Macmillan & Creelman, 2005) assumes that a single, continuous mnesic process, termed familiarity, describes the individuals' decisions based on mnesic information. Given that all items – studied and non-studied – have some degree of familiarity, it is possible to represent them by their respective familiarity distributions, and the ability to discriminate studied items from distractors is defined by the distance between the two distributions, in this case represented by parameter μ. In this framework, an item is declared old or new in a recognition task by comparing its familiarity with an established response criterion, denoted by c. If the item's familiarity value is higher than the response criterion, the item is declared as old, otherwise it is rejected. Within the SDT framework one can distinguish two versions of the model; the Equal-Variance Signal Detection model (EVSDT), which assumes that signal and noise distributions have the same standard deviation, and the Unequal-Variance Signal Detection model (UVSDT), which is a more general version that allows for the signal distribution standard deviation, indexed by σ, to assume a different value from the distractor distribution standard deviation.

The One-High Threshold model (1HTM Blackwell, 1963) assumes that decisions based on mnesic information can be described by two discrete states, a "remember" state and a "guessing" state. Remembering an item during test is a probabilistic event given that only a portion of the studied items will surpass a specific memory threshold, the portion being defined by parameter D_o. The items that are not remembered (which occurs for all distractors) trigger a guessing process defined by parameter g. An extension of this discrete-state approach is the Two-High Threshold model (2HTM Snodgrass & Corwin, 1988) that includes an additional state of "distractor detection" indexed by parameter D_n. This parameter attempts to describe the active rejection of distractors, a phenomenon usually associated with a host of metacognitive strategies (e.g., Strack & Bless, 1994).

The Dual-Process Model (DPSDT Mandler, 1980; Yonelinas, 1997) is a hybrid approach that combines aspects of the previous two model classes, adopting a continuous familiarity process (equivalent to EVSDT) to describe more vague mnesic states, and an additional threshold component that describes episodic remembrance, termed recollection and defined by parameter R. A second hybrid approach is the Finite Mixture Signal Detection Model (MSDT DeCarlo, 2002, 2010), according to which the signal

* Corresponding address: Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, D-79085 Freiburg, Germany.
E-mail address: david.kellen@psychologie.uni-freiburg.de (D. Kellen).

0022-2496/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2010.11.004

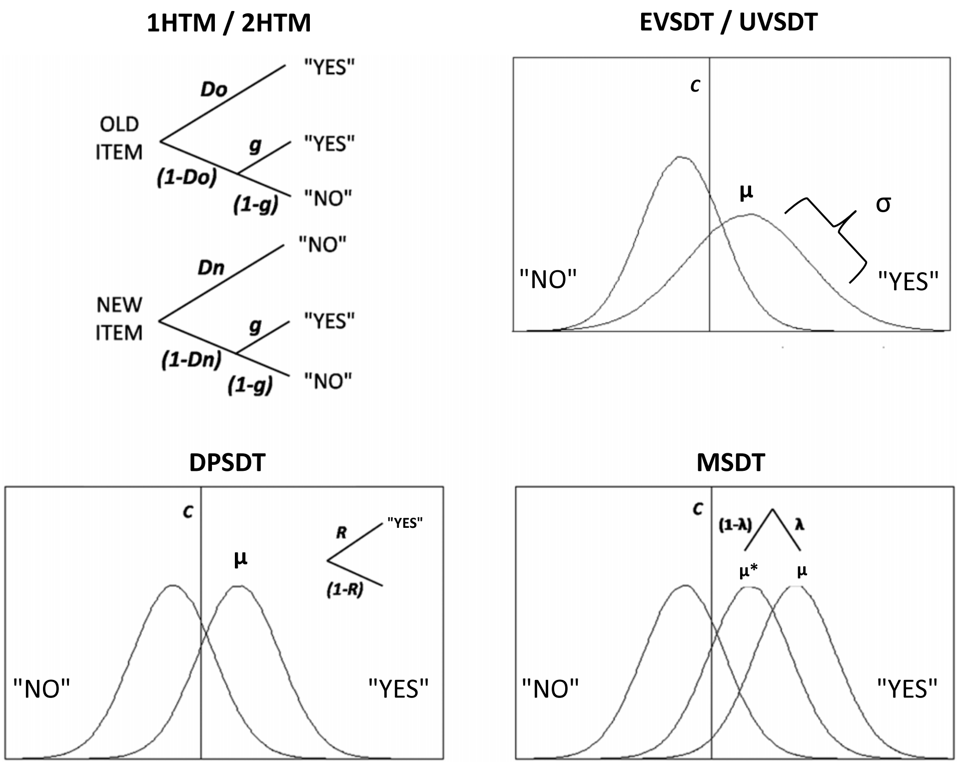

Fig. 1. Representation of the main recognition memory models.

distribution is comprised of a mixture of two latent equal-variance normal distributions—one corresponding to items that were attended during study, with mean μ, and a second distribution for unattended items with mean μ^* and with $\mu^* \leq \mu$. The proportion of attended items among studied items is defined by parameter λ.

Traditionally, recognition memory models are tested by means of their predictions regarding Receiver Operating Characteristic (ROC) functions. ROCs map hits and false alarm probabilities across several levels of response bias. ROCs are informative given that the shape of the expected function varies greatly depending on the assumed processes and underlying probability distributions (see Swets, 1986). For instance, both discrete-state models predict linear ROCs, while SDT predicts curvilinear ones. Both hybrid models can produce intermediate shapes as well as more complex ones (e.g., DeCarlo, 2002).

Although ROC functions are defined as functions obtained through the use of direct bias manipulations, either in terms of changes in item base rates or in the payoff matrix (see Van Zandt, 2000), they are almost always obtained through the use of confidence-rating scales, which are then compiled in order to emulate a bias manipulation (for exceptions, see the review by Bröder & Schütz, 2009). Despite their widespread use in the literature, ROC functions based on confidence ratings are in fact a relatively non-informative method for testing models and theories, given that the candidate models are often indistinguishable by means of the function's shape, that is, the models make similar predictions when extended to the confidence-rating response format appropriately (Bröder & Schütz, 2009; Erdfelder & Buchner, 1998; Klauer & Kellen, 2010; Krantz, 1969; Lockhart & Murdock, 1970; Malmberg, 2002). In order to overcome such limitations, Bröder and Schütz (2009) collected ROCs by manipulating test item base rates, and obtained linear-shaped functions, in contrast to the almost ubiquitous nonlinear shape that has been observed in the past decades and that led to the (possibly premature) dismissal of pure discrete-state approaches (Wixted, 2007). These results make it desirable to explore alternative methods to distinguish between these models. Another advantage of having alternative methods is to avoid what has been called mono-operation bias (Shadish, Cook, & Campbell, 2002, Chap. 3), a bias that is incurred when conclusions rely excessively on a single method, in this case confidence-rating ROCs (see Ratcliff & Starns, 2009).

1. Beyond the ROC: model selection by means of first- and second-choice accuracies

Alternative methods were already considered in the early developments of SDT (Green & Swets, 1966); one such method is based on the predictions that models make for first- and second-choice responses in multiple alternative forced-choice paradigms (Swets et al., 1961). As previously described, according to SDT, responses are based on the test item familiarity or evidence value and the comparison of this value with a previously established response criterion. In the case of multiple alternatives, responses are not based on the comparison of a stimulus with an established response criterion, but on the direct comparison of the alternatives presented at test and their familiarity values. The assumption is that the alternative with the highest evidence value is chosen, and this difference in the nature of the comparison eliminates the need for parameter c. Given that responses are based on these comparisons, individuals order their preferences according to the associated evidence values. If performance is above chance level (that is, on average, correct alternatives have a higher evidence value than incorrect alternatives), two simple predictions result for EVSDT: first, the accuracy of second-choice responses, after an incorrect first choice, must be above chance level. Second, the accuracy of first and second choices should be correlated (see section "Models for First- and Second-Choice Accuracy in 4AFC-2R" for formal derivations of those predictions).

These predictions contrast with the ones that were made by 1HTM; according to this model, on each trial, participants either detect the correct alternative, or else they simply guess. Thus, incorrect first choices occur because the correct alternative did not