

## *Summary*

# Criterion Setting in Recognition Memory and The Manipulations of Response Bias

Cavit Deniz Pala

Middle East Technical University

Aslı Kılıç

Middle East Technical University

Recognition memory is the ability to distinguish whether a stimulus (person, object or situation) is previously experienced. There are many variants of recognition tasks; namely, yes-no recognition, confidence judgement, forced choice, associative recognition, and continuous recognition. Correct recognition of an old item is called a “hit”, when rejecting an old item is a “miss”. “False alarm” is endorsement of a new item, and, finally, rejecting a new item is called “correct rejection”.

### Signal Detection Theory and The Criterion

According to Signal Detection Theory (SDT; Banks, 1970; Green and Swets, 1966; Macmillan and Creelman, 2005; Swets, Tanner and Birdsall, 1961), memory strength, or familiarity, is calculated for each probe item. Old and new items form two overlapping distributions on memory strength scale. Recognition decision is made by using a decision criterion to which memory strength of test items is compared. Sensitivity and response bias are the factors determining the performance according to SDT.

Sensitivity is the extent to which old items are distinguished from the new items.  $d'$ , which is the main sensitivity measure of SDT, is calculated via the following equation:  $d' = z(\text{HR}) - z(\text{FAR})$ , where  $z$  is the inverse of the normal cumulative distribution, HR is hit rate and FAR is false alarm rate (Macmillan and Creelman, 2005, p. 8).  $d' = 0$  is chance performance while increasing positive values indicating better sensitivity.

Main response bias measure of the SDT, Criterion location  $C$ , is calculated using the equation  $C = -.5 [z(\text{HR}) + z(\text{FAR})]$  (Macmillan and Creelman, 2005, p. 29). A value of 0 indicates neutral criterion while increasing negative values signify more lenient criteria, tendency to say “yes” to a probe; and, positive values mean stricter criteria. Likelihood ratio  $\beta$ ,  $\ln(\beta)$ , and  $B''$  are other common response bias measures. FAR is also

used as a response bias measure in the studies in which the locations of old distributions are manipulated. In these studies, new distributions are assumed to be fixed (Franks and Hicks, 2016; Hicks and Starns, 2014). But differentiation models (Criss, 2006, 2010; Kılıç, Criss, Malmberg and Shiffrin, 2017; McClelland and Chappell, 1998; Shiffrin and Steyvers, 1997) suggest that new items are better distinguished among well-learned study items, compared to poorly-learned items; thus the new distributions of strong (well-learned) lists are located more to the left than that of weak lists.

### Studies with Criterion Manipulation

#### Base Rate Manipulations

Base rate (the proportion of old items in a test list) manipulation is one of the most common methods to induce a criterion shift (Aminoff et al. 2012; Cox and Dobbins, 2011; Estes and Maddox, 1995; Franks and Hicks, 2016; Healy and Kubovy, 1977, 1978; Heit, Brockdorff and Lamberts, 2003; Rhodes and Jacoby, 2007). Participants tend to say “yes” in mostly-old test list; however, adopt stricter criteria for mostly-new test lists to achieve optimum performance according to SDT (Macmillan and Creelman, 2005; Swets et al., 1961). Swets et al. (1961) reported the effect of base rate on criterion in a signal detection task. Examining such effects in recognition memory tasks, Healy and Jones (1975) manipulated the base rates of test blocks (25% - 50%) in random order, with the prior information given to the participants about the base rates, and found no reliable criterion shifts between the test blocks. Healy and Kubovy (1977) also reported no changes in criterion in a recognition memory task but they observed criterion shifts in a numerical decision task.

In contrast with these results, there are studies in the literature that reported criterion shifts due to base rate manipulations. Estes and Maddox (1995) observed

criterion shifts when base rate manipulation is supported with item feedback for 3-digit-number and 3-letter stimuli; but failed to observe such shifts when feedback is absent or with word stimuli. Later, Aminoff et al. (2012) and Franks and Hicks (2016, Experiment 2) found that subjects were able to shift their criteria in the test blocks ranged 6-to-9 items within the same test phases. Moreover, Rhodes and Jacoby (2007) managed to induce item-by-item criterion shifts by assigning different base rates to the items appearing on different sides of the screen. They also noted that criterion shifts depend on the presence of feedback.

A special case of base rate manipulations is when test list consists of only old items (pure-old) or only new items (pure-new). Surprisingly, there are many studies in the literature that found null effects of pure list manipulations on hit or false alarm rates (Cox and Dobbins, 2011; Ley and Long, 1987, 1988; Wallace, 1982; Wallace, Sawyer and Robertson 1978), regardless of the presence of prior information. A possible explanation for null results is that participants focused on their subjective mnemonic evidence rather than the composition of test lists. As a matter of fact, Koop, Criss and Malmberg (2015) found criterion shifts due to base rate manipulations with pure lists when feedback is present, but replicated the previous results in the absence of feedback.

### Manipulations of Memory Strength

In another method, locations of old and new distributions (i.e. memory strength) are manipulated to induce criterion shift. When weak and strong items are studied and tested in separate lists, hit rates are higher and false alarms are lower for strong lists than weak lists (strength based mirror effect; Glanzer and Adams, 1985; Ratcliff, Clark and Shiffrin, 1990). It is suggested that this pattern stems from criterion shifts, more specifically, stricter criterion is set for strong lists (Hicks and Starns, 2014; Starns, White and Ratcliff, 2010; Stretch and Wixted, 1998). However, differentiation models (Criss, 2006; Kılıç et al., 2017; McClelland and Chappell, 1998; Shiffrin and Steyvers, 1997) suggest that new probes reveal poor match with strong items, result in fewer false alarms.

Another finding called null list strength effect (Ratcliff et al. 1990; Ratcliff, Sheu and Gronlund, 1992, Experiment 2; Starns et al., 2010) is that sensitivity is not affected by the strength of other items in study list. However, Hirshman (1995) found that criterion is more lenient for weak lists than weak items in mixed lists (consisting of strong and weak items), and criterion is stricter for strong lists than strong items in mixed lists. That means composition of test lists affects criterion.

Morrell, Gaitan and Wixted (2002) investigated the possibility of criterion shifts within a test list via strength manipulation, and reported that the criterion was stable during test phase. Stretch and Wixted (1998), and Verde and Rotello (2007, Experiments 1-4) replicated this result; however, Verde and Rotello (2007, Experiment 5) were able to induce such criterion shifts when subjects are given feedback during test phases. Other studies examining criterion shifts within test lists revealed that as the size of weak or strong test blocks increase, criterion shifts become more likely and their size tends to be larger (Hicks and Starns, 2014; Starns et al., 2010; Verde and Rotello, 2007).

In some studies, strength of new distributions is manipulated rather than old distributions. Adopting this approach, Benjamin and Bawa (2004) used new words from the same categories with old words in a test list, but the other test list included unrelated new words. Stricter criterion was set for the former test list.

### Other Methods

Methods used in criterion manipulations are not limited to base rate and strength manipulations. One of the other methods is using payoff matrices by which the gain and loss balances of “yes” and “no” answers are manipulated. When “yes” answer becomes more advantageous, subjects tend to say “yes” more often; likewise, when gains of “no” answer outweigh gains of “yes” answer, their criteria become more lenient (Curran, DeBuse and Leynes, 2007; Healy and Kubovy, 1978; Van Zandt, 2000).

Selmecezy and Dobbins (2013) used “probabilistic mnemonic cues” that give the participants the probability of the following probe being old. Subjects adopt stricter criterion following the probably-old cues rather than the probably-new cues. As the sensitivity decrease, they tend to rely more on probabilistic cues.

Han and Dobbins (2008) managed to induce criterion shifts by using a covert manipulation, biased feedback. In this method, participants are given “correct” feedback to their misses (strict condition) or false alarms (lenient condition). Participants were not aware of the manipulation; yet, their criteria shifted in accord with the conditions.

Finally, there are many studies in the literature reporting changes in hit and false alarms due to several stimulus properties, such as word frequency (Glanzer and Adams, 1985), or stimulus type (Scimeca, McDonough, and Gallo, 2011); but it remains unclear and is an ongoing debate among the scientists whether this changes stem from criterion shifts or any other mechanisms (e.g. differentiation).

### **Factors in Criterion Shifts**

Literature has conflicting findings regarding the effectiveness of manipulations on response bias. Studies have revealed that additional clues, such as feedback or prior information, are needed to induce reliable criterion shifts. For instance, some studies reported that criterion shifts were observed only when feedback was provided to the participants (Estes and Maddox, 1995; Morrell et al., 2002; Rhodes and Jacoby, 2007; Verde and Rotello, 2007); yet, other studies failed to establish such decisive role of feedback (Hicks and Starns, 2014; Selmeczy and Dobbins, 2013). It is observed that additional clues, such as using different font colors (Hicks and Starns, 2014), semantic categories (Singer, 2009), or response keys (Rhodes and Jacoby, 2007) for different conditions are effective in revealing criterion shifts.

Unlike the auxiliary clues, some factors hinder criterion shifts. For instance, initial criterion placement varies substantially among the subjects (Aminoff et al., 2012; Kantner and Lindsay, 2012); besides, the degree to which they shift their criterion is consistent across different time and tasks (Aminoff et al., 2012; Franks and Hicks, 2016; Kantner and Lindsay, 2012, 2014). Furthermore, the ability to benefit from external cues varies between individuals (Selmeczy and Dobbins, 2013). Finally, it is reported that the degree of criterion shifts is related to critical false alarm rate in Deese/Roediger-McDermott task (Kantner and Lindsay, 2012; Roediger and McDermott, 1995), false identifications in eyewitness task (Kantner and Lindsay, 2014), and hit rates in go-no-go task (Donders, 1969; Kantner and Lindsay, 2014).