aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
1–7

DOI: 10.1177/0963721419872747
www.psychologicalscience.org/CDPS
SAGE

Not All People Are Cut Out for Strategic Criterion Shifting

Michael B. Miller[1] and Justin Kantner[2]
[1]Department of Psychological & Brain Sciences, University of California, Santa Barbara, and
[2]Department of Psychology, California State University, Northridge

Abstract
Decisions are often based on memory, but memories are often vague and ambiguous. Therefore, one must establish some standard for how strong a memory must feel before it is accepted as part of one's past. This standard of evidence, called a *decision criterion*, should be adapted to the demands of a given situation: Sometimes it is critical to accept only strong memories as legitimate, whereas at other times it is more appropriate to act on the basis of relatively weak memory evidence. The ability to shift between these two standards as circumstances warrant is a critical element of successful decision making. We review recent work examining criterion shifting in recognition memory and argue that the true nature of criterion shifting may be seen only at the individual level: Whereas some people shift their criterion quite adaptively, others do not shift at all.

Keywords
criterion shifting, recognition memory, signal detection theory, eyewitness identification, decision making

If all of our memories were strong and vivid, then deciding whether or not something occurred in the past would be straightforward. But many of our memories are vague and uncertain, making it difficult to separate events that we truly experienced from events that, although plausible, never actually occurred. To separate fact from fiction in our memories, we need to establish a threshold for how strong a memory needs to feel before we are willing to accept that the memory is real. This threshold, or criterion, can range from lax, in which we accept memories on the basis of relatively weak or vague evidence, to strict, in which a memory must be very strong before we accept it as a legitimate record of the past.

The decision criterion is a crucial influence on the decisions people make and, by extension, the outcomes they experience. Indeed, neither one's memory acuity nor the strength of a given memory, considered in isolation from the decision criterion, tells one how that memory will be classified and subsequently used to direct behavior. In this review, we highlight the importance of understanding humans' willingness to strategically adapt a decision criterion as circumstances warrant.

Strategically Adapting a Criterion to Suit the Situation

In a typical recognition-memory experiment, participants are presented with a series of items (e.g., words, pictures) and on a later test are asked to discriminate between these previously encountered ("old") items and "new" items that were not part of the initial list. Thus, on each test trial, individuals must judge whether the strength of their memory for an item surpasses their criterion for calling the item "old" (these assumptions are formalized in signal detection theory; Parks, 1966; see Fig. 1). Recent demonstrations of stable individual differences in criterion use suggest that a focus on tendencies inherent to the recognizer is critical for understanding these decisions. Kantner and Lindsay (2012, 2014) showed that on recognition tests containing equal numbers of old and new items and no incentives for biasing responses, there is still wide variation in criterion placement across participants that is stable across time, differences in stimulus materials, and testing situations. Thus, one's standard of evidence for responding "old" appears to be akin to a cognitive trait. In other words, some people tend to be quite strict (or conservative) with memory evidence, whereas others tend to be quite lax (or liberal), with most people falling somewhere in between.

Corresponding Author:
Michael B. Miller, University of California, Santa Barbara, Department of Psychological & Brain Sciences, Santa Barbara, CA 93106-9660
E-mail: miller@psych.ucsb.edu

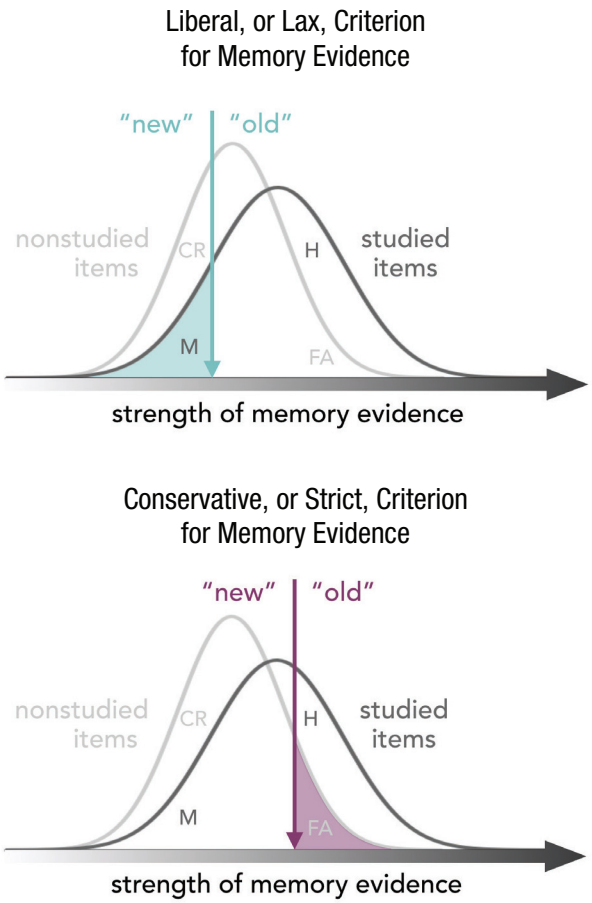

Fig. 1. A standard signal detection model of recognition memory. All studied and nonstudied items will be normally distributed along a continuum of memory strength. Even nonstudied items will vary in memory strength, with some seeming more familiar than others. The criterion (indicated by the vertical arrows) represents the point along that continuum at which one decides the memory is strong enough to respond "old." Correctly responding "old" to a studied item is considered a hit (H), whereas incorrectly responding "old" to a nonstudied item is considered a false alarm (FA). Correctly responding "new" to a nonstudied item is considered a correct rejection (CR), whereas incorrectly responding "new" to a studied item is considered a miss (M). If the goal is to avoid missing any studied items, then an individual should adopt a lax criterion (top panel), which will minimize the number of misses at the risk of increasing the number of false alarms. If the goal is to avoid calling a nonstudied item "old," then an individual should adopt a strict criterion (bottom panel), which will minimize the number of false alarms at the risk of increasing the number of misses.

However, few situations in life present an equal probability or equal subjective valuations of the various decision outcomes. Importantly, the placement of a criterion should depend on the situation, not necessarily on a person's general tendency. For example, police officers recognizing an escaped felon making a menacing move toward his pocket during an otherwise routine traffic stop should not fire their weapons at the felon unless they are extremely confident in their memory in order to avoid shooting a potentially innocent person reaching for his wallet. Likewise, a doctor who sees a patient with unusual symptoms may want to adopt a very lax criterion when trying to remember previous cases with similar symptoms: The memories may be weak and lead to fruitless searches through case files, but the chance of missing a critical sign of disease may be minimized. The willingness to adapt one's standard of evidence to the demands of a situation—to *criterion shift*—is critical for good decision making.

Researchers have identified experimental manipulations that can reliably and robustly induce subjects to adopt a different criterion without affecting their discrimination ability, including revealing the base rate of old items, creating uneven monetary payoffs and penalties for "old" versus "new" responses, or simply warning subjects to avoid either misses (incorrect "new" judgments) or false alarms (incorrect "old" judgments; see Hockley, 2011, for a review). Recent research has examined various properties of criterion shifting, including the amount of effort that may be necessary to adapt a criterion (North, Olfman, Caldera, Munoz, & Light, 2018; Starns & Olchowski, 2015).

Some People Just Do Not Shift at All

However, one thing that has not been appreciated until recently is this: Not all subjects will shift their criteria in response to these inducements (Aminoff et al., 2012; Frithsen, Kantner, Lopez, & Miller, 2018; Kantner, Vettel, & Miller, 2015). Some people shift their criteria well in response to these manipulations, whereas others do not shift at all. In a study by Aminoff and colleagues (2012), 95 subjects completed recognition tests in which the base rate of old items alternated between 30% and 70% across blocks, requiring shifts between strict and lax criteria (respectively) to maximize the proportion of correct responses. Figure 2a displays the range of criterion shifts across subjects. In response to the known changes in the base rate of old items, some subjects appropriately shifted between two extreme criteria (two example subjects are shown on the far left). Other subjects, regardless of whether they were naturally conservative or liberal, did not shift at all (two example subjects are shown on the far right). Critically, these trends do not simply represent random variation across subjects, because subjects were consistent in their criterion shifting across multiple tests. The magnitude of participants' criterion shifts on a word-recognition test strongly predicted those on a face-recognition test (r = .58; see Fig. 2b). There appears to be something systematic and uniquely individualistic about the variations in criterion shifting.

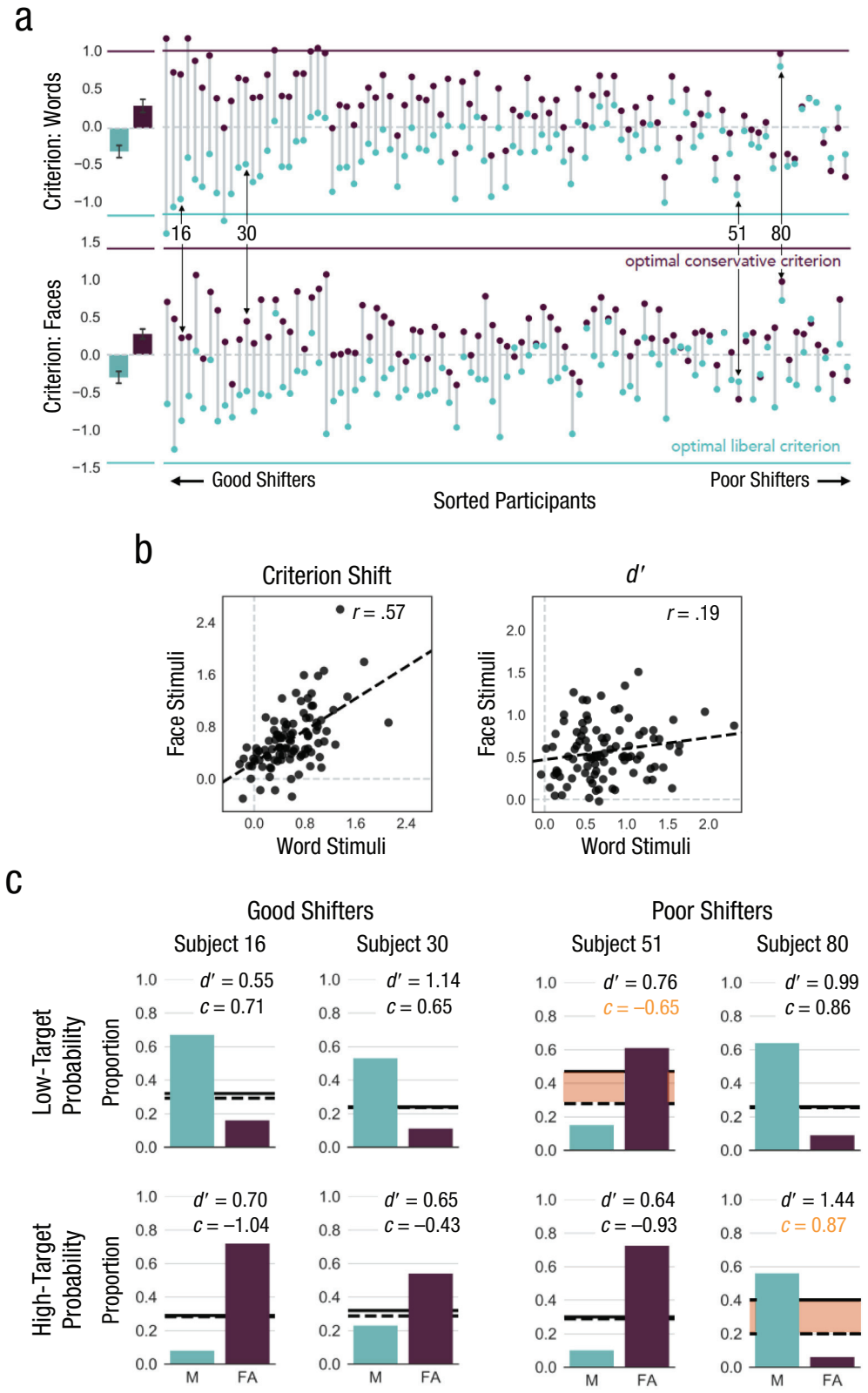

Fig. 2. *(continued on next page)*

Fig. 2. Results from Aminoff and colleagues (2012). The graphs in (a) show the tendency of 95 subjects to shift criterion between a conservative condition (low probability of a target) and a liberal condition (high probability of a target) in a word-recognition task and a face-recognition task. Results are normalized for differences in discrimination (d') and are sorted from the best shifters (to the left) and the worst shifters (to the right) on the task in question. The ordering of subjects in the face-recognition task is based on their criterion-shift ordering in the word-recognition task to give a sense of the consistency of criterion shifting within individuals across tasks. The thick bars to the left represent the group averages, with error bars representing standard errors of the mean. Note that only one individual reached the optimal criterion in both the conservative and liberal conditions of one task: The highest shifter in the word-recognition task exceeded the optimal criterion in the conservative condition (indicated by the upper horizontal line) and the optimal criterion in the liberal condition (indicated by the lower horizontal line). Many individuals did not shift at all. The scatterplots (b; with best-fitting regression lines) show the correlation between (left) criterion shifting in the word- and face-recognition tasks and (right) discrimination (d') in the word- and face-recognition tasks. The bar charts (c) show the consequences of not shifting. Data were derived from the same study as depicted in (a)—but note that the criterion values in (a) are normalized, whereas the criterion values in (c) are not. The proportion of misses (Ms) and false alarms (FAs), along with d' and the criterion value (given as c), is shown separately for 4 individual subjects when target probability was low and when target probability was high. The proportion of misses and false alarms should increase or decrease in response to changes in the base rate of old items, as it did for the two good shifters (Nos. 16 and 30). The broken line represents the optimal overall proportion of errors at that given discrimination level, and the solid line represents the actual overall proportion of errors. Good shifters do not necessarily need to be optimal in the placement of their criteria. They simply need to shift their criteria enough in response to the changes in the base rate of old items to minimize their overall proportion of errors, regardless of their discrimination ability. Poor shifters do not strategically respond to changes in the base rate of old items or do so to such a small extent that their proportion of errors is much higher than it should be. The two poor shifters (Nos. 51 and 80) had a higher proportion of errors—despite higher levels of discrimination—than the good shifters because they did not strategically shift their criteria, remaining either liberal or conservative despite the known changes in the base rate of old items (as shown by the colored area between the solid and dashed lines).

Further studies have demonstrated that recognition-criterion flexibility generalizes across criterion manipulations (Kantner et al., 2015; though see Franks & Hicks, 2016). It also generalizes to domains outside of recognition memory. We recently found significant correlations between criterion shifting on tests of recognition memory and criterion shifting on tests of visual detection and discrimination (Frithsen et al., 2018).

One of the more interesting properties of criterion shifting is that people appear to suboptimally place their criteria. That is, regardless of their ability to discriminate old items from new, people are generally not strict enough when it benefits them to avoid false alarms or lax enough when avoiding misses is more beneficial (see Fig. 2a). This criterion suboptimality (sometimes called *conservatism*, an unfortunate term given that strict criteria are also usually called *conservative*) appears quite resilient even in extreme situations. In Kantner et al. (2015, Experiment 3), subjects in a control condition attempted to identify studied suspects in a simulated security patrol. They were given strong instructions to avoid false alarms at all costs in one scenario (strict-criterion condition) and to avoid misses at all costs in another scenario (lax-criterion condition). As is typically found, most subjects were not nearly as strict as they should have been when asked to minimize false alarms, nor were they as lax as needed when asked to minimize misses. In the experimental condition, subjects were given the same suspect-recognition task but with no preceding study phase (because of a "malfunction"). In this case, one might expect that with no diagnostic memory evidence on which to base their judgments, subjects would base their decisions solely on the instructions to avoid false alarms or misses. Indeed, a small proportion of subjects did so, shifting between appropriately extreme criteria, but many more did not, including many of the same subjects who shifted poorly or not at all in the control condition.

Several explanations for criterion suboptimality have been put forward over the years, including subjects' attempts to align their responses with the base rates of targets and lures (*probability matching*; Thomas & Legge, 1970), trial-to-trial noise in criterion placement (Benjamin, Diaz, & Wee, 2009), and the failure to estimate one or more of three environmental factors: target base rates, the costs and benefits of outcomes, and the confusability of targets and lures (Lynn & Barrett, 2014). We argue that none of these explanations can provide a complete account of shifting behaviors, however, because they were tested on performance that was averaged across a group of subjects. Although these group-level accounts likely offer explanations for why some individuals fall short of the optimal criterion (and some may constitute individual-differences factors in themselves, though they have not been investigated in that way), they fail to consider the consistent performance of individuals at the extreme ends of the criterion-shifting continuum. Indeed, the idea that people are suboptimal criterion shifters is a generalization that masks the subject-level factors that appear to drive memory decision making. We believe the question should be why some people approach optimality in shifting their criteria and other people simply do not shift at all.

The Consequences of Not Shifting

An optimal criterion maximizes the number of positive outcomes (e.g., accurate responses, payoffs, avoidance of critical errors) given the ability of the subject to discriminate between studied items and nonstudied items: The lower the ability to discriminate, the more extreme the optimal criterion becomes (Green & Swets, 1966; Macmillan & Creelman, 2004). The consequences of not shifting a criterion are highly evident. Figure 2c depicts data from an individual (No. 80)

whose recognition decisions tended to be so conservative that this person almost reached the optimal criterion when the probability of an old item was low (Aminoff et al., 2012). But this same subject failed to shift that criterion to a more liberal setting when correctly informed that the recognition test switched to a high-target probability condition—that is, the test items were then much more likely to be old. Because of this failure to criterion shift, this subject had a much higher proportion of errors on the test than a different subject (No. 16) whose discrimination ability was actually much worse but who almost optimally criterion shifted. Thus, this failure to criterion shift is a lost opportunity for correct responses that can undermine a good memory. Conversely, proper criterion shifting can support a weak memory by producing positive outcomes even when memory evidence cannot be relied on.

The failure to criterion shift can also have real-life consequences. As noted, some individuals tend to be quite strict about accepting memories as veridical, allowing them to avoid false alarms. When such people fail to adjust this strategy in situations that call for a more lenient criterion, they may miss something significant, such as when a truck driver fails to take a turnoff that seems familiar or a radiologist fails to flag a spot on a patient's chest X-ray because it seems only vaguely similar to examples of cancerous tumors stored in memory. Analogously, individuals who are inherently lax in accepting memories can commit highly consequential false alarms if they do not shift their criterion in situations that call for a more conservative strategy, such as when a witness to a crime falsely identifies a member of a lineup on the basis of a facial feature shared with the culprit.

The question of appropriate criterion use by eyewitnesses has important implications for policymakers and for society at large (Wixted & Wells, 2017). Mickes and colleagues (2017) recently published a fascinating study in which subjects witnessed a mock videotaped crime. Later, the subjects were asked to identify the culprit in the video out of a lineup, knowing that the suspect was not necessarily in the lineup. After making their decisions, subjects were asked to rate the confidence of their choices from 0% (*not sure at all*) to 100% (*absolutely confident*). Subjects' choices at the 100% confidence level produced extremely few false identifications. However, an entirely different group of subjects was instructed to identify a suspect only if they were 100% confident in their choice; in other words, to be as strict with their recognition criterion as possible to avoid falsely identifying an innocent person at all costs. These subjects were indeed generally conservative with their choices but not nearly as conservative as they should have been, leading to a much higher false-identification rate than in the former group of subjects using that same 100% confidence level. In our interpretation of the data, a number of subjects in the latter group failed to shift to a more conservative criterion despite the instructions to do so. Individual differences in criterion placement on recognition-memory tests have been shown to predict the rate of false identifications on culprit-absent lineups and could potentially be used as part of a test for assessing eyewitness decisions (Baldassari, Kantner, & Lindsay, 2019; Kantner & Lindsay, 2014).

The Willingness to Criterion Shift

What are the characteristics of a good criterion shifter? Whereas some studies have suggested factors that may predict criterion placement and shifting, including age (Cassidy & Gutchess, 2015), sensitivity to the strength of one's own memories (Selmeczy & Dobbins, 2013), and social cues (Cassidy, Dubé, & Gutchess, 2015), what may be more notable are all the factors that do not predict criterion shifting. These include measures of intelligence and executive-functioning skills such as working memory capacity and task-switching ability, as reported by Aminoff and colleagues (2012) in a large-scale study of individual differences that included 95 subjects. Indeed, as the Mickes et al. (2017) study demonstrates, people are capable of determining what a 100% confidence judgment should entail; they simply fail to apply that criterion appropriately when instructed to do so. On the basis of that same large-scale study by Aminoff and colleagues (2012), we suggest that good criterion shifting also does not appear to be strongly related to any personality traits. Though there is still much work needed to identify the characteristics of good criterion shifters, the driving factor may be the willingness to shift rather than the ability to shift.

Interestingly, the degree to which an individual shifts a criterion does not appear to be related to the range of confidence levels that same individual might use on a recognition test. Collecting a confidence rating for old/new recognition judgments is a common method used to construct receiver-operating-characteristic curves, in which each confidence rating available to participants is thought to reflect the use of a separate criterion. Given this assumed relationship, one might expect that an individual who consistently does not shift a criterion would use a very limited range of confidence ratings. For this article (based on a suggestion by John Wixted), we reanalyzed some data we collected on 10 tests of recognition memory, visual discrimination, and visual detection, each including criterion manipulations and the collection of confidence ratings on each old/new judgment (Frithsen et al., 2018). Of these 10 tests, only 1 indicated a significant relationship between the magnitude of a participant's criterion shift and the spread of the participant's confidence ratings

(a small correlation of .26 on a recognition test using a payoff manipulation). More systematic studies will need to be conducted on this relationship, but it would appear that people are capable of establishing appropriately extreme criterion levels (by rating their confidence level), but some of these people are just unwilling to use them when instructed or induced to do so.

The unwillingness to shift may stem from a preference to make judgments using memory evidence—even when it is inconclusive—rather than defaulting to a criterion-based decision rule. In their 1966 book on signal detection theory, Green and Swets noted of sensory discriminations that "the observer tends to avoid extreme criteria" and "probably finds it difficult to believe that he would be performing responsibly if the sensory distinctions he makes are exactly those that he could make by removing the earphones in an auditory experiment or by turning his back on a visual signal" (p. 91). As we found, even in the absence of a study phase, many subjects were unwilling to make decisions solely on the basis of the base rate of targets or the instructions to avoid false alarms or misses (Kantner et al., 2015). In debriefings following the study, many subjects said they based decisions on some sense of familiarity even though they knew there was no basis for doing so. As we noted at the time, it is as if "people would rather attempt to be *correct* than be *correctly biased*" (Kantner et al., 2015, General Discussion, para. 3).

Conclusion

Recent findings strongly support an individual-differences approach to understanding the dynamics of criterion use and their broad implications for decision making. A critical direction for future research, we believe, is to identify the cognitive, neural, and motivational factors that lead one person to adopt appropriate standards of memory evidence and another person to remain either too lax or too strict. Further insight into these factors will drive theory development in memory and decision making and help explain—and perhaps improve—human decision making at the level of the individual.

Recommended Reading

Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Tipper, C., Johnson, A., . . . Miller, M. B. (2012). (See References). The first article (to our knowledge) to demonstrate extreme but stable individual differences in criterion shifting on a recognition-memory test.

Baldassari, M. J., Kantner, J., & Lindsay, D. S. (2019). (See References). A recent report suggesting measures that could be used to predict eyewitness lineup choices.

Hockley, W. E. (2011). (See References). A thorough review of issues related to criterion shifting.

Macmillan, N. A., & Creelman, C. D. (2004). (See References). A comprehensive and readable guide to signal detection theory and its application to psychological processes such as recognition memory.

Action Editor

Randall W. Engle served as action editor for this article.

Acknowledgments

We thank Nikki Marinsek for her work on the figures.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This review was supported by the Institute for Collaborative Biotechnologies through Grant W91NF-09-0001 from the U.S. Army Research Office.

References

Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Tipper, C., Johnson, A., . . . Miller, M. B. (2012). Individual differences in shifting decision criterion: A recognition memory study. *Memory & Cognition*, *40*, 1016–1030.

Baldassari, M. J., Kantner, J., & Lindsay, D. S. (2019). The importance of decision bias for predicting eyewitness lineup choices: Toward a lineup skills test. *Cognitive Research: Principles and Implications*, *4*(1), Article 2. doi:10.1186/s41235-018-0150-3

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review*, *116*, 84–115.

Cassidy, B. S., Dubé, C., & Gutchess, A. H. (2015). Social influences on adaptive criterion learning. *Memory & Cognition*, *43*, 695–708.

Cassidy, B. S., & Gutchess, A. H. (2015). Age effects in adaptive criterion learning. *Journals of Gerontology B: Psychological Sciences & Social Sciences*, *71*, 1015–1023.

Franks, B. A., & Hicks, J. L. (2016). The reliability of criterion shifting in recognition memory is task dependent. *Memory & Cognition*, *44*, 1215–1227.

Frithsen, A., Kantner, J., Lopez, B. A., & Miller, M. B. (2018). Cross-task and cross-manipulation stability in shifting the decision criterion. *Memory*, *26*, 653–663.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: John Wiley & Sons.

Hockley, W. E. (2011). Criterion changes: How flexible are recognition decision processes? In P. A. Higham & J. P. Leboe (Eds.), *Constructions of remembering and metacognition: Essays in honour of Bruce Whittlesea* (pp. 155–166). Basingstoke, England: Palgrave Macmillan.

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition*, *40*, 1163–1177.

Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review*, *21*, 1272–1280.

Kantner, J., Vettel, J. M., & Miller, M. B. (2015). Dubious decision evidence and criterion flexibility in recognition memory. *Frontiers in Psychology*, *6*, Article 1320. doi:10.3389/fpsyg.2015.01320

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological Science*, *25*, 1663–1673.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Mahwah, NJ: Erlbaum.

Mickes, L., Seale-Carlisle, T. M., Wetmore, S. A., Gronlund, S. D., Clark, S. E., Carlson, C. A., . . . Wixted, J. T. (2017). ROCs in eyewitness identification: Instructions versus confidence ratings. *Applied Cognitive Psychology*, *31*, 467–477.

North, L. J., Olfman, D., Caldera, D. R., Munoz, E., & Light, L. L. (2018). Age, criterion flexibility, and item recognition. *Aging, Neuropsychology, and Cognition*, *25*, 390–405.

Parks, T. E. (1966). Signal-detectability theory of recognition-memory performance. *Psychological Review*, *73*, 44–58.

Selmeczy, D., & Dobbins, I. G. (2013). Metacognitive awareness and adaptive recognition biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 678–690.

Starns, J. J., & Olchowski, J. E. (2015). Shifting the criterion is not the difficult part of trial-by-trial criterion shifts in recognition memory. *Memory & Cognition*, *43*, 49–59.

Thomas, E. A., & Legge, D. (1970). Probability matching as a basis for detection and recognition decisions. *Psychological Review*, *77*, 65–72.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*, 10–65.