Individual differences in shifting decision criterion: A recognition memory study

Elissa M. Aminoff · David Clewett · Scott Freeman · Amy Frithsen · Christine Tipper · Arianne Johnson · Scott T. Grafton · Michael B. Miller

Abstract An ability to flexibly shift a decision criterion can be advantageous. For example, a known change in the base rate of targets and distractors on a recognition memory test will lead optimal decision makers to shift their criterion accordingly. In the present study, 95 individuals participated in two recognition memory tests that included periodic changes in the base rate probability that the test stimulus had been presented during the study session. The results reveal a wide variability in the tendency to shift decision criterion in response to this probability information, with some appropriately shifting and others not shifting at all. However, participants were highly reliable in their tendency to shift criterion across tests. The goal of the present study was to explain what factors account for these individual differences. To accomplish this, over 50 variables were assessed for each individual (e.g., personality, cognitive style, state of mind). Using a regression model that incorporated different sets of factors, over 50 % of the variance was accounted for. The results of the analysis describe the total, direct, and mediating effects on criterion shifting from factors that include memory strength, strategy, and inherent characteristics such as a fun-seeking personality, a negative affect, and military rank. The results are discussed with respect to understanding why participants rarely chose an optimal decision-making strategy and provide greater insight into the underlying mechanisms of recognition memory.

Keywords Memory · Decision making · Individual differences · Recognition

E. M. Aminoff · D. Clewett · S. Freeman · A. Frithsen ·
C. Tipper · A. Johnson · S. T. Grafton · M. B. Miller
Department of Psychological & Brain Sciences,
University of California,
Santa Barbara, CA, USA

E. M. Aminoff · D. Clewett · S. Freeman · A. Frithsen ·
C. Tipper · A. Johnson · S. T. Grafton · M. B. Miller
Institute for Collaborative Biotechnologies,
University of California,
Santa Barbara, CA, USA

E. M. Aminoff (✉)
Center for the Neural Basis of Cognition,
Carnegie Mellon University,
115 Mellon Institute, 4400 Fifth Avenue,
Pittsburgh, PA 15213, USA
e-mail: elissa@cnbc.cmu.edu

Introduction

Shifts in decision criteria are common, and, at times, this is the optimal strategy to utilize. Moreover, decision criteria play an important role in recognition memory and have important implications in the explanations of memory performance. However, individual differences in criterion shifting have been largely neglected in the research. In particular, large variations in criterion shifting between individuals could constrain the possible mechanisms underlying criterion placement in recognition memory, or it could suggest that individuals utilize different mechanisms when performing a memory test. In addition, large variations in criterion shifting could also explain some of the inconsistencies that have been reported in the field (Hockley, 2011). Given the strong influence of decision criterion on the underlying mechanism of recognition performance, it is critical to understand why some individuals shift criterion, whereas others do not. This study examines different factors, including inherent characteristics of an individual, that may mediate the willingness to shift a decision criterion and, thus, give rise to individual

differences in the tendency to shift criterion during a recognition memory test.

To decide whether or not a test item has been studied previously, it is thought that individuals weigh the available evidence against a decision criterion (e.g., a threshold of familiarity). Where the criterion is placed on an axis of evidence strength can clearly affect the decision outcome, with a liberal criterion designating more test items as targets and a conservative criterion designating more test items as distractors. Optimal performance on a recognition test may be defined as maximizing correct responses (hits and correct rejections; see Green & Swets, 1966). To maximize correct responses when the base rate of targets is very high (i.e., the test item is likely to be from the study session), the individual should utilize a liberal bias (i.e., respond that most items are old), and vice versa when the base rate is very low. When the base rate of targets knowingly changes during the course of a test, optimal performance requires that the criterion should shift rather than stay static. That is, optimal decision makers will require more evidence in order to recognize an unlikely target and less evidence to recognize a likely target (Green & Swets, 1966; Macmillan & Creelman, 2005).

Considerable research has been dedicated to investigating shifts in decision criterion. While changes in target probability are a reliable manipulation of criterion placement during a recognition memory test, a number of other manipulations are known to affect criterion placement as well, including payoffs, instructions, item distinctiveness, and subjective memorability (Bruno, Higham, & Perfect, 2009; Curran, DeBuse, & Leynes, 2007; Hirshman, 1995; Hockley & Caron, 2007; Rhodes & Jacoby, 2007; Singer & Wixted, 2006; Van Zandt, 2000). Regardless of the method of manipulation, many researchers have noted that individuals tend to be suboptimal in their criterion placement. A much less extreme criterion is adopted than is called for using an optimal decision rule (Macmillan & Creelman, 2005), a phenomenon sometimes referred to as *conservatism* (Benjamin, Diaz, & Wee, 2009). This phenomenon has been a critical component of many hypotheses regarding criterion shifts and placement. For example, probability-matching models suggest that participants place their criterion to respond *old* proportional to the probability of an old item (Parks, 1966; Thomas, 1975; Thomas & Legge, 1970), leading to a criterion much less extreme than optimal when the probability of the old items varies from 50 %. Alternative models, such as range models that depend on the estimated range of the old and new distributions, would make similar predictions (Hirshman, 1995).[1] In contrast, Benjamin,

Diaz, and Wee (2009) have suggested that there is considerable variance in the placement of a criterion over the course of a test and that this criterial noise can account for the general phenomenon of conservatism, thereby negating conservatism as a constraint on models of criterion shifting.

In this study, we highlight another phenomenon that could constrain proposed psychological mechanisms of criterion shifting on a recognition test. As we have observed in our own data sets from previous experiments, there is a wide range of individual variability in the tendency to shift criterion, with some individuals shifting to an optimal placement, other individuals not shifting at all, and others in between these two extremes. If this is generally the case, a model like probability matching would apply only to the subset of individuals who were between optimal shifting and no shifting. This extreme variability would suggest either a mixed model for criterion shifting or an entirely different model that could account for one individual engaging in optimal criterion shifting, but not another individual.

Criterion shifts have been used as an explanation, albeit controversially, for a number of memory-related phenomena (Miller, Guerin, & Wolford, 2011; Miller & Wolford, 1999; Roediger & McDermott, 2000; Wickens & Hirshman, 2000; Wixted & Stretch, 2000). For example, in a popular word-list memory paradigm, the high false alarm rate to related lures has been explained as the activation of associated nonstudy words (Gallo & Roediger, 2002; K. J. Robinson & Roediger, 1997; Roediger, Balota, & Watson, 2001) or, in contrast, as a criterion shift for test words that are strongly related to words on one of the study lists (Miller & Wolford, 1999). The distinction between the two hypotheses is fundamental in that the former proposes a memory process during encoding and the latter proposes a decision-making process during retrieval. Criterion shifting has also been used to explain other well-known memory effects as well, including eyewitness testimony (Mccloskey & Zaragoza, 1985), the mirror effect (Hirshman, 1995), and verbal overshadowing (Clare & Lewandowsky, 2004). Thus, understanding what drives the individual to shift criterion can provide additional insight into the mechanisms underlying memory.

The research regarding whether an individual can flexibly shift criterion within a recognition memory test has been mixed. Some studies have suggested that participants initially set a decision criterion and do not shift, whereas other reports have demonstrated that participants can shift on a trial-by-trial basis, given certain testing parameters, such as subject awareness of base rates (Bruno et al., 2009; Dobbins & Kroll, 2005; Hockley & Niewiadomski, 2007; Miller & Wolford, 1999; Morrell, Gaitan, & Wixted, 2002; Postma, 1999; Rhodes & Jacoby, 2007; Singer, 2009; Stretch & Wixted, 1998; Van Zandt, 2000; Wixted & Stretch, 2000; for a review, see Hockley, 2011). Individual differences in

---

[1] The validity of these various models may depend on the particular manipulation of criterion—for example, a probability matching model for manipulations of base rate and a range model for manipulations of familiarity.

the tendency to shift criterion may provide insight into these mixed results.

To shift criterion, one must exhibit a "willingness" to do so. As Wixted and Stretch (2000) have indicated in previous reports, the tendency to shift a criterion within a recognition test requires an extra mental step in addition to discriminating old from new items. Willingness to shift a decision criterion on a recognition test can be mediated by several factors. One factor that clearly affects the degree to which an individual will shift a decision criterion is how well the individual remembers the studied items (Bruno et al., 2009; Hirshman, 1995). If an individual has a strong memory for studied items (and, therefore, has a high d' on the basis of signal detection theory), the optimal criterion will be less biased (e.g., to liberal or conservative), and the need to shift between conditions will be less necessary. A second factor that affects the degree of criterion shifting is the consideration given to the cues indicating the probability of a target (Postma, 1999; Rhodes & Jacoby, 2007; Van Zandt, 2000). The more the participant relies on the target probabilities, the more likely he or she will be to shift decision criterion accordingly.

Above and beyond those two factors, there may also be variations in the character of the individual that affects his or her willingness to shift a criterion. We propose that the willingness to shift a decision criterion within a recognition test may be partially mediated by inherent characteristics of the individual. For example, it is known that differences in personality traits are related to task-switching tendencies (M. D. Robinson, Wilkowski, Kirkeby, & Meier, 2006; Smillie, Cooper, Tharp, & Pelling, 2009). In this study, we explored a variety of inherent characteristics that may mediate an individual's tendency to shift a criterion, including personality traits and state of mind factors.

To investigate factors that explain the individual variability in shifting criterion, we implemented a recognition memory paradigm in which all properties of the stimuli were kept constant except for an explicit visual cue presented with the test stimulus indicating the probability that it had been presented during the study session (70 % or 30 % likelihood). Memory performance results revealed a large amount of individual variability in shifting criterion between the different probability conditions. Data were analyzed within a framework that examined the direct and mediating relationships between the factors discussed above in the tendency to shift criterion (see Fig. 1). The model indicates that willingness to shift criterion is mediated by memory (measured by d'), reliance on probabilistic cue information (RCI, measured through debriefing questionnaires), and inherent characteristics (e.g., personality, cognitive style—measured through a battery of assessments) of the individual. This model is not exhaustive but highlights some of the key factors in the underlying mechanism of criterion shifting in
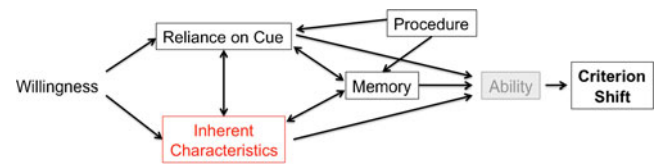

Fig. 1 A model for explaining individual variations in criterion shifting. There are two main factors that mediate the willingness to shift criterion: reliance on cue information and inherent characteristics. Both of these factors have a direct effect on criterion shifting but also have an effect that is mediated by memory performance. Procedural variations (see the Method section) may have a moderating effect on memory and reliance on cue. Ability, beyond the scope of the present study, is the final mediating factor in criterion shifting

recognition memory. We acknowledge that in order to shift, the ability (i.e., the hardware) to shift must be intact. Ability, therefore, is the final mediating factor of shifting criterion. We assume that all of our participants have a similar ability to shift, and therefore, the goal of the present study is to explain the tendency to implement this ability.

We propose that a willingness to shift criterion is mediated by a direct effect of RCI and by a direct effect of inherent characteristics. However, both of these effects can be mediated by the memory of the individual. Using a multistep regression analysis, we tested this proposal. Our study is unique not only for examining individual variability in criterion shifting, but also for investigating specific inherent characteristics that explain this behavior. The results of this analysis reveal why one person may shift criterion and another person may not, which can provide insight into optimal decision making and recognition memory.

Method

This study was part of a larger project investigating the individual variability in the structure and function of brain activity of combat-experienced Army officers, in addition to investigating individual differences in criterion shifting. The fMRI methods and results are beyond the scope of the present analysis, but scanning parameters are available upon request. During the test portion of the session, fMRI scans were acquired.

Participants

Data were analyzed from 95 participants (27–47 years of age, $M = 35$; 5 females; 12 left-handed). The 95 participants included 68 combat-experienced commissioned and noncommissioned Army officers. The remaining 27 were age matched to the officers and were from the University of California, Santa Barbara community. Of the 27, 21 were graduate students or postdoctoral researchers. Informed written consent was obtained from each participant prior to

the experimental sessions. All procedures were approved by the University of California, Santa Barbara Human Subjects Committee. For more details on participants, please refer to the Supplemental Methods.

Stimuli

In this experiment, 360 faces and 360 words were used as stimuli. Faces of varying ethnicity were depicted in black-and-white photographs. Faces were drawn from a variety of online faces databases (Samaria & Harter, 1994; Martinez & Benavente, 1998; CVL Face Database, University of Ljublijana; Weyrauch et al., 2004; Minear and Park, 2004). Words were four to eight letters in length. There were no significant differences between words on the target and distractor lists in imageability (range, 502–655) or frequency (range, 1–382; Kučera–Francis written frequency count), as evaluated through the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). Words were presented in Arial font with a font size of 40. Faces were 2.78 × 3.33 in. Stimuli were balanced across conditions and were counterbalanced across participants and conditions, such that stimuli that were targets for half the participants were the distractors for the other half of the participants.

Experimental procedure

Overall procedure Participants began the experimental session with a short practice study–test session outside the MRI scanner in order to become comfortable with the task. Participants were then positioned in the scanner, after which they studied the first study session (e.g., with face stimuli) not scanned. After a 9-min delay, the test portion began while the participants were scanned using fMRI. Immediately following the conclusion of the first study–test session, the participants began the second study–test session with identical parameters (e.g., with word stimuli). Each participant had separate study and test sessions for each set of stimuli (either faces or words), in which the order (first or second study–test) was counterbalanced.

Study session One hundred eighty stimuli were presented sequentially in the center of the screen on a white background. Words were presented in black font, and faces were without a frame (Fig. 2). Stimuli were presented for 1 s (words) or 1.5 s (faces), with a 1-s intertrial interval. Participants were instructed to remember each stimulus for a later memory test. To facilitate deep encoding of the faces, which were harder to remember, participants judged whether each face was pleasant or unpleasant via a buttonpress. Words were passively viewed.

Test session Each test consisted of the 180 stimuli previously studied and 180 novel stimuli. Stimuli at test were divided

into two conditions: a high-probability condition in which the stimulus had a 70 % probability of being old (i.e., studied previously) and a low-probability condition with a 30 % probability of being old. A color cue (green or red) presented via the font color or a colored frame around the photograph indicated the probability condition (Fig. 2). In the high-probability condition, 70 % of the targets ($N = 126$) were cued, and 30 % of the distractors were cued ($N = 54$). In the low-probability condition, the remaining 30 % of the targets ($N = 54$) were cued, and 70 % of the distractors were cued ($N = 126$). The number of trials was the same for the high- and low-probability conditions, but the proportion of targets to distractors reflected the probabilities of the respective condition. The association between color and probability condition was counterbalanced across participants. Half the participants were told that the color green indicated the high-probability condition and the color red indicated the low-probability condition. The color assigned to a condition was swapped for the other half of the participants. Stimuli were presented at the same size and location as the study session for 1.5 s, with a 1-s intertrial interval.

Participants were instructed to determine whether the stimulus had been previously studied and to press the respective button for an *old* or *new* response. Instructions to the participants included explicit information regarding which color indicated a 70 % likelihood that the stimulus was old and, thus, highly likely that it had been seen at study and which color indicated a 30 % likelihood that the stimulus was old and, thus, most likely a new stimulus and not presented at study. Participants were told that these were accurate probabilities. Yet the instructions did not explicitly tell the participant to incorporate the probability into the memory judgment.

Stimuli were presented in blocks according to probability condition: Six to nine trials of the same probability were presented before the probability switched. Old and new stimuli were intermixed within these blocks. The probability indicated by the color cue did not necessarily correspond to the proportion of old and new stimuli within each block but, rather, to the distribution within the overall test session. Intermixed throughout the test session were 180 fixation trials: a black "+" symbol presented at the center of the screen for 2.5 s necessary for subsequent fMRI analysis. Fixation trials were interspersed with the stimulus trials in a pseudorandom order that was constrained only in a manner to optimize extraction of the hemodynamic response.

Practice To familiarize participants with the task, each participated in a short practice study–test session, which included a study session of 15 words and, separately, 15 faces, and a test session, with color cues, of 30 words and, separately, 30 faces. Only after participants were comfortable with the task did they continue to the rest of the experiment.
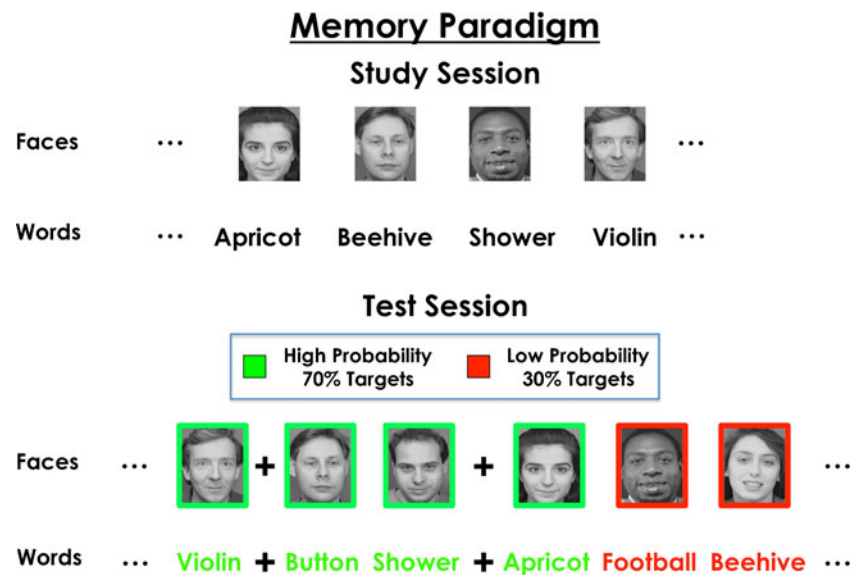
Memory Paradigm

Study Session

Fig. 2 Stimulus presentation. Participants studied 180 of each stimulus set. Studied stimuli were presented in black font and without a colored frame (top). Stimuli at test (*N* = 360) were presented with a color cue via the font or a frame around the picture that indicated the probability that the item had been presented during the study session (bottom). Intermixed with stimulus trials were fixation trials on which a "+" was presented, and there was no task performed on these trials. Trials were presented in blocks of six to nine trials of the same probability before the probability switched. The figure depicts only a portion of each block: only the last four trials (out of the six to nine possible) in the block for the high-probability condition (i.e., green) and the beginning two trials (out of the six to nine possible) in the block for the low-probability condition (i.e., red)

The procedure and parameters were used for a majority of our participants (participants 31–133); however, the first 30 participants (only 20 of which were used in the final analysis) had a slight variation of the sequence of events and durations of presentation (see the Supplemental Methods for details). The variations in study procedure were always partialed out of the regression analyses. No significant direct effects due to these variations on criterion shifting were observed.

Individual difference factors

Participants completed a battery of questionnaires to assess factors that could account for individual differences in criterion shifting, as well as brain structure and function in combat-experienced Army officers. For a complete list of each variable obtained, please refer to the Appendix.

Demographics were self-reported in an in-house questionnaire.

Cognitive style was assessed for visual and verbal abilities using the Santa Barbara Learning Style Questionnaire (Mayer & Massa, 2003), the Object–Spatial Imagery Questionnaire (OSIQ; Blajenkova, Kozhevnikov, & Motes, 2006), the Verbalizer–Visualizer Questionnaire (VVQ; Richardson, 1977), and a vocabulary test (18 items adapted from the Vocabulary scale of the Armed Services Vocational Aptitude Battery as selected from a

test preparation book; Baron's Educational Series 2001). Spatial abilities and executive functioning used in mental planning were assessed with the Card Rotation Test (Ekstrom, French, & Harmann, 1976) and the Paper Folding Test (Ekstrom et al., 1976). The Need for Cognition questionnaire evaluated an individual's tendency to engage in and enjoy thinking (Cacioppo & Petty, 1982). Visual working memory was tested using established paradigms (Luck & Vogel, 1997; Vogel & Machizawa, 2004).

Personality was assessed through the Big Five Inventory (John & Srivastava, 1999), BIS/BAS scales (Carver & White, 1994), the Revised Eysenck Personality Questionnaire (H. J. Eysenck & Eysenck, 1975; S. Eysenck, Eysenck, & Barrett, 1985), and the PANAS mood assessment (Watson & Clark, 1994; Watson, Clark, & Tellegen, 1988).

Mental health was evaluated for psychological and brain trauma via the Beck Depression Inventory II (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), post-traumatic stress disorder (Brewin et al., 2002), and traumatic brain injury (in-house questionnaire developed from the criteria listed in American Academy of Neurology, 1997).

State of mind factors, such as daily habits, were self-reported in an intake questionnaire designed in-house. Mental and physical comfort while in the MRI was assessed postscanning using an in-house questionnaire.

Military factors were assessed through an in-house questionnaire and through the Combat Exposure Scale (Keane et al., 1989).

Behavioral performance analysis

In each test, d' and criterion were calculated for each participant separately for the low- (30 %) and high- (70 %) probability conditions (separately for words and faces).[2] The d' was calculated using the following equation: zHits − zFalse Alarms. Criterion was calculated using the following equation: $-0.5 * (z$Hits + zFalse Alarms). Trials that received either no response or multiple responses were removed from all analyses. The average number of trials removed in the word test was 6.39, and the average number of trials removed in the face test was 7.57. The number of trials excluded for each test was partialed out of the model as a procedural regressor to remove any effect of the excluded trials. Of interest was the difference in criterion for the low- and high-probability conditions for each participant, which would serve as a measure of adaptive criterion shifting (low−high). Optimal criterion was assessed using the following equation: $[\ln(P\text{noise}/P\text{signal})]/d'$ (Macmillan & Creelman, 2005).

While response bias and memory strength are independent processes according to signal detection theory, the measurement of criterion using C is related to d' in that movements between the signal and noise distributions can affect the apparent placement of a criterion even though the likelihood ratio of the responses has not changed (Macmillan & Creelman, 2005). In terms of individual differences in criterion shifting using C, these differences could be entirely accounted for by differences in d', with higher d' leading to lower shifts in criterion. Therefore, we used a criterion score that was normalized for d' in some analyses in which d' was not directly accounted for. A normalized criterion measure was computed on the basis of the residuals in each condition that accounted for the relation between criterion and d' (i.e., yielded from a regression analysis), in which the criterion grand mean was added to the residual for each participant. This normalized criterion was used only in the following: in Fig. 3, illustrating the individual variability; in the correlation of criterion shifting across tests; and in the final model of the five-step hierarchical regression as a measure of general shifting.

Reliance on cue information (RCI) Performance strategy was assessed in a questionnaire through free responses and a rating scale describing the degree to which memory judgment at test typically relied on the probability information (a high rating indicated more reliance on the cue information) or memory strength. For details on ratings, refer to the Supplemental Methods.

Regression analysis

A multistep, hierarchical regression was used to determine the effect of different factors on the tendency to shift criterion. This analysis was run separately for the word and the face data set. All regression analyses included a hierarchical model structure, but within each model, all variables were entered simultaneously.

The first step of the analysis was to reduce the number of inherent characteristic variables that were included in the final regression model. In this first step, linear regressions were run for each category of variables (see the Appendix), using the dependent variable of criterion shifting. In order to account for any mediating effects that might be due to procedural variations, d',[3] and RCI, these factors were entered first into each regression. Any inherent characteristic variable that had a significant ($p < .05$) standardized beta in the final model (for either the word or the face data set) above and beyond the other mediating variables was designated as a "key" variable and was used in the second step of the regression analysis. There were eight key variables that were significantly related to criterion shifting in either the word or the face data set. All variables were tested for collinearity using a tolerance threshold of 0.1 (Cohen, Cohen, West, & Aiken, 2003). We chose this method of variable reduction, rather than a factor analysis or principal components analysis, due to the exploratory nature of this study. We wanted to explore what specific variables had the greatest relation to criterion shifting, rather than examining a factor or component that could be a complex compilation of many variables and, thus, reduce the functionality of the results.

The second step of the regression analysis was designed to determine the total and direct effects of each variable, as well as the mediating effects of memory, RCI, and the inherent characteristics, as depicted in Fig. 1. To accomplish this, each key variable was analyzed in a four-model hierarchical regression predicting criterion shifting. The models were structured with the following variables included: (1) procedural variables (procedure variations and number of

---

[2] Since we have only a simple *old/new* response, we have only one point on an ROC curve. Therefore, we could not empirically derive a slope of the curve, so we utilized an equal variance model, instead of using an unequal variance model and assuming a slope. However, the results did not change depending on whether we used an equal variance model or an unequal variance model using a slope of .70 (as suggested by Macmillan & Creelman, 2005).

[3] The d' in the word regression analyses (all steps) was from the low-probability condition, and the d' in the face regression analyses (all steps) was from the high-probability condition. This choice was determined by the d' value that most strongly correlated with criterion shifting.

trials excluded) and the key variable; (2) added d' (see footnote 3); (3) added RCI; (4) added all the remaining key inherent characteristics. The total effect of the key variable was the zero-order correlation. The direct effect was the standardized beta from the final (fourth) model that included all variables. The direct effect reveals the strength of the relation between the key factor and criterion shift that cannot be accounted for by any other mediating factors. Mediating effects were determined by the difference of the key variable's standardized beta value across the different models: The mediating effect of memory was Model 1 Beta−Model 2 Beta; the mediating effect of RCI was Model 2 Beta−Model 3 Beta; and the mediating effect of the other inherent characteristics was Model 4 Beta−Model 3 Beta. Significance results of the total and direct effects are presented both corrected and uncorrected for multiple comparisons based on controlling for the 12 comparisons in the final model.

The third step in the regression analysis was to determine how much each set of variables (procedure variations, memory, RCI, inherent characteristics, and general criterion shifting) contributed to explaining the variance in criterion shifting. To do this, a five-model hierarchical regression was run with the following model breakdown: (1) procedural variables (procedure variations and number of trials excluded); (2) add d'; (3) add RCI; (4) add all key variables (military rank, age, sleep, caffeine, alcohol, verbal tendencies, negative affect, and a fun-seeking personality); and lastly, (5) add in the normalized criterion from the other memory test. The resulting R^2 changes indicated how much each set of variables accounted for the variance above and beyond that of the previous set of variables.

Results

Overview

To understand individual differences in criterion shifting, data were analyzed in several steps. First, data were analyzed on the individuals to compute their discrimination ability and their decision criterion in both the high- and low-probability conditions. Criterion shifting was the difference between criteria in the high- and low-probability conditions. Additional analyses and details exploring memory test performance can be found within the Supplemental Results. A series of regression analyses were then computed to examine individual differences in the tendency to shift criterion. First, a regression analysis on each category of variables was performed to identify the key factors within each category (e.g., age within the category of demographics and verbal ability within the category of cognitive style) that accounted for individual differences in criterion shifting. The

second step included the key variables within one hierarchical regression to examine total, direct, and mediating effects of each variable, including memory, RCI, and the inherent characteristics. The third step was a hierarchical regression to determine how much each category of variables (e.g., demographics, state of mind) accounted for individual differences in criterion shifting.
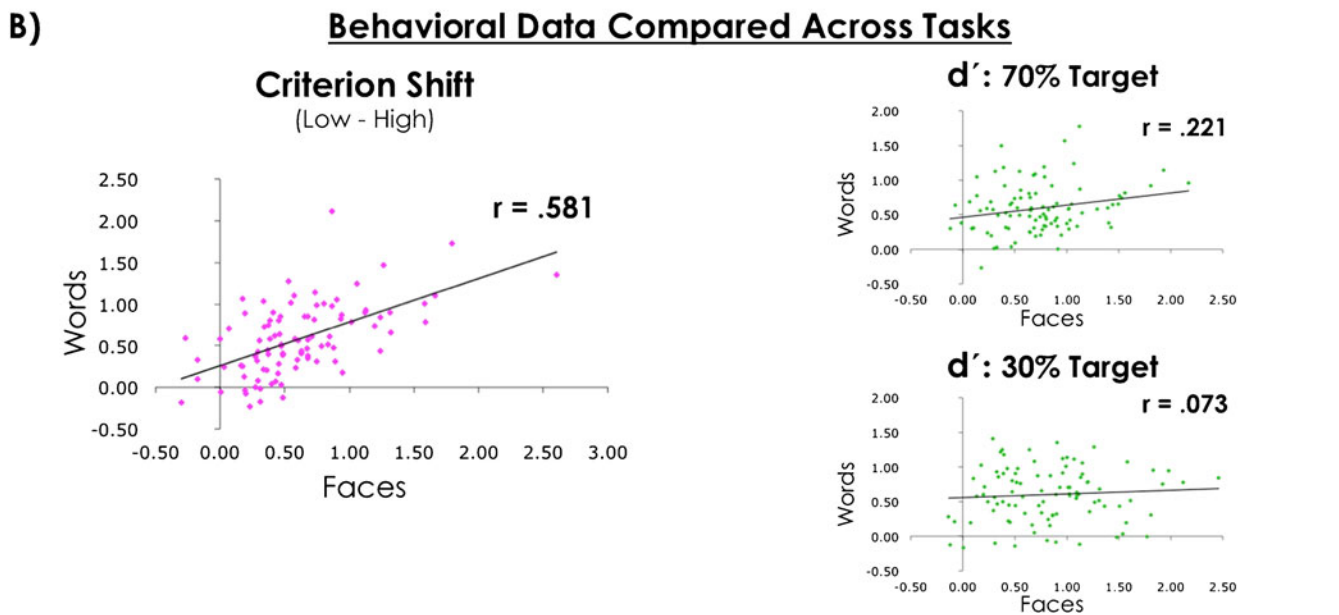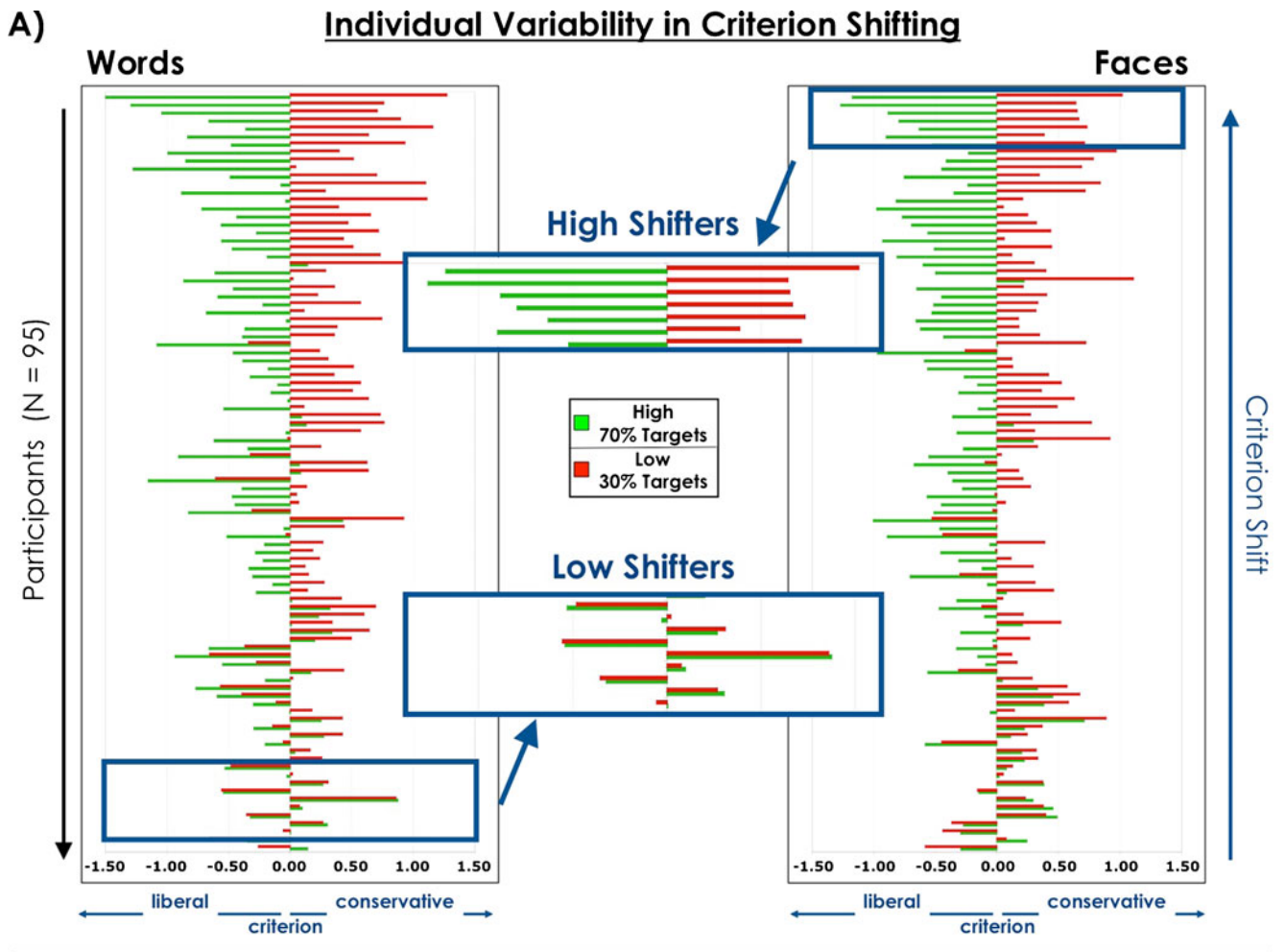
Memory test performance

Average performance was calculated for the high- and low-probability conditions for each separate test (i.e., faces and words; see Table 1). ANOVA results demonstrated that the main manipulation of target probability was successful, such that participants, on average, applied a more liberal criterion in the high-probability condition and a more conservative criterion in the low-probability condition [words, $F(1, 92) = 109.59$, $p < .001$; faces, $F(1, 92) = 87.72$, $p < .001$]. To examine criterion shifting regardless of memory performance, d' was used as a covariate in the ANOVA.

Although criterion shifting was observed across the different probability conditions, the extent needed to reach optimal performance was never achieved. Although optimal performance was never achieved, some participants were closer than others (for details, see the Supplemental Results). Figure 3a illustrates the variability across individuals in criterion shifting. As is highlighted in the figure, some participants were high shifters and shifted a lot, whereas others were low shifters and shifted minimally, if at all. For details on the criterion shift distribution, see the Supplemental Results.

Even though there was a large range of variability in criterion shifting across participants in each memory test, high consistency was found across tasks for the same individual. Participants who shifted criterion in one task were likely to shift criterion in the other task ($r = .581$, $p < .001$, correlation run on normalized C [see the Method section]; see Fig. 3b). Thus, participants maintained a stable level of

Table 1 Average group results (with standard errors)

Test	Words		Faces	
Condition	High	Low	High	Low
Hit Rate	0.74 (.01)	0.54 (.02)	0.72 (.01)	0.51 (.02)
False Alarm Rate	0.48 (.02)	0.26 (.02)	0.51 (.02)	0.29 (.01)
d'	0.73 (.05)	0.83 (.05)	0.59 (.04)	0.60 (.04)
Criterion	−0.32 (.04)	0.30 (.04)	−0.31 (.04)	0.28 (.04)
Criterion Corrected	−0.32 (.04)	0.30 (.04)	−0.31 (.04)	0.28 (.04)
Reaction Time	1.02 (.01)	1.03 (.01)	1.09 (.01)	1.10 (.01)
Strategy	2.42 (.09)		2.73 (.10)	

A) Individual Variability in Criterion Shifting

Words

Faces

Participants (N = 95)

High Shifters

Low Shifters

High 70% Targets

Low 30% Targets

Criterion Shift

-1.50 -1.00 -0.50 0.00 0.50 1.00 1.50

liberal   conservative
criterion

-1.50 -1.00 -0.50 0.00 0.50 1.00 1.50

liberal   conservative
criterion

B) Behavioral Data Compared Across Tasks

Criterion Shift
(Low - High)

r = .581

d′: 70% Target

r = .221

d′: 30% Target

r = .073

flexibility in decision criterion across tasks. In contrast, participant's recognition performance (i.e., d') was not strongly consistent across tasks (high-probability condition, $r = .221$, $p < .03$; low-probability condition, $r = .073$, n.s.). The

◄ **Fig. 3** Individual variability in criterion shifting. **A** Each pair of bars (green for the high condition, 70 % targets; red for the low condition, 30 % targets) represents the normalized criterion score from each of the 95 participants. The more liberal the criterion, the more negative the value (toward the left); the more conservative the criterion, the more positive the value (toward the right). Some participants shifted a lot between target conditions (e.g., high shifters), whereas others applied the same criterion regardless of target probability (e.g., low shifters). Amount of criterion shift is ordered as increasing from bottom to top. The data are from the words and the faces separately; please note that participants are ordered on the basis of amount of criterion shifting, which slightly varies across tasks. **B** The correlation of the tendency to shift across both memory tests, and the correlation of d' across both memory tests in the high- and low-probability conditions

selective stability in criterion shifting suggests an inherent strategy irrespective of recognition performance. Uncovering factors that describe individual variability could be fruitful in predicting future behavior.

Explaining individual variability: Regression analyses

Step 1: Variable reduction There were 50 variables describing each individual, which were analyzed in categorical regressions to identify the key variables related to criterion shifting. Eight variables were identified as significantly related to criterion shifting: demographics (military rank and age), state of mind (amount of sleep [determined by arrival time], caffeine consumption, and alcohol habits), cognitive style (verbal tendencies), and personality (a negative affect and a fun-seeking personality) (see Table 2). See Table S1 in the Supplemental Results for detailed results of each categorical regression.

Step 2: Hierarchical regression with key variables A four-model hierarchical regression using criterion shift as the dependent variable was performed for each factor. From this regression, the total, direct, and mediating effects could be calculated (Table 3). For both the word and the face data set, memory performance significantly explained the variance in criterion shifting even when all other effects were removed (i.e., the direct effect; words, beta = −.361, $p < .001$; faces, beta = −.320, $p < .001$). The greater the d', the less the criterion was shifted across conditions, which was the correct strategy indicated by the equation for reaching an optimal criterion shift. RCI also had a strong relation to criterion shifting (total effect: words, beta = .252, $p < .014$; faces, beta = .267, $p < .009$). However, this relation was largely mediated by memory performance, and thus the direct effect was not significant with the words but was significant with the faces (direct effect: faces, beta = .201, $p < .025$).

When examining the inherent characteristics of the individual, we found military rank to have a strong relation to

criterion shifting in both the word and the face data set (total effect: words, beta = .302, $p < .003$; faces, beta = .233, $p < .023$). The relation in the word data set was not mediated strongly by any other factors and remained significant in the direct effect (beta = .248, $p < .039$). In the face data set, the relation between rank and criterion shifting was mediated by other inherent characteristics, and although it still had a high direct effect (beta = .163), the effect was no longer significant. In order to examine what other characteristics mediated the effect of rank on criterion shifting in the face data set, we examined the correlations between the factors (Table S2). The correlation between rank and sleep was high, $r = −.544$, $p < .001$, and rank and verbal ability were highly correlated, $r = .264$, $p < .01$. Thus, these variables likely share explained variance that took away from the direct effect of rank in the face data set. In contrast, in the word data set, rank had a significant amount of unique explained variance, which resulted in a significant direct effect. For additional analysis regarding the relation of military rank to criterion shifting, refer to the Supplemental Results.

In examining the other inherent characteristics, although no other variables consistently reached significance in both the word and the face data sets, two personality variables had unmediated effects on criterion shifting: a negative affect and a fun-seeking personality. Having negative personality affect consistently decreased the likelihood of criterion shifting (words, beta = −.248, $p < .015$; faces, beta = −.110, n.s.). In contrast, a fun-seeking personality positively affected the tendency to shift (words, beta = .142, n.s.; faces, beta = .345, $p < .001$). Other variables had strong effects but were largely mediated by other variables (refer to the Supplemental Results for details).

Step 3: Explained variance by each set of variables In order to determine how much each set of variables adds to explaining the variance of criterion shifting, we ran a five-step hierarchical regression (see Table 4). Model 1, procedural variables, did not yield a significant R^2 value (words, .036; faces, .002). Model 2, d', did account for a significant amount of the variance (R^2: words, .145, $p < .001$; faces, .089, $p < .004$). Model 3, RCI, added to the explained variances in the face data set, $R^2 = .068$, $p < .008$, but not in the word data set, $R^2 = .011$. Model 4, inherent characteristics, yielded significant changes in R^2 (words, .147, $p < .03$; faces, .285, $p < .001$). Model 5, general shifting ability (i.e., criterion shifting from the other memory test), yielded a significant change in R^2 for both words (.178, $p < .001$) and faces (.176, $p < .001$). In total, 51.7 % ($p < .001$) of the variance could be explained in the word data set, and 62 % ($p < .001$) of the variance could be explained in the face data set.

Table 2 Standardized beta scores yielded from the categorical regressions for each of the key variables (bold indicates $p < .05$)

Category	Variable	Words	Faces
Demographic	military rank	**0.347**	**0.247**
	age	0.028	**0.251**
State of Mind	sleep (arrival time)	**−0.445**	−0.111
	caffeine	0.212	**0.296**
	alcohol	−0.003	**−0.266**
Cognitive Style	verbalizer (VVQ-W)	0.077	**0.278**
Personality	negative (PANAS)	−0.234	**−0.321**
	fun-seeking (BAS)	0.162	**0.461**

Discussion

The goal of this study was to observe individual variations in the tendency to shift a decision criterion during a recognition test and to explore various inherent characteristics that may mediate an individual's willingness to shift a decision criterion. We found that manipulating the probability that the test item was old (either 70 % or 30 %) effectively modulated the placement of the decision criterion such that, on average, a liberal criterion was applied in the high-probability condition and a conservative criterion was applied in the low-probability condition. However, the degree to which individuals shifted

their criteria between these two conditions was varied, with some individuals shifting a lot and others not shifting at all. This tendency to shift within an individual was found to be highly consistent across the two memory tasks and, therefore, likely reflects a general tendency in the flexibility of criterion shifting. Using a multistep regression analysis, we found that certain inherent characteristics have a strong effect on the tendency to shift criterion.

Optimal performance on a recognition test may be defined as maximizing correct responses (hits and correct rejections). In order to maximize correct responses when the base rate of targets is very high, the individual should utilize an extremely liberal response bias, and vice versa when the base rate of targets is very low (Green & Swets, 1966). Therefore, an optimal criterion, C, would be a direct function of the probability of the target and the individual's memory accuracy (the more accurate the memory, the less extreme the criterion needs to be to achieve optimal performance). The results of the present study demonstrate that while the individuals' behavioral performance, on average, indicated that they appropriately move toward the optimal criteria, very few actually reached the optimal criterion in any condition, and no individual reached an optimal criterion shift.

Suboptimal performance could, in part, be due to the psychological mechanisms underlying the criterion placement.

Table 3 Results of total, direct, and mediating effects from the hierarchical regression including all key variables. Total and direct effects are standardized beta values, with bold values indicating $p < .052$, * indicates that significance withstands Bonferroni correction for 12 comparisons. Mediating effects are differences in standardized betas across the different models

		Words					Faces				
		total effect	direct effect	med. char.	med. RCI	med. mem	total effect	direct effect	med. char.	med. RCI	med. mem
Procedure	variations	−.028	.103				−.035	−.106			
	excluded trials	.190	.136				.031	.049			
Memory	d′	**−.404***	**−.361***	.008	−.039		**−.299***	**−.320***	.033	−.017	
Strategy	RCI	**.252**	.093	.022		.115	**.267**	**.201**	.068		.019
Demographics	military rank	**.302***	**.248**	.080	.008	−.021	**.233**	.163	.066	.002	−.009
	age	−.038	−.014	.017	−.002	−.043	**.216**	**.181**	.059	−.025	−.003
State of Mind	sleep (arrival)	**−.235**	−.017	−.196	−.007	−.041	−.108	.073	−.216	.024	.002
	caffeine	.090	.075	.033	.010	−.012	.185	.158	−.006	.023	−.009
	alcohol	−.003	.038	.017	.011	−.066	−.163	−.110	−.058	−.002	.002
Cognitive style	verbalizer (VVQ-W)	.116	−.002	.108	.017	.000	**.308***	**.174**	.101	.001	−.008
Personality	negative (PANAS)	**−.248**	−.176	−.045	.013	−.027	−.110	−.055	−.081	.019	001
	fun-seeking (BAS)	.142	.126	.028	−.002	.013	**.329***	**.345***	.042	.037	−.015

Table 4 Variance accounted for by each set of variables; R^2 is the amount of variance accounted for over and above the preceding model. Bolded values indicate $p < .05$

	Words R^2	Faces R^2
Procedural Variations	.036	.002
Memory	**.145**	**.089**
Strategy	.011	**.068**
Individual Characteristics	**.147**	**.285**
General Shifting Ability	**.178**	**.176**
Total Explained	**.517**	**.620**

Given that our paradigm involved manipulating the probability of the targets, a probability-matching model would be most likely (Craig, 1976; Parks, 1966; Thomas & Legge, 1970; Vulkan, 2000). The model claims that participants set the criterion to respond *old* to match the probably of an old item occurring on the test. Only with perfect sensitivity would this achieve optimal performance. Typically, probability matching is the preferred, or most often chosen, strategy unless instructions indicate that another strategy may be more optimal (Koehler & James, 2009). Although this model may explain why individuals typically utilize a suboptimal criterion, it does not explain why some participants shift between extreme criterion placements, while other participants do not shift at all.

We suggest a model to explain the individual variations in criterion shifting on a recognition test (Fig. 1). In the model, we first account for variance in criterion shifting that may be due to individual differences in memory accuracy. Previous studies have shown that as the accuracy, or subjective accuracy, in memory strengthens, the range in biases narrows (Bruno et al., 2009; Stretch & Wixted, 1998). Therefore, our aim was to keep the overall d' relatively low in order to induce criterion shifts and to promote variance between individuals. The more one can rely on the strength of familiarity to discriminate old items from new items, the less one has to consider the base rate of old items. Indeed, our results indicated that the strongest direct effect across both memory tasks was individual differences in d'. Memory accuracy clearly has a direct effect on the tendency to shift criterion, but it may also be a mediating factor for other individual differences.

Our model highlights another important variable in the tendency to shift a criterion, and that is the individual's willingness to make the extra effort to shift a decision criterion. According to signal detection theory, a recognition decision involves two components: discriminating targets from distractors and establishing a criterion on the basis of the evidence for calling a test item a target (Green & Swets, 1966; Macmillan & Creelman, 2005). It has been suggested before that individuals are more focused on the discrimination

than they are on the placement of a criterion (Macmillan & Creelman, 2005; Wickelgren, 1967). Therefore, shifting a criterion requires a willingness to make that extra effort to pay attention to the placement of the criterion. Wixted and Stretch (2000) have suggested that individuals are generally unwilling to make that extra mental effort (particularly within a test run) but may do so if the conditions for shifting (e.g., changes in the base rate) are easily identifiable. We ensured that all participants were made aware of the probability conditions through practice and instruction and that changes in the conditions were easily identifiable (e.g., by changing the color of the font). Nevertheless, we hypothesized that some participants may be more willing to shift their criterion than others.

However, an individual's willingness is not readily measurable. Therefore, we treated willingness as a latent variable and examined a couple of measurable variables that may directly mediate an individual's willingness to shift. One of those variables is measured by the individual's reported reliance on the cue information. Even though participants may have been aware of the probability conditions, they may have chosen to ignore that information during the test. Indeed, a large proportion of the participants reported they did not use that information in their recognition judgments. We hypothesized that the more the participants relied on the probability (or cue) information, the more likely it was that there would be a tendency to shift criterion. This measure also indirectly indicated how much the participants followed instructions, with those that relied more heavily on the cue likely following the instructions more intently. Although reliance on cue information had a large total effect, the direct effect was much smaller in comparison, due to much of the variance being accounted for by memory performance (as seen in the word data set). In the face data set, the reliance on cue information explained a unique portion of variance that was exclusive from the variance accounted for by memory performance. In general, this reliance on probability information was expected to have a stronger direct effect, but this effect may have been dampened due to the self-report nature of the variable. Moreover, the reliance on cue had a stronger effect in the face data set than in the word data set. Memory performance was also worse for the faces than for the words. It may be that as memory performance declines and, possibly, as perceived difficulty of task increases, individuals rely more on cue information in order to shift their decision criterion, as is demonstrated in Bruno et al. (2009).

Another variable that may mediate an individual's willingness to shift a decision criterion includes a set of inherent characteristics that cause one individual to be more willing than another to make the extra mental effort necessary to shift a criterion. Indeed, we found a number of inherent characteristics that were strongly related to criterion shifting. A fun-

seeking personality trait and a negative personality had strong effects on the tendency to shift criterion. A personality with a negative affect is described as distressed and unpleasant, including mood states such as anger, contempt, disgust, guilt, fear, and nervousness (Watson et al., 1988). A larger negative affect was associated with less of a tendency to shift criterion across both memory tasks, although the effect was larger for the word task. The traits that describe a negative affect are typically thought of as ones that are ruminating traits and less cognitively flexible, which may explain the negative relation with criterion shifting. It is interesting, however, that mental health disorders associated with these same characteristics, such as depression and posttraumatic stress disorder, were not related to criterion shifting, which suggests that there is something unique that the negative affect captures beyond just these shared characteristics. Fun-seeking, an approach characteristic, is described as belonging to one who acts on the spur of the moment, tries new and different things because it will be fun, and craves excitement and new sensations (Carver & White, 1994). Fun-seeking was positively correlated with criterion shifting in both memory tasks, although the effect was stronger with the face data. In fact, the direct effect of a fun-seeking personality on criterion shifting during the recognition of faces was stronger than the direct effect of differences in d' for faces. One can speculate that the fun-seeking trait reflects a readiness to change behavior and, thus, lends itself to a flexibility needed for criterion shifting.

The explanation for the relationship between other inherent characteristics and criterion shifting is less clear but, nevertheless, intriguing. For example, military rank was strongly correlated with criterion shifting: The higher-ranking officers were more likely to shift criterion than the low ranking officers and nonmilitary controls. This positive effect was also found when tested with only the military participants. Moreover, other factors, such as age, combat experience, and years of education, could not account for this relation. It is interesting to speculate on why rank would have such a strong effect on criterion shifting. Is it possible that people achieve a higher rank because they are more adaptive and better shifters, or does the training necessary for the higher ranks cultivate and train people to be more adaptive? Future studies will be necessary to uncover what mediating characteristics within rank are most strongly associated with criterion shifting.

Interestingly, other factors, such as working memory capacity and years of education, did not affect the tendency to shift criterion. We expect that certain executive functioning skills will play a role in criterion shifting, through direct and mediating effects, but a full battery of executive skills was beyond the scope of this study. Nevertheless, the few skills we did measure were not related.

Moreover, some of the unexplained variance may be due to differences in the subjective values that individuals place on hits and false alarms. We calculated the optimal likelihood ratio at criterion by dividing the probability of distractors by the probability of targets (Green & Swets, 1966) without regard to the value of distractors and targets, since there were no payoffs. However, the subjective value of distractors and targets could easily vary among individuals, with some individuals valuing the acquisition of hits more than the avoidance of false alarms, while other individuals may have been the opposite and more risk aversive. Indeed, subjective value is a prominent component of behavioral and neural models underlying choice (Kable & Glimcher, 2009; Rangel, Camerer, & Montague, 2008), and it undoubtedly has a role in recognition memory as well. Future studies that can independently assess the subjective utilities of individuals may be able to account for more of the variance in criterion shifting. Furthermore, it is well know that manipulations in instructions, payoffs, and feedback can influence criterion shifting (Postma, 1999; Rhodes & Jacoby, 2007; Van Zandt, 2000). Future studies are needed to determine how individual differences in response to these other factors can explain individual differences in the tendency to shift criterion or, on the other hand, how these experimental procedures may override the relation between the inherent characteristics and criterion shifting highlighted in this study.

The theoretical construct of criterion shifting has played a prominent role in the explanation of a number of memory phenomena. Specifically, the idea that participants can vary the amount of memorial evidence that they are willing to accept in order to judge a test item as *old* has been used to explain the effects of misleading postevent information (McCloskey & Zaragoza, 1985), the high false alarm rate for critical lures in the DRM false memory paradigm (Miller & Wolford, 1999), the reduced false alarms via a distinctive heuristic (Gallo, Weiss, & Schacter, 2004), the increased false alarm rate with delayed text retrieval (Singer, Gagnon, & Richards, 2002; Singer & Wixted, 2006), and the effect of verbal overshadowing (Clare & Lewandowsky, 2004). In all of these studies, criterion shifts were used to argue against changes in the underlying memory trace or shifts in distributions of familiarity. However, the most influential account of criterion shifting may be its application to the mirror effect (Hockley & Niewiadomski, 2007; Ratcliff, Clark, & Shiffrin, 1990; Stretch & Wixted, 1998). The mirror effect is the well-known phenomenon that performance on new items mirrors performance on old items; that is, stimuli that elicit high hit rates also elicit low false alarm rates (Glanzer & Adams, 1985). At the time, this was used to eliminate any models of recognition that may have been based on a unidimensional model of memory strength. But various studies were able to demonstrate that shifts in decision criterion that might accompany changes

in memory strength could produce that same pattern of results (Hirshman, 1995; Hockley & Niewiadomski, 2007; Ratcliff et al., 1990; Stretch & Wixted, 1998).

Those early studies on the mirror effect, though, claimed that participants did not shift their criteria dynamically within a set of test items (Hirshman, 1995; Morrell et al., 2002; Stretch & Wixted, 1998), but more recent studies have demonstrated that participants will shift criteria dynamically within a test set given the right conditions (Bruno et al., 2009; Dobbins & Kroll, 2005; Hockley & Niewiadomski, 2007; Postma, 1999; Rhodes & Jacoby, 2007; Singer et al., 2002; Singer & Wixted, 2006; Van Zandt, 2000; Verde & Rotello, 2007). We found that, even under conditions that promote criterion shifts within a recognition test, some participants will shift appropriately, while others will not shift at all. All participants were made fully aware of the base rate conditions, but an individual's willingness to shift varies considerably. We found a number of inherent characteristics that may mediate an individual's willingness to shift a decision criterion that could affect his or her optimal performance. Future models of recognition performance will need to consider these individual differences. Moreover, individual variability in criterion shifting must be taken into account when considering models of memory. Some individuals can dynamically shift, while others do not.

Appendix Complete list of variables (*key variable)

Variables entered first in all regression analyses
Procedure
Procedural variations
Number of trials excluded
Memory
d': Low probability for the word data set
d': High probability for the face data set
Strategy
Reliance on cue information (RCI)
Variable categories
Demographics
Age*
Gender
Years of education
Handedness
Military rank* [Rank of military officers ranged from Staff Sergeants (E6) through Majors (O4). Rank was coded

in a numerical scale reflecting the military hierarchy, and nonmilitary participants were coded as a zero.]
Performance
Reaction time: high-probability faces
Reaction time: high-probability words
Reaction time: low-probability faces
Reaction time: low-probability words
State of mind
Arrival time*
Scan time
Hours of sleep night before
Hours since last meal
Exercise during current day
Caffeine consumption*
Smoking habits
Alcohol habits*
Mental comfort/anxiety in MRI session
Physical comfort in MRI session
Cognitive
Paper-folding task
Card rotation task
Vocabulary task
OSIQ-S: spatial visual ability
OSIQ-O: object visual ability
VVQ-W: verbal cognitive style*
VVQ-P: visual cognitive style
Need for cognition
SBCSQ: visual ability
SBCSQ: verbal ability
Mental health
BDI: depression
Posttraumatic stress disorder
History of concussion (lifetime)
History of concussion (within 5 years)
Personality
PANAS: shyness
PANAS: fatigue
PANAS: serenity
PANAS: surprise
PANAS: positive
PANAS: negative*
BIS: inhibition
BAS: reward
BAS: fun-seeking*
EPQR: psychoticism
EPQR: lying
Big 5: conscientiousness
Big 5: agreeableness
Big 5: openness
Extraversion: z score average of the EPQR and Big 5 scores
Neuroticism: z score average of the EPQR and Big 5 scores
Military
Combat experience

Time in army
Length of deployment
Time since last deployment

References

American Academy of Neurology. (1997). Practice parameter: The management of concussion in sports (summary statement) (Report of the Quality Standards Subcommittee). *Neurology, 48,* 581–585.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory measuring depression. *Archives of General Psychiatry, 4,* 561–571.

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

Blajenkova, O., Kozhevnikov, M., & Motes, M. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology, 20,* 239–265.

Brewin, C. R., Rose, S., Andrews, B., Green, J., Tata, P., McEvedy, C., & Foa, E. B. (2002). Brief screening instrument for post-traumatic stress disorder. *The British Journal of Psychiatry, 181,* 158–162.

Bruno, D., Higham, P. A., & Perfect, T. J. (2009). Global subjective memorability and the strength-based mirror effect in recognition memory. *Memory & Cognition, 37,* 807–818.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116–131.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67,* 319–333.

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 739–755.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Craig, A. (1976). Signal recognition and the probability-matching decision rule. *Perception & Psychophysics, 20,* 157–162.

Curran, T., DeBuse, C., & Leynes, P. A. (2007). Conflict and criterion setting in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 2–17.

Dobbins, I. G., & Kroll, N. E. (2005). Distinctiveness and the recognition mirror effect: Evidence for an item-based criterion placement heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1186–1198.

Ekstrom, R. B., French, J. W., & Harmann, H. H. (1976). *Manual for kit of factor referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck personality questionnaire.* London: Hodder & Stoughton.

Eysenck, S., Eysenck, H., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences, 6,* 21–29.

Gallo, D. A., & Roediger, H. L., III. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47,* 469–497.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language, 51,* 473–493.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20.

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

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 302–313.

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 honor of Bruce Whittlesea* (pp. 155–166). Houndmills: Palgrave Macmillan.

Hockley, W. E., & Caron, A. M. (2007, November). *Opposing strength-based mirror effects for words versus pictures: Evidence for within-list criterion changes.* Paper presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA.

Hockley, W. E., & Niewiadomski, M. W. (2007). Strength-based mirror effects in item and associative recognition: Evidence for within-list criterion changes. *Memory & Cognition, 35,* 679–688.

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford.

Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: Consensus and controversy. *Neuron, 63,* 733–745.

Keane, T., Fairbank, J., Caddell, J., Zimering, R., Taylor, K., & Mora, C. (1989). Clinical evaluation of a measure to assess combat exposure. *Psychological Assessment, 1,* 53–55.

Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition, 113,* 123–127.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Martinez, A. M., & Benavente, (1998). The AR Face Database. CVC Technical Report #24.

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology, 95,* 833–846.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology. General, 114,* 1–16.

Miller, M. B., Guerin, S. A., & Wolford, G. L. (2011). The strategic nature of false recognition in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1228–1235.

Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review, 106,* 398–405.

Minear, M., Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behaviour Research Methodology Instrumentation Computer, 36,* 630–633.

Morrell, H. E., Gaitan, S., & Wixted, J. T. (2002). On the nature of the decision axis in signal-detection-based models of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1095–1110.

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

Postma, A. (1999). The influence of decision criteria upon remembering and knowing in recognition memory. *Acta Psychologica, 103,* 65–76.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience, 9,* 545–556.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: 1. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.

Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate,

awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 305–320.

Richardson, A. (1977). Verbalizer–visualizer: A cognitive style dimension. *Journal of Mental Imagery, 1,* 109–126.

Robinson, K. J., & Roediger, H. L., III. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8,* 231–237.

Robinson, M. D., Wilkowski, B. M., Kirkeby, B. S., & Meier, B. P. (2006). Stuck in a rut: Perseverative response tendencies and the neuroticism–distress relationship. *Journal of Experimental Psychology. General, 135,* 78–91.

Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001). Spreading activation and the arounsal of false memories. In H. L. Roediger III, J. S. Naine, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 3–16). Washington, DC: American Psychological Association Press.

Roediger, H. L., III, & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science, 9,* 123–127.

Samaria, F. & Harter, A. (1994). "Parameterisation of a stochastic model for human face identification". 2nd IEEE Workshop on Applications of Computer Vision. Sarasota (Florida).

Singer, M. (2009). Strength-based criterion shifts in recognition memory. *Memory & Cognition, 37,* 976–984.

Singer, M., Gagnon, N., & Richards, E. (2002). Strategies of text retrieval: A criterion shift account. *Canadian Journal of Experimental Psychology, 56,* 41–57.

Singer, M., & Wixted, J. T. (2006). Effect of delay on recognition decisions: Evidence for a criterion shift. *Memory & Cognition, 34,* 125–137.

Smillie, L. D., Cooper, A. J., Tharp, I. J., & Pelling, E. L. (2009). Individual differences in cognitive control: The role of psychoticism and working memory in set-shifting. *British Journal of Psychology, 100,* 629–643.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 1379–1396.

Thomas, E. A. C. (1975). Criterion adjustment and probability matching. *Perception & Psychophysics, 18,* 158–162.

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

Van Zandt, T. (2000). ROC curves and confidence judgements in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 582–600.

Verde, M., & Rotello, C. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition, 35,* 254–262.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428,* 748–751.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys, 14,* 101–118.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule—expanded form.* Iowa City: University of Iowa.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54,* 1063–1070.

Weyrauch, B., Huang, J., Heisele, B., Blanz, V. (2004). Component-based face recognition with 3D morphable models. First IEEE Workshop on Face Processing in Video, Washington D.C.

Wickelgren, W. A. (1967). Strength theories of disjunctive visual detection. *Perception & Psychophysics, 2,* 331–337.

Wickens, T. D., & Hirshman, E. (2000). False memories and statistical design theory: Comment on Miller and Wolford (1999) and Roediger and McDermott (1999). *Psychological Review, 107,* 377–383.

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review, 107,* 368–376.