Title

Who Gives a Criterion Shift? Behavioral and Neural Mechanisms of a Stable Cognitive Trait

Author

Layher, Evan

Publication Date

2021

UNIVERSITY OF CALIFORNIA

Santa Barbara

Who Gives a Criterion Shift?

Behavioral and Neural Mechanisms of a Stable Cognitive Trait

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Psychological and Brain Sciences

by

Evan A. Layher

Committee in charge:
Professor Michael B. Miller, Chair
Professor Michael Gazzaniga
Professor Barry Giesbrecht
Professor Emily Jacobs

March 2022

The dissertation of Evan A. Layher is approved.

Michael Gazzaniga

Barry Giesbrecht

Emily Jacobs

Michael B. Miller, Committee Chair

December 2021

Who Gives a Criterion Shift?

Behavioral and Neural Mechanisms of a Stable Cognitive Trait

by

Evan A. Layher

ACKNOWLEDGEMENTS

Many people helped me throughout graduate school, and I am very grateful to have met and worked with so many great individuals—my successes in graduate school would not be possible without all the help I received over the years! Mike Miller, thank you for accepting me into the program and giving me an opportunity to pursue *many* research projects in your lab. I really appreciate your help and guidance and always felt you genuinely cared about my future and well-being. Thank you to all my former and current lab mates for being there anytime I needed help. Justin Kantner, although our time in the Miller lab never overlapped, you first introduced me to criterion shifting research and am glad that the introduction transformed into a full-fledged dissertation. Nikki Marinsek, you set the bar high for defining a successful graduate student in the Miller lab and your hard work and creativity inspired me to push myself early in the program. Ben Turner, your statistical expertise proved very useful, and I appreciate your help during my first couple of years as a graduate student. Tyler Santander, thank you for filling the void when Nikki graduated, and Ben moved halfway across the world—your guidance and assistance helped me out *tremendously* and I am a much better researcher because of you. Courtney Durdle and Sara Leslie, I am grateful that you joined the lab and willing collaborated with me on research projects and put trust in me to provide you with guidance early on in your graduate school careers. Christina Boardman, you are the workhorse that is keeping EEG research alive in the Miller lab and I greatly appreciate your efforts towards that. Selin Bekir, Luna Li, and Henri Skinner, I know the lab will be in good hands and believe you will all be successful throughout graduate school, thank you for making my last couple of years in the program more special.

One aspect that I will always cherish about my graduate school experience is the collaborative nature amongst researchers in the department. I am grateful to have collaborated with Miguel Eckstein's lab on an fMRI project in which some of the findings became Experiment 6 in this dissertation (thank you Miguel and Puneeth Chakravarthula). The Barry Giesbrecht lab proved extremely helpful with the dumpster-fire that is EEG/fMRI and the subsequent implementations of much more manageable EEG-only experiments (thank you Barry, Tom Bullock, Diya Das, James Elliott, Jordan Garrett, Mary MacLean, and Anne Milner). Emily Jacob's lab opened my eyes to the necessity of studying hormonal effects on the brain, I greatly appreciate being able to collaborate on many dense-sampling fMRI projects and learning so much from the lab (thank you Emily, Laura Pritschet, Caitlin Taylor, and Shuying [Shuuuuuuuying!] Yu). Scott Grafton's lab provided technical and emotional support (being the basement neighbor to the Miller lab), thank you for being available to answer my questions, train me on neurostimulation (though you may not want to take credit for that), and provide me with guidance on statistics and graph-making in R [special shout out to Allison Shapiro] (thank you Scott, Allison, Viktoriya Babenko, Neil Dundon, Michelle Marneweck, Gold Okafor, Liz [The COAT] Rizor, and Lukas Volz). Thank you to the countless research assistants from all the aforementioned labs who assisted me with experimental design, data collection, and analyses. I truly could not have collected so much data without the dozens of research assistants who helped me over the years.

Other individuals I want to thank include Mario Mendoza for the countless hours of help in the Brain Imaging Center over the years and being the best phlebotomist that the department has to offer—I am very thankful for all your time and assistance with various MRI projects. Mike Gazzaniga, thank you for working with me on writing projects during my first couple of years in the program, I learned a lot from you. Craig Abbey, thank you for teaching me maximum likelihood estimation, which provides much better fits to my ROC curves. Madeleine Gross, I am thankful to have met you in stats class during my first year in the program—you proved to be *immensely* helpful in my final push to collect data for my dissertation and I will always be grateful for that.

I also want to thank my Ph.D. cohort who partook in the journey of graduate school with me. Youngki Hong, you are a *great* friend, and I am glad to have lived with you for most of my graduate school career—thank you for always being there for me. Spencer Mermelstein and honorary cohort member Maggie Cazares, thank you for feeding me throughout graduate school—I will always value our friendship.

Lastly, thank you to my family for always supporting me no matter what. Without your support I definitely would have chosen a different career path before even beginning a Ph.D program, but I am grateful for your continued encouragement. When things got tough, you always knew that situations would eventually work themselves out and that new opportunities were always on the horizon. I love you all.

VITA OF EVAN A. LAYHER
JANUARY 2022

EDUCATION

Bachelor of Arts and Science in Psychology and Neurobiology, Physiology, & Behavior, University of California, Davis, September 2012 (with honors)

Doctor of Philosophy in Psychological and Brain Sciences, University of California, Santa Barbara, January 2022

PROFESSIONAL EMPLOYMENT

2015-2021: Teaching Assistant, Department of Psychological and Brain Sciences, University of California, Santa Barbara

2012-2015: Junior Specialist, Department of Psychiatry, University of California, Davis

PUBLICATIONS

Layher, E., Dixit, A., & Miller, M. B. (2020). Who gives a criterion shift? A uniquely individualistic cognitive trait. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2075–2105.

Taylor, C. M., Pritschet, L., Olsen, R. K., **Layher, E.**, Santander, T., Grafton, S. T., & Jacobs, E. G. (2020). Progesterone shapes medial temporal lobe volume across the human menstrual cycle. *NeuroImage*, *220*, 117125.

Pritschet, L., Santander, T., Taylor, C. M., **Layher, E.**, Yu, S., Miller, M. B., Grafton, S. T., & Jacobs, E. G. (2020). Functional reorganization of brain networks across the human menstrual cycle. *NeuroImage*, *220*, 117091.

Layher, E., Santander, T., Volz, L. J., & Miller, M. B. (2018). Failure to Affect Decision Criteria During Recognition Memory With Continuous Theta Burst Stimulation. *Frontiers in Neuroscience*, *12*, 1–14.

Ragland, J. D., **Layher, E.**, Hannula, D. E., Niendam, T. A., Lesh, T. A., Solomon, M., Carter, C. S., & Ranganath, C. (2017). Impact of schizophrenia on anterior and posterior hippocampus during memory for complex scenes. *NeuroImage: Clinical*, *13*, 82–88.

Ragland, J. D., Ranganath, C., Harms, M. P., Barch, D. M., Gold, J. M., **Layher, E.**, Lesh, T. A., MacDonald, A. W., Niendam, T. A., Phillips, J., Silverstein, S. M., Yonelinas, A. P., & Carter, C. S. (2015). Functional and Neuroanatomic Specificity of Episodic Memory Dysfunction in Schizophrenia. *JAMA Psychiatry*, *72*(9), 909-916.

AWARDS

Los Angeles Clippers Community Hero, 2021
Graduate Student Fellow, Kavli Summer Institute in Cognitive Neuroscience, 2017 & 2018
Graduate Research Fellowship Honorable Mention, 2016

ABSTRACT

Who Gives a Criterion Shift?

Behavioral and Neural Mechanisms of a Stable Cognitive Trait

by

Evan A. Layher

Individuals should *strategically* shift decision criteria when there are disproportionate likelihoods or consequences for falsely identifying versus missing target items. Despite being explicitly aware of the advantages for criterion shifting, people on average do not shift extremely, leading many theories to conclude that people are generally suboptimal at placing decision criteria. However, assessments of individual differences reveal that some people do shift criteria quite well while others fail to shift entirely. These individual differences are remarkably consistent across time, tasks, and decision domains, yet cannot be adequately explained by other cognitive or personality measures—the degree to which people shift a criterion is a stable, uniquely individualistic cognitive trait. Individuals who inadequately shift criteria are *capable* of shifting to greater extents but appear *unwilling* to do so. Understanding criterion shifting tendencies at the individual level is vital since assessments of group averages fail to capture the true nature of this behavior. These individual differences carry important implications for investigating the neural mechanisms that underly the placement of a decision criterion.

The role of the decision criterion is often neglected in neuroimaging studies. For instance, widespread frontoparietal activity is consistently observed in recognition memory tests that compare studied (“target”) versus unstudied (“nontarget”) responses. However, there are conflicting accounts that ascribe various aspects of frontoparietal activity to mnemonic evidence versus decisional processes. According to Signal Detection Theory, recognition judgments require individuals to *decide* whether the memory strength of an item exceeds a decision criterion for reporting previously studied items. Yet, most fMRI studies fail to manipulate both memory strength *and* decision criteria, making it difficult to appropriately identify frontoparietal activity associated with each process. Systematic manipulations of criteria and discriminability revealed that maintaining a conservative versus liberal decision criterion drastically affects frontoparietal activity in target versus nontarget response contrasts, whereas changes in discriminability showed virtually no differences. Findings from dense-sampling fMRI data revealed multiple frontoparietal networks associated with inhibiting prepotent responses whereas the default mode network is relatively more engaged when participants *provide* a prepotent response. This supports a response bias account of recognition memory indicating that widespread frontoparietal activity observed during recognition memory tests is largely attributable to decisional processes. Attempts to modulate decision criteria using neurostimulation have unfortunately failed to provide a causal link between frontoparietal activity and criterion placement, despite the robust fMRI correlates suggesting that such a relationship should exist.

TABLE OF CONTENTS

Introduction 1

Chapter I: Criterion shifting is a uniquely individualistic cognitive trait 8

- A. Experiments 1 & 2: Test-retest reliability of criterion shifting during recognition memory 18
- B. Experiment 3: Individual characteristics and generalizability of criterion shifting 35
- C. Experiments 4 & 5: Differentiating criterion shifting from confidence ratings 52

Chapter II: Neural correlates underlying the decision criterion 85

- A. Experiment 6: Decision criteria greatly affects fMRI activity, not discriminability ... 89
- B. Experiments 7 & 8: Dense-sampling fMRI reveals dissociable criterion networks .. 116

Chapter III: Failed attempts to modulate decision criteria via neurostimulation 152

- A. Experiments 9 & 10: TMS of prefrontal cortex fails to modulate decision criteria .. 155
- B. Experiment 11: Diffuse tDCS across the insula fails to modulate decision criteria .. 179

Chapter IV: General Discussion 190

LIST OF FIGURES AND TABLES

Figure 1. Experiment 1 recognition memory task (payment manipulation) 23

Figure 2. Experiment 2 recognition memory task (base rate manipulation) 25

Figure 3. Experiment 1 Pearson correlations of criterion shifting across 10 sessions 28

Figure 4. Experiment 1 individual differences in criterion shifting across 10 sessions 29

Figure 5. Experiment 2 Pearson correlations of criterion shifting across 10 session 32

Figure 6. Experiment 2 individual differences in criterion shifting across 10 sessions 33

Figure 7. Experiment 3 recognition memory and visual detection tasks 40

Figure 8. Experiment 3 individual differences in criterion shifting across decision domains . 48

Figure 9. Experiments 4 and 5 recognition memory tasks with confidence ratings 58

Figure 10. Experiment 4 individual differences in criterion shifting across conditions 62

Figure 11. Experiment 5 individual differences in criterion shifting across conditions 65

Figure 12. Experiment 6 recognition memory and visual detection fMRI tasks 93

Figure 13. Experiment 6 Pearson correlations across decision domains 98

Figure 14. Experiment 6 whole-brain recognition memory T > NT response contrasts 100

Figure 15. Experiment 6 whole-brain visual detection T > NT response contrasts 101

Figure 16. Experiment 6 mean beta values across 12 ROIs in T > NT response contrasts .. 106

Figure 17. Experiment 6 mean beta values across 12 ROIs in T > NT item contrasts 107

Figure 18. Experiment 6 parameter estimates for mean beta values 108

Figure 19. Experiments 7 and 8 depiction of idealized SDT models across 16 conditions ... 117

Figure 20. Experiment 7 recognition memory fMRI task 123

Figure 21. Experiment 7 ROC curves 125

Figure 22. Experiment 7 whole-brain covariate T > NT response and item contrasts 128

Figure 23. Experiment 8 recognition memory fMRI task..138

Figure 24. Experiment 8 ROC curves..140

Figure 25. Experiment 8 whole-brain recognition memory T > NT response contrasts 143

Figure 26. Experiment 8 whole-brain recognition memory T > NT item contrasts 144

Figure 27. Experiment 8 whole-brain covariate T > NT response and item contrasts 144

Figure 28. Experiments 9 and 10 recognition memory rTMS task.......................................159

Figure 29. Experiments 9 and 10 rTMS research paradigm 163

Figure 30. Experiments 9 and 10 whole-brain recognition memory H > CR contrast 165

Figure 31. Experiments 9 and 10 rTMS stimulation sites 167

Figure 32. Experiments 9 and 10 mean criterion values..168

Figure 33. Experiments 9 and 10 mean discriminability values...169

Figure 34. Experiments 9 and 10 parameter estimates for discriminability values...............171

Figure 35. Experiments 9 and 10 parameter estimates for criterion values..........................172

Figure 36. Experiment 11 tDCS current flow model...184

Figure 37. Experiment 11 mean SDT values...186

Figure 38. Experiment 11 parameter estimates of SDT values 187

Table 1. Experiment 1 behavioral SDT means 26

Table 2. Experiment 2 behavioral SDT means 30

Table 3. Experiment 3 behavioral SDT means and Pearson correlations...............................49

Table 4. Experiment 3 Pearson correlations of criterion shifting with extraneous measures..51

Table 5. Experiment 6 behavioral SDT means 97

Table 6. Experiment 6 fMRI local maxima in T > NT response contrasts 101

Table 7. Experiment 6 parameter estimates for mean beta values..108

Table 8. Experiment 7 behavioral SDT means 124

Table 9. Experiment 7 fMRI local maxima in covariate T > NT contrasts 128

Table 10. Experiment 8 behavioral SDT means 140

Table 11. Experiment 8 fMRI local maxima in covariate T > NT contrasts 145

Table 12. Experiments 9 and 10 fMRI local maxima in H > CR contrast............................165

Table 13. Experiments 9 and 10 parameter estimates of discriminability values.................172

Table 14. Experiments 9 and 10 parameter estimates of criterion values 174

Table 15. Experiment 11 parameter estimates of SDT values..187

Introduction

Strategic criterion shifting occurs when a person *knowingly* alters a decision strategy when the known prevalence of a target item changes or when the relative rewards or consequences of different response types change. Shifting decision criteria can improve decisional outcomes, particularly when there is uncertainty in the detected signal. A common example of this is when you see a person who looks familiar (the signal) but are unsure whether you know them (a target) or not (a nontarget). The ideal goal is to greet a known acquaintance (a hit) and ignore a stranger (a correct rejection), but the uncertainty in your memory prevents you from knowing the correct course of action. Fortunately, there usually is other information at your disposal that can help minimize the chances of either potentially greeting a stranger (a false alarm) or failing to greet a known acquaintance (a miss). For instance, if you believe the person is a co-worker and you are in the workplace, then you should establish a *liberal* criterion by greeting the person even when your memory is vague, since the chances of such an encounter are high (a strategy to avoid *misses*). However, if you are on vacation in Tahiti you should establish a *conservative* criterion by only greeting a potential co-worker when your memory is strong since this encounter is much less likely to occur (a strategy to avoid *false alarms*). Remarkably, in situations where criterion shifting is *clearly* advantageous, some individuals will readily shift decision criteria while others fail to shift entirely, which can detrimentally impact decisional outcomes (Aminoff et al., 2012, 2015; Kantner et al., 2015; Frithsen, et al., 2018; Layher et al., 2018; Miller & Kantner 2019). Extreme variability in criterion shifting across participants is well-documented, but currently no studies have systematically characterized the stability of criterion shifting tendencies *within* individuals over time. Yet, stable differences in criterion shifting tendencies across individuals may represent a fundamental aspect of those individual's decision-making strategies and carry theoretical implications for signal detection models of recognition memory.

Analyses of criterion shifting reported here focus on individual differences, which can reveal aspects of data that may contradict previous hypotheses that draw conclusions from group averages (Miller & Kantner, 2019). For example, a longstanding observation of group-averaged data shows that people are generally suboptimal[1] at placing a criterion (i.e. people do not shift criteria extreme enough given the circumstances), leading to several hypotheses that attempt to explain this phenomenon (Ulehla, 1966; Parks 1966; Thomas & Legge, 1970; Kubovy, 1977; Hirshman, 1995; Maddox & Bohil, 2005; Benjamin et al., 2009; Lynn & Barret, 2014). One hypothesis advocates that participants will probability match during test blocks that include a base rate manipulation (Parks 1966; Thomas & Legge, 1970). That is, if 70% of items are targets, participants will respond "target" 70% of the time, even though the best strategy for maximizing accuracy is to always respond "target" unless there is strong evidence that an item is a nontarget. Aminoff and colleagues (2012) employed a base rate manipulation during recognition memory tests where participants received explicit instructions informing them that target (previously studied) items would appear either 70%

---

[1] The term "optimal" performance is meant to describe a criterion that maximizes payoffs or the proportion correct at any level of discriminability, given the assumptions of signal detection theory (see Macmillan & Creelman, 2005). While the classification of an "optimal" criterion will vary depending on the theoretical model (see Lynn & Barret, 2014), the underlying claim that strategic criterion shifting is a stable cognitive trait is unrelated to whether certain individuals actually implement a model's definition of "optimal" criteria. This simple device is implemented to demonstrate the inherent disadvantages of not shifting a criterion in response to changes in payoffs or base rates (i.e. when discriminability is held constant, individuals who appropriately shift criteria will achieve better outcomes, in regards to the intended goals of the task, than those who do not shift criteria).

(liberal condition) or 30% (conservative condition) of the time. Group-averaged results from this study suggested that probability matching is indeed a plausible explanation. However, when examining the data at an individual level, this hypothesis seems less plausible because some individuals actually do shift criteria quite well (i.e. almost always respond “target” in the liberal condition and almost never respond “target” in the conservative condition), while others fail to shift entirely (i.e. respond “target” at equal rates across both criterion conditions). It is necessary to account for these individual differences in order to gain a full understanding of the nature of criterion shifting tendencies. In **Chapter I**, five experiments are presented to characterize criterion shifting tendencies at the individual level.

The decision criterion is oftentimes neglected when considering neural mechanisms underlying decisional evidence, such as in recognition memory. Neuroimaging studies of recognition memory have revealed widespread frontoparietal activity associated with contrasts comparing studied (“target”) versus unstudied (“nontarget”) responses. Some attribute these patterns of activity to mnemonic evidence, given that “target” responses confer greater memory strength on average than “nontarget” responses (Wagner et al., 2005; Vilberg & Rugg, 2009; Criss et al., 2013; Gilmore et al., 2015; McDermott et al., 2017). Others argue that frontoparietal activity is associated with decisional processes since recognition judgments require individuals to *decide* whether items are “targets” versus “nontargets” (O’Connor et al., 2010; Jaeger et al., 2013; Aminoff et al., 2015; King & Miller, 2017; Kim, 2020). In a Signal Detection Theory (SDT) framework, recognition memory judgments encompass both evidential *and* decisional processes—participants must determine whether the memory strength elicited by an item is strong enough (i.e. exceeds the decision criterion) to warrant a “target” response (Macmillan and Creelman, 2005). However, most neuroimaging studies of recognition memory fail to manipulate both memory strength *and* decision criteria, making it difficult to determine which aspects of frontoparietal activity are associated with mnemonic evidence versus decisional processes. Three experiments reported in **Chapter II** investigate fMRI neural correlates associated with decision criteria and memory strength at both the group and individual level.

One limitation of fMRI findings is that results are strictly correlational. To provide a *causal* link between criterion shifting and neural activity, **Chapter III** discusses three experiments that target the right prefrontal cortex (PFC) through various neurostimulation techniques in attempts to alter criterion shifting performance. There is evidence that suggests the PFC plays a role in maintaining a conservative decision criterion. Patients with frontal lobe lesions tend to establish more liberal decision criteria as evidenced by increased false alarm rates during recognition memory (Parkin et al., 1996; Schacter et al., 1996; Swick & Knight, 1999; Verfaellie, et al., 2004; Callahan, et al., 2011; Biesbroek, et al., 2014). A tendency to set liberal decision criteria is also observed in other patient populations associated with frontal lobe damage or dysfunction, including Alzheimer's disease (Budson, el al., 2006; Waring et al., 2008; Beth, et al., 2009; Deason et al., 2017) and schizophrenia (Moritz et al., 2008). PFC processes can also be disrupted through drug administration, such as with **Δ**9-tetrahydrocannabinol (THC) (Bossong et al., 2012), which demonstrated increased false alarm rates during recognition memory (Doss et al., 2018). Taken together, these studies strongly suggest that a dysfunctional PFC impairs the ability to set conservative criteria. Therefore, regions within the right PFC were targeted to attempt to make individuals implement a more liberal criterion, which would provide a causal link between the PFC and maintaining a conservative criterion. In sum, the 11 reported experiments provide a comprehensive report on individual differences in criterion shifting behavior and the neural mechanisms that underly maintaining a conservative versus liberal criterion during recognition memory.

Method

Participant recruitment

Participants across the 11 experiments enrolled in the experiments via the University of California Santa Barbara (UCSB) paid research participation website. The experiments received approval from the UCSB Human Subjects Committee Institutional Review Board (IRB) or Western IRB (Experiment 11), and all participants provided written informed consent.

Signal detection theory

Unless otherwise specified, data analyses implemented an equal-variance SDT model to compute discriminability (d'), criterion placement (c), and criterion shifting (C) (Macmillan & Creelman, 2005). For each test condition, summation of the number of hit (H), miss (M), correct rejection (CR), and false alarm (FA) trials allowed for computations of hit rate (HR), false alarm rate (FAR), percent correct (PC), and SDT measures through the following equations:

$$HR = \mathrm{H} / (\mathrm{H} + \mathrm{M})$$

$$FAR = \mathrm{FA} / (\mathrm{CR} + \mathrm{FA})$$

$$d' = \mathrm{z}(HR) - \mathrm{z}(FAR)$$

$$c = -0.5 * [\mathrm{z}(HR) + \mathrm{z}(FAR)]$$

$$C = c(\text{conservative}) - c(\text{liberal})$$

$$PC = (\text{H} + \text{CR}) / (\text{H} + \text{M} + \text{CR} + \text{FA}),$$

where z represents the density of the standard normal distribution (Macmillan & Creelman, 2005; Stanislaw & Todorov, 1999). To prevent infinite normalized values, rare occurrences of *HR*s and *FAR*s of 0% and 100% were adjusted by adding or subtracting, respectively, 1 divided by the total number of trials within a test condition (see Macmillan & Kaplan, 1985). Equal-variance SDT models assume that the variance of the target and lure distributions are equal. However, recognition memory experiments reveal that the target distribution typically has greater variance than the lure distribution indicating that unequal-variance SDT models provide more accurate measures of discriminability and criterion placement (Egan, 1958; Mickes et al., 2007). The challenge with implementing an unequal-variance SDT model is that it requires many criterion manipulations or confidence ratings to accurately assess the degree to which the variance of the target and lure distributions are unequal (Macmillan & Creelman, 2005). Therefore, an equal-variance SDT model determined measures of discriminability and criterion placement, since most of the reported experiments include criterion shift tasks with only two or three criterion manipulations.

Criterion placement and discriminability are behaviorally independent processes; however, a statistical relationship exists in SDT between the optimal criterion placement of an ideal observer and the extent of discriminability when a biased decision criterion is advantageous (i.e. the more uncertain the discrimination, the more extreme the criterion should be) (Macmillan & Creelman, 2005). Therefore, it is important to control for potential changes in criterion placement that simply arise from changes in discriminability by residualizing *c* against *d'* across all participants within each test condition and session to obtain normalized c (c_n) values, which ensures statistical independence (see Aminoff et al., 2012). This computation consists of correlating c with d' and adding the residuals of c to the grand mean of c to obtain c_n values. This ensures no linear relationship between c_n and d' values (i.e. $r = 0$) across participants within the specific test conditions of each session (e.g. conservative criterion condition in session 1 of Experiment 1). This correction is advantageous because it removes the correlation between c and d' while maintaining the same group average for c (i.e. mean c_n = mean c). Normalized C (C_n) values are obtained by taking c_n in the conservative condition and subtracting c_n in the liberal condition.

Linear Mixed Models

Analyses for various measures in Experiments 6-11 included additive linear mixed models, implemented with the lme4 packages (Bates et al., 2015) in R, to assess significant differences across many conditions within a single model. Deviation contrasts specified fixed effects and each model included a random effect specified on the model intercept across subjects to account for baseline variation in the measure of interest. Linear mixed models do not yield p-values for parameter estimates due to inherent difficulties in estimating denominator degrees of freedom. However, the restricted maximum likelihood approach to model estimation yields a posterior distribution over the parameters, allowing construction of empirical confidence intervals via simulation. Each linear mixed model included 1,000 iterations of posterior simulation to approximate 95% CIs around each parameter estimate. Any CI whose range includes zero is considered non-significant. Effect size approximations of Cohen's d are derived from dividing contrast parameter estimates by the square root of the total random effects variance of the model (Westfall et al., 2014).

Materials

Stimuli consisted of face images drawn from the 10k US Adult Faces database (Bainbridge et al., 2013), except for Experiments 3, 6, and 8. The stimulus sets for Experiments 3 and 6 contained two versions of 1,024 scene images. One version contained a single person whereas an edited version did not include a person. These scene stimuli derived from a center-cropped 500x500-pixel portion of images found on several open-source online databases. Stimuli for Experiment 8 consisted of center-cropped 330x330-pixel portions of 20,480 unique scene images from the SUN database (Xiao, et al., 2010). Participants conducted all tasks at a computer or within an MRI scanner using MATLAB version R2016B that incorporated open-source code from Psychophysics Toolbox, v3 (Brainard, 1997).

Chapter I: Criterion shifting is a uniquely individualistic cognitive trait

Although the within-subject stability of *criterion shifting* tendencies is poorly understood, test-retest recognition memory studies suggest that *criterion placement* tendencies are stable over time (Kantner & Lindsay, 2012, 2014). In recognition memory, criterion placement is the threshold of familiarity strength that must be exceeded to recognize items. Criterion placement, like criterion shifting, is quite variable across individuals (Aminoff et al., 2012, 2015; Kantner et al., 2015; Kantner & Lindsay, 2012, 2014; Frithsen et al., 2018; Layher et al., 2018; Miller & Kantner 2019). Despite large between-subject variability, Kantner and Lindsay (2012, 2014) proposed that the within-subject consistency of criterion placement over time makes it a stable cognitive trait. Some individuals regularly recognize stimuli based on weak familiarity evidence while others routinely require strong memory evidence before recognizing items. Criterion shifting, on the other hand, is a *shift* in the placement of a decision threshold to require more or less evidence before identifying a target when the circumstances surrounding a decision change (e.g. when recognizing a co-worker in the workplace versus a foreign vacation spot). The consistency and extent to which a person shifts a criterion is likely unrelated to an individual's criterion placement tendencies (though empirical reports of this relationship are lacking) because placing and shifting a criterion are separate behaviors. For example, two individuals might be quite adept at regularly establishing a *neutral* criterion by missing and falsely identifying items at equal rates. However, one individual might adaptively shift between conservative and liberal criteria when the situation calls for it, while the other may continuously maintain a neutral criterion even when criterion shifting becomes advantageous.

Criterion shifting can be categorized into two putative classes where an individual may either (1) knowingly shift a decision criterion based on known changes in the circumstances surrounding a decision (Egan, 1958; Banks, 1970; Healy & Kubovy, 1978; Rotello et al., 2005; Aminoff et al., 2012) or (2) unknowingly shift a decision criterion, such as through reinforcement learning[2] (Wixted & Gaitan, 2002; Han & Dobbins, 2008, 2009). An example of criterion shifting through reinforcement learning comes from Han and Dobbins (2008) who covertly altered feedback conditions by rewarding one error type (either false alarms or misses), but not the other. Over time, participants *unknowingly* shifted towards a more liberal criterion when feedback encouraged false alarms and established a

---

[2] There are many ways a criterion could be influenced unknowingly, which may include semantic similarity, sequential effects of test items, multidimensional representations, emotion-laden stimuli, word frequency, etc.

more conservative criterion when misses resulted in positive feedback. Strategic criterion shifting, on the other hand, occurs immediately and does not require feedback, but participants must be explicitly aware of the advantages for shifting criteria. For example, making participants aware that target items are more likely to appear than nontarget items during a test block (known as a base rate manipulation) will immediately cause many participants to shift to a liberal criterion, whereas participants who are unaware of such information will tend not to shift (Rhodes & Jacoby, 2007). The experiments reported here specifically investigate the stability of *strategic* criterion shifting tendencies where participants always received explicit instructions that informed them of the advantages for switching between conservative and liberal criteria.

Previous studies demonstrate that individual tendencies in strategic criterion shifting are largely consistent across different task types (Aminoff et al., 2012; Kantner et al., 2015; Frithsen et al., 2018). Aminoff and colleagues (2012) found a strong relationship in the degree to which individuals shift criteria during recognition memory tests for word versus face stimuli ($r(93) = .58$), which greatly exceeded the relationships in discriminability between the two tasks ($r(93) = .07$ in the conservative criterion condition and $r(93) = .22$ in the liberal criterion condition). Criterion shifting tendencies also appear stable across bias manipulations, regardless of whether a base rate manipulation is employed, or individuals are incentivized to shift via a payoff manipulation (e.g. participants earn money for correct responses, but lose money for either false alarms or misses) (Kantner et al., 2015; Frithsen et al., 2018). Frithsen and colleagues (2018) additionally found that criterion shifting tendencies are generally consistent across decision domains regardless of whether participants make recognition memory, visual detection, or visual discrimination judgments. However, the strength in the relationship of criterion shifting tendencies across tasks is sometimes mixed. For instance, Frithsen and colleagues (2018) found strong correlations between the extent of criterion shifting during a recognition memory test for *words* versus a visual detection test for identifying the presence of a white blob on a noisy background (*r*(47) = .53) and a visual discrimination test for determining the orientation of a Gabor patch on a noisy background (*r*(49) = .64). However, a weak relationship occurred between a recognition memory test for *faces* and a visual detection test for spotting white blobs on noisy backgrounds (*r*(49) = .17). Franks and Hicks (2016) found no significant relationship in the extent of criterion shifting between recognition memory tests that employed a base rate manipulation versus a manipulation that varied the known memory strength of studied items. However, manipulating the memory strength of items produces the strength-based mirror effect where an increase in discriminability results in both an increased hit rate *and* decreased false alarm rate (Hirshman 1995; Starns & Olchowski, 2015). The underlying cause of the strength-based mirror effect is strongly debated where some argue that the familiarity strength of novel items remains constant and an increase in discriminability results in a criterion shift towards being more conservative (i.e. require stronger memory evidence), which decreases the false alarm rate (Stretch & Wixted, 1998a; Starns, White, & Ratcliff, 2010). Others argue that the familiarity strength of novel items *can* change and observed decreases in false alarm rates when discriminability increases can be attributed to memory processes instead of strategic criterion shifts (Shiffrin & Steyvers, 1997; Criss 2006). Therefore, it is inconclusive as to whether using the strength-based mirror effect as a criterion manipulation is even valid since observed changes in false alarm rates may not actually be a result of strategic criterion shifting. Nevertheless, it helps raise the question as to why there are occasional inconsistent findings in the cross-task stability of criterion shifting; is the stability of criterion shifting a task specific phenomenon (Franks & Hicks, 2016), or are there alternative explanations for these mixed results?

One potential explanation for the observed inconsistencies in the cross-task stability of criterion shifting might be attributed to differences in demand characteristics between the two tasks (Kantner et al., 2015). Kantner and colleagues (2015) found that completely removing the study phase from recognition memory tests (i.e. participants received instructions that the study phase "malfunctioned" and thus could not encode any of the study images, but were still asked to perform the test phase anyways with an instruction induced criterion manipulation) dramatically affected the extent of criterion shifting (as it should), but only for a subset of participants. Since participants could not reliably use memory evidence to inform their decisions during the test phase (since they did not actually study any images), the best strategy is to maximally shift criteria by always responding "old" or always responding "new" depending on which response is more advantageous. However, many participants seemed unwilling to adopt this extreme strategy when task instructions required making a memory-based decision. Some individuals may have attempted to use other irrelevant perceptual cues to make recognition judgments or may have felt compelled to vary response types due to demand characteristics. This suggests that systematic design differences across tasks could alter response strategies and differentially affect how people integrate decision evidence with a criterion. For instance, Franks and Hicks (2016) observed no relationship in the extent of criterion shifting when comparing across recognition memory tests that implemented a base rate manipulation with a *blocked* design versus a strength-based manipulation with an *unblocked* design. Even if altering the memory strength of items is considered a valid criterion manipulation, a blocked design allows participants to shift and *maintain* a decision criterion throughout a test block whereas an unblocked design requires individuals to shift criteria on a trial-by-trial basis. Some individuals may be less willing to change decision strategies on a trial-by-trial basis compared to changing strategies once per test block (see Stretch & Wixted, 1998a). These differing task designs may have disrupted the stability of criterion shifting between the two bias manipulations that may otherwise be quite strong if both tasks incorporated the same design structure. Task design inequalities that could also differentially affect individual criterion shifting tendencies may result from other disparate design features such as differences in stimulus complexity or presentation times. For example, Frithsen and colleagues (2018) observed the weakest correlation in the extent of criterion shifting when comparing between a recognition memory test for faces and a visual detection test for the presence of a white blob on a noisy background. The critical difference in the design of the two tasks is that the face stimuli appeared for 1,500 ms whereas the white blob stimuli appeared for less than 350 ms. If task design differences are the culprit for this weak relationship, then homogenizing the presentation times across tests should improve the stability of criterion shifting between the two decision domains. For example, Aminoff and colleagues (2012) implemented identical recognition memory task designs that either used word or face stimuli and found a strong relationship in the extent of criterion shifting across the two tasks ($r(93) = .58$). The effect of differing task designs on criterion shifting tendencies is poorly understood but should be considered when assessing criterion shifting stability across tasks. Nevertheless, strong relationships in criterion shifting should be observed across all test types and decision domains when demand characteristics are equivalent, if criterion shifting tendencies are truly a stable cognitive trait.

In order to provide evidence that strategic criterion shifting tendencies are a stable cognitive trait, it is important to empirically demonstrate that individual criterion shifting tendencies (1) are stable over multiple testing sessions, (2) generalize across decision domains when demand characteristics are held constant, and (3) are not epiphenomenal of another trait or simply reflect a lack of motivation to perform well on the tasks. Most studies comparing the cross-task stability of criterion shifting occur within a single research visit (Aminoff et al., 2012; Kantner et al., 2015; Frithsen et al., 2018). However, individual criterion shifting tendencies may change over time. Franks and Hicks (2016) provide the only report that criterion shifting in recognition memory is stable over time, but these results only included two time points across two days. Experiments 1 and 2 assess the stability of criterion shifting over longer time periods, where participants conducted test-retest recognition memory tasks on 10 separate days across six weeks. Experiment 3 examines the test-retest reliability of criterion shifting tendencies across decision domains by comparing performance on recognition memory and visual detection tests (with equivalent demand characteristics) across two separate testing sessions. In both sessions, participants also conducted a test-retest battery of other cognitive tasks and questionnaires to determine whether other factors are related to individual criterion shifting tendencies. Aminoff and colleagues (2012) attempted to search for factors, such as cognitive style, personality traits, and executive functioning skills, that may explain individual differences in the extent of criterion shifting. Despite administering many questionnaires that assess a wide variety of cognitive and personality characteristics, only a few measures showed significant relationships with the extent of criterion shifting. These included positive relationships with a fun-seeking personality and verbal cognitive style as well as a negative relationship with characteristics of a negative affect. However, no published studies have attempted to replicate these findings. Experiment 3 attempts to both replicate some of these previously-reported relationships and probe several novel factors, such as performance on tasks assessing risk aversion, response inhibition, working memory, and task-switching ability. Additional assessments investigated whether a motivation to perform well on these tasks related to criterion shifting tendencies.

Another intriguing possibility is that individual differences in strategic criterion shifting tendencies are related to individual decision strategies for reporting confidence in recognition memory judgments. In lieu of criterion manipulations, many recognition memory studies implement confidence ratings to measure criterion shifts because confidence judgments require individuals to establish multiple decision criteria to differentiate between various levels of memory strength (Macmillan & Creelman, 2005; Yonelinas & Parks, 2007). Since many recognition studies implement confidence ratings to assess criterion shifts, it is reasonable to predict that both encompass similar decision processes. The ability to provide a confidence rating to a recognition memory judgment is believed to represent an individual's meta-awareness for the amount of familiarity strength that an item elicits (Koriat, 2007). Confidence ratings are oftentimes used to assess measures of meta-awareness, such as metacognitive sensitivity (how accurately one can distinguish between correct and incorrect responses) and metacognitive bias (the propensity to make judgments with the highest levels of confidence regardless of performance) (Fleming & Lau, 2014). Previous studies demonstrated that individuals typically have high metacognitive sensitivity, where judgments made with high confidence are generally more accurate than low confidence judgments (Mickes et al., 2007; Mickes et al., 2011). However, Mickes and colleagues (2011) found individual differences in metacognitive bias—some reserve the most confident responses for the strongest (or weakest) of memories while others make high confidence responses more frequently, even when discriminability is held constant. Similar to strategic criterion shifting, obtaining appropriate confidence ratings from participants requires very little training and occurs immediately. Given the commonalities between rating confidence and strategic criterion shifting, it is possible that criterion shifting tendencies are driven by an individual's meta-awareness of familiarity strength in test items. If this is the case, then individual tendencies to shift criteria should be directly related to measures of metacognitive sensitivity or metacognitive bias. For example, people with high metacognitive sensitivity might be more *capable* of shifting criteria to large extents relative to those who struggle with distinguishing between strong versus weak memory evidence. Individuals with high metacognitive bias have a lax standard for the amount of memory evidence needed (or lack thereof) to make recognition judgments with high confidence, which could result in smaller criterion shifts if participants shift criteria based on the level of confidence in "old" or "new" responses. Alternatively, strategic criterion shifting tendencies and tendencies to report meta-awareness in test items via confidence ratings may represent completely different decision process. Miller and Kantner (2019) conducted post hoc analyses on previously reported data in which participants both shifted criteria and rated confidence but found no relationship between the extent of criterion shifting and metacognitive bias. However, no studies have systematically assessed this relationship *a priori*. Experiments 4 and 5 assessed whether individual differences in strategic criterion shifting tendencies related to metacognitive sensitivity or metacognitive bias.

Taken together, Experiments 1–5 sought to demonstrate that individual tendencies in the extent of strategic criterion shifting are stable across time and decision domains while being unrelated to other factors such as performance on other cognitive tasks, personality traits, motivation to perform well, metacognitive sensitivity, or metacognitive bias. Stable differences in criterion shifting tendencies may reflect individual differences in people's *willingness* to ignore uncertain evidence in favor of a decision strategy that optimizes decisional outcomes (Green & Swets, 1966; Aminoff et al., 2012; Kantner et al., 2015; Miller & Kantner, 2019) as opposed to individual differences in people's *ability* to shift criteria. Since all individuals should theoretically be *capable* of shifting criteria to great extents, it is not expected that strategic criterion shifting tendencies are related to other abilities or characteristics.

Method

Statistical analysis

Effect size measures of Cohen's *d* and Pearson *r* correlation coefficients are reported with 95% CIs. Any CI spanning zero is considered nonsignificant. When assessing whether multiple correlation coefficients are statistically significant, the false discovery rate (FDR) is controlled for using the method described by Benjamini and Hochberg (1995). Averaged group results are presented with *SD* values that are adjusted for within-subject variables using the method described in Morey (2008). For non-significant Pearson correlation coefficients, the BayesFactor package (Morey et al., 2015) in R computed Bayes factors that assess the strength of evidence for the null versus alternative hypotheses (BF_{01}) using uninformed uniform priors. BF_{01} values greater than three are considered strong evidence for the null hypothesis (see Jeffreys, 1961).

Experiments 1 & 2: Test-retest reliability of criterion shifting in recognition memory

In determining whether the tendency to strategically criterion shift is a stable cognitive trait, Experiments 1 and 2 assessed the test-retest reliability of criterion shifting during recognition memory tests for faces across 10 sessions in the span of six weeks. In Experiment 1, a payoff manipulation incentivized criterion shifting, where participants received payment at the end of each session based *entirely* on an individual's performance. Participants earned five cents for each correct response, lost 10 cents for critical errors (either false alarms or misses), but received no penalty for noncritical errors. The likelihood of encountering old and new images remained equal in Experiment 1, but criterion shifting in this paradigm is advantageous for avoiding costly critical errors. When the critical error is a false alarm, participants should maintain a conservative criterion, but a liberal criterion becomes advantageous when the critical error is a miss. Perfect accuracy during an Experiment 1 session would result in a payment of $30, but participants could easily earn $15 by simply maximizing responses (i.e. always choosing the response that went unpenalized if incorrect). Experiment 2 followed similar procedures as Experiment 1 except a base rate manipulation induced criterion shifts and everyone earned $10 per session regardless of performance. Franks and Hicks (2016) showed that a base rate manipulation during recognition tests for words produced a modest test-retest relationship in the extent of criterion shifting across two testing sessions separated by 48 hours ($r(109) = .38$).

Experiments 1 and 2 investigated whether test-retest relationships in criterion shifting tendencies during recognition memory is sustained across many testing sessions.

Although criterion shifting tendencies are consistent across payoff and base rate manipulations when conducted within the same testing session (Kantner et al., 2015; Frithsen et al., 2018), the different criterion manipulations across the two experiments allowed for investigations of whether the extra monetary incentive to shift in Experiment 1 affected the stability of criterion shifting over *time*. For instance, participants in Experiment 1 who initially inadequately shift criteria may learn to shift to greater extents in subsequent sessions since doing so results in greater payouts. Participants in Experiment 2 lacked this additional monetary incentive to shift criteria, which could affect the long-term stability of criterion shifting. For example, some participants may shift criteria more extremely across sessions to improve accuracy while others may become less concerned about shifting criteria and more focused on completing the task as quickly as possible. If strategic criterion shifting tendencies are a stable cognitive trait, then there should be strong test-retest relationships in the extent of criterion shifting over time regardless of the type of criterion manipulation.

For strategic criterion shifting tendencies to be considered a stable trait, the test-retest reliability of criterion shifting should be as strong as other performance measures that are believed to reflect cognitive traits. Both experiments included a neutral criterion condition where participants either received no penalty for any errors (Experiment 1) or the likelihood of encountering a target or nontarget remained equal at test (Experiment 2). This allowed investigations of whether the test-retest reliability of criterion shifting is as stable as criterion placement in situations where criterion shifting yields no advantage, which Kantner and Lindsay (2012, 2014) identified as a stable cognitive trait (though this is by no means a standard for what should constitute a “stable cognitive trait”). If criterion shifting proves to be as stable as criterion placement in the neutral criterion condition, then the tendency to strategically shift criteria should also be considered a stable cognitive trait.

An additional factor that may affect the stability of criterion shifting tendencies is the strength of discriminability (i.e. how well participants can distinguish between studied and novel test images). According to SDT, as the strength of discriminability increases the need to criterion shift decreases (Macmillan & Creelman, 2005). Though most people fail to shift criteria to an extent that maximizes accuracy or payoffs during recognition memory tests (Aminoff et al., 2012, 2015; Kantner et al., 2015; Frithsen et al., 2018), studies of the strength-based mirror effect reveal that changes in the level of discriminability can affect the extent of criterion shifting (Glanzer & Adams, 1985; Franks & Hicks 2016), regardless of whether these changes are believed to result from strategic criterion shifts (Stretch & Wixted, 1998a) or not (Criss, 2006). In both experiments the strength of discriminability differed after the first five sessions to assess whether the extent of criterion shifting becomes less (or more) stable as discriminability improves. In situations where a neutral criterion is most advantageous, SDT predicts that changes in discriminability should not impact the stability of criterion placement since an ideal observer should always maintain an unbiased criterion regardless if memory strength is strong or weak (Macmillan & Creelman, 2005). To be considered a cognitive trait, criterion shifting tendencies should be as stable as neutral criterion placement tendencies regardless of the criterion manipulation or level of discriminability.

Method

Participants

Thirty-nine participants successfully completed all 10 test-retest sessions on separate days across six weeks in Experiment 1 (10 males; ages 18–28, *M* = 20, *SD* = 1.7). A separate sample of 39 participants completed Experiment 2 within a six-week span (11 males; ages 18–37, *M* = 20, *SD* = 3.1). Additional participants failed to complete all 10 sessions in Experiment 1 (five) and Experiment 2 (seven) and are excluded from all analyses. Participants in Experiment 1 earned anywhere from \$5–\$30 per session depending on performance (see ***Procedure***) whereas participants in Experiment 2 received \$10 per session. Participants in both experiments earned an additional \$50 bonus for completing all 10 sessions.

Procedure

The recognition memory task consisted of three blocks where each of the 10 self-paced sessions lasted for 20-60 minutes. A block consisted of a 100-image study phase followed by a 200-image test phase. Each session contained three different test phase conditions (conservative, liberal, and neutral) where instructions prior to each test phase (unless otherwise specified) *explicitly* informed participants of an advantage for establishing a conservative, liberal, or neutral decision criterion, respectively. A discriminability manipulation occurred after session five where the number of times each image appeared during the study phase changed. In the low discriminability condition, each study image appeared once whereas study images in the moderate discriminability condition appeared five times.

During each study phase, participants passively viewed a sequence of 100 unique face images on a black background in the center of a computer screen for 300 ms followed by a 200 ms blank screen presentation. The quick presentation time induced low discriminability making it more advantageous to shift criteria. During the test phases, each image appeared in the center of the screen with text displayed above the image to remind participants of the criterion condition. The numbers “0” or “1” appeared at the bottom of the screen to indicate the keyboard button corresponding to an “old” (studied) or “new” (unstudied) response, which randomly changed on a trial-by-trial basis. Images remained on screen until the participant made a response. Participants received feedback at the end of each session indicating the amount of money earned (Experiment 1) or the percentage of correct trials obtained (Experiment 2) for the entire session.

In each test phase of Experiment 1, participants received five cents for correctly responding “old” to a studied image (a hit) and “new” to an unstudied image (a correct rejection). Incorrect responses consisted of penalized critical errors and penalty-free noncritical errors. In the conservative condition, participants lost 10 cents for responding “old” to an unstudied image (a false alarm) but did not lose money for responding “new” to a studied image (a miss). In the liberal condition, participants lost 10 cents for a miss, but received no penalty for a false alarm. In the neutral condition, participants did not lose any money for false alarms or misses. Participants conducted the conservative, liberal, and neutral test phases in three separate blocks, each of which included 100 studied and 100 novel images. The conservative, liberal, and neutral test blocks appeared in a pseudo-random order across sessions and subjects. All participants conducted the low discriminability condition for sessions 1–5 and the moderate discriminability condition for sessions 6–10 (**Figure 1**).

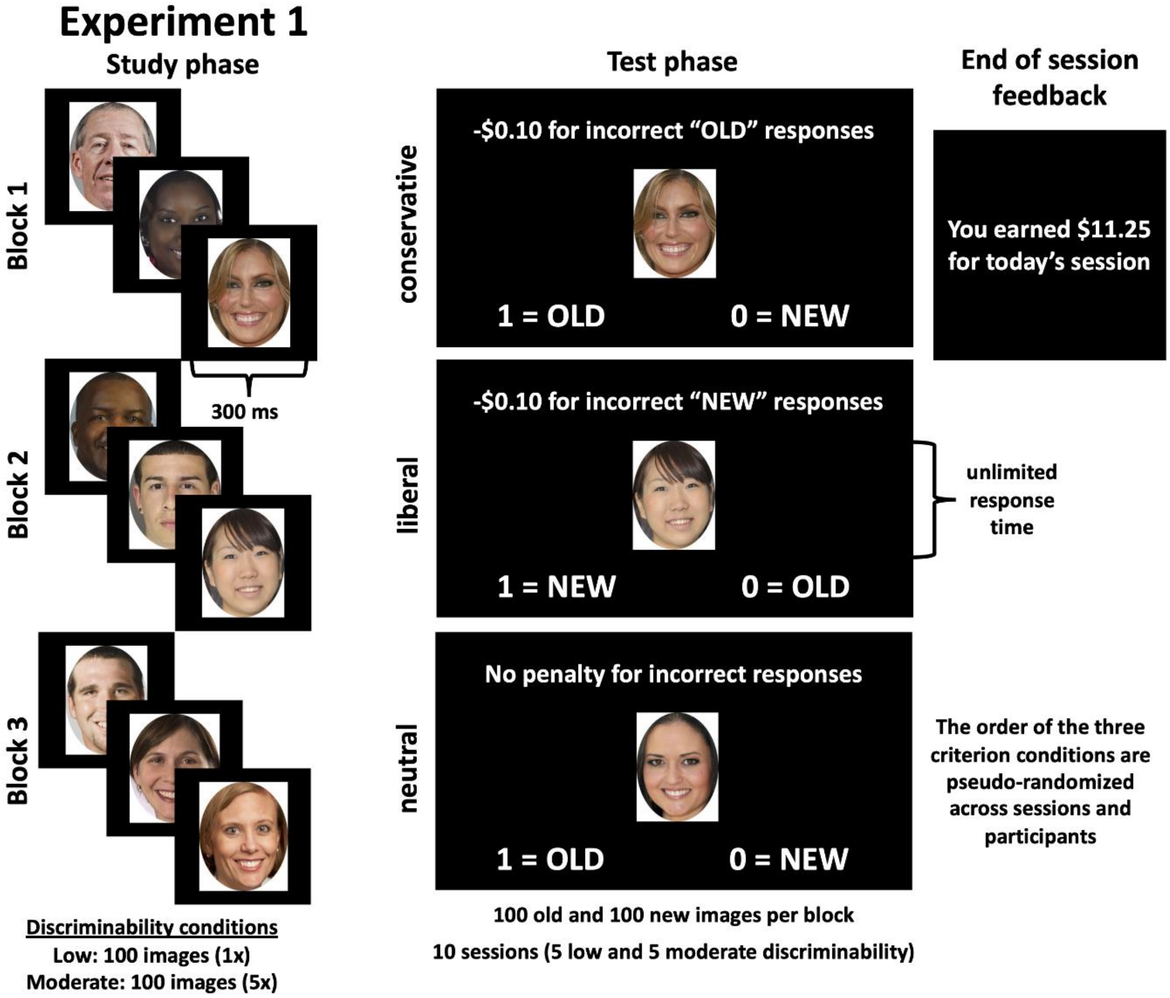


Figure 1: Experiment 1 recognition memory task. A 100-image study block preceded a 200-image test block for each criterion condition (conservative, liberal, and neutral). A payoff manipulation induced criterion shifts where participants earned 5 cents for correct responses, lost 10 cents for critical errors, but did not lose money for noncritical errors. At the end of each session, participants received feedback on total money earned for that session.

In each test phase of Experiment 2, a studied item appeared 25% (conservative), 75% (liberal), or 50% (neutral) of the time during a test block. Importantly, the probability manipulations in Experiment 2 and the payoff manipulations in Experiment 1 required the same conservative, liberal, and neutral criterion placements for *optimal* performance according to an equal-variance SDT model (see Macmillan & Creelman, 2005). Unlike Experiment 1, participants did NOT receive information about the 50% likelihood of encountering a studied item in the neutral condition to ensure that explicit instructions did not affect an individual’s criterion placement tendencies. The neutral criterion block always occurred first, whereas test blocks 2 and 3 consisted of 2 parts: a 100-image conservative (25 studied, 75 novel images) and a 100-image liberal (75 studied, 25 novel images) test mini-block. Instructions appeared before each mini-block to indicate the likelihood of encountering a studied item. The mini-block presentation order appeared pseudorandomly so that each session consisted of the two possible order types (conservative before liberal, or vice versa) and the block orders switched every other session (conservative in test block two or three). Twenty participants conducted the low discriminability condition in sessions 1–5 and moderate discriminability condition in sessions 6–10, whereas 19 participants conducted the low and moderate discriminability conditions in the reverse order (**Figure 2**). Experiments 1 and 2 included the same 6,000 unique face images (600 per session) and each participant received a completely randomized assignment and presentation order for the target and nontarget images.

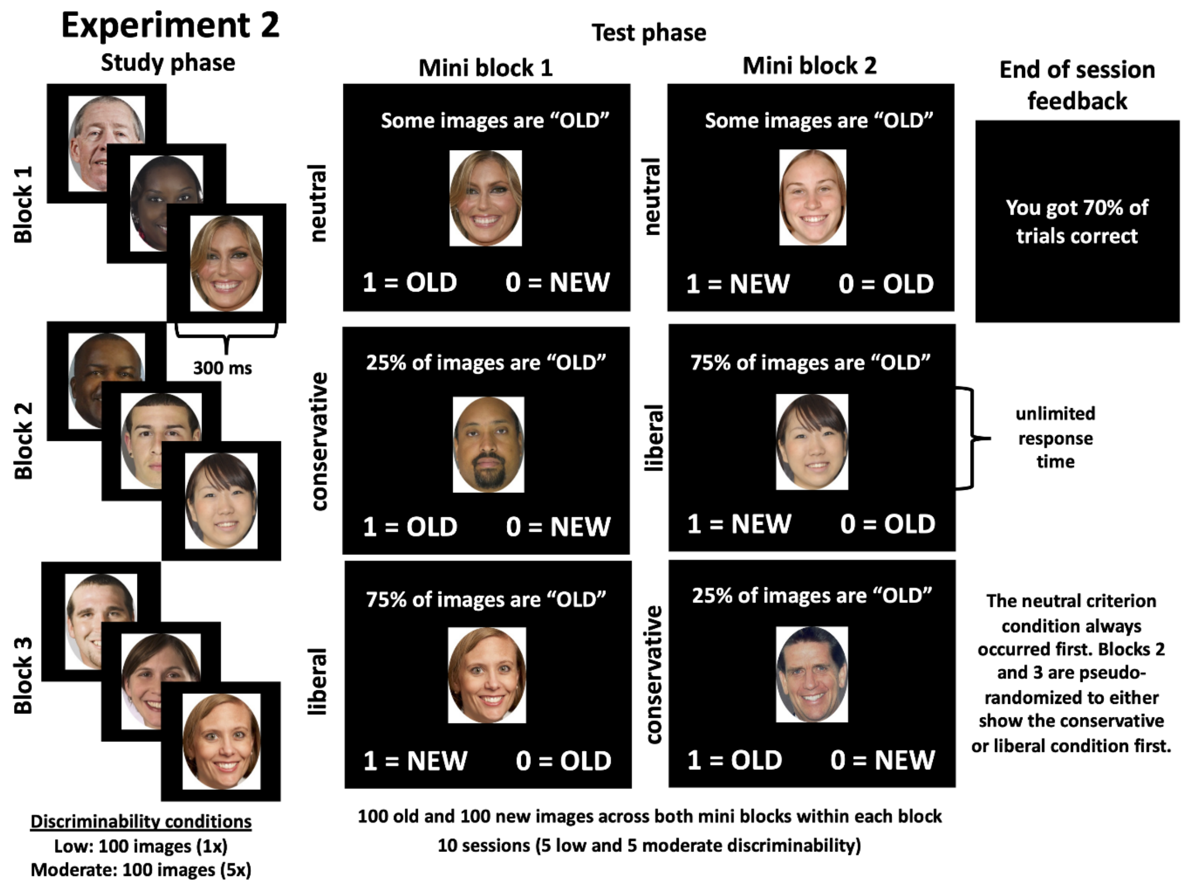


Figure 2: Experiment 2 recognition memory task. A 100-image study block preceded two 100-image test mini-blocks for each criterion condition (conservative, liberal, and neutral). A base rate manipulation induced criterion shifts where "old" images appeared either 25% (conservative), 50% (neutral), or 75% (liberal) of the time. At the end of each session, participants received feedback on the percentage of correct trials for that entire session.

Statistical Analysis

Pearson *r* correlation coefficients for the nine session-to-session comparisons and all 45 pairwise comparisons across the 10 sessions for values of C_n and neutral c_n are reported as well as the relationship between C_n and neutral c_n across the 10 sessions.

Results and discussion

Experiment 1

Mean *d'* across all low discriminability sessions (*M* = 0.41, *SD* = 0.43) remained lower compared to the mean *d'* of the moderate discriminability sessions (*M* = 0.87, *SD* = 0.59), *d* = 1.00, 95% CI [0.88, 1.12], confirming that viewing stimuli once versus five times during the study phase effectively modulated discriminability. On average, participants in the low discriminability condition shifted between c_n in the conservative (*M* = 1.03, *SD* = 0.63) and liberal criterion conditions (*M* = -1.00, *SD* = 0.71), *d* = 2.94, 95% CI [2.65, 3.23], as well in the moderate discriminability condition between the conservative (*M* = 1.00, *SD* = 0.56) and liberal criterion conditions (*M* = -0.98, *SD* = 0.70), *d* = 3.05, 95% CI [2.76, 3.35]. Surprisingly, mean C_n did not significantly differ between the low (*M* = 2.03, *SD* = 1.00) and moderate (*M* = 1.98, *SD* = 0.77) discriminability conditions, *d* = 0.05, 95% CI [-0.15, 0.25], even though SDT predicts that lower discriminability will lead to greater criterion shifts (see Macmillan & Creelman, 2005). **Table 1** shows a complete list of mean c_n, C_n, *d'*, and *PC* values for each session as well as for the low and moderate discriminability conditions combined across sessions.

Experiment 1: SDT means

	c_n					
Session	**Conservative**	**Neutral**	**Liberal**	C_n	***d'***	***PC***
1	0.75 (0.59)	0.22 (0.41)	0.76 (0.67)	1.5 (0.9)	0.36 (0.28)	55% (4)
2	1.06 (0.53)	0.00 (0.44)	-1.05 (0.66)	2.1 (0.8)	0.39 (0.30)	55% (4)
3	1.06 (0.60)	0.00 (0.61)	-1.04 (0.59)	2.1 (0.5)	0.45 (0.33)	56% (4)
4	1.10 (0.60)	0.13 (0.45)	-1.04 (0.60)	2.1 (0.6)	0.38 (0.34)	55% (4)
5	1.18 (0.53)	0.13 (0.50)	-1.10 (0.71)	2.3 (0.6)	0.46 (0.33)	55% (4)
6	0.93 (0.57)	0.07 (0.54)	-0.86 (0.65)	1.8 (0.6)	0.88 (0.44)	62% (6)

7	1.05 (0.48)	0.16 (0.38)	-1.08 (0.73)	2.1 (0.5)	0.89 (0.43)	61% (6)
8	1.00 (0.46)	-0.02 (0.39)	-0.97 (0.60)	2.0 (0.5)	0.87 (0.45)	61% (5)
9	0.99 (0.54)	0.09 (0.41)	-0.99 (0.66)	2.0 (0.6)	0.86 (0.41)	61% (6)
10	1.04 (0.57)	0.08 (0.45)	-0.98 (0.59)	2.0 (0.6)	0.84 (0.47)	61% (6)
low	1.03 (0.63)	0.10 (0.53)	-1.00 (0.71)	2.0 (1.0)	0.41 (0.43)	55% (6)
mod	1.00 (0.56)	0.08 (0.47)	-0.98 (0.70)	2.0 (0.8)	0.87 (0.59)	61% (8)

Table 1: Experiment 1 mean and standard deviation values (in parentheses) for c_n (three criterion conditions), C_n, d', and *PC* for each of the 10 sessions and collapsed across the low and moderate (mod) discriminability conditions.

In the neutral criterion condition, c_n remained fairly consistent across the 10 sessions (*r*(37) session-to-session range: .45-.74, *Mdn* = .54; all pairwise comparisons range: .18-.74, *Mdn* = .51), which is comparable to all of the test-retest relationships of *c* reported by Kantner and Lindsay (2012, 2014) (range: .31-.81, *Mdn* = .64). Correlations of C_n (*r*(37) session-to-session range: .58-.85, *Mdn* = .75; all pairwise comparisons range: .38-.85, *Mdn* = .68) remained high despite the discriminability manipulation that occurred after the first five sessions. **Figure 3** shows matrices of Pearson correlations for all 45 pairwise comparisons of neutral c_n and C_n. No significant relationships existed between C_n and neutral c_n across any of the 10 sessions after FDR correction (*r*(37) range: -.14-.34, *Mdn* = .01; BF_{01} range: 0.56-4.96, *Mdn* = 4.20), providing support to the assumption that criterion shifting and placement are independent behaviors. **Figure 4** displays c_n values for each criterion condition across all 10 sessions, ordered from left to right based on who shifted criteria to the greatest extent during session 10.

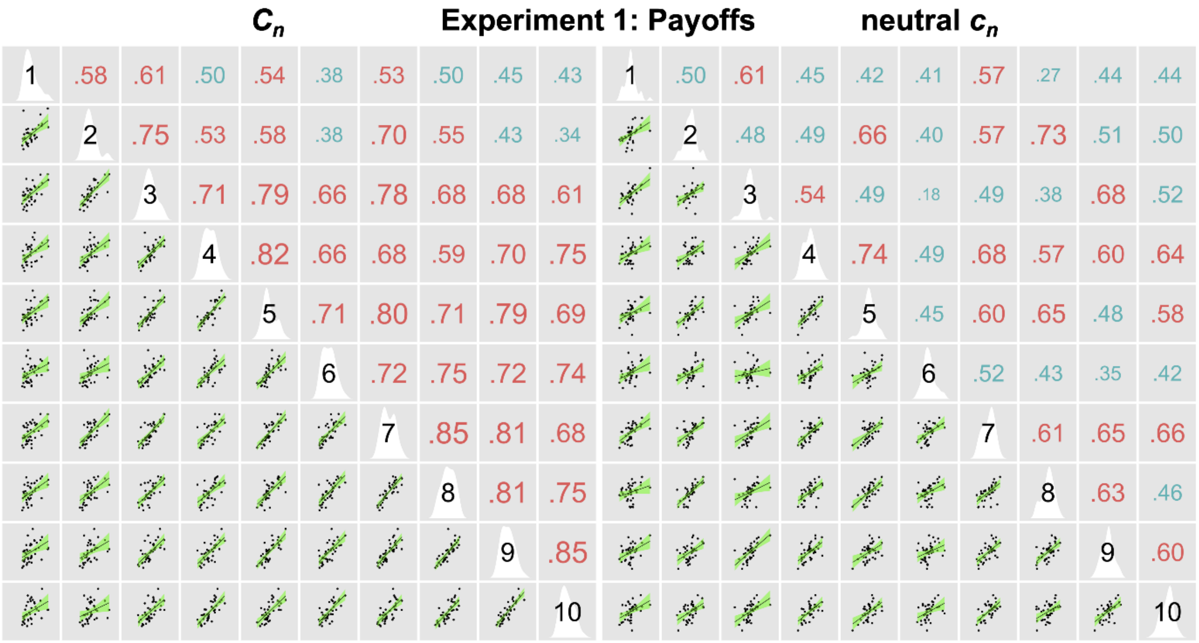


Figure 3: Experiment 1 Pearson correlation matrices comparing all 45 pairwise comparisons of C_n (left) and neutral c_n (right). The left side of each matrix displays the regression line for each comparison whereas the right side shows Pearson r values (red values are $q < .001$). The diagonal displays the session number along with the distribution of values (in white) for each session.

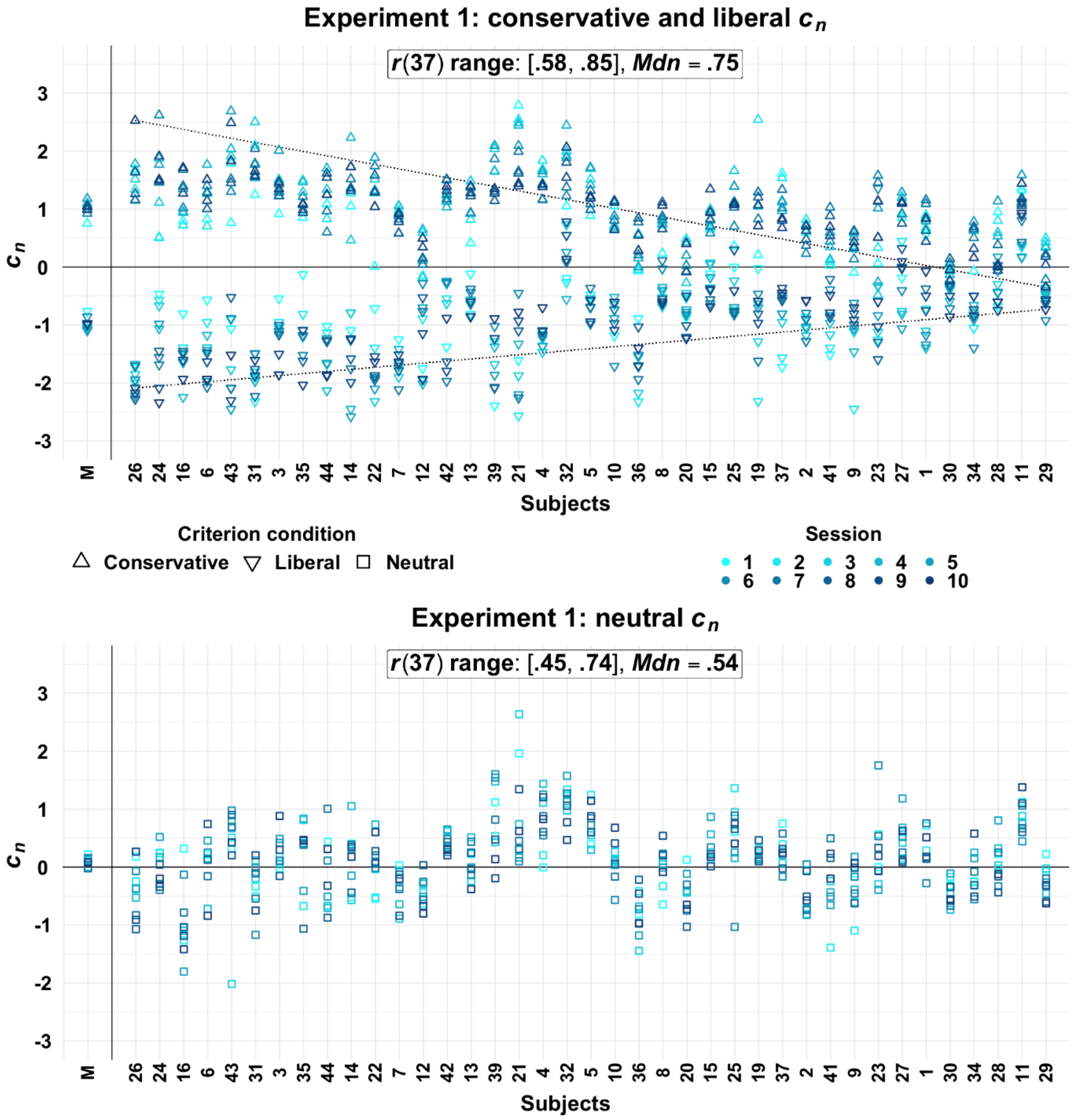


Figure 4: Experiment 1 c_n values for each participant and the mean (M) in the conservative and liberal criterion conditions (top) as well as the neutral criterion condition (bottom) across the 10 sessions. The extent of criterion shifting is depicted by the distance between the triangles representing the conservative and liberal c_n values. Participants are ordered from left to right based on who shifted criteria the most during session 10. Participants on the left have a large spread between conservative and liberal c_n values, but the magnitude of the spread steadily decreases as you view subjects from left to right. The dotted lines emphasize this criterion shifting trend by connecting the session 10 conservative and liberal c_n values of the leftmost and rightmost subjects. The range and median session-to-session Pearson correlation coefficients are shown for C_n (top) and neutral c_n (bottom) below the graph titles.

Experiment 2

Although some participants conducted the low discriminability condition in sessions 1-5 and others in sessions 6-10, statistics are reported with all 39 subjects together. Mean differences are computed within the discriminability conditions regardless of session number and correlation coefficients are computed across session number regardless of the order of the discriminability conditions.

Mean d' remained lower in the low discriminability sessions ($M = 0.27$, $SD = 0.46$) relative to the moderate discriminability sessions ($M = 0.85$, $SD = 0.66$), $d = 1.16$, 95% CI [1.04, 1.28]. On average, participants in the low discriminability condition shifted between c_n in the conservative ($M = 0.86$, $SD = 0.59$) and liberal criterion conditions ($M = -0.34$, $SD = 0.73$), $d = 1.83$, 95% CI [1.59, 2.06], as well as in the moderate discriminability condition between the conservative ($M = 0.82$, $SD = 0.56$) and liberal criterion conditions ($M = -0.53$, $SD = 0.53$), $d = 2.61$, 95% CI [2.34, 2.88]. As in Experiment 1, mean C_n did not significantly differ in the low ($M = 1.20$, $SD = 0.85$) versus moderate ($M = 1.35$, $SD = 0.89$) discriminability conditions, $d = 0.15$, 95% CI [-0.05, 0.35]. **Table 2** shows a complete list of mean c_n, C_n, d', and PC values for each session as well as for the low and moderate discriminability conditions combined across sessions.

Experiment 2: SDT means

	c_n					
Session	**Conservative**	**Neutral**	**Liberal**	C_n	d'	PC
1	0.60 (0.51)	0.16 (0.33)	-0.36 (0.39)	1.0 (0.9)	0.67 (0.51)	67% (8)
2	0.70 (0.42)	0.18 (0.30)	-0.27 (0.42)	1.0 (0.5)	0.63 (0.50)	66% (10)
3	0.83 (0.49)	0.23 (0.34)	-0.38 (0.54)	1.2 (0.6)	0.62 (0.51)	66% (9)
4	0.81 (0.43)	0.34 (0.30)	-0.34 (0.52)	1.2 (0.5)	0.62 (0.50)	66% (10)

5	0.81 (0.41)	0.37 (0.24)	-0.30 (0.60)	1.1 (0.5)	0.51 (0.48)	64% (9)
6	0.85 (0.45)	0.25 (0.25)	-0.41 (0.56)	1.3 (0.5)	0.58 (0.59)	66% (10)
7	0.91 (0.63)	0.34 (0.25)	-0.49 (0.72)	1.4 (0.7)	0.55 (0.55)	66% (10)
8	1.04 (0.57)	0.37 (0.26)	-0.56 (0.70)	1.6 (0.7)	0.54 (0.56)	66% (10)
9	0.91 (0.57)	0.32 (0.46)	-0.61 (0.66)	1.5 (0.6)	0.49 (0.51)	66% (9)
10	0.95 (0.66)	0.33 (0.37)	-0.61 (0.70)	1.6 (0.7)	0.40 (0.47)	64% (10)
low	0.86 (0.59)	0.38 (0.32)	-0.34 (0.73)	1.2 (0.9)	0.27 (0.46)	62% (12)
mod	0.82 (0.56)	0.20 (0.34)	-0.53 (0.53)	1.4 (0.9)	0.85 (0.66)	70% (11)

Table 2: Experiment 2 mean and standard deviation values (in parentheses) for c_n (three criterion conditions), C_n, *d'*, and *PC* for each of the 10 sessions and collapsed across the low and moderate (mod) discriminability conditions.

Similar to Experiment 1, correlation coefficients for neutral c_n ($r(37)$ session-to-session range: .68-.82, *Mdn* = .76; all pairwise comparisons range: .10-.82, *Mdn* = .59) are comparable to the test-retest relationships of *c* reported by Kantner and Lindsay (2012, 2014) (range: .31-.81, *Mdn* = .64). Strong correlation coefficients of C_n persisted across sessions ($r(37)$ session-to-session range: .71-.89, *Mdn* = .83; all pairwise comparisons range: .11-.90, *Mdn* = .67) despite counterbalancing the order in which participants conducted the low and moderate discriminability conditions. **Figure 5** shows matrices of Pearson correlations for all 45 pairwise comparisons of neutral c_n and C_n. Across the 10 sessions, only 1 significant relationship existed between C_n and neutral c_n after FDR correction (session 8: $r(37) = -.45$, CI = -.67, -.15). However, no obvious relationships existed when considering all 10 sessions together ($r(37)$ range: -.45-.19, *Mdn* = -.14; BF_{01} range: 0.10-4.99, *Mdn* = 2.33). **Figure 6** displays c_n values for each criterion condition across all 10 sessions, ordered from left to right based on who shifted criteria to the greatest extent during session 10.

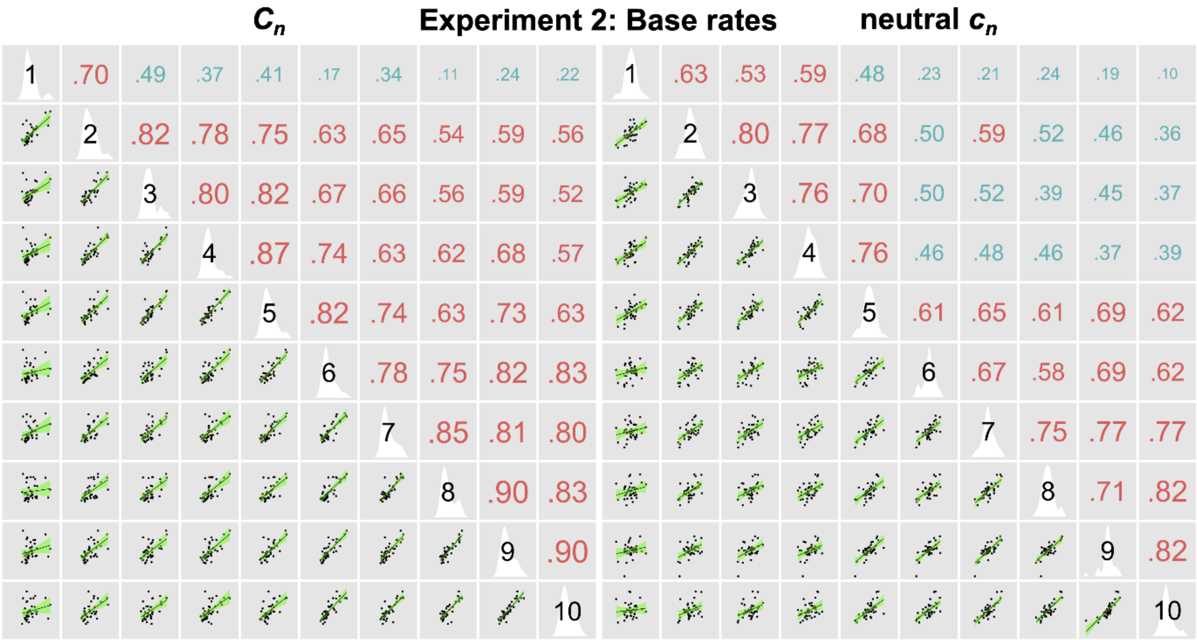


Figure 5: Experiment 2 Pearson correlation matrices comparing all 45 pairwise comparisons of C_n (left) and neutral c_n (right). The left side of each matrix displays the regression line for each comparison whereas the right side shows Pearson r values (red values are $q < .001$). The diagonal displays the session number along with the distribution of values (in white) for each session.

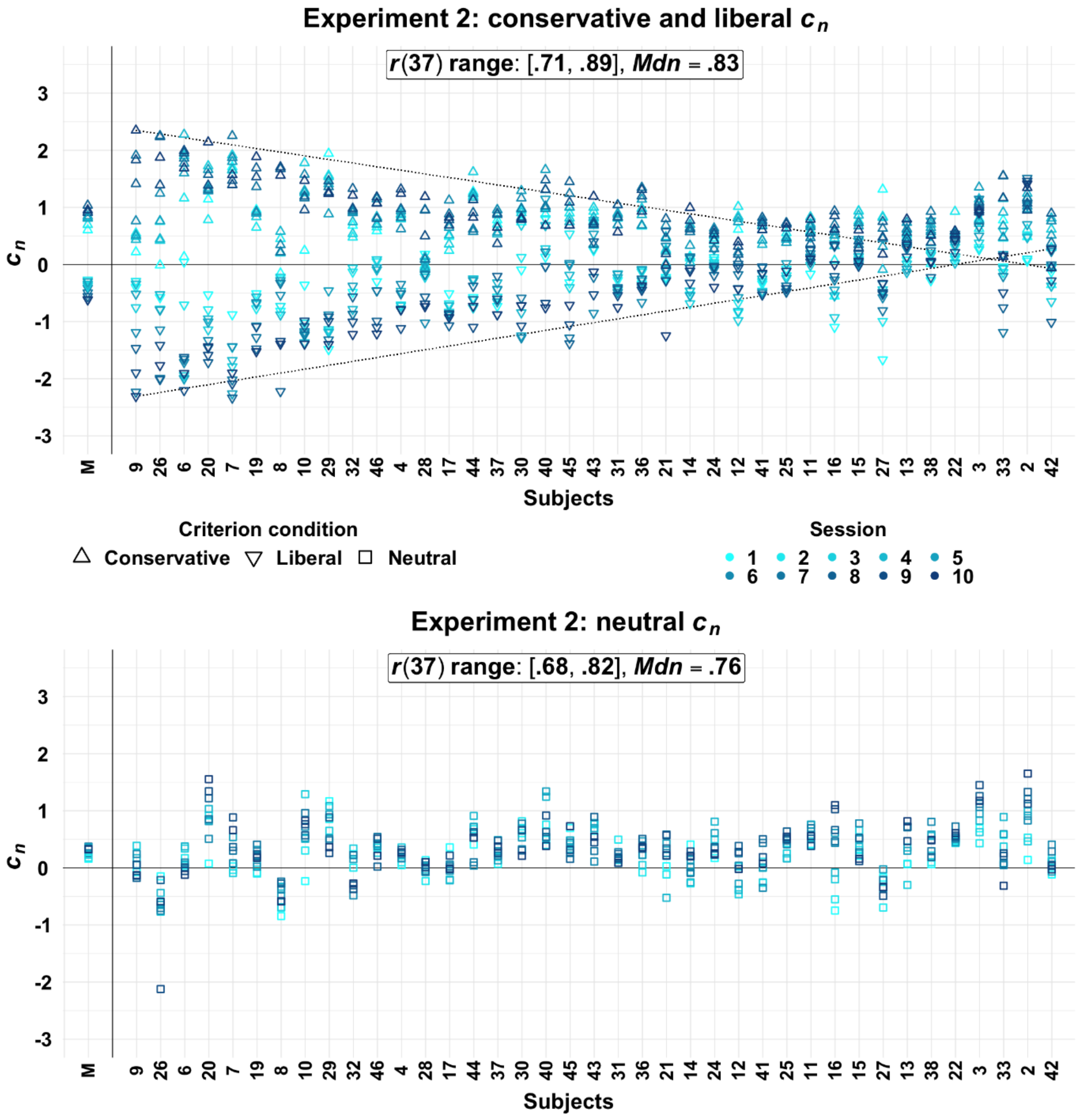


Figure 6: Experiment 2 c_n values for each participant and the mean (M) in the conservative and liberal criterion conditions (top) as well as the neutral criterion condition (bottom) across the 10 sessions. The extent of criterion shifting is depicted by the distance between the triangles representing the conservative and liberal c_n values. Participants are ordered from left to right based on who shifted criteria the most during session 10. Participants on the left have a large spread between conservative and liberal c_n values, but the magnitude of the spread steadily decreases as you view subjects from left to right. The dotted lines emphasize this criterion shifting trend by connecting the session 10 conservative and liberal c_n values of the leftmost and rightmost subjects. The range and median session-to-session Pearson correlation coefficients are shown for C_n (top) and neutral c_n (bottom) below the graph titles.

Comparing across Experiments 1 and 2, participants on average shifted criteria to a greater extent in Experiment 1 (*M* = 2.00, *SD* = 1.08) versus Experiment 2 (*M* = 1.27, *SD* = 1.02), *d* = 0.69, 95% CI [0.55, 0.84]. This likely occurred because the monetary incentive probably encouraged *some* individuals to shift criteria to greater extents in Experiment 1, since doing so increased total payout. Participants in Experiment 2 lacked this extra incentive to shift criteria since everyone received the same payment regardless of performance. However, it is important to note that some individuals in Experiment 1 did not shift to great extents (even by the 10th session; see the rightmost subjects in the top graph of **Figure 4**), while some individuals in Experiment 2 consistently shifted to large extents even though doing so did not affect the amount of money received (see the leftmost subjects in the top graph of **Figure 6**). This suggests that there could be individual differences in the factors that motivate individuals to shift criteria to greater extents. Future studies must confirm this prediction that monetary incentives motivate *some* individuals to shift criteria to larger extents since it is possible that people may generally shift criteria to lesser extents in response to a base rate versus payoff manipulation for reasons that are unrelated to motivating factors.

Experiments 1 and 2 revealed that strategic criterion shifting tendencies during recognition memory are stable across multiple sessions regardless of the criterion manipulation (payoff or base rates) or the strength of discriminability. Some participants consistently shifted criteria to large extents, others regularly shifted criteria to moderate degrees, while some individuals hardly shifted criteria at all. The stability of criterion shifting showed no relationship with neutral criterion placement tendencies indicating that placing and shifting a criterion are independent behaviors.

Experiment 3: Individual characteristics and generalizability of criterion shifting

Experiments 1 and 2 showed that criterion shifting is stable over time, at least during recognition memory tests. To further these findings, Experiment 3 tested whether the test-retest reliability of the extent of criterion shifting is stable both within and across decision domains. Frithsen and colleagues (2018) revealed that the extent of criterion shifting is largely consistent when making recognition memory judgments versus visual detection or visual discrimination judgments, but the strength of these relationships can sometimes vary. One possible explanation for this discrepancy is that differing demand characteristics across tasks may sometimes affect the stability of criterion shifting (Kantner et al., 2015). When two tasks have different designs, it *might* differentially affect how individuals strategically adapt a decision criterion. If a weak relationship exists between the extent of criterion shifting between two tasks with differing designs *and* decision domains, then it is impossible to know whether criterion shifting strategies are truly domain-specific or are simply affected by the particular task designs. In order to isolate the decision domain, the recognition memory and visual detection tests in Experiment 3 were virtually identical. The setup remained the same between the two tests with the only difference being whether participants reported if a scene appeared during an initial study phase (recognition memory) or if an image contained a person or not (visual detection). This allowed for assessments of the test-retest stability of criterion shifting tendencies across multiple decision domains while controlling for potential demand characteristic effects.

In addition to performing recognition memory and visual detection tests, participants also performed a battery of cognitive tests to assess whether consistencies in criterion shifting tendencies can be explained by other cognitive abilities. Although no published studies compare the extent of criterion shifting with other task measures, criterion shifting tendencies are predicted to be unrelated to performance on other cognitive tasks since these tendencies are likely a result of an individual’s *willingness* to shift criteria as opposed to an *ability* to do so (Kantner et al., 2015; Miller & Kantner, 2019). With this assumption, everyone should be *capable* of shifting criteria to great extents without the need of any particular skill that may otherwise be required for other cognitive tasks. However, one could argue that strategically shifting criteria might be associated with other cognitive abilities because there are many cognitive factors that go into a criterion shift. For example, participants who are more risk averse may shift to greater extents to simply avoid critical errors detrimental to decisional outcomes. Individuals with exceptional working memory might be more skilled at maintaining a consistent criterion during the entire length of a test block. Response inhibition is likely necessary for inhibiting prepotent familiarity responses in favor of more optimal decision strategies based on the criterion manipulation. Individuals may require more general task-switching ability in order to adequately shift between conservative and liberal decision criteria. To test whether these cognitive abilities show a relationship with criterion shifting, participants performed four additional standardized tasks that assessed risk aversion, response inhibition, working memory, and task-switching ability during each session. Although there are countless cognitive abilities that could possibly be associated with individual criterion shifting tendencies, these four cognitive abilities could reasonably be related to criterion shift strategies given the commonalities between these abilities and aspects of the decision processes that underly criterion shifting.

The extent to which an individual shifts a criterion might also be related to how motivated a person is to perform well during the tasks or may be associated with other personality characteristics. Following the criterion shifting tasks, participants completed a motivation questionnaire to assess whether a relationship exists between self-reported motivation to perform well on the tasks and the extent of criterion shifting. After completing all cognitive tasks, participants conducted additional personality and cognitive style questionnaires. Aminoff and colleagues (2012) previously conducted a large-scale study to assess whether individual differences in criterion shifting during recognition memory tests are related to any personality or cognitive characteristics. Despite collecting over 25 different standardized questionnaire measures that assess personality and cognitive characteristics, Aminoff and colleagues (2012) only found three questionnaire measures that significantly correlated with the extent of criterion shifting: positive relationships with a fun-seeking personality and verbal cognitive style as well as a negative relationship with traits associated with a negative affect. Therefore, questionnaire assessments were limited to these three measures for replication purposes. As stated in a preregistration (https://osf.io/jkfp6), criterion shifting is predicted to be stable within and across decision domains while being unrelated to performance on other tasks and questionnaire measures.

Participants

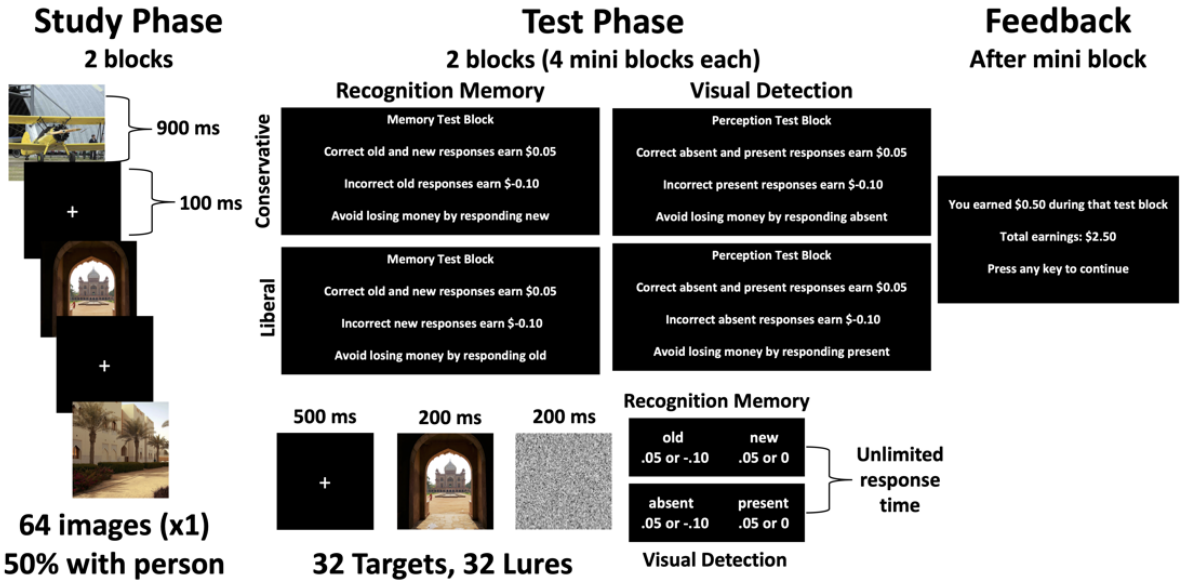
One hundred and seventy-two participants successfully completed both test-retest sessions (53 males; ages 18–30 years, *M* = 20, *SD* = 2.0). Exclusion of five additional participants occurred due to computer malfunctions (two) or incomplete datasets (three) and are not included in any analyses. Participant payment relied *entirely* on task performance (see ***Procedures***) unless participants earned less than $10 across all tasks during a session. Total payment for each session ranged from $10-$30 depending on performance.

Procedure

Participants completed two self-paced sessions on different days within the same week. Each session lasted between 45 and 90 minutes and included five computer tasks and four questionnaires. The computer tasks included (1) recognition memory and visual detection criterion shifting paradigms, (2) a Balloon Analogue Risk Task (BART), (3) a Go/No-go response inhibition task, (4) an *N*-Back working memory task, and (5) a Task-Switching paradigm. The questionnaires consisted of (1) the Effort/Importance section of the Intrinsic Motivation Inventory (IMI) (Ryan, 1982), (2) the Behavioral Inhibition System and Behavioral Activation System (BIS/BAS) scales (Carver & White, 1994), (3) the Positive and Negative Affect Schedule – Expanded Form (PANAS-X) (Watson & Clark, 1999), and (4) a modified version of the Verbalizer-Visualizer Questionnaire (VVQ) (Richardson, 1977). Participants first conducted the recognition memory and visual detection criterion shifting paradigms followed by the IMI questionnaire. This allowed for assessments of whether self-reports of motivation to perform well on the criterion shifting paradigms related to individual differences in criterion shifting tendencies. Afterwards, participants conducted the other four computer tasks in a randomized order followed by the remaining three questionnaires. The questions within each questionnaire appeared in a random order. Although randomized across participants, the order of the additional tasks, questionnaires, and questions within each questionnaire remained the same for each participant across both sessions. At the end of each session, participants received payment based entirely on the amount of money earned during the criterion shifting paradigms and the BART.

Recognition memory and visual detection criterion shifting paradigms

After conducting a short practice task, participants completed two cycles of a study phase followed by two recognition memory and two visual detection test mini-blocks (**Figure 7**). During each study phase, participants viewed a randomized sequence of 64 unique scene stimuli, half of which contained a single person and half contained no people. Scenes appeared in the center of a computer screen on a black background for 900 ms followed by a 100 ms presentation of a white crosshair. During recognition memory test mini-blocks, participants decided whether or not test images appeared during the study phase by responding “old” or “new,” respectively. For visual detection blocks, participants determined whether a person appeared in the image or not by responding “present” or “absent,” respectively. A payoff manipulation incentivized criterion shifting where participants earned five cents for correct responses, lost 10 cents for critical errors, but received no penalty for noncritical errors. In the conservative criterion condition, critical errors consisted of incorrect “old” responses during recognition memory tests and incorrect “present” responses during visual detection tests (false alarms). Critical errors in the liberal criterion condition included incorrect “new” and “absent” responses (misses). This created four test mini-block conditions that each appeared once in a random order after every study phase: (1) conservative recognition memory, (2) liberal recognition memory, (3) conservative visual detection, and (4) liberal visual detection. Prior to each test block participants received explicit instructions detailing the task type (recognition memory or visual detection) and criterion condition (conservative or liberal). Each test mini-block contained a randomized sequence of 64 stimuli (32 targets and 32 nontargets) that appeared for 200 ms followed by a 200 ms noise mask to destroy the perceptual afterimage. Afterwards, participants made an old/new or present/absent judgment by using the “f” and “j” keys on a keyboard, in which pseudorandom assignment across participants mapped the keys to each response type. On each trial, text appeared below each response type to remind participants of the critical and noncritical errors. Participants made responses with unlimited time and a 500 ms white crosshair presentation followed each response. After each test mini-block, participants received feedback detailing the amount of money earned on that mini-block as well as the running total of money earned during the task. Across both sessions, a single version of each stimulus (randomly assigned for each participant) appeared once during the test blocks (512 with a person, 512 without).



Memory Test Block
Correct old and new responses earn $0.05
Incorrect old responses earn $-0.10
Avoid losing money by responding new

Perception Test Block
Correct absent and present responses earn $0.05
Incorrect present responses earn $-0.10
Avoid losing money by responding absent

Memory Test Block
Correct old and new responses earn $0.05
Incorrect new responses earn $-0.10
Avoid losing money by responding old

Perception Test Block
Correct absent and present responses earn $0.05
Incorrect absent responses earn $-0.10
Avoid losing money by responding present

You earned $0.50 during that test block
Total earnings: $2.50
Press any key to continue

Figure 7: Experiment 3 recognition memory and visual detection (perception) tasks. After each study phase, participants conducted four test mini-blocks, one for each task and criterion condition combination. To control for demand characteristics, the recognition memory and visual detection tests maintained the exact same structure except participants either responded as to whether an image appeared during the study phase or if a person appeared in the image, respectively. Participants received feedback on the amount of money earned after each mini-block.

BART

The BART is a standardized computer task that assesses risk-taking behavior (Lejuez et al., 2002). Participants conducted 20 trials of a version of the BART where simulated balloons needed to be "pumped" to earn money. Participants earned one cent per pump and could collect money at any time during a trial unless the balloon popped. The balloon could pop anywhere from the first to the 128th pump determined by a random number generator with equal probability. Participants pressed the "f" or "j" key to either pump the balloon or collect money on a trial, depending on the pseudorandom assignment across subjects. Participants always saw the running total of money collected during the task as well as the amount of potential earnings that could be collected on a given trial. Trials ended when the balloon popped or the participant collected the money. Participants wore headphones which made pumping noises on each pump, a cash register sound when collecting money (with an accompanying green screen portraying the amount of money earned on that trial), and a popping noise when the balloon popped (with an accompanying red screen displaying the word "POP!").

Go/No-go

The go/no-go task assessed inhibitory control and consisted of five, 40-trial blocks where participants needed to press the "j" key during common "go" trials while not responding to rare "no-go" trials. Participants viewed random sequences of yellow and cyan squares that either served as "go" or "no-go" trials based on pseudorandom assignment across subjects. Stimuli appeared for 800 ms followed by a 500 ms white crosshair presentation where each block contained 32 "go" trials and 8 "no-go" trials. These parameters are within the recommended ranges for truly evoking response inhibition activity to a prepotent motor response on “no-go” trials (Wessel, 2017). Prior to the main task, participants conducted a 20-trial practice task.

N-Back

Paradigm. The *n*-back task assessed working memory performance and consisted of five, 40-trial blocks and followed similar procedures as the 2-back task described by Kane, Conway, Miura, and Colflesh (2007). Participants viewed a sequence of letters and needed to decide if each letter matched the case-insensitive letter that appeared two trials early, known as a 2-back trial (e.g. the third letter in the sequences *“B-F-B”* and *“B-F-b”* are 2-back trials). Each trial lasted for 2,500 ms in which a white letter on a black background appeared in the center of a computer screen for 500 ms followed by a 2,000 ms screen that displayed the response types. Participants needed to respond “yes” on 2-back trials and “no” on other trial types and could make a response at any point during the 2,500 ms trial. Participants made responses with the “f” and “j” keys, in which pseudorandom assignment across subjects determined the mapping between keys and response types. After each trial, a white crosshair appeared for 500 ms. Of the 40 trials in each block, eight constituted 2-back trials. Prior to the 2-back task, participants conducted a 20-trial practice block that provided feedback on performance to ensure comprehension of the instructions.

Stimuli. Upper and lowercase versions of the following eight letters made up the stimulus set: B, F, K, H, M, Q, R, and X. Each block consisted of a randomly generated 40-trial sequence that met the following five conditions: (1) 20% of trials are 1-back, (2) 20% of trials are 2-back, (3) 20% of trials are 3-back, (4) stimuli could not constitute both a 1-, 2-, and/or 3-back, and (5) each of the eight stimuli appeared at least three times, but no more than seven times in the sequence.

Task Switching

Paradigm. The task included five, 40-trial blocks of a modified task-switching paradigm described by Rogers and Monsell (1995). On each trial, a randomly ordered number/letter pairing (e.g. “U2” or “2U”) appeared within one square of a 2 x 2 grid. Participants pressed the “f” and “j” keys to either respond “yes” or “no” (depending on pseudorandom assignment across subjects) to one of the two following questions: “Is the letter a vowel?” or “Is the number odd?” The stimulus remained on screen until the participant responded. Afterwards a 300 ms presentation of a green crosshair or red “x” appeared to indicate a correct or incorrect response, respectively. The next trial appeared in a new square that moved in either a clockwise or counterclockwise fashion depending on pseudorandom assignment across participants. The question to be answered depended on whether the stimulus appeared in the top row or bottom row of squares (also pseudorandomly assigned across participants). Thus, the task switched every two trials.

Stimuli. The stimulus set consisted of four odd numbers (3, 5, 7, 9), even numbers (2, 4, 6, 8), vowels (A, E, O, U) and consonants (G, K, M, R). Letter/number pairings occurred randomly on a trial-by-trial basis under the condition that odd numbers always paired with consonants and even numbers always paired with vowels. This pattern ensured that the answer to one of the two question types is always “yes” while the other is always “no.”

Questionnaires

Intrinsic Motivation Inventory. The IMI is a standardized 45-item post-task questionnaire intended to assess the subjective experiences that a participant felt during a recently completed task (Ryan, 1982). To conduct the IMI, participants read a statement (e.g. “I tried very hard on this activity.”) and rate how true they believe the statement pertains to them on a scale from 1 (not at all true) to 7 (very true). Although the IMI consists of seven subscales, participants only conducted the five items from the Effort/Importance subscale to assess perceived effort and motivation during the recognition memory and visual detection criterion shifting paradigms.

Behavioral Inhibition System and Behavioral Activation System scales. The BIS/BAS scales are a standardized 24-item questionnaire that assesses an individual’s motivation to avoid aversive outcomes and approach desired outcomes (Carver & White, 1994). During the questionnaire participants rate how true or false a statement (e.g. “I will often do things for no other reason than that they might be fun.”) pertains to them on a scale from 1 (very true for me) to 4 (very false for me). The BIS/BAS scales consist of four subscales, but only the four items from the BAS fun-seeking subscale were analyzed to attempt to replicate the finding of Aminoff and colleagues (2012) that showed the BAS fun-seeking score is positively associated with the extent of C_n during a recognition memory test.

Positive and Negative Affect Schedule – Expanded Form. The PANAS-X is a standardized 60-item questionnaire that assesses a person’s recent feelings and emotions (Watson & Clark, 1999). The questionnaire requires participants to read a word or phrase (e.g. “afraid”) and rate the extent to which they felt that way during the past few weeks on a scale from 1 (very slightly or not at all) to 5 (extremely). The PANAS-X consists of 13 scales, but only the 10 items from the Negative Affect scale were analyzed to attempt to replicate the finding of Aminoff and colleagues (2012) that the Negative Affect scale score is negatively correlated with the extent of C_n during a recognition memory test.

Verbalizer-Visualizer Questionnaire (modified). The VVQ is a standardized 15-item questionnaire that assesses an individual's preference to represent knowledge in a visual or verbal manner (Richardson, 1977). A modified version of the VVQ was implemented in which some of the 15-items derived from the Individual Differences Questionnaire (IDQ) (Paivio, 1971), which is an 86-item questionnaire that formed the basis of the original VVQ (Richardson, 1977). Although the original VVQ requires individuals to respond "true" or "false" to whether a statement (e.g. "I enjoy learning new words.") applies to the participant, the questionnaire was modified to also include how strongly an individual agreed or disagreed with each statement on a scale from 1 (strongly disagree) to 7 (strongly agree). This modified version of the VVQ is the same questionnaire that Aminoff and colleagues (2012) implemented (though the authors simply refer to the modified version as the VVQ). The implementation of the VVQ was designed to potentially replicate the finding of Aminoff and colleagues (2012) that the extent of C_n during a recognition memory test is positively associated with the verbalizer score (seven items) on this modified version of the VVQ.

Statistical Analysis

From the extraneous tasks and questionnaires, eight additional individual difference measures were obtained to assess whether various cognitive and personality factors could explain variance in criterion shifting tendencies. For the BART, risk-taking behavior is assessed via mean adjusted pumps: the average number of balloon pumps on trials where the participant chose to collect money (see Lejuez & Brown, 2002). Computations of d' for the go/no-go and *n*-back tasks assessed performance of response inhibition and working memory, respectively. Task-switching ability was assessed from the time cost measure, which is computed from the difference in reaction time when responding to switch trials (where the task of the current trial differs from the previous trial) versus same trials (where the task remained the same between the previous and current trial). To be consistent with Rogers and Monsell (1995), trials with reaction times less than 100 ms and trials that immediately followed an error were excluded. An assessment of motivation to perform well on the criterion shifting tasks came from the IMI questionnaire Effort/Importance subscale (IMI: effort/importance). Although participants completed the entire BIS/BAS scales, PANAS-X, and modified VVQ surveys, analyses focused on the three measures that Aminoff and colleagues (2012) found to be significantly related to C_n during a recognition memory test. The fun-seeking score of the BIS/BAS scales (BAS: fun-seeking), the PANAS-X negative affect score (PANAS: negative), and the verbalizer score on the modified VVQ (VVQ: verbal) were correlated with criterion shifting performance (see **Appendix** for questionnaire items and scoring details).

For each of the eight additional measures, the mean values for each session, Cohen's *d* effect sizes for mean differences between the two sessions, and Pearson correlation coefficients to assess the test-retest reliability of each measure are reported. To assess if any of the eight measures relate to individual differences in criterion shifting tendencies, Pearson *r* correlations were conducted between each of the eight additional individual difference measures against both recognition memory C_n and visual detection C_n during the two sessions. Assessments of the strength of evidence for the null versus alternative hypotheses occurred using Bayes factors with uninformed uniform priors.

Results and discussion

Since the extent of criterion shifting can be affected by changes in *d'* (Macmillan & Creelman, 2005), attempts were made to make mean discriminability in the recognition memory and visual detection paradigms approximately equal across both tasks and sessions. In the recognition memory task, mean *d'* did not significantly differ between session 1 (*M* = 0.95, *SD* = 0.53) and session 2 (*M* = 1.02, *SD* = 0.57), *d* = 0.14, 95% CI [-0.01, 0.29]. In the visual detection task, mean *d'* slightly increased from session 1 (*M* = 1.12, *SD* = 0.63) to session 2 (*M* = 1.34, *SD* = 0.61), *d* = 0.32, 95% CI [0.17, 0.47]. Mean *d'* remained higher on average during the visual detection task relative to recognition memory for both session 1 (*d* = 0.27, CI = 0.12, 0.42) and session 2 (*d* = 0.51, CI = 0.36, 0.67), despite efforts to make mean discriminability equivalent across decision domains. On the recognition memory tests, participants on average shifted between c_n in the conservative (*M* = 0.72, *SD* = 0.57) and liberal criterion conditions (*M* = -0.76, *SD* = 0.59), *d* = 2.69, 95% CI [2.48, 2.89], as well as across visual detection tasks between the conservative (*M* = 0.84, *SD* = 0.52) and liberal criterion conditions (*M* = -0.43, *SD* = 0.58), *d* = 2.41, 95% CI [2.22, 2.61].

Average C_n during recognition memory did not significantly differ between session 1 (*M* = 1.40, *SD* = 0.45) and session 2 (*M* = 1.56, *SD* = 0.50), *d* = 0.18, 95% CI [-0.03, 0.39], as well as during visual detection between session 1 (*M* = 1.22, *SD* = 0.55) and session 2 (*M* = 1.31, *SD* = 0.43), *d* = 0.11, 95% CI [-0.10, 0.32]. Mean C_n remained marginally higher during recognition memory compared to visual detection in session 1 (*d* = 0.21, CI = 0.00, 0.42) and session 2 (*d* = 0.29, CI = 0.08, 0.50).

The extent of C_n from session 1 to session 2 remained very consistent for both the recognition memory ($r(170) = .75$, CI = .67, .80) and visual detection ($r(170) = .65$, CI = .55, .73) tests. Strong relationships in C_n also persisted across the two tasks during session 1 ($r(170) = .68$, CI = .59, .75) and session 2 ($r(170) = .78$, CI = .72, .83), despite small differences in mean *d'* and C_n across decision domains. Correlations even remained strong when comparing C_n in session 1 of the recognition memory test to session 2 of the visual detection test ($r(170) = .65$, CI = .55, .73) and vice versa ($r(170) = .57$, CI = .46, .67). Although C_n remained strongly consistent across tasks, *d'* only showed weak correlations between the two tasks during session 1 ($r(170) = .14$, CI = -.01, .29; $BF_{01} = 1.03$) and session 2 ($r(170) = .17$, CI = .02, .31; $BF_{01} = 0.57$). This provides evidence that the cross-task stability of criterion shifting cannot simply be attributed to discriminability performance alone. Overall, the strong correlations in C_n across sessions and tasks suggest that criterion shifting is a stable, domain-general process. **Figure 8** displays conservative and liberal c_n values for each participant across both sessions and tasks, ordered from left to right based on who shifted criteria the most during session 2 of the recognition memory task.

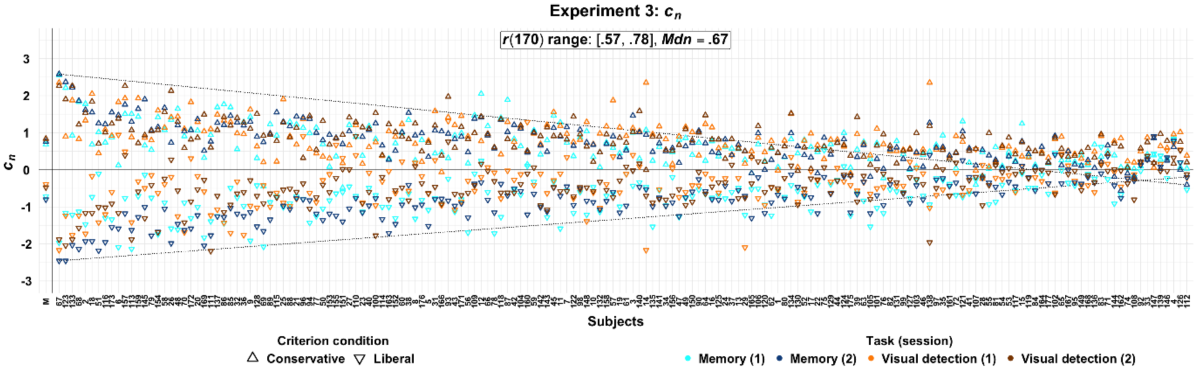


Figure 8: Experiment 3 c_n values for each participant and the mean (M) in the conservative and liberal criterion conditions for the recognition memory (blue) and visual detection (orange) tests across both sessions. The extent of criterion shifting is depicted by the distance between the triangles representing the conservative and

liberal c_n values. Participants are ordered from left to right based on who shifted criteria the most during session 2 of the recognition memory test. The dotted lines emphasize this criterion shifting trend by connecting the session 2 recognition memory conservative and liberal c_n values of the leftmost and rightmost subjects. The range and median Pearson correlation for C_n across sessions and tasks (6 measures total) is shown below the graph title.

For the eight additional task and questionnaire measures, reports of mean performance during both sessions, a Cohen's *d* effect size measure for mean differences across sessions, and test-retest Pearson *r* correlation coefficients are included in **Table 3**. It is possible that previously reported weak correlations between the extent of criterion shifting and other task measures could be a result of low test-retest reliability of the other measures. In Experiment 3 the test-retest reliability remained moderately strong for most of the additional measures (*r*(170) range: .47-.83), but it is possible that imperfect reliability could still attribute to attenuation in the relationships between these measures and the extent of C_n.

Experiment 3: Test-retest measures

measure	**session 1**	**session 2**	**Cohen's *d***	**Pearson *r***
Recognition C_n	1.40 (0.45)	1.56 (0.50)	0.18 [-0.03, 0.39]	.75 [.67, .81]
Visual detection C_n	1.22 (0.55)	1.31 (0.43)	0.11 [-0.10, 0.32]	.65 [.55, .73]
BART: pumps	34.72 (8.31)	40.83 (8.31)	0.42 [0.20, 0.63]	.68 [.59, .75]
Go/no-go: *d'*	3.48 (0.51)	3.25 (0.51)	0.32 [0.10, 0.53]	.47 [.34, .58]
N-back: *d'*	1.66 (0.54)	2.03 (0.54)	0.35 [0.13, 0.56]	.76 [.69, .82]
Switching: time cost	377 ms (121)	260 ms (121)	0.60 [0.39, 0.82]	.61 [.51, .70]
IMI: effort/importance	29.59 (3.12)	28.83 (3.12)	0.15 [-0.06, 0.37]	.63 [.53, .71]
BAS: fun-seeking	12.60 (1.40)	12.14 (1.40)	0.20 [-0.01, 0.41]	.63 [.54, .72]
PANAS: negative	21.18 (3.84)	20.99 (3.84)	0.03 [-0.19, 0.24]	.74 [.67, .80]
VVQ: verbal	30.40 (2.54)	30.85 (2.54)	0.07 [-0.14, 0.29]	.83 [.78, .87]

Table 3: Experiment 3 session 1 and 2 means with standard deviation values (in parentheses), Cohen's *d* effect sizes of mean differences between sessions 1 and 2 with 95% CIs (in brackets), and test-retest Pearson correlation coefficients with 95% CIs (in brackets).

To test whether criterion shifting is related to the eight additional task and questionnaire measures, Pearson correlations were conducted between each measure and C_n in both the recognition memory and visual detection tasks across both sessions (**Table 4**). No statistically significant relationships existed after FDR correction. These null findings are furthered by computing Bayes factors to assess the amount of support for the null versus alternative hypotheses (BF_{01}) for each comparison. Of the 32 comparisons, 28 showed greater than three times support for the null versus alternative hypothesis ($BF_{01} > 3$), which is considered strong evidence for the null hypothesis (Jeffreys, 1961). Two of the other four comparisons include negative relationships during session 2 between recognition memory C_n and the BAS fun-seeking score ($r(170) = -.16$, CI = -.30, -.01; $BF_{01} = 1.10$) and modified VVQ verbalizer score ($r(170) = -.14$, CI = -.28, .01; $BF_{01} = 1.94$). Both of these findings are in opposition to the positive relationships observed by Aminoff and colleagues (2012), suggesting that a fun-seeking personality and verbal cognitive style are not actually predictive of criterion shifting tendencies. The only comparison to show support for the alternative hypothesis is a negative relationship between recognition memory C_n and the IMI Effort/Importance subscale score during session 2 ($r(170) = -.18$, CI = -.33, -.04; $BF_{01} = 0.57$). Interestingly, a negative relationship between the IMI Effort/Importance subscale score and visual detection C_n during session 2 only slightly supported the null hypothesis ($r(170) = -.16$, CI = -.30, -.01; $BF_{01} = 1.24$). However, relationships between the IMI Effort/Importance subscale score and C_n during session 1 strongly supported the null hypothesis for both the recognition memory ($r(170) = -.06$, CI = -.20, .09; $BF_{01} = 7.39$) and visual detection (*r*(170) = -.06, CI = -.21, .09; BF_{01} = 10.40) tests, suggesting that there is not a consistent relationship between motivation to perform well during the criterion shifting tasks and the extent of criterion shifting itself. Taken together, these standardized measures of risk aversion, response inhibition, working memory, task-switching, motivational effort, and personality attributes cannot explain the vast individual differences in the extent of criterion shifting.

Results from Experiment 3 revealed that strategic criterion shifting tendencies over time are a stable, domain general process. Individuals who consistently shift criteria to large extents during recognition memory tests also regularly shift criteria to large degrees during visual detection tests. Relationships in discriminability across the two decision domains remained weak indicating that the cross-task stability in performance is specific to criterion shifting and not simply a function of overall discriminability performance. Measures from other tasks and questionnaires could not explain individual differences in criterion shifting tendencies. It is possible that other personality or cognitive characteristics not tested in Experiment 3 or by Aminoff and colleagues (2012) are associated with the extent of criterion shifting, but currently no obvious relationships are known. These findings indicate that strategic criterion shifting tendencies are a uniquely individualistic cognitive trait.

Experiment 3: Pearson *r* correlations between C_n and other measures

	recognition memory		**visual detection**	
measure vs. C_n	**session 1**	**session 2**	**session 1**	**session 2**
BART: pumps	-.08 [-.22, .07] (6.32)	.00 [-.15, .15] (10.47)	.03 [-.12, .18] (9.78)	-.05 [-.20, .10] (8.23)

Go/no-go: *d'*	-.07 [-.22, .08] (6.83)	-.03 [-.18, .12] (9.51)	-.11 [-.26, .04] (3.77)	-.08 [-.22, .08] (6.50)
N-back: *d'*	-.01 [-.16, .14] (10.44)	.01 [-.13, .16] (10.28)	-.01 [-.16, .14] (10.40)	.01 [-.14, .16] (10.40)
Switching: time cost	.09 [-.06, .23] (5.52)	.02 [-.13, .17] (9.98)	.03 [-.12, .18] (9.61)	-.02 [-.16, .13] (10.28)
IMI: effort/importance	-.06 [-.21, .09] (7.39)	**-.18 [-.33, -.04]** **(0.57)**	-.01 [-.16, .14] (10.40)	**-.16 [-.30, -.01]** **(1.24)**
BAS: fun-seeking	-.06 [-.20, .09] (8.06)	**-.16 [-.30, -.01]** **(1.10)**	-.06 [-.21, .09] (7.93)	-.06 [-.21, .09] (7.68)
PANAS: negative	.03 [-.12, .18] (9.47)	.01 [-.14, .15] (10.45)	.08 [-.07, .23] (5.84)	-.04 [-.19, .11] (9.21)
VVQ: verbal	-.05 [-.20, .10] (8.22)	**-.14 [-.28, .01]** **(1.94)**	.01 [-.14, .16] (10.43)	-.05 [-.20, .10] (8.30)

Table 4: Experiment 3 Pearson correlation coefficients and 95% CIs (in brackets) between C_n in both sessions and tasks against the eight additional task and questionnaire measures. Bayes factors supporting the null versus alternative hypotheses (BF_{01}) are presented in parentheses below each correlation coefficient. BF_{01} scores greater than 3 are considered strongly supportive of the null hypothesis (Jeffreys, 1961). Values that include BF_{01} values below 3 are in bold, though no significant relationships existed after FDR correction.

Experiments 4 & 5: Differentiating criterion shifting from confidence ratings

Experiment 4

Findings from Experiment 3 revealed that criterion shifting tendencies are stable within and across decision domains without any obvious relationship to certain cognitive abilities or a motivation to perform well on the criterion shifting tasks. In Experiment 4, these findings are furthered by assessing whether strategic criterion shifting strategies during recognition memory tests are related to an individual's meta-awareness of the memory strength elicited by test items as measured by metacognitive bias and metacognitive sensitivity via confidence ratings (Fleming & Lau, 2014). Confidence ratings are typically implemented in recognition memory studies to assess criterion shifts (Macmillan & Creelman, 2005; Yonelinas & Parks, 2007), making it reasonable to predict that strategic criterion shifting and decision strategies for reporting confidence are strongly related. In the first three experiments, the tasks required extreme criterion shifts in order to maximize payoffs (Experiments 1 and 3) or accuracy (Experiment 2). In these extreme situations, people should only choose the riskier option (i.e. respond "old" when a conservative criterion is advantageous or "new" when a liberal criterion is propitious) when they have high confidence that it is the correct choice. If people use meta-awareness of the familiarity strength elicited by test items to strategically shift criteria, then criterion shifting tendencies should be strongly related to how often people report high confidence in "old" and "new" responses (i.e. metacognitive bias). However, some people may be more adept than others at discerning between different levels of familiarity strength (i.e. have higher metacognitive sensitivity), which may allow them to shift criteria to greater extents since they can better differentiate between stronger and weaker memory evidence. To assess whether metacognitive bias or metacognitive sensitivity could explain individual differences in criterion shifting tendencies the paradigm directly compared performance on recognition memory tests that either required strategic criterion shifting or confidence ratings.

Participants on *average* tend to have high metacognitive sensitivity, demonstrating that confidence ratings scale well with the accuracy of old/new judgments (Mickes et al., 2007; 2011). However, Mickes and colleagues (2011) showed that there are individual differences in metacognitive bias, particularly when scaling strong memories. Some people use extreme criteria for making old/new judgments with the highest level of confidence while others are less judicious. Metacognitive bias may be related to individual criterion shifting tendencies if people shift criteria based on the level of confidence in a recognition judgment. However, people may implement completely different decision strategies when rating confidence versus strategically shifting a criterion. For instance, a criterion shift may reflect an individual's willingness to make a strategic old/new response *before* making a recognition judgment whereas a confidence rating could represent an assessment of an old/new response *after* the judgment is made. Mickes and colleagues (2017) provide some evidence that people employ different decision strategies for rating confidence versus shifting criteria, since individuals tended to establish more conservative criteria when making high confident "old" judgments on a multipoint scale compared to binary decisions for responding "old" when specifically instructed to only do so when there is 100% confidence. Miller and Kanter (2019) failed to find any significant relationships between metacognitive bias and the extent of criterion shifting during recognition memory tests when conducting post hoc analyses on previously reported data. This alludes to the possibility that metacognitive bias might be a poor indicator of strategic criterion shifting tendencies, but no studies have systematically compared this relationship during recognition memory tests *a priori*.

In Experiment 4, participants conducted two different recognition memory tests on separate days that required either a confidence judgment on a six-point scale (confidence ratings session) or a binary old/new response (binary response session). During the confidence ratings session, participants responded to each test image on a six-point scale ranging from high confidence “new” to high confidence “old.” In the binary response session participants conducted the exact same task except instead of responding on a six-point scale, participants received instructions to only respond “old” with “high confidence” (otherwise respond “new”) in the conservative criterion condition or only respond “new” with “high confidence” (otherwise respond “old”) in the liberal criterion condition. This made the instructions as similar as possible across the two tasks and gave participants the best opportunity for establishing the same criteria for high confident responses, regardless of whether participants responded on a six-point scale or made binary old/new judgments.

Given the results of Mickes and colleagues (2017) and Miller and Kantner (2019), it is predicted that there is no relationship between metacognitive bias and strategic criterion shifting. It is also predicted that individuals will establish more extreme criteria when rating high confidence on a multipoint scale relative to the extent of criterion shifting to explicit instructions. Such a finding would indicate that individuals who do not adequately shift criteria are at least *capable* of shifting criteria to greater extents but might simply be *unwilling* to do so. Additionally, no relationship was predicted between metacognitive sensitivity and the extent of criterion shifting since meta-awareness of memory strength is not predicted to affect an individual’s *willingness* to shift a criterion.

Method

Participants

One hundred and seventy participants (45 males; ages 18–34, *M* = 19, *SD* = 1.8) completed both sessions on separate days within the same week. Three additional participants only completed one of the two sessions and are excluded from all analyses. Participants earned $10 for completing each session.

Although no relationship was predicted between the extent of criterion shifting and metacognitive bias or metacognitive sensitivity, it is important to ensure that a modest effect is statistically powered if a relationship does exist. An *a priori* power analysis revealed that data collection on 123 participants provides 80% power for detecting a Pearson correlation of $r = .25$, which is a non-negligible relationship that could help explain individual criterion shifting tendencies. Initially, all participants conducted the confidence ratings session first because presenting the tasks in this order gave participants the best opportunity to implement similar decision strategies for rating confidence and strategically shifting criteria. However, an unexpected relationship did exist which prompted collection of additional participants who conducted the two tasks in the reverse order, to rule out potential order effects. A second *a priori* power analysis revealed that 46 participants provide 95% power to find an effect of $r = .50$, a value derived from the initial sample (see **Results and discussion**).

Procedure

The two self-paced sessions lasted for 20-45 minutes where participants conducted four cycles of studying 75 unique face stimuli followed by three, 50-image test mini-blocks (**Figure 9**). In one of the two test sessions, participants made recognition judgments on a six-point confidence scale (high confidence new, medium confidence new, low confidence new, low confidence old, medium confidence old, or high confidence old) for all test blocks (the confidence ratings session). The other session required a binary old/new judgment (the binary response session), but under three different conditions. Participants received instructions to either (1) simply respond “old” or “new” (neutral condition), (2) only respond “old” when there is high confidence an image is old (conservative condition), or (3) only respond “new” when there is high confidence an image is new (liberal condition).

In the study phase, participants passively viewed a sequence of images in the center of a computer screen on a black background for 700 ms followed by a 100 ms presentation of a white crosshair. After viewing a test image for 700 ms, participants made a response with unlimited time, followed by a 300 ms crosshair presentation. During confidence ratings sessions, participants made high confidence responses using the “1” and “9” keys, medium confidence responses with “2” and “8” keys, and low confidence responses with the “3” and “7” keys. The familiarity strength corresponding to the response types either increased or decreased from left-to-right depending on the pseudorandom assignment (i.e. high confidence “new” to high confidence “old,” or vice versa). The response screen displayed all six keys with the corresponding response values. For binary response sessions, participants made old/new judgments via the “1” or “0” keys where a response screen reminded participants of the mapping between keys and response types. The conservative and liberal criterion conditions included the phrase “high confidence only” below the old or new response text, respectively. The stimuli included 1,200 unique face images (600 per session) and each participant received a completely randomized assignment and presentation order for the target and lure images.

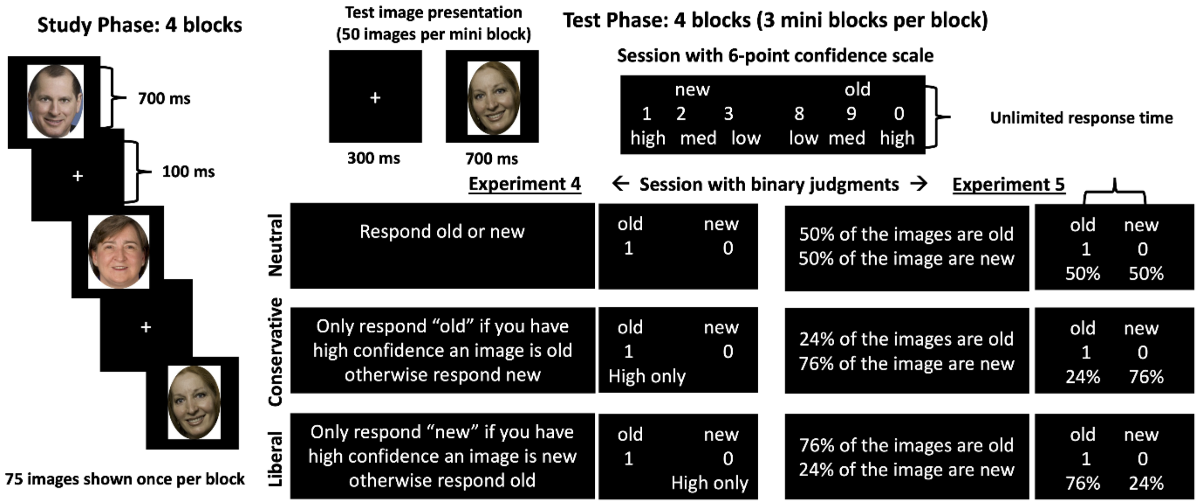


Figure 9: Experiments 4 and 5 recognition memory tasks. Participants conducted two task sessions on separate days. In one session participants made confidence judgments on a 6-point scale. The other session required making binary old/new judgments with a conservative, liberal, or neutral criterion manipulation. In Experiment 4, participants received instructions to only respond old or new with "high confidence" in the conservative and liberal criterion conditions, respectively, whereas a base rate manipulation induced criterion shifts in Experiment 5.

Statistical Analysis

Although traditional measures of metacognitive bias do not account for accuracy (e.g. the percentage of "high confidence" responses made throughout an entire recognition test regardless if responses are correct or not; see Fleming and Lau, 2014), it is important to make direct comparisons between strategic criterion shifts and metacognitive bias. Therefore, different confidence levels were classified as decision criteria and used to compute conservative, liberal, and neutral c_n by treating different confidence ratings as binary old/new responses. Computations of a "conservative" c_n occurred by only treating high confident old responses as "old" and all other response types as "new" (i.e. the criterion for "high confidence old" responses). Calculations for a "liberal" c_n occurred by only treating high confident new responses as "new" and all other response types as "old" (i.e. the criterion for all responses except "high confidence new"). This allowed for computations of a measure of "metacognitive C_n" (i.e. "conservative" c_n minus "liberal" c_n, which is the measure used by Miller and Kantner, 2019) for the confidence ratings session that is computed the same way as C_n in the binary response session. For measures of "neutral" c_n, all old responses were considered "old" and all new responses as "new," regardless of the confidence level assigned to each response.

To measure metacognitive sensitivity, computations of the area under the type-2 receiver operator characteristic (AUROC2) curve occurred, which is more robust to influences of decision bias compared to other common measures, such as type-2 *d'* (an analogous measure to *d'*) (see Fleming & Lau, 2014).

Results and discussion

As specified in a preregistration (https://osf.io/4wnjm), collection of an initial dataset of 122 subjects occurred where all participants first conducted the confidence ratings session. Since no relationship was expected in C_n between the two tasks, it seemed best to allow participants to familiarize themselves with the confidence ratings structure so that decision strategies for identifying old and new items with "high confidence only" could easily be transferred to the binary response session. However, an unexpected relationship existed in C_n across the two tasks ($r(120) = .60$, CI = .47, .70). To test for potential order effects, data collection ensued from an additional 48 participants who conducted the two sessions in the reverse order (as specified in a subsequent preregistration: https://osf.io/ae2rp), but a modest relationship in C_n persisted ($r(46) = .39$, CI = .12, .61). Since the relationship in C_n across the two tasks could not be completely attributed to an order effect, data from all 170 participants were combined for subsequent analyses.

Mean *d'* did not significantly differ between the confidence ratings session (*M* = 0.56, *SD* = 0.26) and the neutral criterion condition of the binary response session (*M* = 0.56, *SD* = 0.22), *d* = 0.00, 95% CI [-0.21, 0.22]. However, a strong relationship in *d'* existed between the two sessions (*r*(168) = .49, CI = .37, .60). Similarly, c_n did not significantly differ between the confidence ratings session (*M* = 0.21, *SD* = 0.24) and the neutral criterion condition of the binary response session (*M* = 0.27, *SD* = 0.19), *d* = 0.15, 95% CI [-0.06, 0.37], and a strong relationship in c_n existed between the two sessions (*r*(168) = .62, CI = .52, .71). This confirms that the different test instructions across the two tasks did not substantially affect discriminability and neutral criterion placement. As in Experiments 1 and 2, no significant relationship existed in the binary response session between C_n and neutral c_n (*r*(168) = .08, CI = -.08, .22; BF_{01} = 3.52), providing more evidence that placing and shifting a criterion are independent decision processes.

Unexpectedly, a strong relationship existed in C_n (*r*(168) = .53, CI = .41, .63) between the two tasks, suggesting that metacognitive bias is predictive of the extent of criterion shifting in this paradigm. In the confidence ratings session, participants on average drastically shifted between the “conservative” c_n (*M* = 1.28, *SD* = 0.54) and “liberal” c_n values (*M* = -1.21, *SD* = 0.69), *d* = 3.97, 95% CI [3.60, 4.34]. Participants shifted criteria to a large extent in the binary response session between c_n in the conservative (*M* = 0.72, *SD* = 0.40) and liberal criterion conditions (*M* = -0.27, *SD* = 0.47), *d* = 2.00, 95% CI [1.73, 2.26]. However, metacognitive C_n (*M* = 2.49, *SD* = 0.65) proved to be much greater than C_n when making binary old/new judgments (*M* = 0.99, *SD* = 0.65), *d* = 1.64, 95% CI [1.40, 1.89]. Even though the instructions for reporting high confidence remained similar across the two tasks, virtually all participants established much more extreme criteria for high confident responses when asked to report on a six-point scale. **Figure 10** illustrates individual differences in conservative and liberal c_n values across the two sessions in order from left to right based on the largest to smallest metacognitive C_n value. Finally, assessments of whether differences in metacognitive sensitivity in the confidence ratings session could predict the extent of C_n during the binary response session showed no relationship between AUROC2 (M = .57, SD = .04) in the confidence ratings session and C_n in the binary response session (r(168) = .07, CI = -.08, .22; BF_{01} = 7.17).

Experiment 4 revealed that participants generally establish much more extreme criteria for high confident “old” and “new” responses when reporting on a six-point scale versus making binary old/new judgments. However, a strong relationship existed between C_n in the binary response session and metacognitive C_n in the confidence ratings session. Individuals who maintain extreme criteria for the highest levels of confidence on a six-point scale also shifted criteria to a large extent (though to a much lesser degree), while those who established less extreme criteria for high confident responses also shifted criteria to a smaller extent. This indicates that individuals may implement similar decision strategies for making confidence judgments and strategically shifting a criterion, which is contradictory to the findings of Miller and Kantner (2019). However, Miller and Kantner (2019) examined data that included either a base rate or payoff manipulation, which does not cue participants to respond based on a level of confidence. Since the criterion manipulation in Experiment 4 involved instructions to respond based on confidence levels, participants may have treated the criterion manipulation as a type of confidence judgment (i.e. a response on a two-point confidence scale). It is possible that individuals only use similar decision processes for strategic criterion shifting and metacognitive bias when the criterion manipulation explicitly instructs participants to respond based on confidence levels.

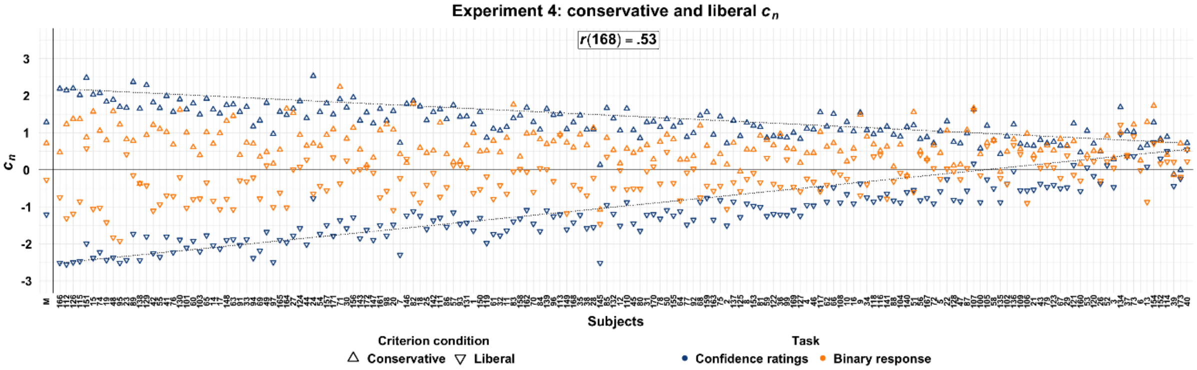


Figure 10: Experiment 4 c_n values for each participant and the mean (M) in the conservative and liberal criterion conditions of the binary response session (orange) and “conservative” and “liberal” c_n values computed from the confidence ratings session (dark blue). The extent of criterion shifting is depicted by the distance between the triangles representing the conservative and liberal c_n values. Participants are ordered from left to right based on the largest to smallest metacognitive C_n value. The dotted lines emphasize individual differences in metacognitive C_n by connecting the “conservative” and “liberal” c_n values from the confidence rating sessions of the leftmost and rightmost subjects.

Experiment 5

Experiment 4 showed a strong relationship between the degree to which individuals use “high confidence” judgments on a six-point scale versus shifting criteria to instructions that require responding “old” or “new” with “high confidence only.” Since this result contradicts the findings of Miller and Kantner (2019), a subsequent test of whether the extent of criterion shifting in response to a criterion manipulation without reference to confidence levels would also be related to metacognitive bias. The instruction manipulation was changed in Experiment 4 to a base rate manipulation in Experiment 5. Again, no relationship was predicted between the extent of criterion shifting and metacognitive bias (see the preregistration: https://osf.io/tqc42).

Procedure

The procedures of Experiment 5 matched those of Experiment 4 except the binary response session induced criterion shifts with a base rate manipulation instead of instructions. Participants received information prior to each test block about the likelihood of encountering old and new items. An old item appeared either 24% (conservative criterion condition), 50% (neutral criterion condition), or 76% (liberal criterion condition) of the time during a test block (**Figure 9**). Text appeared below each response type to indicate the likelihood of encountering an old or new image during a test block. The session order was counterbalanced across participants, and the stimuli included 1,200 unique face images (600 per session) which differed from the stimulus set of Experiment 4.

Participants

One hundred and twenty-nine participants (37 males; ages 18–48, *M* = 22, *SD* = 4.5) successfully completed both sessions within a week. Four additional participants failed to complete both sessions and are excluded from all analyses. Participants received $10 for completing each session.

Results and discussion

Similar to Experiment 4, no significant differences existed in mean *d'* between the confidence ratings session (*M* = 0.58, *SD* = 0.17) and the neutral criterion condition in the binary response session (*M* = 0.56, *SD* = 0.18), *d* = 0.08, 95% CI [-0.16, 0.33], and a strong relationship in *d'* existed between the two tasks (*r*(127) = .61, CI = .48, .70). The c_n values on the recognition tests also showed no significant differences between the confidence ratings session (*M* = 0.18, *SD* = 0.18) and the neutral criterion condition in the binary response session (*M* = 0.15, *SD* = 0.11), *d* = 0.10, 95% CI [-0.15, 0.34], while showing a strong relationship between sessions (*r*(127) = .64, CI = .52, .73). As in Experiments 1, 2, and 4, no significant relationship existed in the binary response session between C_n and neutral c_n (*r*(127) = -.14, CI = -.31, .03; BF_{01} = 1.44).

In the confidence ratings session, participants on average dramatically shifted between "conservative" c_n (*M* = 1.26, *SD* = 0.53) and "liberal" c_n values (*M* = -1.24, *SD* = 0.74), *d* = 3.92, 95% CI [3.50, 4.34]. Participants also shifted criteria in the binary response session on average between c_n in the conservative (*M* = 0.42, *SD* = 0.21) and liberal criterion conditions (*M* = -0.06, *SD* = 0.24), *d* = 1.68, 95% CI [1.39, 1.96]. Unlike Experiment 4, no relationship existed between metacognitive C_n (*M* = 2.50, *SD* = 0.82) and C_n in the binary response session (*M* = 0.48, *SD* = 0.82), *r*(127) = -.06, 95% CI = [-.23, .11]; BF_{01} = 3.87. This suggests that strategic criterion shifting tendencies are unrelated to metacognitive bias in this paradigm. Similar to Experiment 4, a large mean difference in C_n existed between the two tasks, *d* = 2.51, 95% CI [2.18, 2.84]. Metacognitive sensitivity as measured by AUROC2 (*M* = .57, *SD* = .04) in the confidence ratings session showed no relationship with C_n in the binary response session (*r*(127) = .03, CI = -.14, .21; BF_{01} = 8.44). **Figure 11** displays individual differences in conservative and liberal c_n values across the two sessions in order from left to right based on the largest to smallest metacognitive C_n.

Experiment 5 revealed that individual criterion shifting tendencies in response to a base rate manipulation are unrelated to individual differences in metacognitive bias during recognition memory tests. This finding is in line with those of Miller and Kantner (2019), who also found no relationship between criterion shifting and metacognitive bias from post hoc data analyses. The extent of strategic criterion shifting only appears to relate to metacognitive bias when the criterion manipulation specifically requires a response based on a level of confidence. When a criterion manipulation does not include instructions to respond based on confidence, individuals seem to implement different decision strategies for rating judgements with high confidence and strategically shifting a criterion.

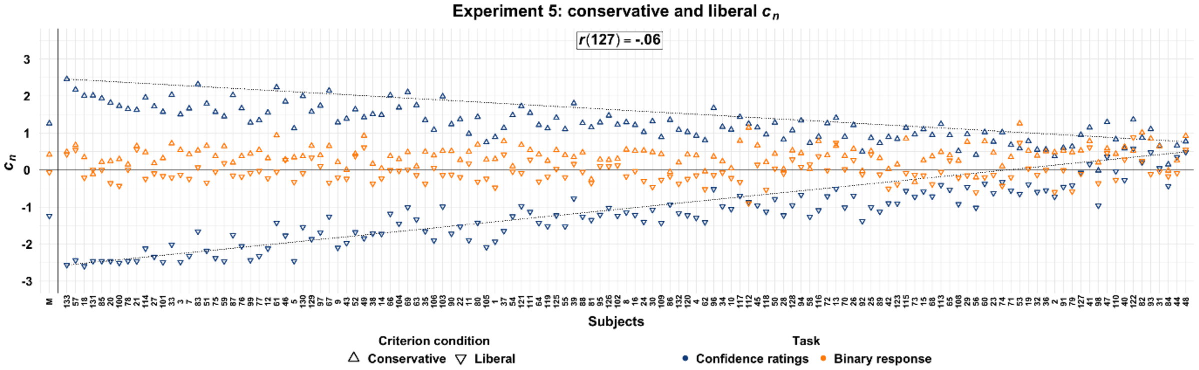


Figure 11: Experiment 5 c_n values for each participant and the mean (M) in the conservative and liberal criterion conditions of the binary response tests (orange) and "conservative" and "liberal" c_n values computed from the confidence ratings session (dark blue). The extent of criterion shifting is depicted by the distance between the triangles representing the conservative and liberal c_n values. Participants are ordered from left to right based on the largest to smallest metacognitive C_n value. The dotted lines emphasize individual differences in metacognitive C_n by connecting the "conservative" and "liberal" c_n values from the confidence rating sessions of the leftmost and rightmost subjects.

General Discussion

The tendency to strategically shift criteria should be considered a stable cognitive trait if it (1) shows strong test-retest reliability, (2) generalizes across tasks, and (3) cannot be explained by other cognitive factors. These findings demonstrate that criterion shifting during recognition memory is quite stable across many testing sessions and is as strong as the stability of criterion placement—a stable cognitive trait (Kantner & Lindsay, 2012; 2014). The test-retest reliability of criterion shifting in these experiments is even comparable to measures believed to reflect stable traits in other cognitive domains. For example, Xu and colleagues (2018) administered test-retest working memory tasks across 30 days and determined that the strong consistency in performance indicates that visual working memory capacity is a stable trait. The high session-to-session correlation coefficients observed by Xu and colleagues (2018) (*r*(77) range: .64-.86, *Mdn* = .77) are as strong as the session-to-session correlations obtained for C_n during recognition memory tests in Experiment 1 (*r*(37) range: .58-.85, *Mdn* = .75), Experiment 2 (*r*(37) range: .71-.89, *Mdn* = .83), and Experiment 3 (*r*(170) = .75). It should be noted, however, that these paradigms purposely implemented tests with low discriminability and fairly extreme criterion manipulations. These situations require some of the largest criterion shifts to maximize payoffs (Experiments 1 and 3) or accuracy (Experiment 2), and it is possible that the stability of criterion shifting is weaker when discriminability is much higher and/or the criterion manipulations are less extreme (e.g. when targets or nontargets appear 55% as opposed to 75% of the time). Nevertheless, the strong test-retest stability of criterion shifting during recognition memory tests that encourage large criterion shifts, appears to qualify as a trait-like feature.

The stability of individual tendencies to strategically criterion shift is not limited to a single task, but generalizes across stimuli sets (Aminoff et al., 2012), bias manipulations (Kantner et al., 2015; Frithsen et al., 2018), and decision domains (Frithsen et al., 2018). In Experiment 3, these findings were furthered by illustrating that the test-retest reliability of C_n extends across decision domains during recognition memory and visual detection tests (*r*(170) range: .57-.78, *Mdn* = .67). Importantly, the relationship in *d'* between the two decision domains remained weak (*r*(170) range: .14-.17, *Mdn* = .16), showing that the cross-task stability in performance is specific to criterion shifting and not discriminability. Although there are occasional inconsistencies in the cross-task stability of criterion shifting

(Franks & Hicks, 2016; Frithsen et. al, 2018), it appears that weak relationships occur when there are large disparities in demand characteristics (e.g. when there are differences in the experimental designs that may affect an individual's decision strategy). However, when demand characteristics are carefully controlled for, the stability of criterion shifting across tasks and decision domains is generally quite strong, suggesting that differing task designs may lead to occasional inconsistencies in criterion shifting stability and not simply the decision domains themselves (as suggested by Franks and Hicks, 2016). Future research needs to investigate the underlying factors that can lead to differing demand characteristics and why such differences may sometimes affect the stability of criterion shifting, but strategic criterion shifting tendencies appear to be a domain general process.

Though the extent to which individuals strategically shift criteria proved stable over time and decision domains, Experiment 3 investigated whether these stable individual differences reflect an epiphenomenon of other cognitive or personality traits. Aminoff and colleagues (2012) first attempted to identify traits associated with individual criterion shifting tendencies during recognition memory tests by correlating the extent of C_n with many standardized measures of cognitive and personality characteristics. Despite collecting over 25 cognitive and personality measures, Aminoff and colleagues (2012) found that the extent of C_n only significantly related to one cognitive measure (a positive relationship with the modified VVQ verbal score) and two personality measures (a positive relationship with the BAS fun-seeking score and a negative relationship with the PANAS-X negative affect score). However, in Experiment 3 these findings were not replicated as no relationship between the extent of C_n on recognition memory and visual detection tests with the modified VVQ verbal score, BAS fun-seeking score, or PANAS-X negative affect score. To expand on Aminoff and colleagues (2012) efforts to identify characteristics associated with individual criterion shifting tendencies, Experiment 3 sought to examine relationships between the extent of C_n and performance on other cognitive tasks. For example, people who perform worse on working memory tests might be less able to maintain the strategy goals necessary to strategically shift criteria, resulting in little to no shifting. However, no significant relationship existed between the extent of C_n on recognition memory and visual detection tests with standardized test measures that assess risk aversion, response inhibition, working memory, and task-switching ability. Self-reports of motivation to perform well on the tasks also showed no relationship with the extent of C_n, nor did measures of metacognitive sensitivity. Interestingly, in Experiment 4 a strong relationship was observed between the extent of C_n when making binary responses versus metacognitive C_n during recognition memory tests ($r(168) = .53$), but this only occurred when the criterion manipulation included instructions to respond based on levels of confidence. When the criterion manipulation did not cue participants into responding with confidence in Experiment 5, no significant relationship existed between the extent of C_n during recognition memory tests with a base rate manipulation and metacognitive C_n. This suggests that people implement different decision strategies for conveying meta-awareness of the uncertainty in familiarity strength via high confident responses versus strategically shifting criteria. However, it is possible that the criterion manipulation was not extreme enough to appropriately align with people's criteria for responding with high confidence. For instance, if the base rate manipulation was more extreme (e.g. 95% of test items are targets or nontargets), then people may have shifted criteria to greater extents and a significant relationship might be observed between the extent of C_n and metacognitive C_n. Still, the findings in Experiment 5 match those of Miller and

Kantner (2019) suggesting that people use different strategies when establishing decision criteria for rating recognition memory judgments with “high confidence” versus making decisions in situations where extreme criterion shifts promote better decisional outcomes. Although Experiment 3 and Aminoff and colleagues (2012) tested many factors that could potentially relate to criterion shifting tendencies, there still are countless numbers of other measures that may explain individual differences in criterion shifting tendencies. As of right now, there currently are no known measures that can reliably predict the extent to which an individual will strategically shift criteria except for criterion shifting performance itself on another task. However, it is possible that other characteristics are associated with criterion shifting tendencies that have yet to be tested. Despite this, the strong stability of criterion shifting across time, tasks, and decision domains, coupled with the fact that individual differences cannot be easily attributed to many other factors, demonstrates that the tendency to strategically criterion shift appears to be a uniquely individualistic cognitive trait.

Although criterion shifting tendencies are quite stable within people, there are vast individual differences across people (Aminoff et al., 2012, 2015; Kantner et al., 2015; Frithsen et al., 2018; Layher et al., 2018; Miller & Kantner 2019). Individual differences in strategic criterion shifting do not appear to be a result of an *inability* for certain people to shift criteria. In Experiments 4 and 5, almost all individuals used much more extreme criteria for responding “old” and “new” with high confidence on a six-point scale compared to the extent of criterion shifting even when the criterion manipulation specifically instructed participants to respond with “high confidence only.” Mickes and colleagues (2017) made a similar finding by showing that participants establish a much more conservative criterion when responding with the highest level of confidence on a multipoint scale compared to when instructions state to only respond “old” when there is 100% confidence. This shows that individuals are indeed *capable* of shifting criteria to more extreme extents if they simply adopt the same extreme criteria for responding with high confidence as they do for strategic criterion shifting. However, it appears that extreme differences in strategic criterion shifting are a result of individual differences in a *willingness* to disregard uncertain evidence in favor of a decision strategy that maximizes accuracy or payoffs[3] (Green & Swets, 1966; Aminoff et al., 2012; Kantner et al., 2015; Miller & Kantner. 2019). Considering strategic criterion shifting tendencies as a stable cognitive trait is likely due to individual differences in a willingness to shift a criterion will better inform theories of criterion placement and shifting. Future experiments must examine the nuances of these individual difference to gain a full understanding of how people adapt decision criteria to particular situations and outline specific considerations when investigating these decision strategies at an individual level.

An individual differences approach can elucidate the nature of a phenomenon

Suboptimality

---

[3] Green and Swets (1966, p. 91) nicely describe the potential thought process of an individual who is unwilling to shift criteria to extreme extents: “The observer tends to avoid extreme criteria: when the optimal β is relatively large, his actual criterion is not so high as the optimal criterion, and when the optimal β is relatively small, his criterion is not so low as the optimal criterion. Although this pattern is consistent with studies of decision making under uncertainty which do not involve ambiguous sensory information, the significance of its appearance here is not totally clear. It may be suspected that the subject’s natural disinclination to make the same response on all trials is strengthened by his awareness that the experimenter’s principle interest is in a sensory process. He 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.”

The fact that criterion shifting tendencies are stable within participants, yet variable across people, emphasizes the importance of understanding criterion shifting tendencies at an individual level. For several decades, most studies of criterion shifting drew conclusions from group-averaged data, neglecting the vast individual differences in shifting behavior. One generalized conclusion drawn from group averages is that people are quite suboptimal at appropriately adapting a decision criterion to a particular situation (Ulehla, 1966; Parks, 1966; Thomas & Legge, 1970; Kubovy, 1977; Hirshman, 1995; Maddox & Bohil, 2005; Benjamin, Diaz, & Wee, 2009; Lynn & Barret, 2014). Although the classification of an "optimal" criterion is strongly debated (Lynn & Barret, 2014), it is important to convey that decisional outcomes can dramatically vary depending on how well individuals adapt decision criteria to a particular situation. For example, if only 25% of items on a recognition test contain previously studied images (targets), then the optimal criterion is conservative, but the magnitude of the optimal placement will depend on how well a person can discriminate between old and new images. If a person is completely unable to distinguish between old and new images, a maximally conservative criterion is optimal because responding "new" every time will result in a correct response rate of 75% (no false alarms, but no hits either). However, if items are highly familiar, then the optimal conservative criterion is much less extreme because an individual can be correct more than 75% of the time by responding "old" to very familiar items even if it results in an occasional false alarm.

Many theories of suboptimal criterion shifting posit that people fail to shift criteria extreme enough because individuals probability match (Parks 1966; Thomas & Legge, 1970), erroneously estimate signal and noise distributions (Ulehla, 1966; Kubovy, 1977), poorly integrate decisional evidence with decisional outcomes (Lynn & Barrett, 2014), or are unable to maintain a stable criterion during a test block (Benjamin et al. 2009). These are reasonable explanations based on group averages, but many of these theories inadequately describe performance at an individual level because some people do consistently shift criteria quite well, some shift to modest degrees, while others hardly shift at all. Even when individuals inadequately shift criteria across situations, there are instances where people will consistently establish an appropriately conservative (or liberal) criterion in one situation but fail to shift when a liberal (or conservative) criterion is advantageous in another situation. For example, subject 2 in Experiment 2 deploys such conservative criteria in the conservative conditions (when 25% of test image are old) and subsequently makes a correct response 73% of the time on average. Yet, this subject fails to shift criteria in the liberal conditions (when 75% of test images are old) resulting in being correct only 44% of the time on average (see **Figure 6**). These instances are at odds with theories that suggest people poorly estimate signal and noise distributions (Ulehla, 1966; Kubovy, 1977) or misestimate decisional parameters given the strength of discriminability (Lynn & Barrett, 2014) because it is unreasonable to believe these individuals are quite skilled at such estimations in some situations (e.g. when a conservative criterion is advantageous), but are grossly inept in others (e.g. when a liberal criterion is optimal). A theory of suboptimal criterion shifting must account for these individuals who strategically adapt a conservative criterion but fail to shift to a liberal criterion (and vice versa). That is, the degree to which people are suboptimal at adapting a decision criterion depends on the individual *and* the situation.

Assessments of individual differences may not necessarily falsify previous hypotheses of suboptimal criterion shifting based on group averages, but these assessments certainly better inform them. For instance, Benjamin and colleagues (2009) suggest that a participant’s criterion will fluctuate throughout a test block creating “criterial noise” that results in measurements of criterion placement that are suboptimal. Criterial noise is a plausible phenomenon that might be occurring at an individual level. However, this hypothesis needs refinement to include the possibility that the amount of criterial noise may vary considerably across people. That is, some individuals may have a lot of criterial noise which may lead to relatively small criterion shifts, while others who shift to large extents may do so with little to no criterial noise. Any account of suboptimal criterion strategies must consider these consistent individual differences in order to fully understand the nature of strategic criterion shifting.

Improving criterion shifting through awareness, feedback, and motivation

Findings from *group averages* reveal that criterion shifting is improved when people are made aware of the advantages for shifting, provided with feedback on criterion shifting performance, and presented with motivating factors to shift (Rhodes & Jacoby, 2007). If people are unaware of the advantages for shifting criteria, criterion shifts are generally not observed (Rhodes & Jacoby, 2007; Verde & Rotello, 2007). When people are made aware of the advantages for shifting, the extent of criterion shifting increases *on average*, but analyses of individual differences reveal that this is not true for everyone (Aminoff et al., 2012, 2015; Kantner, et al., 2015; Frithsen et al., 2018; Layher et al., 2018; Miller & Kantner 2019). The current experiments revealed that these individual differences in criterion shifting behavior are remarkably consistent across multiple testing sessions. People who shift to large extents during one testing session do not simply regress back to the mean on subsequent sessions.

Rather, awareness of the advantages for criterion shifting impacts the extent of shifting differently for each person.

To increase awareness of the advantages for shifting criteria, some studies provide corrective feedback, which improves criterion shifting performance at a group level (Rhodes & Jacoby, 2007; Kantner, et al., 2015). However, the extent to which this is true may vary considerably across individuals. For instance, participants received performance feedback at the end of each test block (Experiment 3) or testing session (Experiments 1 and 2), which may have cued some participants to shift to greater extents on subsequent sessions to increase total payout (Experiments 1 and 3) or accuracy (Experiment 2). Although some people shifted to greater extents after the first session, several others shifted to similar or lesser degrees during successive sessions (e.g. subjects 19 and 37 in Experiment 1; see **Figure 4**). It seems that criterion shifting tendencies are unaffected by feedback for some individuals. However, feedback may more effectively alter individual criterion shifting tendencies if participants directly benefit from shifting to greater extents. Kantner and colleagues (2015) found that corrective feedback made individuals shift criteria more extremely when a payoff manipulation induced criterion shifting versus a paradigm where participants simply received instructions to shift. Since a criterion manipulation with instructions does not affect a participant's total payment, some individuals may be unwilling to shift criteria more extremely in response to feedback. However, when shifting to greater extents leads to a greater payout, feedback seems to be more effective at altering criterion shifting tendencies for *some* individuals. Future studies must assess how feedback under different circumstances affects *individual* criterion shifting tendencies as feedback will likely make some individuals consistently shift to greater extents, others will likely be completely unaffected by feedback, and some may only be affected by feedback under certain conditions (e.g. when there is a direct benefit for shifting criteria).

Another factor that may affect the extent of criterion shifting is an individual's motivation to shift criteria, which proved to be unrelated to a person's self-reported motivation to perform well during recognition memory and visual detection tests that incentivized criterion shifting. On average, participants shifted criteria to a greater extent in response to the payoff manipulation in Experiment 1 ($M_{(C_n)}$ = 2.00) compared to the base rate manipulation in Experiment 2 ($M_{(C_n)}$ = 1.27), even though both manipulations required the same degree of criterion shifting for *optimal* performance in an SDT framework. The payoff manipulation in Experiment 1 may have motivated *some* individuals to shift to a greater extent compared to the base rate manipulation in Experiment 2, since the extent of criterion shifting directly impacted payment. However, some individuals in Experiment 1 continuously shifted to a small degree across all 10 sessions despite receiving relatively low payouts, while many individuals shifted to a large extent during Experiment 2 even though doing so did not affect total payment. This suggests that there might be individual differences in the factors that motivate people to criterion shift to greater extents, but within-subject paradigms are needed to ensure that this finding is not due to other factors (e.g. individuals may simply shift criteria to lesser extents in response to a base rate versus payoff manipulation regardless of motivating factors). Kantner and colleagues (2015) observed individual differences in motivating factors for shifting criteria during recognition memory tests in a paradigm where a study phase preceded a test phase in one condition, but "malfunctioned" and did not actually present any images in another condition. In the latter case, participants should be highly motivated to shift criteria since there is no reliable memory evidence to guide the decision. Some individuals in the test condition *without* a study phase appropriately maximized responses by always responding “old” or “new” depending on the criterion manipulation, but others failed to adequately shift criteria despite never actually viewing any images before the test phase! The extent of criterion shifting was completely unaffected by the presence of a study phase or not for some individuals. Post-study debriefings suggest that these participants still attempted to use perceptual features, such as skin tone, to guide decisions despite being told that such features are not diagnostic of whether an image is old or new. Some individuals seem more motivated to attempt to provide correct responses instead of consistently choosing the response that maximizes accuracy or payoffs, even under conditions of complete uncertainty. Overall, there are individual differences and several nuances in the degree that awareness, feedback, and motivation affect the extent of criterion shifting. Assessments of group averages are insufficient for identifying ways to improve criterion shifting performance because the influence of these factors on the extent of criterion shifting seems specific to the individual.

Consequences of not criterion shifting

Failing to adequately shift decision criteria can be quite consequential, particularly at the individual level. To illustrate this, performance and payment outcomes from Experiment 3 are presented, where participants earned five cents for each correct response, lost ten cents for critical errors, but received no penalty for noncritical errors during two sessions of recognition memory and visual detection tests. On average, participants earned a total bonus of $26.32 across the four tests and attained a mean d' = 1.10. Given the payout structure and SDT model, a person with a d' = 1.10 who shifts criteria to an extent that maximizes total payment would earn $29.30. The fact that participants *on average* only fall short of the maximum payout (given the mean level of d') by 11% suggests that the consequences of suboptimal criterion shifting are relatively minor. However, when examining individual performance, it becomes quite clear that not shifting criteria carries major consequences. For instance, Experiment 3 subject 4 (E3-4) and E3-123 both attained relatively low mean d' scores across both tasks and sessions ($M_{(d')} = 0.52$ and $M_{(d')} = 0.36$, respectively). E3-4 on average did not shift criteria across the four tests ($M_{(C_n)} = -0.05$), whereas E3-123 shifted criteria quite well ($M_{(C_n)} = 3.38$). Even though E3-4 garnered more correct responses than E3-123, this individual only earned a bonus of $9.75 while E3-123 earned $24.25. By simply shifting criteria, E3-123 earned 2.5 times more money than E3-4 despite worse discriminability performance! However, classifying E3-4 as being generally suboptimal at placing a criterion is an inaccurate depiction of this individual’s behavior. On average, E3-4 maintained a conservative criterion ($M_{(c_n)} = 0.62$) when false alarms resulted in a ten-cent loss, and earned $10.85 across all conservative conditions. In the liberal conditions, E3-4 established extremely suboptimal criteria ($M_{(c_n)} = 0.67$) resulting in a *loss* of $1.10. This individual can appropriately adopt a conservative criterion when the situation calls for it but maintains that same conservative criterion when a liberal criterion is advantageous. Theories attempting to explain suboptimal criterion shifting behavior must account for this phenomenon. The relationship between the extent of criterion shifting and total payment in Experiment 3 is not limited to these select subjects but extends across all participants to a modest degree ($r(170)$ range: .37-.44, *Mdn* = .42). The relative consequences for inadequate criterion shifting may generalize to real world scenarios.

There are many situations where extreme criterion shifts are necessary for avoiding consequential errors. One real world example comes from radiologists who must assess whether a mammogram shows signs of breast cancer. If a radiologist falsely identifies an abnormal mammogram, then the patient must endure unnecessary worry while undergoing additional costly examinations. However, if a radiologist misses an abnormal growth, then a breast cancer diagnosis and subsequent treatment will be delayed increasing the likelihood of major surgery (e.g. a mastectomy) or even death. Studies reveal vast individual differences in the rate that radiologists recall patients for further testing and a more conservative criterion is associated with an increased miss rate (as expected) (Yankaskas, Cleveland, Schell, & Kozar 2001; Gur et al., 2004). Yankaskas and colleagues (2001) examined patient recall rates across 31 practices and found a large range from 1.9% to 13.4%. This means that some radiologists establish very conservative criteria for identifying an abnormal mammogram while others set much more liberal criteria. Although Yankaskas and colleagues (2001) could not assess the extent of criterion shifting since there is not a second criterion condition to compare against, it is presumed that these radiologists needed to shift their decision criteria when identifying an abnormal mammogram where errors result in extreme consequences relative to everyday decisions where the consequences of an error are negligible. The fact that the patient recall rate is so variable suggests that at the very least there are vast individual differences in the end result of criterion shifts, even when examining a group of experts.

Examining individual criterion shifting tendencies can prove challenging in situations where there are insufficient observations from each individual. For instance, an eyewitness to a crime who needs to select potential suspects from a lineup should consider the potential costs of falsely identifying an innocent person versus missing the perpetrator. If an identification will simply lead to further questioning of the suspect, then the eyewitness should establish a liberal criterion, since questioning an innocent person is only a minor inconvenience. However, if an eyewitness' statement could substantially impact whether a suspect is arrested, then a more conservative criterion should be maintained to avoid incarcerating a potentially innocent person. Assessing whether eyewitnesses establish appropriate decision criteria is challenging because typically there is only one observation for each person. This means that analyses of individual differences cannot be conducted, and conclusions are limited to group-averaged results. For example, Mickes and colleagues (2017) conducted photo lineup recognition tests and found that people *on average* establish less extreme criteria when making a binary response with a criterion manipulation versus the criterion set for the highest level of confidence in judgments made on a multipoint scale. However, it is likely that *some* people will appropriately establish extreme criteria when instructed to do so, but individual differences are impossible to evaluate in this paradigm because each participant only contributes a single observation.

Strategic criterion shifting

Experiments 1–5 specifically measured the stability of *strategic* criterion shifting tendencies where individuals explicitly received information indicating an advantage for shifting criteria. Information about the testing conditions appeared on every trial and participants could respond with unlimited time. Strategic criterion shifting occurs when people proactively set a goal to either avoid false alarms or misses depending on the situation. Criterion shifting stability may differ in situations where participants are not explicitly informed of the testing conditions (Verde & Rotello, 2007), are provided with false information (Selmeczy & Dobbins, 2013), are affected by sequential dependencies from prior responses (Malmberg & Annis, 2012), or are in situations where a time pressure is imposed (Ratcliff et al., 2016). Future research needs to investigate the relationship, if any, between individual differences in strategic criterion shifting and criterion shifting tendencies in situations where participants are not explicitly informed of the advantages for doing so or when speeded responses are required. In the case of strategic criterion shifting, many individuals consistently fail to adequately shift criteria despite explicitly knowing the advantages for doing so.

Why do some people give a criterion shift, but not others?

There are many similarities between criterion shifts and confidence ratings that inform our understanding of why there are vast individual differences in strategic criterion shifting tendencies. Criterion manipulations served as the original method for obtaining cumulative hit and false alarm rates to create receiver operator characteristic (ROC) plots (Swets et al., 1955; Tanner et al., 1956), which illustrate the hit and false alarm rate for all possible decision criteria at a specific level of discriminability. Later, the implementation of confidence judgments provided an analog to strategic criterion shifts for recognition memory tests (Egan, 1958). Minimal instructions are required for people to accurately rate confidence during memory tests suggesting that people regularly assess the level of confidence associated with memories over the course of a lifetime (Mickes et al., 2011). Although participants typically have high metacognitive sensitivity when scaling the strength of memories to varying levels of confidence, there are limits in the degree to which this is achieved, particularly as discriminability increases (Stretch & Wixted, 1998b). The highest levels of confidence should be reserved for the strongest and weakest memories, which should result in virtually no false alarms or misses, respectively. However, when participants report confidence ratings on a 20-point scale, many individuals will respond “old” with the highest confidence ratings more often than any other confidence level (Criss, 2009; Mickes et al., 2011). Mickes and colleagues (2011) believe this occurs because people struggle to finely scale confidence with strong memories. Thus, it appears that the criterion for “old” responses with the highest confidence level is less extreme than it theoretically should be. Interestingly, there are individual differences in the degree to which people finely scale strong memories suggesting that some individuals are more adept than others at establishing more extreme criteria for the highest levels of confidence (i.e. individual differences in metacognitive bias; see Figure 8 from Mickes and colleagues, 2011).

When examining group averages, ROC curves produce similar curvilinear shapes regardless of whether confidence ratings are acquired or a criterion manipulation is implemented (Macmillan & Creelman, 2005; Koen & Yonelinas, 2011; Dube & Rottello, 2012). This suggests that individual tendencies to finely scale confidence might be related to individual differences in the extent of criterion shifting. Afterall, when extreme criteria are advantageous people should only respond with the riskier option when there is high confidence in the decisional evidence. In Experiment 4, a strong relationship persisted between the extent of criterion shifting and metacognitive bias when the criterion manipulation included instructions to respond with high confidence only. However, in Experiment 5 no such relationship existed when the criterion manipulation did not make any reference to confidence levels. When not explicitly cued to respond with confidence, participants seem to adopt a decision strategy for shifting criteria that is unrelated to decision processes for assessing high confidence in a recognition judgment. Some individuals will have very low metacognitive bias, but not shift criteria much while others will have high metacognitive bias yet shift criteria to large extents. Assessing confidence in a response and adapting a decision criterion to explicit instructions are two separate behaviors that appear largely independent of each other. While confidence judgments may represent a meta-assessment of the *varying strength* of memory evidence, criterion shifting appears to represent a mode of response that can strategically vary for any *single strength* of evidence. That is, each test item will elicit a degree of memory strength that a participant can convey through a confidence judgement, but the choice to identify an item as “old” or “new” depends on the situation.

Another key finding from Experiments 4 and 5 is that almost all participants adopted much more extreme criteria when responding with the highest levels of confidence compared to strategic criterion shifting. This suggests that these individuals are *capable* of shifting criteria to greater extents if they simply implement the same stringent criterion thresholds for rating recognition judgments with “high confidence” as they do for strategically shifting criteria when making an old/new judgment. Instead, it appears that individuals are simply *unwilling* to disregard uncertain evidence in favor of a decision strategy based on known circumstances surrounding a decision (Aminoff et al., 2012; Kantner, et al., 2015; Miller and Kantner 2019). As Kantner and colleagues (2015) state, “people would rather attempt to be *correct* than be *correctly biased*.” When reporting on a six-point confidence scale, participants can convey the level of uncertainty in familiarity strength while still making old/new judgments as they typically would on a recognition test. However, when forced to only respond “old” or “new” when there is high confidence, participants must identify relatively familiar items as “new” and relatively unfamiliar items as “old” in order to maintain the stringent criteria established for high confidence responses on a six-point scale. That is, memory evidence that elicits “low” or “medium” confidence for either an “old” or “new” judgment must be completely disregarded which may feel unnatural or seem incorrect for some individuals even when this is the best decision strategy. Therefore, individual differences in criterion shifting might be a result of stable differences in peoples’ *willingness* to disregard uncertain evidence in favor of a decision strategy based on knowledge of payoffs, probabilities, or instructions (Aminoff et al., 2012; Kantner, et al., 2015; Miller and Kantner 2019).

Ultimately, it is important to understand why one individual is willing to shift a criterion, but not another. It is possible that individual criterion shifting tendencies are a response strategy learned across a lifetime of experience, similar to how Mickes and colleagues (2011) suggest that accurately scaling confidence judgements to recognition responses are learned over the course of one’s life. One bit of data from Aminoff and colleagues (2012) provides some intriguing insight into this hypothesis, albeit inconclusive and speculative; in their study, the subject pool consisted of 68 combat-experienced commissioned and non-commissioned officers from the U.S. Army Fort Irwin National Training Center along with 27 age-matched, non-military participants from the community. Participants were categorized into nine different hierarchical levels of military rank, with non-military participants ranked at the bottom. Interestingly, military rank turned out to be one of the few factors that significantly correlated with the extent of criterion shifting across both recognition memory tasks. Higher military rank associated with greater criterion shifts and this relationship could not be explained by other factors such as age or education level. It is possible that individuals who have learned to be more adaptive with their response strategies are better suited for the decision-making demands of high-ranking military officials. Conversely, the experience of making decisions in those high-ranking positions may have led to more adaptive response strategies. This single data point cannot provide definitive answers, but it should encourage future studies to more systematically assess why some individuals are more willing to shift criteria than others.

Conclusion

Individual tendencies in strategic criterion shifting appear to represent a stable cognitive trait. These tendencies seem to result from individual differences in people's willingness to disregard uncertain evidence in favor of a response that avoids a critical error (Aminoff et al., 2012; Kantner, et al., 2015; Miller and Kantner 2019). For example, when conducting a difficult recognition memory test that requires extreme criterion shifts, it is perfectly rational to simply look away from the screen and just choose the response that promotes better decisional outcomes. However, many people are reluctant to make such extreme shifts and may feel compelled to make responses based on memory, even a very poor one. These demand characteristics may not simply be an artifact of laboratory studies, but likely occur in real life situations where people may feel obliged to make decisions based on uncertain memory evidence. Evidently some individuals are completely comfortable with disregarding weak evidence and will shift criteria to extreme extents to optimize decisional outcomes. Other individuals appear to have a standard criterion and would rather rely on memory evidence to make decisions while completely disregarding other situational information. Most individuals fall somewhere in between where individuals are both uncomfortable with completely abandoning memory evidence and ignoring situational information resulting in less extreme criterion shifts.

Data availability

Datasets for Experiments 1–5 can be accessed through the Open Science Framework: https://osf.io/4k2hb/

Chapter II: Neural correlates underlying the decision criterion

The fact that criterion shifting is a uniquely individualistic cognitive trait brings important implications for studies investigating the neural mechanisms of recognition memory. Many neuroimaging studies fail to account for the decision criterion when drawing conclusions about recognition memory neural correlates. Recognition memory fMRI contrasts of target versus nontarget (T > NT) responses reveal similar patterns of frontoparietal activity regardless if responses are correct or incorrect, leading some researchers to conclude that *the subjective experience of remembering* itself drives the effect (Wagner et al., 2005; Dennis et al., 2015; McDermott et al., 2017). If differences in subjective memory strength are associated with frontoparietal activity, then the recruitment of this network should be invariant to “target” responses requiring strong memory evidence (i.e. establishing a conservative decision criterion) versus “target” responses requiring relatively weak memory evidence (i.e. maintaining a liberal decision criterion). However, if the decision criterion predominantly drives this frontoparietal network, then responding under a conservative versus liberal criterion should drastically affect T > NT response contrasts. Aminoff and colleagues (2015) proposed that T > NT response contrasts are affected by response bias, where inhibiting versus providing prepotent responses increases frontoparietal activity. When a conservative criterion is maintained the prepotent response is “nontarget” whereas a “target” response is preponderant under a liberal criterion. Thus, a response bias account predicts greater frontoparietal activity for T > NT responses under a conservative criterion, and the reverse should be true when a liberal criterion is utilized (i.e. greater activity for NT > T responses). Individuals who appropriately shift decision criteria during recognition memory tests show widespread frontoparietal activity in the correct-only target (hit) versus nontarget (correct rejection) response contrast (H > CR) when maintaining a conservative criterion—but not a liberal criterion (Aminoff et al., 2015; King & Miller, 2017). This demonstrates that the decision criterion modulates the H > CR contrast, but a response bias account alone is insufficient for explaining these findings: maintaining a liberal criterion did not reveal significant differences in the reverse contrast (CR > H). One limitation of the work by Aminoff and colleagues (2015) and Miller and King (2017) is that these studies did not manipulate discriminability—differences in memory strength in the H > CR contrast are not necessarily equivalent when participants respond under a conservative versus liberal criterion. Therefore, it is necessary to implement both criterion *and* discriminability manipulations to circumvent this potential confound.

A third manipulation that may further distinguish frontoparietal activity associated with mnemonic evidence versus decision criteria is to assess fMRI activity across decision domains. Individual criterion shifting strategies are consistent across decision domains whereas discriminability performance is virtually unrelated (Frithsen et al., 2018; Layher et al., 2020). This suggests that neural mechanisms underlying decision criteria may be conserved across decision domains regardless of the type of evidence. Experiment 6 implements manipulations of both discriminability and decision criteria during recognition memory *and* visual detection tests during fMRI scanning to differentiate frontoparietal activity associated with evidence strength versus decision criteria across memory and perceptual domains.

Although Experiment 6 attempts to further findings from Aminoff and colleagues (2015) and Miller and King (2017) by including a discriminability and task manipulation, it is important to consider the neural correlates associated with many levels of criteria and discriminability. SDT assumes familiarity strength is a continuum, which means there is theoretically an infinite number of ways to place a criterion across an infinite amount of discriminability levels (Macmillan and Creelman, 2005). Therefore, it is important to investigate fMRI activity associated with *many* levels of criteria and discriminability to understand how the neural mechanisms underlying both processes interact during recognition memory tests. The downfall of this approach is that it creates many conditions and would require thousands of trials to attain sufficient statistical power given the low signal-to-noise ratio of the BOLD signal. Experiments 7 and 8 circumvent this problem by implementing a dense-sampling approach where a single participant conducts recognition memory tasks during fMRI across many sessions. While a dense-sampling approach within a single individual cannot provide appropriate inferences to the population, it can bring insights into nuances of fMRI activity that are otherwise missed at the group-level (Poldrack, 2017). The individual assessments of Experiments 7 and 8 complement the group-level analyses of Experiment 6 to provide a holistic picture of fMRI correlates associated with the decision criterion and discriminability during recognition memory.

Method

Prescreening

A major issue with investigating neural mechanisms underlying a conservative versus liberal decision criterion is that many individuals fail to adequately shift their decision criterion even when they are explicitly told to do so (Aminoff et al., 2012, 2015; Kantner et al., 2015; Frithsen, et al., 2018; Layher et al., 2018; Miller & Kantner 2019; Layher et al., 2020). If a participant fails to shift, then it is impossible to investigate *within*-subject neural correlates associated with multiple criterion placements. Therefore, Experiments 6-8 included an initial prescreen procedure to exclude participants who fail to adequately shift decision criteria, with the hope that individuals who shift during the prescreen will also shift during the fMRI experiments.

MRI data acquisition and fMRI preprocessing

A 64-channel head and neck coil within a Siemens 3T PRISMA MRI scanner at UCSB acquired all imaging data. Functional image acquisition occurred via a T_2*-weighted multiband echo planar imaging sequence (72 oblique slices; TR = 720 ms; voxel size = 2 mm^3; FoV = 208 mm^2; TE = 37 ms; flip angle = 52^o; multiband factor = 8). To correct for magnetic field inhomogeneities, a T_2*-weighted gradient recall echo (GRE) field map scan was collected with the same slice count and dimensions as the functional scans (TE$_1$ = 4.92 ms; TE$_2$ = 7.38 ms). Structural images aided in functional image registration and included a T_1-weighted magnetization-prepared rapid gradient echo (MPRAGE) sequence (208 sagittal slices; TR = 2,500 ms; voxel size = 0.94 mm^3; FoV = 241 mm^3; TE = 2.22 ms; flip angle =

7^{o}) and a T_2-weighted Sampling Perfection with Application optimized Contrasts using different flip angle Evolution (SPACE) sequence (208 sagittal slices; TR = 3,200 ms; voxel size = 0.94 mm3; FoV = 241 mm^3; TE = 566 ms; flip angle = 120^{o}).

The FMRI Brain Software Library (FSL), v6.0.4 (Jenkinson et al., 2012) performed initial fMRI preprocessing in which functional scans underwent motion correction, field map unwarping, temporal high pass filtering (0.01 Hz), prewhitening, and spatial smoothing using a 5 mm^3 full-width at half-maximum (FWHM) isotropic Gaussian kernel. Registration of functional scans to subject-specific anatomical images occurred via the Advanced Normalization Tools (ANTs) software.

Experiment 6: Decision criteria greatly affects fMRI activity, not discriminability

Method

Participants

Thirty healthy adult participants (11 males; ages 18-32, *M* = 21, *SD* = 3.0; 3 left-handed) from UCSB completed the fMRI experiment and earned $20/hour plus monetary bonuses based on task performance. Selection of the fMRI participants derived from a sample of one hundred and forty-four subjects (60 males; ages 18-35, *M* = 21, *SD* = 2.8) who completed an initial prescreen computer task and earned $10/hour in addition to monetary bonuses.

Procedure

Participants completed an initial prescreen computer task that consisted of shortened and modified versions of the recognition memory and visual detection tests used in the fMRI experiment. To be eligible for the fMRI experiment, participants needed to have no MRI contraindications, adequately shift decision criteria (*C* > 0.7 in either the recognition memory or visual detection test, which is approximately the cutoff that Aminoff and colleagues (2015) implemented to designate the “High Shifters” group for fMRI analyses) and perform above chance on both tasks (*d’* > 0). Eligible participants received an invitation to partake in the fMRI experiment on a first come first serve basis until a total of 30 eligible individuals agreed to participate.

The fMRI experiment included recognition memory and visual detection tests with discriminability and criterion manipulations in a fully-crossed 2 (task domain: recognition memory vs. visual detection) x 2 (discriminability condition: low vs. moderate) x 2 (criterion condition: conservative vs. liberal) factorial design creating 8 test conditions (**Figure 12**). The research paradigm consisted of a modified version of the recognition memory and visual detection task implemented in Experiment 3.

For the recognition memory task, participants completed two cycles of a study block followed by four test blocks during fMRI scanning. Each study block consisted of 256 unique scene images—half of which appeared once (for low discriminability at test) whereas the other half appeared six times (for moderate discriminability), yielding 896 total presentations. Participants passively viewed each study item sequentially and continuously for 720 ms (1 TR) in a randomized order for subsequent recognition tests. Half of the images contained a person whereas the other half did not (split evenly between images presented once vs. six times).

Each test block encompassed eight mini-blocks (one per test condition) of 16 trials (8 target and 8 nontarget images). Every test trial began with a white crosshair displayed on a black background for 320 ms, followed by the presentation of a scene image for 200 ms, then a noise mask appeared for 200 ms to destroy the perceptual afterimage. Afterwards, participants viewed a screen displaying the two possible response types and needed to respond within 2,160 ms (3 TRs). Participants held MRI-compatible two-button response boxes in each hand and made responses with their left or right pointer finger. During recognition memory tests, participants decided whether an image appeared in the study phase (“old,” target) or not (“new,” nontarget); visual detection tests required participants to determine whether an image contained a person (“present,” target) or not (“absent,” nontarget). The response type corresponding to a left or right button press randomly changed on a trial-by-trial basis to prevent participants from knowing which button to press until *after* stimulus presentation. During low discriminability recognition test mini-blocks, “old” images only appeared once during the study phase whereas “old” images in the moderate discriminability condition appeared six times. To manipulate discriminability during visual detection tests, 15 researchers prior to the experiment independently rated the difficulty of finding a person in each scene image. Classification of scenes into the low or moderate discriminability condition occurred by taking a median-split of the mean difficulty ratings. A payment manipulation induced criterion shifts where participants earned five cents for each correct response, lost 10 cents for a critical error, but received no penalty for a noncritical error. In the conservative criterion condition, a critical error consisted of incorrectly responding “old” or “present” (false alarms) during recognition memory or visual detection tests, respectively, whereas incorrect “new” and “absent” responses (misses) served as critical errors in the liberal criterion condition. The assignment of images to each task type, criterion condition, and discriminability condition as well as the image version (person present or absent) occurred randomly across participants with the exception that images assigned to the low versus moderate discriminability conditions of the visual detection tests remained fixed.

Prior to each test mini-block, an instruction screen appeared for 7,200 ms (10 TRs) informing participants of the task type and criterion manipulation for the upcoming trials. The top of the instruction screen displayed “MEMORY TEST” or “TARGET DETECTION TEST” to indicate the task type while the payout structure for hits, correct rejections, and the critical error appeared on separate lines in the middle of the screen. At the bottom of the instruction screen, participants received explicit instructions to avoid making critical errors (e.g. in the conservative criterion condition of visual detection tests: “You will be penalized for saying a person is present when a person is actually absent. Avoid making false alarms by choosing absent.”). During each test trial, the top of the screen displayed the message “avoid false alarms” or “avoid misses” when presented with the two response options in the conservative and liberal conditions, respectively, to remind participants of the critical error. Participants did not receive explicit instructions as to whether a mini-block corresponded to the low or moderate discriminability conditions. Following each mini-block, a feedback screen appeared for 3,600 ms (5 TRs) displaying the number of correct responses, non-critical errors, and critical errors made during that mini-block as well as money earned for that mini-block and the running total for the entire experiment. Each functional test block scan included a white crosshair on a black screen for the first 7,200 ms (10 TRs) and the final 14,400 ms (20 TRs). A variable number of jitter trials (randomly determined) displayed a crosshair for 720 ms (1 TR) and appeared randomly after various instruction, test trial, and feedback displays throughout each test block, with a maximum of two consecutive jitter trials (i.e. a crosshair displayed for up to 1,440 ms or 2 TRs). The number of jitter trials displayed during each test block across all participants ranged from 82 to 135. Each study block lasted for about 11 minutes whereas each test block took between 9 and 10 minutes, depending on the number of jitter trials. The entire fMRI task lasted for approximately 100 minutes.

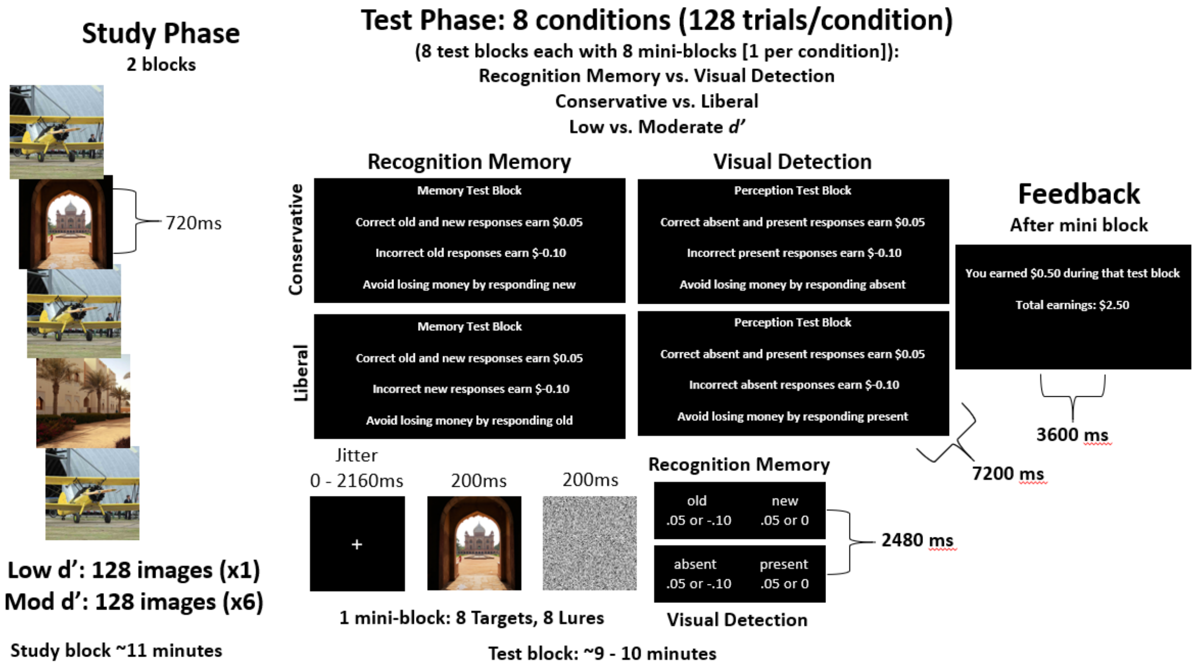


Figure 12: Experiment 6 recognition memory and visual detection (perception) tasks that occurred during fMRI scanning. Participants conducted two cycles of an 11-minute study block, followed by four, 9-to-10-minute test blocks. To control for demand characteristics, the recognition memory and visual detection tests maintained the exact same structure except participants either responded as to whether an image appeared during the study phase or if a person appeared in the image, respectively. Participants received feedback on the amount of money earned after each mini-block.

fMRI analysis

Event-related general linear models (GLM) implemented in FSL identified within-subject activity related to T > NT response and item contrasts across the eight test conditions. Each test block contained 16 regressors of interest (target and nontarget responses, or items, in each condition of the 2x2x2 design) as well as regressors for instructions, feedback, and trials with no responses. The default settings of FMRIB's Linear Optimal Basis Sets

(FLOBS) toolkit were implemented to model hemodynamic response function (HRF) convolution for each regressor. The time window for HRF convolution on each test trial started at image onset and ended when the participant made a response. Additional nuisance regressors included six head motion parameters derived from motion correction realignment.

Whole-brain group contrasts of statistical *Z*-maps with voxel-wise thresholding at $Z = 3.1$ and cluster correction using Gaussian Random Field Theory ($p < .05$), implemented in the FMRI Expert Analysis Tool (FEAT), determined statistically significant activity related to T > NT response and item contrasts within each of the eight test conditions, as well as across conditions. Additional region of interest (ROI) analyses were conducted for T > NT response and item contrasts. ROI centroids were identified according to 12 peak cortical voxels reported by Aminoff and colleagues (2015): specifically, the H > CR contrast in the conservative condition of the recognition memory tests for words. These included regions in the insula, inferior frontal gyrus (IFG), middle frontal gyrus (MFG), medial frontal gyrus (MeFG), inferior parietal lobule (IPL), superior parietal lobule (SPL), precuneus (Pc), and posterior cingulate (PoC). Using the MNI152 standard template, mean beta values were computed from spheres with 5 mm radii around each peak voxel (81 voxels/ROI). Analyses in the main text primarily focus on T > NT contrasts *between* criterion, discriminability, and task conditions.

For ROI analyses, an additive linear mixed model assessed the extent to which mean beta values across the 12 ROIs, separately for responses and items, are affected by task type (recognition memory vs. visual detection [RM > VD]), criterion (conservative vs. liberal [CON > LIB]), discriminability (moderate vs. low [MOD > LOW]), and targetness (target vs. nontarget [T > NT]). Modeling of a four-way interaction occurred between task, criterion, discriminability, and targetness contrasts, along with all marginal three-way and two-way interactions. Specification of crossed random effects on the model intercept accounted for baseline variation in mean beta values across subjects and ROIs. The fixed effects models took the following form:

$$\hat{y} = b_0 + b_1(RM > VD) + b_2(CON > LIB) + b_3(MOD > LOW) + b_4(T > NT) + b_5(RM > VD * CON > LIB) + b_6(RM > VD * MOD > LOW) + b_7(CON > LIB * MOD > LOW) + b_8(RM > VD * T > NT) + b_9(CON > LIB * T > NT) + b_{10}(MOD > LOW * T > NT) + b_{11}(RM > VD * CON > LIB * MOD > LOW) + b_{12}(RM > VD * CON > LIB * \ T > NT) + b_{13}(RM > VD * MOD > LOW * \ T > NT) + b_{14}(CON > LIB * MOD > LOW * \ T > NT) + b_{15}(RM > VD * CON > LIB * MOD > LOW * \ T > NT) + \ \varepsilon.$$

Results

Behavioral findings

Discriminability manipulations in both tasks proved successful as mean *d'* in the recognition memory tests remained significantly higher for the moderate (*M* = 1.04, *SD* = 0.29) versus low (*M* = 0.33, *SD* = 0.20) discriminability conditions (*M* Δ = 0.70, 95% CI [0.61, 0.80], *d* = 2.17), as well as in the visual detection tests between the moderate (*M* = 1.56, *SD* = 0.27) and low (*M* = 0.33, *SD* = 0.23) discriminability conditions (*M* Δ = 1.23, 95% CI [1.14, 1.33], *d* = 4.75). Mean *d'* did not significantly differ between the recognition memory and visual detection tests in the low discriminability condition (*M* Δ = 0.00, 95% CI [-0.09, 0.10], *d* = 0.02) but participants on average obtained a substantially higher *d'* in the moderate discriminability condition for the visual detection versus recognition memory tests (*M* Δ = 0.52, 95% CI [0.43, 0.62], *d* = 1.48), despite efforts to make levels of discriminability similar across decision domains.

Mean *C* indicated that participants shifted decision criteria to large extents between criterion conditions in the recognition memory low (*M* = 1.50, *SD* = 0.27) and moderate (*M* = 1.37, *SD* = 0.21) discriminability conditions, as well as in the visual detection low (*M* = 1.37, *SD* = 0.23) and moderate (*M* = 1.00, *SD* = 0.27) discriminability conditions. Participants shifted criteria to a somewhat larger extent between the recognition memory and visual detection tests for both the low (*M* Δ = 0.12, 95% CI [0.04, 0.21], *d* = 0.18) and moderate discriminability conditions (*M* Δ = 0.37, 95% CI [0.28, 0.46], *d* = 0.63). Additionally, mean *c* across all conditions remained higher for visual detection (*M* = 0.32, *SD* = 1.00) versus recognition memory (*M* = -0.06, *SD* = 1.16) tests, especially in the liberal criterion conditions (*M* Δ = 0.51, 95% CI [0.43, 0.58], *d* = 1.21) relative to the conservative conditions (*M* Δ = 0.26, 95% CI [0.18, 0.34], *d* = 0.64). Thus, participants maintained a relatively more conservative criterion throughout all conditions of the visual detection versus recognition memory tests. Mean *c* and *d'* values across all conditions are listed in **Table 5**.

Very strong relationships existed between the extent of criterion shifting in the recognition memory and visual detection tasks in both the low (*r*(28) = .84, 95% CI [.69, .92]) and moderate (*r*(28) = .85, 95% CI [.70, .92]) discriminability conditions (**Figure 13**, top). In contrast, relatively weak correlations were observed in *d'* across the two task domains in the low (*r*(28) = .27, 95% CI [-.10, .58]) and moderate (*r*(28) = .36, 95% CI [.00, .64]) discriminability conditions (**Figure 13**, bottom). This indicates that behavioral similarities between the recognition memory and visual detection tests are largely specific to the decision criterion and *not* discriminability.

Experiment 6: SDT measures					
Condition		**Recognition Memory**		**Visual Detection**	
Discriminability	**Criterion**	***c***	***d'***	***c***	***d'***
Low	**Conservative**	**0.84 (0.41)**	**0.36 (0.29)**	**1.20 (0.36)**	**0.37 (0.30)**
	Liberal	**-0.66 (0.37)**	**0.31 (0.30)**	**-0.17 (0.38)**	**0.30 (0.32)**
Moderate	**Conservative**	**0.47 (0.41)**	**1.21 (0.36)**	**0.63 (0.27)**	**1.76 (0.31)**
	Liberal	**-0.90 (0.36)**	**0.86 (0.37)**	**-0.37 (0.32)**	**1.36 (0.32)**

Table 5: Experiment 6 mean and standard deviation values (in parentheses) for *c* and *d'* across criterion, discriminability, and task conditions.

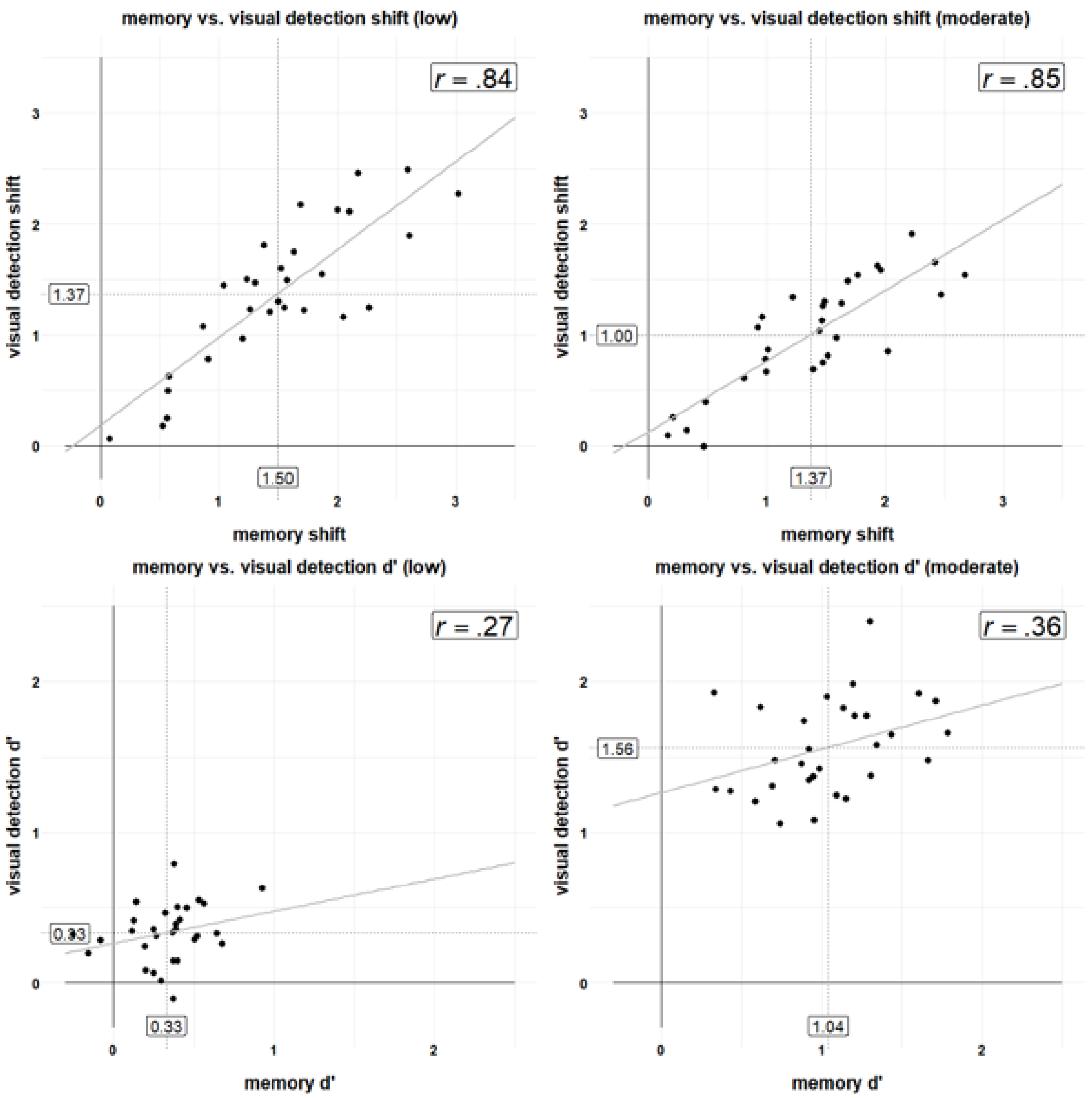


Figure 13: Experiment 6 Pearson correlations between performance in the recognition memory versus visual detection tests. Relationships between criterion shifting (top) across decision domains remained much stronger than discriminability (bottom) for both the low (left) and moderate (right) discriminability conditions

Target > Nontarget response contrast

Whole-brain GLM analyses in the recognition memory task revealed widespread frontoparietal activity in the T > NT response contrast when participants maintained a conservative criterion, but not when maintaining a liberal criterion, in both the low and moderate discriminability conditions (**Figure 14**). In fact, under a liberal criterion, the reverse contrast (NT > T responses) in both discriminability conditions revealed significant activity in frontal regions including the right anterior insula, IFG, and MeFG, suggesting that recruitment of these areas is particularly well-described by a response bias account. The T > NT response contrast subtracted between the conservative versus liberal (CON > LIB) criterion conditions revealed widespread frontoparietal activity including bilateral regions in the insula, anterior cingulate cortex (ACC), IFG, MFG, MeFG, SPL, and Pc in both discriminability conditions. When comparing across the moderate versus low (MOD > LOW) discriminability conditions in the recognition memory task, there were no significant differences in the T > NT response contrast regardless of whether participants maintained a conservative or liberal criterion. These results strikingly reveal that changes in the placement of decision criteria during recognition memory tests drastically affect the T > NT response contrast, whereas changes in discriminability do not.

In the visual detection task, whole-brain analyses of T > NT response contrasts also revealed greater frontoparietal activity when participants maintained a conservative, but not a liberal criterion—though to a much lesser spatial extent relative to the recognition memory tests (**Figure 15**). The MOD > LOW contrast did not reveal any significant differences in activity for either criterion condition, except for sparse differences within the visual cortex specifically in the conservative condition. When comparing the T > NT response contrasts in the recognition memory versus visual detection (RM > VD) tasks, only sparse differences in activity were observed, but no consistent patterns existed across criterion or discriminability conditions (e.g. greater activity in the right Pc in the low discriminability, conservative condition and less activity in the right IFG of the moderate discriminability, liberal condition). These results suggest that the hallmark frontoparietal activity in the T > NT response contrast may represent domain-general neural mechanisms associated with criterion placement, at least to a certain extent. A list of local maxima in the (CON > LIB) * (T > NT) response contrast in the low and moderate discriminability conditions for the recognition memory and visual detection tests are shown in **Table 6**.

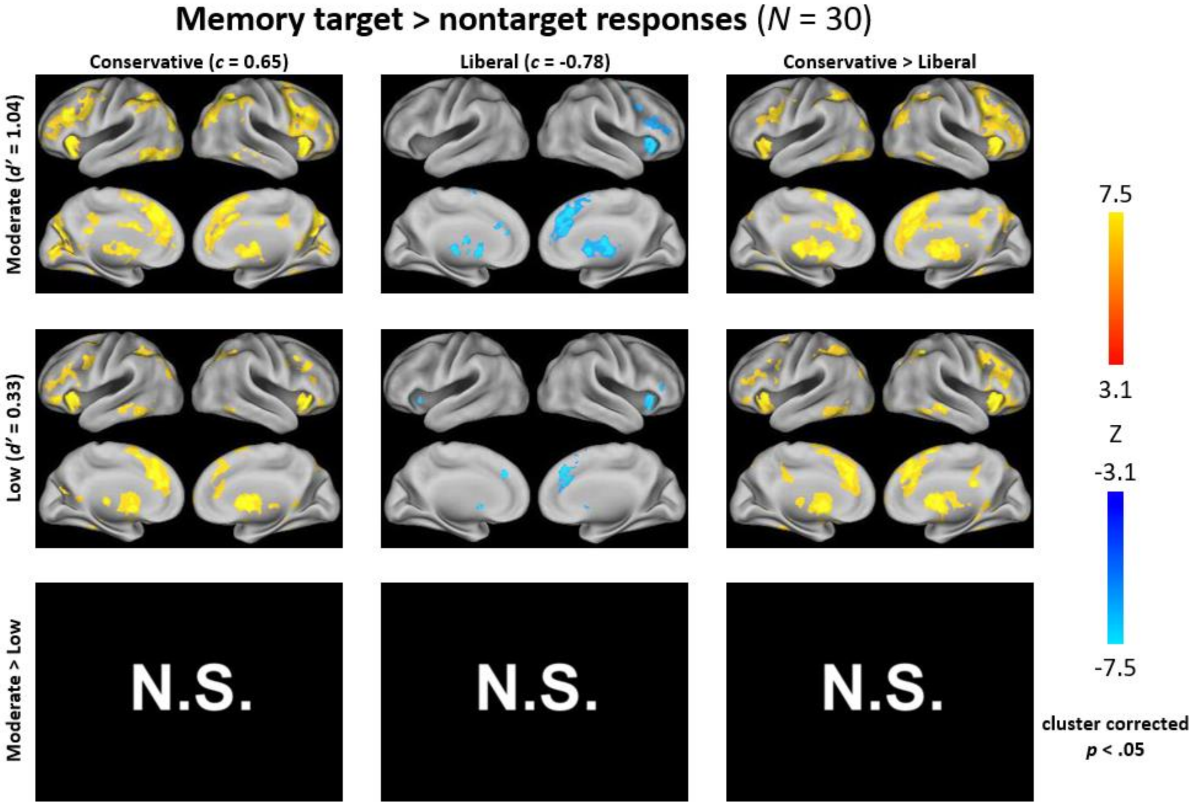


Figure 14: Experiment 6 whole-brain statistical *Z*-maps of T > NT response contrasts in the recognition memory task across criterion and discriminability conditions. Statistically significant activity with thresholding at $Z > 3.1$ and cluster corrected at $p < .05$, are displayed in orange (T > NT) and blue (NT > T). Images containing "N.S." represent conditions in which no significant activity occurred at the whole-brain level.

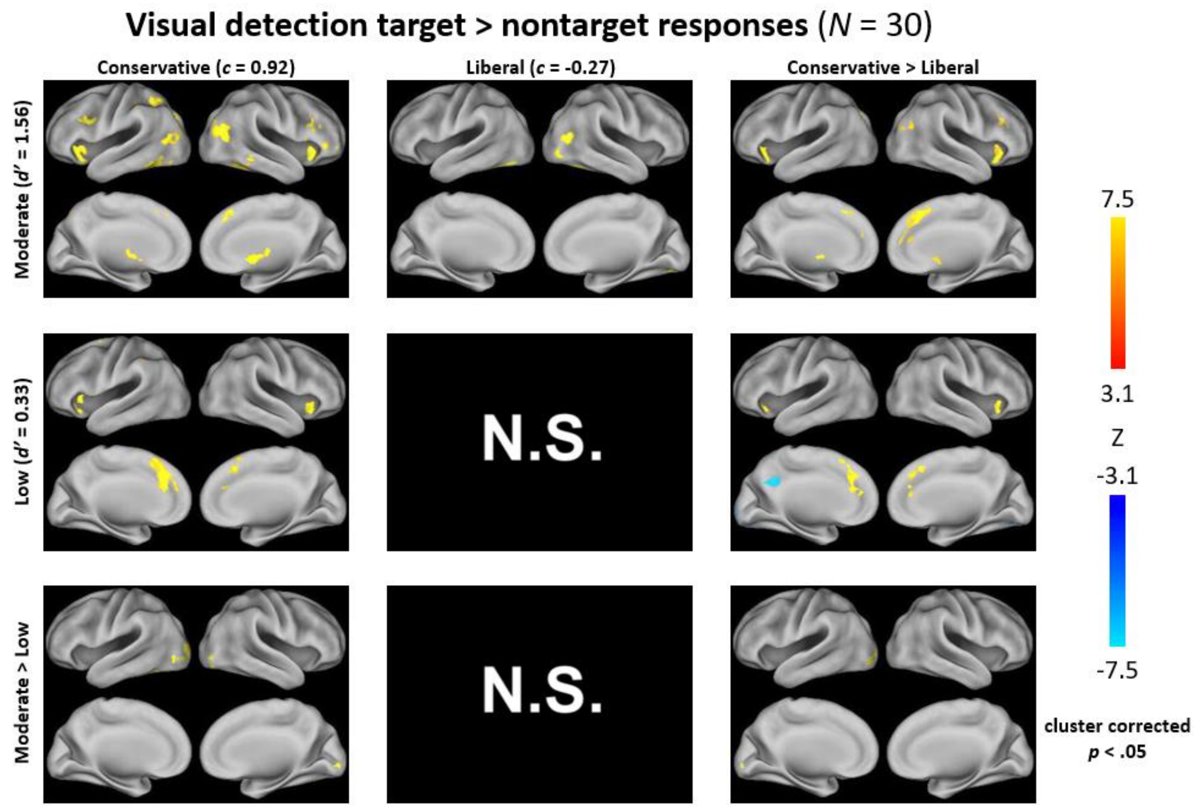


Figure 15: Experiment 6 whole-brain statistical Z-maps of T > NT response contrasts in the visual detection task across criterion and discriminability conditions. Statistically significant activity with thresholding at $Z > 3.1$ and cluster corrected at $p < .05$, are displayed in orange (T > NT) and blue (NT > T). Images containing "N.S." represent conditions in which no significant activity occurred at the whole-brain level.

Experiment 6: fMRI local maxima (CON > LIB) * (T > NT) responses

Recognition Memory Moderate Discriminability

Cluster	*Z*-value	X	Y	Z	Location	BA
1	6.74	0	20	44	Left Medial Superior Frontal Gyrus	8
1	6.55	4	36	24	Right Anterior Cingulate Gyrus	9
1	6.45	4	38	20	Right Anterior Cingulate Gyrus	9
1	6.31	34	22	-6	Right Insula	13
1	6.31	-10	2	2	Left Thalamus	48
1	6.26	-14	6	14	Left Caudate	48
Recognition Memory Low Discriminability						
1	5.26	6	-34	26	Right Posterior Cingulate Gyrus	23
1	4.12	-4	-30	28	Left Posterior Cingulate Gyrus	23
1	4.06	10	-38	44	Right Median Cingulate Gyrus	31
1	3.95	8	-32	38	Right Median Cingulate Gyrus	31
1	3.93	0	-34	40	Left Median Cingulate Gyrus	31

1	3.79	-12	-40	44	Left Median Cingulate Gyrus	31
2	5.09	34	-50	44	Right Inferior Parietal Lobule	39
2	4.80	10	-76	50	Right Precuneus	7
2	4.59	30	-86	40	Right Superior Occipital Gyrus	19
2	4.43	30	-74	44	Right Superior Occipital Gyrus	7
2	4.37	52	-54	54	Right Inferior Parietal Lobule	39
2	4.31	20	-70	64	Right Superior Parietal Lobule	7
3	6.58	-36	18	-4	Left Insula	13
3	6.49	40	24	-8	Right Insula	47
3	6.41	56	14	34	Right Precentral Gyrus	44
3	6.20	-6	6	-2	Left Caudate	48
3	6.18	-10	0	4	Left Thalamus	50
3	6.17	10	16	2	Right Caudate	48
Visual Detection Moderate Discriminability						
1	4.56	-28	24	-8	Left Insula	47
1	4.51	-30	28	-2	Left Insula	13
1	4.25	-26	22	-14	Left Insula	47
2	4.14	44	-64	26	Right Middle Occipital Gyrus	39
2	4.12	38	-74	30	Right Middle Occipital Gyrus	39
2	4.00	38	-80	38	Right Middle Occipital Gyrus	39
2	3.86	46	-76	36	Right Angular Gyrus	39
2	3.58	46	-72	24	Right Middle Temporal Gyrus	19
2	3.57	46	-76	24	Right Middle Occipital Gyrus	19
3	4.35	48	24	26	Right Inferior Frontal Gyrus	9
3	3.96	54	34	16	Right Inferior Frontal Gyrus	46
3	3.68	48	40	16	Right Inferior Frontal Gyrus	46
3	3.65	52	26	18	Right Inferior Frontal Gyrus	9
3	3.54	42	26	24	Right Inferior Frontal Gyrus	9
3	3.34	44	28	16	Right Inferior Frontal Gyrus	9
4	5.17	28	18	-14	Right Insula	13
4	4.93	30	20	-10	Right Insula	13
4	4.63	34	18	2	Right Insula	13
4	4.53	42	20	-8	Right Insula	13
4	3.54	44	24	2	Right Inferior Frontal Gyrus	45
4	3.53	46	28	-14	Right Inferior Frontal Gyrus	47
5	4.73	10	8	-6	Right Caudate	48
5	4.50	-16	-2	18	Left Caudate	48
5	4.17	-4	0	2	Left Thalamus	50
5	4.10	-10	6	6	Left Caudate	48
5	3.96	10	6	2	Right Thalamus	48

5	3.90	-8	-4	14	Left Thalamus	50
6	4.60	-12	-72	56	Left Precuneus	7
6	4.33	-24	-70	36	Left Superior Occipital Gyrus	7
6	4.01	-34	-86	30	Left Middle Occipital Gyrus	19
6	3.95	-6	-70	60	Left Precuneus	7
6	3.88	-40	-84	28	Left Middle Occipital Gyrus	19
6	3.82	-28	-76	32	Left Middle Occipital Gyrus	39
7	5.34	6	26	44	Right Medial Superior Frontal Gyrus	8
7	4.76	8	30	36	Right Median Cingulate Gyrus	8
7	4.69	8	44	16	Right Anterior Cingulate Gyrus	10
7	4.58	0	34	36	Left Medial Superior Frontal Gyrus	8
7	4.54	4	36	42	Right Medial Superior Frontal Gyrus	8
7	4.37	2	40	34	Left Medial Superior Frontal Gyrus	8
8	5.09	-14	-50	-44	Left Cerebellum	37
8	4.77	12	-78	-26	Right Cerebellum	18
8	4.75	20	-44	-46	Right Cerebellum	37
8	4.71	-8	-74	-32	Left Cerebellum	18
8	4.54	18	-60	-26	Right Cerebellum	37
8	4.51	8	-74	-46	Right Cerebellum	18
Visual Detection Low Discriminability						
1	4.09	-32	26	-4	Left Insula	13
1	4.07	-30	22	-10	Left Insula	47
1	4.05	-32	28	-8	Left Inferior Frontal Gyrus	47
1	3.98	-40	20	-2	Left Insula	13
1	3.44	-44	14	-12	Left Insula	47
1	3.37	-38	18	-12	Left Insula	47
2	4.67	38	22	-8	Right Insula	13
2	4.63	36	22	-2	Right Insula	13
2	4.48	30	24	-12	Right Inferior Frontal Gyrus	47
3	4.81	0	32	38	Left Medial Superior Frontal Gyrus	8
3	4.62	2	38	36	Left Medial Superior Frontal Gyrus	8
3	4.60	-4	44	14	Left Anterior Cingulate Gyrus	32
3	4.56	-4	30	30	Left Anterior Cingulate Gyrus	32
3	4.48	0	20	50	Left Supplementary Motor Area	8
3	4.43	6	40	16	Right Anterior Cingulate Gyrus	32
4	-4.28	42	-6	-10	Right Superior Temporal Gyrus	13
4	-3.98	52	6	-10	Right Superior Temporal Gyrus	22
4	-3.30	56	-14	-12	Right Middle Temporal Gyrus	22
4	-3.29	52	-8	-4	Right Superior Temporal Gyrus	22
5	-4.17	-6	-56	30	Left Precuneus	31

5	-3.96	-8	-48	32	Left Posterior Cingulate Gyrus	23
6	-4.62	-8	-98	2	Left Calcarine Fissure	18
6	-4.35	-26	-80	-14	Left Lingual Gyrus	19
6	-4.18	-18	-90	-16	Left Lingual Gyrus	18
6	-3.87	-12	-94	-2	Left Calcarine Fissure	18
6	-3.66	-12	-100	-6	Left Inferior Occipital Gyrus	18
6	-3.19	-36	-82	-4	Left Inferior Occipital Gyrus	18
7	-4.66	20	-84	-10	Right Lingual Gyrus	18
7	-4.56	24	-78	-4	Right Fusiform Gyrus	18
7	-4.55	26	-80	-12	Right Fusiform Gyrus	19
7	-4.47	34	-64	-8	Right Inferior Occipital Gyrus	37
7	-4.34	26	-72	-4	Right Fusiform Gyrus	19
7	-3.67	40	-70	-14	Right Inferior Occipital Gyrus	19

Table 6: Experiment 6 fMRI local maxima in the statistical *Z*-maps of the (CON > LIB) * (T > NT) response contrast in the low and moderate discriminability conditions for both the recognition memory and visual detection tests (see also **Figures 14** and **15**). Negative *Z*-values represent the reverse contrast of (CON > LIB) * (NT > T).

Target > Nontarget item contrast

Given the striking finding that changes in decision criteria robustly affected frontoparietal activity in the T > NT response contrasts—but *not* changes in discriminability—assessments of the T > NT item contrast (regardless of response type) sought to provide another means of identifying activity changes related to the discriminability manipulations. Given that target and nontarget items were randomly and evenly distributed across conditions, mean target strength should be equivalent between the conservative and liberal conditions. If frontoparietal activity is associated with target evidence strength, then greater activity should be observed in the moderate versus low discriminability conditions regardless of the criterion condition. The T > NT item contrasts in both tasks revealed sparse activations, but only when participants maintained a conservative criterion within the moderate discriminability condition (particularly in the right insula and MeFG). Since differences in activity between T > NT item contrasts across discriminability conditions were specific to the conservative condition, this again supports the notion that criterion placement plays a major role in frontoparietal differences between target and nontarget items. However, an interaction may exist where greater discriminability enhances frontoparietal activity, specifically when maintaining a conservative criterion, though whole-brain T > NT item contrasts revealed virtually no significant differences in (MOD > LOW) * (CON > LIB) subtractions across tasks.

ROI analyses

Whole-brain analyses of T > NT response contrasts revealed much more widespread frontoparietal activity in the recognition memory versus visual detection tests. However, virtually no differences existed at the whole-brain level when comparing across decision domains. One possibility is that these comparisons are underpowered, given the high-dimensionality of the data and the need for strict multiple comparisons correction. Therefore, more focal analyses were conducted based on 12 ROIs identified as criterion-sensitive regions during recognition memory tests for words (Aminoff et al., 2015). In the recognition memory task, all 12 ROIs revealed greater activity in the T > NT response contrast in CON > LIB comparisons across both discriminability conditions (**Figure 16**). Additionally, the (CON > LIB) * (T > NT) contrast in the recognition memory task remained higher for all ROIs relative to the visual detection task ($b = 5.89$, 95% CI [3.65, 8.16], $SD = 1.16$, $t = 5.09$, $d = 0.64$). However, mean beta values in the (CON > LIB) * (T > NT) response contrast of the visual detection task revealed significantly greater activity for both discriminability conditions in bilateral MFG and insula as well as left MeFG. The moderate discriminability condition of the visual detection task also revealed significantly greater activity in the (CON

> LIB) * (T > NT) response contrast for the right SPL and IPL, as well as left IFG. These findings reveal similar activation patterns across frontoparietal networks that generalize across task domains to a degree, though frontoparietal activity appears to be more strongly activated during recognition memory versus visual detection tests.

Comparisons of items, regardless of response type, for the (CON > LIB) * (T > NT) contrasts showed virtually no differences in mean beta values for the 12 ROIs between the recognition memory and visual detection task *within* each discriminability condition (b = 0.20, 95% CI [-1.43, 1.83], SD = 0.83, t = 0.24, d = 0.02) (**Figure 17**). Interestingly, the (MOD > LOW) * (CON > LIB) * (T > NT) contrast showed greater activity in both tasks (b = 3.60, 95% CI [1.30, 5.92], SD = 1.17, t = 3.08, d = 0.37), particularly in the insula, MeFG, MFG, IPL, and SPL. This again suggests that greater discriminability in both tasks increases frontoparietal activity when comparing across target and nontarget items, *specifically* when individuals maintain a conservative criterion. All comparisons included in the linear mixed models for T > NT responses and items are shown in **Table 7** and **Figure 18**.

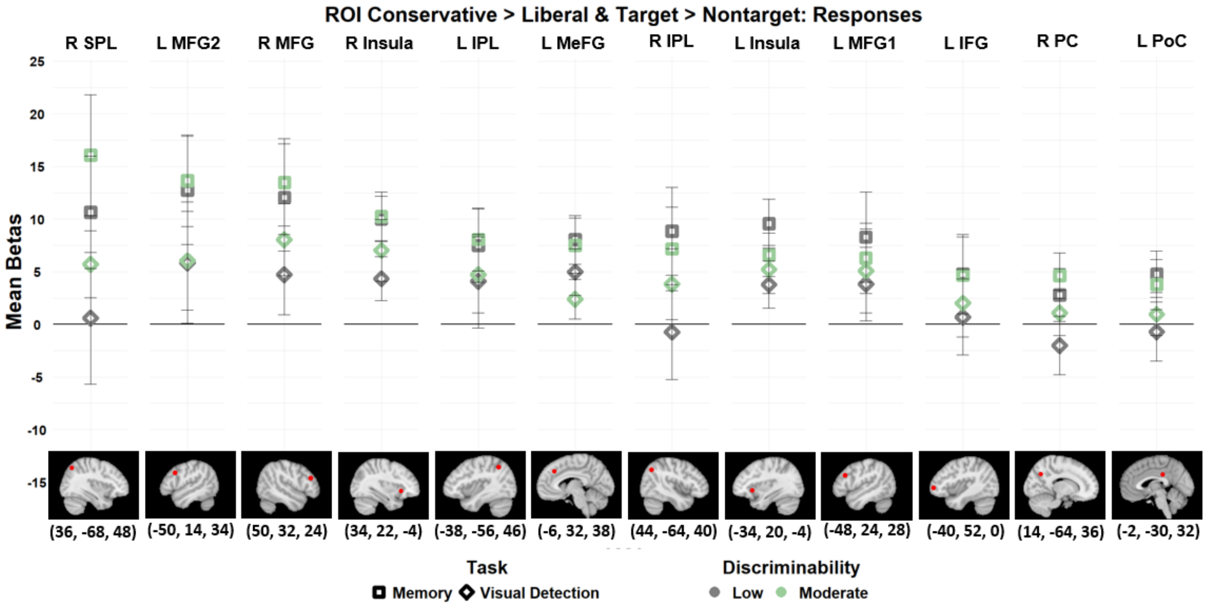


Figure 16: Experiment 6 ROI analyses comparing mean beta values across the T > NT response contrast between the conservative versus liberal conditions for the low (gray) and moderate (green) discriminability conditions in both the recognition memory (square) and visual detection (diamond) tasks. ROIs are ordered left to right based on the highest to lowest values in the moderate discriminability condition of the recognition memory task. The standard brain template coordinates of each ROI are listed at the bottom of the figure along with illustrations depicting the ROI location. Each point is fitted with 95% CIs. L = left; R = right; IFG = inferior frontal gyrus; IPL = inferior parietal lobules; MFG = middle frontal gyrus; MeFG = medial frontal gyrus; PC = precuneus; PoC = posterior cingulate.

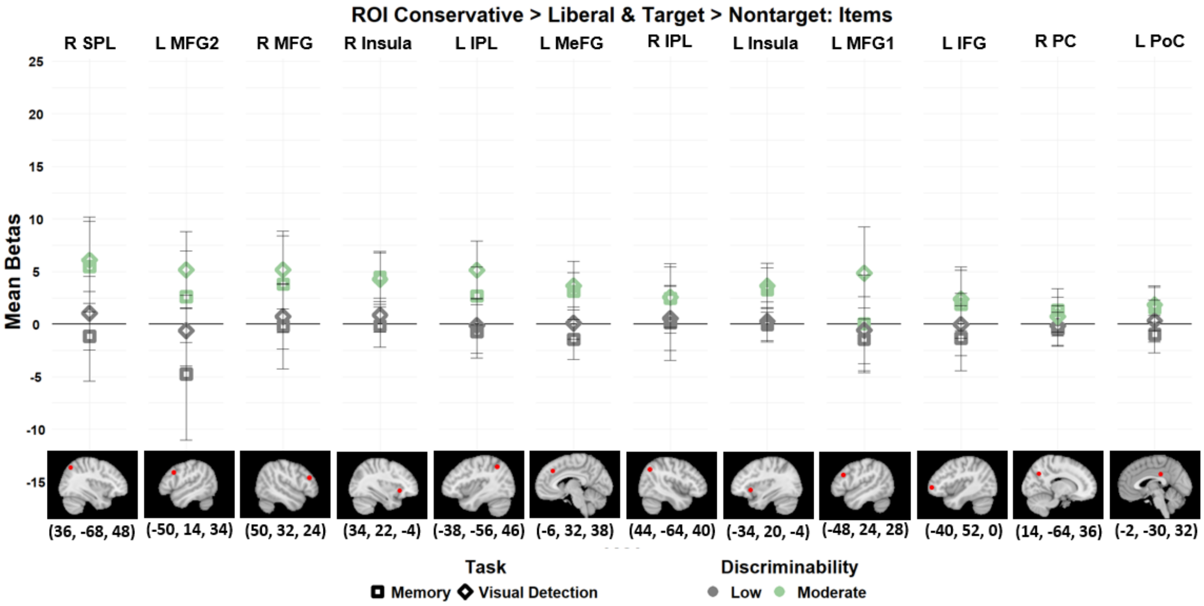


Figure 17: Experiment 6 ROI analyses comparing mean beta values across the T > NT item contrast between the conservative versus liberal conditions for the low (gray) and moderate (green) discriminability conditions in both the recognition memory (square) and visual detection (diamond) tasks. ROIs are ordered left to right based on the highest to lowest values in the moderate discriminability condition of the recognition memory task. The standard brain template coordinates of each ROI are listed at the bottom of the figure along with illustrations depicting the ROI location. Each point is fitted with 95% CIs. L = left; R = right; IFG = inferior frontal gyrus; IPL = inferior parietal lobules; MFG = middle frontal gyrus; MeFG = medial frontal gyrus; PC = precuneus; PoC = posterior cingulate.

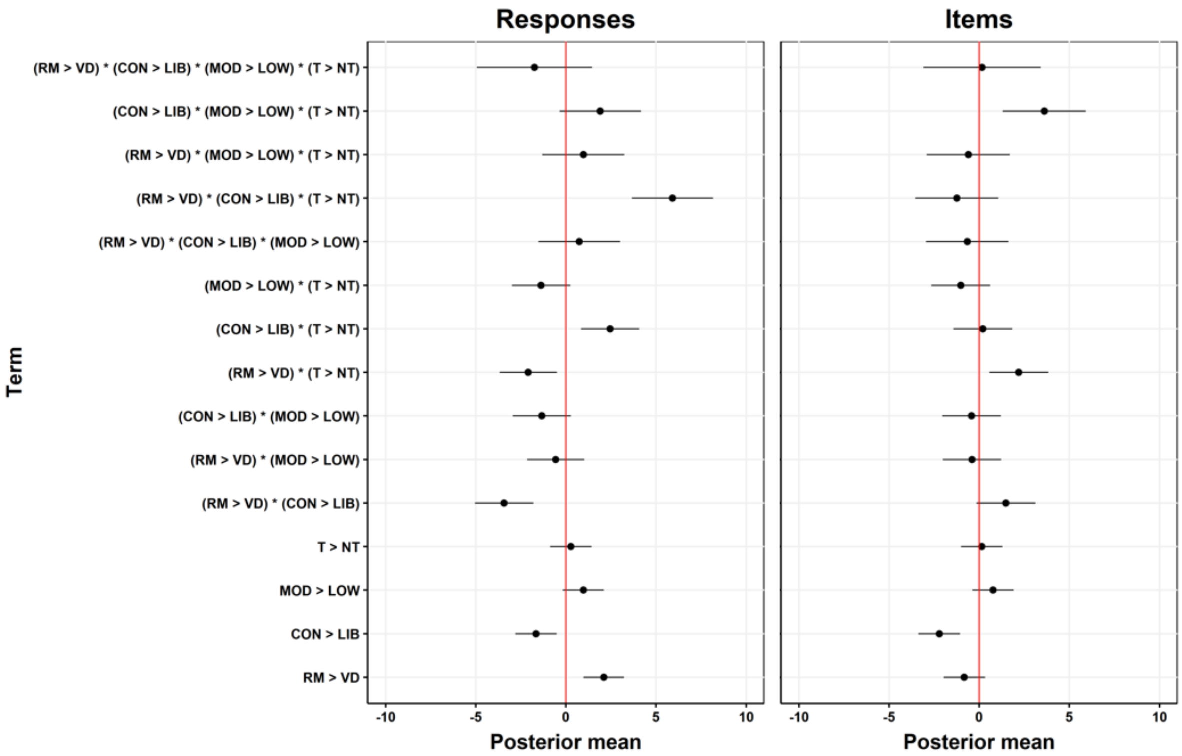


Figure 18: Experiment 6 posterior mean of parameter estimates across fixed effects of mean beta values in the 12 ROIs. Each parameter estimate is fitted with 95% confidence intervals for response (left) and item (right) contrasts. Estimates not intersecting zero are statistically significant.

Experiment 6 Model-Level Statistics: ROI Mean Beta Values (Responses)

Term	**Estimate (95 CI)**	***SE***	***t***	**Effect Size (*d*)**
Intercept	3.71 (1.25, 6.17)	1.25	2.96	0.41
RM > VD	2.09 (0.98, 3.22)	0.58	3.62	0.23
CON > LIB	-1.66 (-2.80, -0.51)	0.58	-2.87	0.18
MOD > LOW	0.96 (-0.18, 2.11)	0.58	1.66	0.11
T > NT	0.28 (-0.87, 1.42)	0.58	0.48	0.03
(RM > VD) * (CON > LIB)	-3.42 (-5.04, -1.81)	0.82	-4.17	0.37
(RM > VD * (MOD > LOW)	-0.56 (-2.16, 1.02)	0.82	-0.68	0.06

(CON > LIB) * (MOD > LOW)	-1.33 (-2.95, 0.27)	0.82	-1.62	0.14
(RM > VD) * (T > NT)	-2.08 (-3.69 -0.49)	0.82	-2.54	0.23
(CON > LIB) * (T > NT)	2.45 (0.84, 4.06)	0.82	2.99	0.27
(MOD > LOW) * (T > NT)	-1.38 (-3.00, 0.24)	0.82	-1.69	0.15
(RM > VD) * (CON > LIB) * (MOD > LOW)	0.72 (-1.52, 3.00)	1.16	0.62	0.08
(RM > VD) * (CON > LIB) * (T > NT)	5.89 (3.65, 8.16)	1.16	5.09	0.64
(RM > VD) * (MOD > LOW) * (T > NT)	0.94 (-1.31, 3.24)	1.16	0.81	0.10
(CON > LIB) * (MOD > LOW) * (T > NT)	1.90 (-0.36, 4.16)	1.16	1.64	0.21
(RM > VD) * (CON > LIB) * (MOD > LOW) * (T > NT)	-1.72 (-4.95, 1.44)	1.64	-1.05	0.19
Random Effect: (Intercept | Subject) + (Intercept | ROI)				
Subjects	30			
ROIs	12			
(Intercept | Subject) (*SD)*	3.33			
(Intercept | ROI) (*SD)*	3.52			
N	5760			

Experiment 6 Model-Level Statistics: ROI Mean Beta Values (Items)

Term	**Estimate (95 CI)**	***SE***	***t***	**Effect Size (*d*)**
Intercept	5.02 (2.18, 7.87)	1.45	3.45	0.51
RM > VD	-0.83 (-1.98, 0.33)	0.58	-1.41	0.08
CON > LIB	-2.22 (-3.37, -1.07)	0.58	-3.79	0.23
MOD > LOW	0.76	0.58	1.30	0.08

	(-0.38, 1.92)			
T > NT	0.14 (-1.01, 1.29)	0.58	0.24	0.02
(RM > VD) * (CON > LIB)	1.47 (-0.15, 3.11)	0.83	1.78	0.15
(RM > VD * (MOD > LOW)	-0.40 (-2.02, 1.21)	0.83	-0.48	0.04
(CON > LIB) * (MOD > LOW)	-0.41 (-2.06, 1.20)	0.83	-0.50	0.04
(RM > VD) * (T > NT)	2.20 (0.56, 3.83)	0.83	2.66	0.22
(CON > LIB) * (T > NT)	0.20 (-1.43, 1.83)	0.83	0.24	0.02
(MOD > LOW) * (T > NT)	-1.02 (-2.66, 0.61)	0.83	-1.24	0.11
(RM > VD) * (CON > LIB) * (MOD > LOW)	-0.67 (-2.95, 1.62)	1.17	-0.58	0.07
(RM > VD) * (CON > LIB) * (T > NT)	-1.24 (-3.55, 1.06)	1.17	-1.06	0.12
(RM > VD) * (MOD > LOW) * (T > NT)	-0.61 (-2.91, 1.70)	1.17	-0.52	0.06
(CON > LIB) * (MOD > LOW) * (T > NT)	3.60 (1.30, 5.92)	1.17	3.08	0.37
(RM > VD) * (CON > LIB) * (MOD > LOW) * (T > NT)	0.16 (-3.09, 3.41)	1.65	0.10	0.01
Random Effect: (Intercept | Subject) + (Intercept | ROI)				
Subjects	30			
ROIs	12			
(Intercept | Subject) (*SD)*	4.43			
(Intercept | ROI) (*SD)*	3.93			
N	5760			

Table 7: Experiment 6 model-level statistics for mean beta values for target (T) and nontarget (NT) responses (top) or items (bottom) across recognition memory (RM) and visual detection (VD) tasks in the conservative

(CON) and liberal (LIB) conditions as well as the low (LOW) and moderate (MOD) discriminability conditions (see also **Figure 18**).

Discussion

Despite decades of research unequivocally and reliably associating widespread frontoparietal activity with T > NT response contrasts during recognition memory, the debate remains as to whether activity in these regions can best be ascribed to memory versus decisional processes. Some theories predict that activity in T > NT response contrasts is associated with the subjective experience of familiarity (Gilmore et al., 2015; McDermott et al., 2017), including processes such as mnemonic evidence accumulation (Wheeler and Buckner, 2003; Kahn et al., 2004; Wagner et al., 2005), the buffering of retrieved content (Wagner et al., 2005; Vilberg and Rugg, 2009), or memory-related attentional processes (Cabeza et al., 2008; Ciaramelli et al., 2008; Ciaramelli et al., 2020), which should be affected by changes in discriminability regardless of the decision criterion. Others suggest that expectations of an item to be old versus new (O'Connor et al., 2010; Jaeger et al., 2013) or the placement of a decision criterion (Aminoff et al., 2015; King & Miller 2017) is linked to activity in these contrasts, which should be affected by decision strategies independently of memory strength. Here we directly manipulated discriminability, criterion placement, and decision domains to better assess which aspects of frontoparietal activity are associated with each manipulation when comparing between responses that exceed the decision criterion (target) versus those that do not (nontarget).

Evidence strength and response bias accounts both predict greater frontoparietal activity in T > NT response contrasts when a conservative criterion is maintained: per SDT (Macmillan & Creelman, 2005), target responses confer greater strength on average, *and* require inhibiting prepotent nontarget responses. These accounts diverge when a liberal criterion is maintained because target responses still carry greater evidence strength; however, prepotent target responses must be inhibited to make a nontarget response. One challenge with examining the neural mechanisms underpinning a conservative versus liberal criterion is that some individuals will not shift criteria despite being explicitly aware of the advantages for doing so (Aminoff et al., 2012, 2015; Kantner et al. 2015; Frithsen et al., 2018; Layher et al., 2018; Miller & Kantner, 2019, Layher et al., 2020). Failing to strategically shift precludes the ability to investigate differential activity related to multiple criterion placements within-subjects. Aminoff and colleagues (2015) revealed no significant differences in the H > CR contrast across criterion conditions when participants failed to shift during recognition memory tests, demonstrating that a criterion manipulation alone does not significantly impact frontoparietal activity. We therefore carefully prescreened participants to exclude those who did not adequately shift criteria, ensuring that criterion-related contrasts reflected changes in decision-making behavior.

Whole-brain GLM analyses revealed that the adaptation of conservative versus liberal criteria drastically altered frontoparietal activity in T > NT response contrasts, both during recognition memory and visual detection tasks. Previous studies revealed robust frontoparietal activity in the H > CR contrast during recognition memory tests *specifically* when participants maintained a conservative criterion when the likelihood of encountering “old” items decreased (Aminoff et al., 2015; King & Miller 2017). Our results extend these findings by revealing that this pattern of widespread frontoparietal activity is also observed (1) when a payment manipulation is implemented (with equal probability of encountering target vs. nontarget items), (2) at different levels of discriminability, and (3) in both recognition memory *and* visual detection tasks. Additionally, our results revealed significant activations in the right insula, MFG, and MeFG in the NT > T response contrast when participants maintained a liberal criterion during recognition memory, which supports a response bias account. However, in the visual detection task, whole-brain models revealed no significant frontoparietal activity in the NT > T response contrasts when participants maintained a liberal criterion—despite the strong relationship in criterion shifting performance between decision domains. Additionally, the (CON > LIB) * (T > NT) response contrasts revealed more widespread frontoparietal activity for recognition memory versus visual detection tests, though whole-brain analyses revealed virtually no significant differences. However, ROI analyses revealed significantly greater activity in the (CON > LIB) * (T > NT) response contrasts across frontoparietal regions in the recognition memory versus visual detection tests, suggesting that the task domain may modulate frontoparietal activity. It is possible that the added demands of recognizing images versus visual detection alone, engages these frontoparietal networks to greater extents when a conservative versus liberal criterion is maintained. Nonetheless, the T > NT response contrasts across criterion conditions elicited similar frontoparietal networks across task domains, even though activity tended to be greater for recognition memory versus visual detection tests.

In stark contrast to the robust differences in frontoparietal activity associated with changes in criterion placement, varying levels of discriminability revealed virtually no significant differences in activity across T > NT response contrasts in either the recognition memory or visual detection tasks. Broader assessments of T > NT *item* contrasts revealed sparse activity in the right anterior insula and MeFG in the moderate discriminability conditions of both decision domains, but this *only* occurred when participants maintained a conservative criterion. ROI analyses also revealed virtually no significant differences in the T > NT item contrast between tasks, although an interaction of (MOD > LOW) * (CON > LIB) comparisons showed a generalized increase in frontoparietal activity. Together, these results suggest that maintaining a conservative criterion engages a frontoparietal network that may be modulated by changes in discriminability, regardless of the decision domain or response type. Other than this potential interaction between discriminability and maintaining a conservative criterion, these results suggest that the frontoparietal network classically observed in T > NT contrasts is rather insensitive to changes in discriminability (when controlling for the decision criterion) in both recognition memory and visual detection tasks.

While some studies report greater frontoparietal activity in memory tests at higher versus lower levels of discriminability (Wheeler and Buckner 2003; Criss et al., 2013; Ciaramelli et al., 2020), these studies generally do not include a criterion manipulation, making it difficult to rule out a response bias explanation. Furthermore, when studies attempt to manipulate both discriminability and decision criteria (e.g. Ciaramelli et al., 2020), there is no prescreen procedure to identify individuals who adequately shift, nor is there a large enough sample size to exclude individuals who fail to shift criteria from fMRI analyses, which detrimentally impacts accurate assessment of frontoparietal activity associated with criterion placement. Thus, the biggest hurdles for dissociating task-related activity due to strength of evidence versus criterion placement in T > NT response (or item) contrasts are prescreening out participants who do not adequately shift criteria and obtaining a large enough sample to overcome the apparent insensitivities of fMRI for detecting activity related to changes in discriminability. Our results clearly reveal that T > NT response contrasts are robustly modulated by appropriately adopting a conservative versus liberal criterion—but *not* when target strength is modulated between near-chance versus moderate levels of discriminability.

One limitation of our findings is that participants tended to shift criteria to large degrees, which may have caused individuals to be more attuned to the decision strategy rather than evidence strength, relative to tests that do not include a criterion manipulation. However, there are trait-like individual differences in how people place a decision criterion in recognition memory tests that do not include a criterion manipulation (Kantner and Linsday, 2012, 2014). Some people will regularly establish a conservative criterion, whereas others consistently maintain a liberal criterion, even when there is no advantage or instructions to do so. Thus, participants almost always exhibit some inherent bias in their decision strategies that must be accounted for when comparing across response types.

Importantly, we are *not* proposing that frontoparietal activity observed in T > NT response contrasts of recognition memory and visual detection tests is entirely attributable to the decision criterion. A response bias account alone is insufficient: widespread frontoparietal activity is more robust when comparing T > NT responses under a conservative criterion relative to NT > T responses when a liberal criterion is maintained. Maintaining a conservative criterion may require greater cognitive control for discerning relatively stronger versus weaker target evidence, whereas responding “target” under a liberal criterion may be less cognitively demanding since the decision may be a simpler assessment of whether an item elicits *any* decisional evidence or not. Additionally, changes in discriminability appear to modulate the strength of frontoparietal activity in T > NT item contrasts across decision domains, but only when participants maintain a conservative—but not a liberal—criterion.

Our results unambiguously demonstrate that frontoparietal activity in T > NT response contrasts is predominantly sensitive to changes in criterion placement rather than changes in discriminability, which future studies must account for. Recruitment of this frontoparietal network is dependent on the decision criterion in a domain-general manner, though recognition memory may modulate frontoparietal regions to a greater extent relative to visual detection tests. It will be critical for future studies to systematically assess the effects of decision evidence *and* criteria at many levels of discriminability (from near-chance to near-perfect performance) and criterion placement (from very conservative to very liberal) to better dissociate the neural substrates associated with these intertwining cognitive processes.

Experiments 7 & 8: Dense-sampling fMRI reveals dissociable criterion networks

Findings from Experiment 6 revealed at the group-level that widespread frontoparietal activity during recognition memory tests in the T > NT response contrast is greatly affected by the decision criterion but *not* changes in discriminability. However, the results from Experiment 6 might be underpowered given the relatively small sample size (*N* = 30) and the apparent insensitivities of fMRI for detecting BOLD signal changes in the T > NT response contrast as a function of differing levels of discriminability. Additionally, deriving results from group averages may miss key features of frontoparietal activity associated with discriminability changes, particularly if there are individual differences in the underlying neural mechanisms. Manipulations of decision criteria and discriminability in Experiment 6 only occurred at two levels each, and there may be nuanced interactions between the two processes since decisional evidence exists on a continuum, according to SDT (Macmillan & Creelman, 2005).

To better assess the role of frontoparietal activity across many levels of decision criteria and discriminability at the individual level, Experiments 7 and 8 implemented dense-sampling fMRI recognition memory paradigms in which one participant in each experiment underwent repeated testing across many sessions. In both experiments, decision criteria and discriminability during recognition memory tests were manipulated to four levels each in a fully-crossed design creating 16 test conditions. If a participant successfully establishes four distinct levels of criteria across four levels of discriminability, then it is possible to better assess fMRI activity associated with the conservativeness of a decision criteria, familiarity strength, or both processes (see **Figure 19** for an ideal SDT schematic of the 16 test conditions).

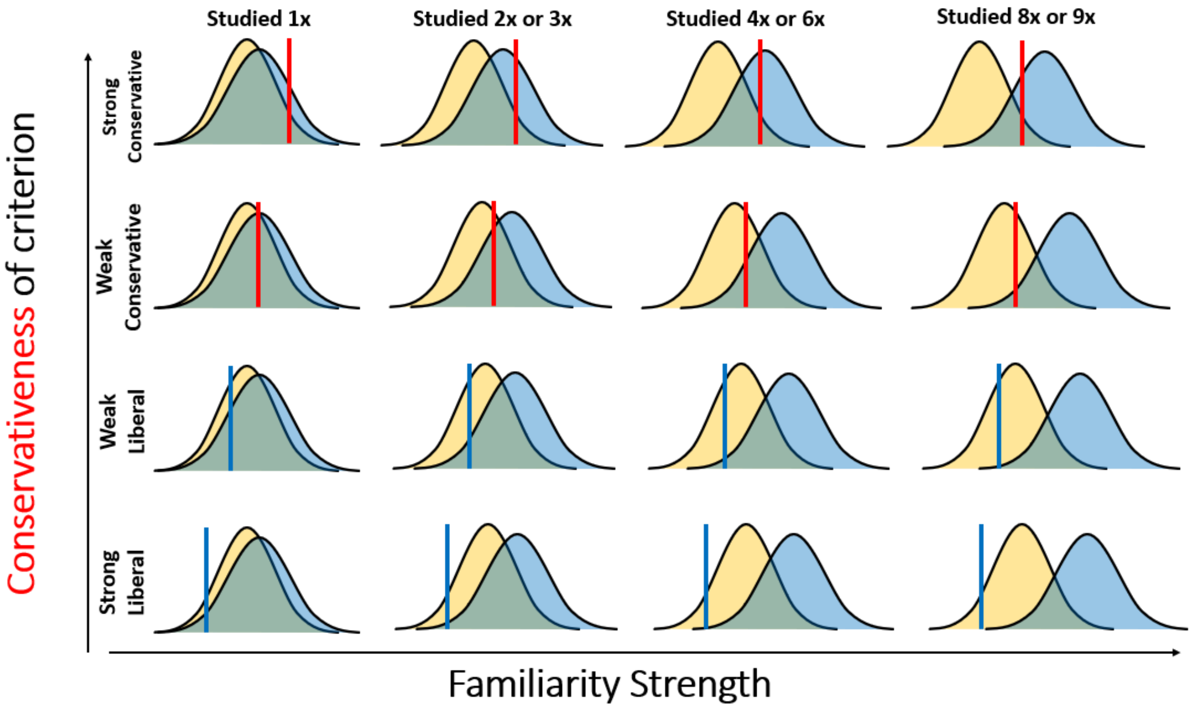


Figure 19: Experiments 7 and 8, ideal SDT depiction of distributions and criteria for differing familiarity strength (x-axis) and criterion placement (y-axis). Conditions are dependent on performance where the ideal

scenario requires participants to achieve four distinct levels of discriminability that are equivalent across the criterion conditions, while also maintaining four distinct levels of criteria that are equal across the discriminability conditions.

Method

Procedure

The dense-sampling fMRI experiments implemented a recognition memory paradigm that consisted of a fully-crossed 4x4 design in which decision criteria and discriminability were manipulated to four degrees each. A payment manipulation induced criterion shifting where each correct response resulted in earning four cents, a critical error resulted in a monetary loss, while a noncritical error resulted in no penalty. False alarms constituted the critical error in the conservative conditions whereas misses resulted in a critical error in the liberal conditions. To manipulate criteria to four different levels, the penalty amount differed between an eight-cent loss versus a loss of one cent. Since a correct response resulted in a four-cent gain, it is advantageous to only weakly bias a decision criterion when the critical error is only a one cent loss, whereas an extreme bias is advantageous when the critical error penalty is a loss of eight cents. Thus, the four criterion conditions are categorized as strong conservative (-8 cents for false alarms), weak conservative (-1 cent for false alarms), weak liberal (-1 cent for misses), and strong liberal (-8 cents for misses).

Manipulations of discriminability occurred by varying the number of times images were presented during the study phase. In Experiment 7 images appeared either once, twice, four times, or eight times during the study phase, whereas images in Experiment 8 appeared either once, three times, six times, or nine times. Familiarity strength of images during the test phase should be greater on average with an increased presentations during the study phase. This created a total of 16 test conditions (four criterion conditions per discriminability condition and vice versa). However, these 16 test conditions are entirely based on the participant's performance. Since many individuals fail to shift criteria when explicitly told to do so (Aminoff et al., 2012, 2015; Kantner et al. 2015; Frithsen et al., 2018; Layher et al., 2018; Miller & Kantner, 2020, Layher et al., 2020), prescreen computer tasks, which mimicked the fMRI tasks, were implemented to identify individuals who appropriately shift criteria to four different levels for each of the four discriminability conditions given the manipulations.

Prescreen

Since the paradigm required participants to maintain four different criteria at four different levels of discriminability, performance on an initial prescreen recognition memory task determined eligibility for the fMRI experiments. Participants first conducted a single session of the recognition memory test at a computer and were told that they may be invited back for additional prescreen sessions, depending on their performance, which may ultimately lead to an invitation to participate in the fMRI experiment. Individuals who appropriately established four distinct criteria (collapsed across discriminability conditions) and four distinct levels of discriminability (collapsed across criterion conditions) received an invitation to conduct three more prescreen sessions. After the fourth (Experiment 7) or eighth (Experiment 8) prescreen session, participants received an invitation to participate in the fMRI study if individuals appropriately spread out both criteria and discriminability across all 16 test conditions. To be eligible for the fMRI experiment, participants needed to have a difference of at least $c = 0.3$ between each criterion condition and a difference of $d' = 0.25$ across each discriminability condition.

Unequal-Variance Signal Detection Theory Model

Since the dense-sampling fMRI experiments implemented four criterion manipulations and many trials per condition, parameters for unequal-variance SDT models could be more accurately estimated. While the equal-variance SDT model implemented in Experiments 1-6 assumes that the evidence strength of *both* target (old) and nontarget (new) items follows a standard normal distribution, the unequal-variance SDT model used in Experiments 7 and 8 allows the standard deviation of the old distribution to vary. ROC curves were generated via maximum likelihood estimation through the ROC Toolbox (Koen & Yonelinas, 2016) based on the hit and false alarm rates from each criterion condition within the four discriminability conditions. Maximum likelihood estimation occurred separately for each of the four discriminability conditions, which produced unequal-variance SDT measures of discriminability (d_a), criterion (c_2), and old item distribution variance (V_o). For covariate fMRI analyses (see ***fMRI analysis*** below), unequal-variance measures of d_a and c_2 were attained through the following equations:

$$d_a = \sqrt{(2 / (1 + s^2))} * [z(HR) - s * z(FAR)]$$

$$c_2 = (-s / (1 + s)) * [z(HR) + z(FAR)],$$

where z represents the density of the standard normal distribution and s is the quotient of the standard deviation of the new versus old item distributions (Macmillan & Creelman, 2005; Stanislaw & Todorov, 1999). The s value derived from the average V_o ($M_{(Vo)}$) across the four discriminability conditions through the equation: $s = 1 / \sqrt{M_{(Vo)}}$.

fMRI analysis

Event-related GLMs implemented in FSL identified activity related to T > NT response and item contrasts across the 16 test conditions. Each test block contained 8 regressors of interest (old and new responses, or items, for each of the four mini-block conditions within a test block, as well as regressors for instructions, feedback, and trials with no responses). The default settings of FLOBS modelled HRF convolution for each regressor. The time window for HRF convolution spanned from image onset to offset. Additional nuisance regressors included six head motion parameters derived from motion correction realignment.

Whole-brain contrasts of statistical Z-maps with voxel-wise thresholding at $Z = 3.1$ and cluster correction using Gaussian Random Field Theory ($p < .05$), implemented in FEAT, determined statistically significant activity related to the T > NT response and item contrasts averaged within each of the 16 test conditions. Additionally, covariate analyses across all test blocks examined how whole-brain activity varies as function of criterion placement and discriminability. For the covariate analyses, c_2 and d_a values were computed for each test mini-block and mean centered across all test blocks. The mean centered values were included as co-variates for the mean T > NT response and item contrasts.

Experiment 7

Method

Participants

One healthy adult participant (male; age 22; right-handed) from UCSB completed the 16-session fMRI experiment and earned $20/hour plus monetary bonuses based on task performance. This participant was selected from a sample of forty-two subjects (12 males;

ages 18-25, *M* = 20, *SD* = 1.9) who completed an initial prescreen computer task and earned $10/hour in addition to monetary bonuses.

Procedure

The fMRI experiment included 16 sessions conducted on separate days over the span of a month. The recognition memory task during each session included 512 unique face stimuli, which summed to a total of 8,192 unique stimuli for the entire experiment. On each session the participant completed four cycles of a study block followed by four test mini-blocks for a total of 16 test mini-blocks (one per condition). During each study block, 64 unique images appeared either once, twice, four times, or eight times for a total of 240 stimulus presentation. Each study phase stimulus appeared for 200 ms in the center of the screen on a black background, followed by a crosshair presentation for 520 ms. Jitter trials of a crosshair presented on a black background lasting for 720 ms (1 TE) appeared randomly throughout the study phase. Zero to three jitter trials followed each study stimulus presentation and a total of 197 jitter trials appeared in each study block. Each study block scan included 7.2 seconds (10 TRs) before *and* after the study block presentation and lasted for approximately 5.5 minutes.

A test block scan followed each study block, in which four test mini-blocks appeared. Each mini-block included an instruction screen informing the participant of the payout structure for the upcoming test trials. The instruction screen appeared for 5.04 seconds (7 TRs) and listed the amount of money earned for correct responses (4 cents) and the amount of money lost for incorrect old and new responses (either 0, -1, or -8 cents). Afterwards a block of 32 trials appeared (16 old and 16 new) in a randomized order. Each test stimulus appeared for 2.5 seconds followed by a 380 ms presentation of a crosshair (4 TRs total). At the end of each test mini-block a feedback screen appeared for 5.04 seconds (7 TRs) to inform the participant of the amount of money earned on the previous test mini-block as well as the running total of money earned for the session. In between instruction, test stimuli, and feedback displays, zero to three jitter trials of a crosshair presented on a black background appeared for 720 ms (1 TR). A total of 203 jitter trials randomly appeared throughout the entire test block, which was preceded and ended by 7.2 seconds (10 TRs) of a crosshair presentation. Each test block lasted for approximately 9.5 minutes. The order of the 16 test conditions (one per mini-block) appeared randomly for each session with the exception that each of the four discriminability conditions appeared in each test block (in order to keep the length of the study phase the same throughout the experiment). The entire task took about an hour to complete (see that task design in **Figure 20**).

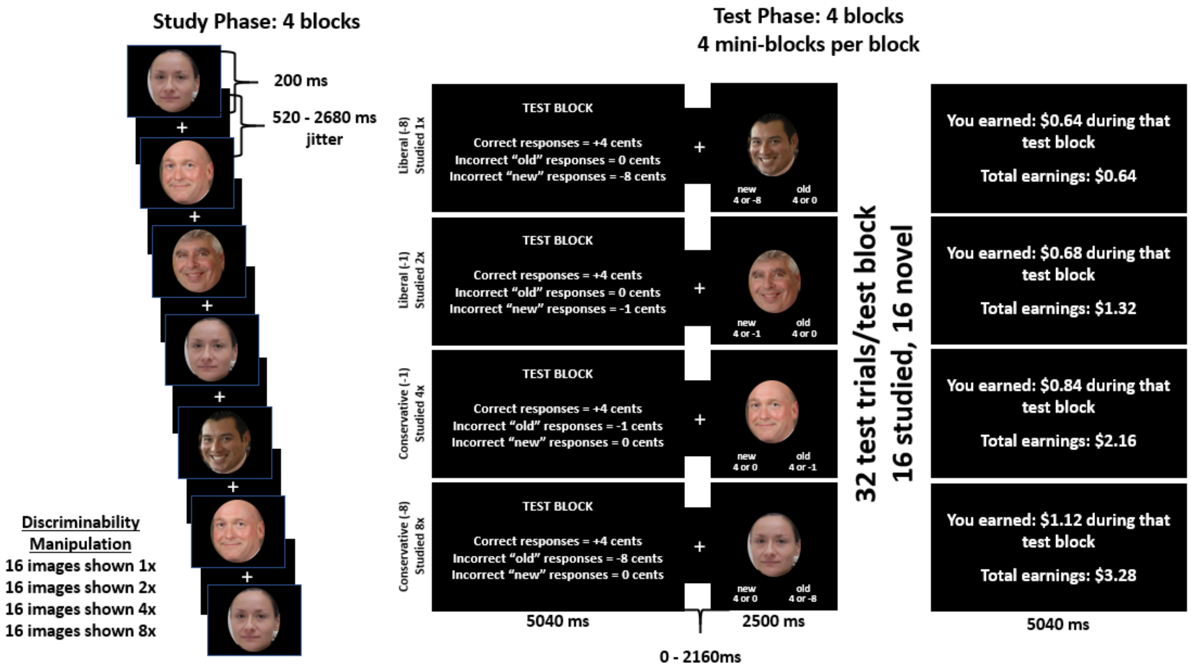


Figure 20: Experiment 7 recognition memory task that occurred during fMRI scanning across 16 sessions. The participant conducted four cycles of a ~5.5-minute study block, followed by a ~9.5-minute test block that included four test mini-blocks. During each session, the participant conducted 16 test mini-blocks (one per condition) in a randomized order with the exception that each test block needed to contain one mini-block per

discriminability condition in order to keep the length of the study phase consistent. The participant received feedback on the amount of money earned after each mini-block as well as the running total for the entire session.

Results

Behavioral findings

Manipulations of discriminability across the 16 sessions proved successful with maximum likelihood estimates of d_a increasing from the studied once (0.49), twice (0.80), four times (1.21), and eight times (1.96) conditions. The participant also successfully shifted decision criteria across the criterion conditions in each of the four discriminability conditions, which ranged from $c_2 = -1.23$ to $c_2 = 1.69$ (see **Table 8** for maximum likelihood estimates of c_2 and d_a values across all 16 conditions). Across the four discriminability manipulations, the participant on average shifted between the strong conservative and liberal conditions ($M = 2.62$) as well as the weak conservative and liberal conditions ($M = 1.20$). ROC curves derived from maximum likelihood estimates are presented in **Figure 21**.

Experiment 7 unequal-variance SDT measures					
	c_2				d_a
	Strong Liberal	**Weak Liberal**	**Weak Conservative**	**Strong Conservative**	
Studied 1x	-1.23	-0.14	1.16	1.61	0.49
Studied 2x	-0.88	-0.15	1.03	1.64	0.80
Studied 4x	-1.14	-0.21	1.08	1.69	1.21
Studied 8x	-0.66	0.20	1.26	1.64	1.96

Table 8: Experiment 7 maximum likelihood estimates for c_2 across the 16 conditions and d_a for each of the four discriminability conditions.

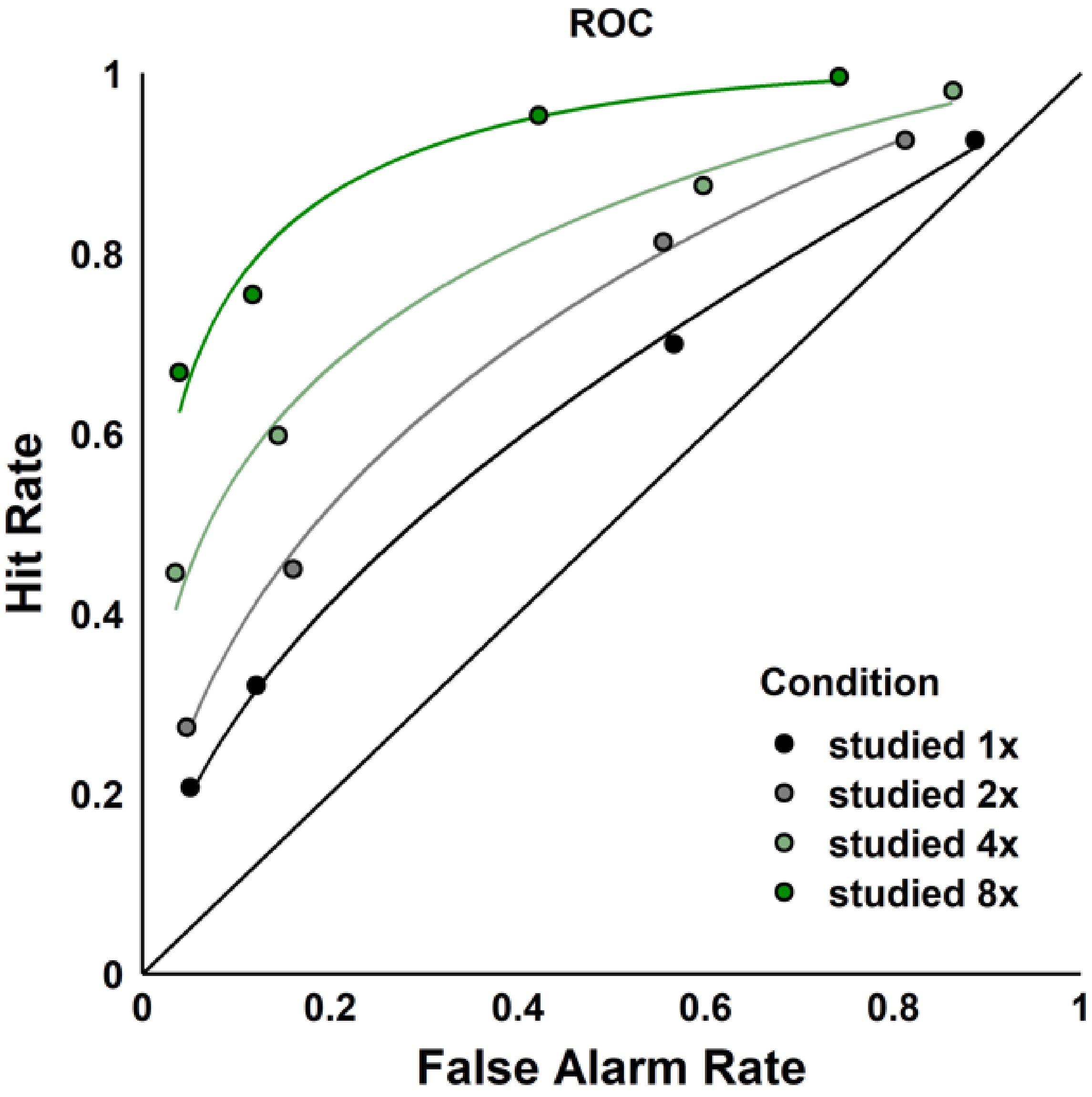


Figure 21: Experiment 7 ROC curves based on maximum likelihood estimates representing the hit and false alarm rates across criterion conditions from strong conservative (left) to strong liberal (right) across the four discriminability conditions.

T > NT response and item contrasts

Whole-brain GLM analyses computed separately for the 16 test conditions showed virtually no significant differences in the T > NT response *or* item contrast with the exception for significantly greater activity in the left insula of the strong conservative T > NT response contrast for items presented once during the study phase (maximum *Z*-value = 4.03). Contrasts conducted within each test condition are likely underpowered since each condition only consisted of 16 test mini-blocks and therefore the fMRI results for this experiment focus mainly on the covariate fMRI analyses, which assesses activity across all test blocks.

The covariate fMRI analyses revealed widespread frontoparietal activity in T > NT response contrast associated with varying levels of c_2. As the criterion becomes more conservative, many frontoparietal regions become more active including areas in the insula, IFG, MFG, superior frontal gyrus (SFG), IPL, and SPL (**Figure 22**, top left). These findings are consistent with group average T > NT response contrasts when comparing between conservative and liberal criteria (Aminoff et al. 2015; King & Miller 2017; Experiment 6). Interestingly, as the criterion becomes more *liberal* in the T > NT response contrast, other frontoparietal regions increase in activity including areas in the MeFG, SFG, temporal gyrus, and Pc. These results expand on previous group-level findings by showing that a different frontoparietal network might be associated with maintaining a liberal criterion (as opposed to a conservative criterion) in the T > NT response contrast.

When examining whole-brain activity associated with varying levels of discriminability, only sparse regions within the parietal cortex became more active as d_a increased, including areas in the IPL, SPL, and Pc (**Figure 22**, top right). While these findings suggest that these parietal regions may be sensitive to changes in recognition memory strength in the T > NT response contrast, these spatially sparse activations pale in comparison to the robust widespread frontoparietal activity associated with changes in the decision criterion. This supports the group-level conclusion from Experiment 6 that fMRI measures are rather insensitive to detecting brain activity associated with differing levels of discriminability.

Examinations of whole-brain T > NT *item* contrasts as a function of c_2 revealed spatially sparse cortical activations associated with maintaining a more liberal criterion in regions such as the angular gyrus (AG), IPL, and SPL, but no significant activations related to an increasingly conservative criterion (**Figure 22**, bottom left). These regions largely overlap with regions in the T > NT *response* contrast as c_2 decreases. The T > NT item contrast associated with varying levels of d_a revealed spatially sparse activations including regions within the IFG, MFG and temporal cortex (**Figure 22**, bottom right). As in Experiment 6, widespread frontoparietal activity in T > NT contrasts seems to be largely associated with the criterion and subsequent response types as opposed to discriminability and whether an item actually is old or new. **Table 9** provides a complete list of local maxima fMRI activations associated with c_2 and d_a in both T > NT response and item contrasts.

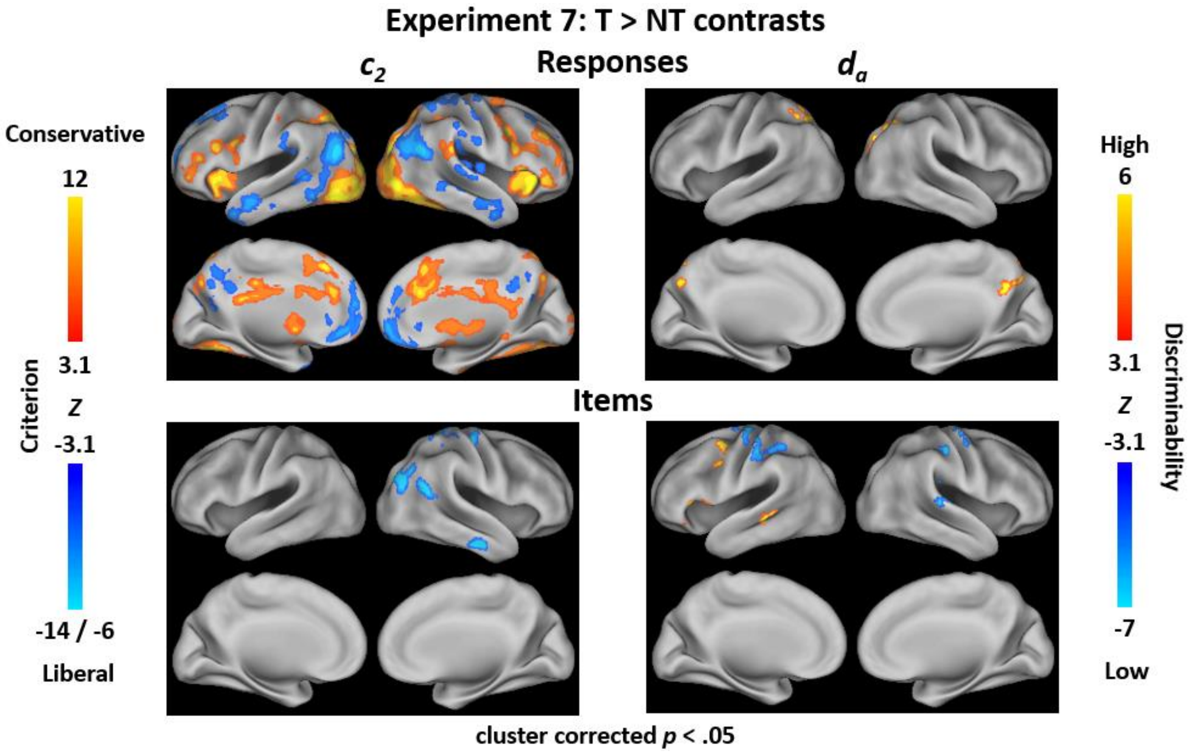


Figure 22: Experiment 7 whole-brain statistical *Z*-maps of T > NT response (top) and item (bottom) contrasts as a function of c_2 (left) and d_a (right) with thresholding at $Z > 3.1$ and cluster correction ($p < .05$). Values in orange and yellow represent brain areas that increase in activity as the criterion becomes more *conservative* (c_2 increases; left) or as discriminability *increases* (d_a increases; right). Values in blue represent brain areas that increase in activity as the criterion becomes more *liberal* (c_2 decreases; left) or as discriminability *decreases* (d_a decreases; right). Numbers on the color scales represent the minimum and maximum *Z*-value in each condition (when two values are separated by a slash, the first and second numbers represent the *Z*-value for the response and item contrasts, respectively).

Experiment 7: fMRI local maxima covariate T > NT contrasts

T > NT response contrast as a function of c_2

Cluster	*Z*-value	X	Y	Z	Location	BA
1	8.19	12	-70	42	Right Precuneus	7
2	5.64	20	6	66	Right Superior Frontal Gyrus	6
2	4.46	22	-2	56	Right Superior Frontal Gyrus	6
2	3.56	18	-6	66	Right Superior Frontal Gyrus	6
3	5.96	-30	-74	-50	Left Cerebellum	19
3	5.14	-22	-72	-44	Left Cerebellum	19
4	6.76	64	-34	24	Right Superior Temporal Gyrus	40
4	6.56	64	-42	18	Right Superior Temporal Gyrus	22
4	5.13	48	-40	24	Right Supramarginal Gyrus	22
4	5.00	56	-40	28	Right Supramarginal Gyrus	40

4	3.65	68	-32	32	Right Supramarginal Gyrus	40
5	7.82	-12	56	-18	Left Superior Frontal Gyrus	11
5	5.05	-14	66	-18	Left Superior Frontal Gyrus	11
5	4.29	-24	62	-10	Left Middle Frontal Gyrus	10
5	4.27	-28	48	-18	Left Middle Frontal Gyrus	10
5	3.79	-2	64	-22	Left Gyrus Rectus	11
5	3.41	-32	56	-14	Left Middle Frontal Gyrus	10
6	5.82	-10	2	4	Left Thalamus	48
6	5.82	-8	4	-2	Left Thalamus	48
6	4.98	-16	12	-8	Left Thalamus	49
6	4.78	-18	6	-14	Left Olfactory Cortex	49
7	6.64	-40	40	-4	Left Inferior Frontal Gyrus	47
7	6.30	-40	32	12	Left Inferior Frontal Gyrus	46
7	6.13	-46	44	6	Left Inferior Frontal Gyrus	46
7	5.45	-44	38	14	Left Inferior Frontal Gyrus	46
7	4.00	-42	54	4	Left Middle Frontal Gyrus	10
8	7.82	-6	-76	-30	Left Cerebellum	18
8	5.59	10	-74	-30	Right Cerebellum	18
8	4.59	12	-78	-16	Right Cerebellum	18
8	4.35	6	-76	-36	Right Cerebellum	18
8	3.35	-16	-82	-20	Left Cerebellum	18
9	7.98	-44	2	22	Left Precentral Gyrus	44
9	6.38	-46	22	24	Left Inferior Frontal Gyrus	44
9	5.92	-56	22	24	Left Inferior Frontal Gyrus	45
9	5.00	-50	4	12	Left Rolandic Operculum	6
9	4.86	-52	6	32	Left Precentral Gyrus	6
9	4.53	-58	10	32	Left Precentral Gyrus	6
10	8.97	-34	22	0	Left Insula	13
10	7.52	-44	14	-6	Left Insula	13
10	6.67	-26	22	-14	Left Insula	47
10	3.72	-36	4	2	Left Insula	13
11	8.86	-44	-44	44	Left Inferior Parietal Lobule	40
11	8.85	-22	-66	50	Left Superior Parietal Lobule	7
11	8.15	-38	-46	34	Left Angular Gyrus	39
11	8.03	-32	-62	38	Left Middle Occipital Gyrus	7
11	7.96	-26	-64	44	Left Inferior Parietal Lobule	7
11	7.60	-30	-64	50	Left Superior Parietal Lobule	7
12	10.60	0	26	42	Left Medial Superior Frontal Gyrus	8
12	9.79	6	30	28	Right Anterior Cingulate Gyrus	8
12	7.15	0	-32	22	Left Posterior Cingulate Gyrus	23

12	6.52	-4	36	22	Left Anterior Cingulate Gyrus	32
12	6.03	0	8	26	Left Anterior Cingulate Gyrus	24
12	5.76	2	-12	28	Right Median Cingulate Gyrus	23
13	11.70	44	16	10	Right Inferior Frontal Gyrus	44
13	10.80	34	26	-4	Right Insula	13
13	10.50	30	24	-6	Right Inferior Frontal Gyrus	13
13	9.34	48	40	-6	Right Inferior Frontal Gyrus	47
13	9.20	44	6	18	Right Rolandic Operculum	44
13	8.04	38	12	-4	Right Insula	13
14	10.40	-38	-76	-12	Left Fusiform Gyrus	19
14	9.49	-34	-74	-2	Left Middle Occipital Gyrus	19
14	9.18	-28	-92	12	Left Middle Occipital Gyrus	18
14	9.11	-22	-94	16	Left Middle Occipital Gyrus	18
14	8.84	-36	-88	0	Left Middle Occipital Gyrus	18
14	8.76	-34	-88	-8	Left Inferior Occipital Gyrus	18
15	11.50	46	-78	-2	Right Inferior Occipital Gyrus	19
15	11.20	36	-78	20	Right Middle Occipital Gyrus	39
15	9.60	28	-80	26	Right Middle Occipital Gyrus	39
15	9.53	44	-64	-8	Right Inferior Temporal Gyrus	37
15	9.52	34	-66	-14	Right Fusiform	37
15	9.12	46	-46	-18	Right Inferior Temporal Gyrus	37
16	-5.26	-62	-32	38	Left Supramarginal Gyrus	40
16	-4.79	-52	-32	20	Left Superior Temporal Gyrus	40
16	-4.26	-64	-30	28	Left Supramarginal Gyrus	40
17	-6.71	-26	52	14	Left Middle Frontal Gyrus	10
18	-5.89	66	-28	-4	Right Middle Temporal Gyrus	21
18	-5.68	66	-18	-4	Right Superior Temporal Gyrus	22
18	-4.98	60	-44	-6	Right Middle Temporal Gyrus	37
18	-4.20	56	-30	-4	Right Middle Temporal Gyrus	21
18	-4.10	64	-40	0	Right Middle Temporal Gyrus	21
19	-7.62	56	-8	-22	Right Middle Temporal Gyrus	21
19	-6.49	52	0	-32	Right Inferior Temporal Gyrus	20
19	-3.47	44	-4	-34	Right Inferior Temporal Gyrus	20
20	-6.56	0	-56	42	Left Precuneus	31
20	-6.18	12	-44	34	Right Median Cingulate Gyrus	23
20	-5.60	-8	-70	20	Left Calcarine Fissure	17
20	-5.52	-2	-66	34	Left Precuneus	31
20	-5.13	-4	-50	32	Left Posterior Cingulate Gyrus	23
20	-5.04	-8	-74	26	Left Cuneus	18
21	-7.62	-14	40	46	Left Superior Frontal Gyrus	8

21	-7.20	-24	32	50	Left Middle Frontal Gyrus	8
21	-6.92	-30	26	54	Left Middle Frontal Gyrus	8
21	-6.03	-36	18	48	Left Middle Frontal Gyrus	8
21	-5.67	-20	34	38	Left Superior Frontal Gyrus	8
21	-5.62	-22	22	38	Left Superior Frontal Gyrus	8
22	-7.50	-56	-2	-20	Left Middle Temporal Gyrus	21
22	-7.16	-44	12	-34	Left Temporal Pole	38
22	-6.41	-48	-10	-24	Left Inferior Temporal Gyrus	21
22	-6.29	-42	6	-42	Left Inferior Temporal Gyrus	38
22	-5.37	-36	16	-38	Left Temporal Pole	38
22	-4.50	-28	20	-38	Left Temporal Pole	38
23	-13.10	56	-60	24	Right Angular Gyrus	39
23	-5.87	38	-74	40	Right Middle Occipital Gyrus	39
23	-4.72	52	-62	48	Right Angular Gyrus	39
24	-10.70	-34	-66	28	Left Middle Occipital Gyrus	39
24	-10.70	-40	-70	26	Left Middle Occipital Gyrus	19
24	-9.79	-30	-74	38	Left Middle Occipital Gyrus	39
24	-9.54	-36	-64	20	Left Middle Occipital Gyrus	19
24	-9.51	-48	-66	22	Left Middle Temporal Gyrus	39
24	-9.14	-26	-78	44	Left Superior Parietal Lobule	7
25	-7.08	28	-44	68	Right Postcentral Gyrus	5
25	-6.74	48	-28	14	Right Superior Temporal Gyrus	40
25	-6.54	40	-18	68	Right Precentral Gyrus	6
25	-6.37	20	-38	76	Right Postcentral Gyrus	5
25	-6.32	54	-16	10	Right Superior Temporal Gyrus	40
25	-6.23	40	-26	16	Right Heschl Gyrus	41
26	-9.44	2	58	-10	Left Medial Superior Frontal Gyrus	10
26	-8.07	-4	52	10	Left Anterior Cingulate Gyrus	10
26	-7.86	-2	54	-4	Left Medial Superior Frontal Gyrus	10
26	-7.58	30	26	50	Right Middle Frontal Gyrus	8
26	-7.10	14	44	32	Right Superior Frontal Gyrus	9
26	-6.79	0	50	22	Left Medial Superior Frontal Gyrus	9
T > NT response contrast as a function of d_a						
1	4.95	32	-74	34	Right Middle Occipital Gyrus	39
1	4.67	34	-80	22	Right Middle Occipital Gyrus	39
1	4.25	32	-62	36	Right Middle Occipital Gyrus	39
1	4.17	26	-64	40	Right Superior Occipital Gyrus	7
1	4.06	26	-66	46	Right Superior Occipital Gyrus	7
1	3.78	38	-62	48	Right Angular Gyrus	39
2	4.42	-4	-76	32	Left Cuneus	18

2	4.38	4	-70	36	Right Precuneus	7
2	4.34	2	-60	36	Right Precuneus	31
2	4.20	6	-56	32	Right Precuneus	31
2	4.17	12	-74	54	Right Superior Parietal Lobule	7
2	4.16	6	-76	38	Right Cuneus	7
3	5.38	-26	-62	62	Left Superior Parietal Lobule	7
3	4.56	-30	-66	50	Left Superior Parietal Lobule	7
3	4.45	-18	-70	52	Left Superior Parietal Lobule	7
3	4.40	-26	-62	38	Left Middle Occipital Gyrus	39
3	4.34	-44	-54	58	Left Inferior Parietal Lobule	39
3	4.33	-40	-54	54	Left Inferior Parietal Lobule	39
T > NT item contrast as a function of c_2						
1	-4.69	-20	-78	56	Left Superior Parietal Lobule	7
1	-4.22	-30	-74	38	Left Middle Occipital Gyrus	39
1	-4.20	-26	-80	46	Left Superior Parietal Lobule	7
1	-3.74	-24	-80	50	Left Superior Parietal Lobule	7
1	-3.72	-32	-76	50	Left Inferior Parietal Lobule	39
1	-3.58	-24	-84	38	Left Superior Occipital Gyrus	19
2	-4.79	56	-12	-28	Right Inferior Temporal Gyrus	21
2	-3.49	48	-10	-26	Right Inferior Temporal Gyrus	21
3	-4.13	36	-20	50	Right Postcentral Gyrus	4
3	-3.90	30	-20	48	Right Postcentral Gyrus	4
3	-3.60	38	-18	66	Right Precentral Gyrus	6
3	-3.44	24	-26	50	Right Postcentral Gyrus	4
4	-5.61	22	-80	54	Right Superior Parietal Lobule	7
4	-5.13	30	-70	62	Right Superior Parietal Lobule	7
4	-4.55	26	-42	70	Right Postcentral Gyrus	5
4	-4.17	16	-32	72	Right Postcentral Gyrus	1
4	-4.11	22	-38	74	Right Postcentral Gyrus	1
4	-4.08	42	-68	52	Right Angular Gyrus	39
5	-5.31	48	-72	28	Right Middle Occipital Gyrus	39
5	-4.99	40	-76	38	Right Middle Occipital Gyrus	39
5	-4.88	56	-60	24	Right Angular Gyrus	39
5	-4.80	60	-60	26	Right Angular Gyrus	39
5	-4.26	40	-80	32	Right Middle Occipital Gyrus	39
5	-4.26	58	-66	14	Right Middle Temporal Gyrus	19
T > NT item contrast as a function of d_a						
1	5.14	-50	22	0	Left Inferior Frontal Gyrus	45
1	3.84	-46	28	-12	Left Inferior Frontal Gyrus	47
1	3.81	-46	22	-10	Left Inferior Frontal Gyrus	47

2	5.25	-46	-40	0	Left Middle Temporal Gyrus	21
2	4.39	-58	-30	-6	Left Middle Temporal Gyrus	21
3	4.36	-38	0	52	Left Precentral Gyrus	6
3	4.23	-42	2	56	Left Precentral Gyrus	6
3	4.22	-42	0	34	Left Precentral Gyrus	6
3	3.78	-38	0	44	Left Precentral Gyrus	6
4	-4.35	64	-28	6	Right Superior Temporal Gyrus	22
4	-4.28	54	-26	10	Right Superior Temporal Gyrus	41
4	-3.72	54	-16	12	Right Rolandic Operculum	40
4	-3.72	64	-22	10	Right Superior Temporal Gyrus	41
4	-3.67	46	-30	14	Right Superior Temporal Gyrus	40
4	-3.52	46	-28	20	Right Rolandic Operculum	40
5	-4.93	34	-28	38	Right Postcentral Gyrus	1
5	-4.30	46	-22	40	Right Postcentral Gyrus	1
5	-4.04	46	-24	46	Right Postcentral Gyrus	1
6	-4.00	44	-2	62	Right Middle Frontal Gyrus	6
6	-3.97	44	-14	66	Right Precentral Gyrus	6
6	-3.90	28	-8	72	Right Superior Frontal Gyrus	6
6	-3.87	40	-8	66	Right Precentral Gyrus	6
6	-3.52	36	-18	66	Right Precentral Gyrus	6
6	-3.46	46	-10	54	Right Precentral Gyrus	6
7	-6.74	-46	-18	62	Left Postcentral Gyrus	4
7	-5.96	-36	-40	68	Left Postcentral Gyrus	1
7	-5.87	-54	-28	56	Left Postcentral Gyrus	1
7	-5.76	-40	-18	68	Left Postcentral Gyrus	6
7	-5.48	-42	-40	62	Left Postcentral Gyrus	1
7	-5.30	-46	-38	62	Left Postcentral Gyrus	1

Table 9: Experiment 7 fMRI local maxima in the T > NT response (top) and item (bottom) contrasts as a function of increasing c_2 or d_a (see also **Figure 22**). Negative Z-values represent areas that become more active in T > NT contrasts as c_2 or d_a *decreases*.

Discussion

Experiment 7 sought to examine how fMRI activity is modulated by *many* manipulations of criteria and discriminability as opposed the two manipulations of criteria and discriminability implemented in Experiment 6. The dense-sampling fMRI findings revealed widespread frontoparietal activations in the T > NT response contrast associated with maintaining a conservative criterion, which is consistent with group-level findings

(Aminoff et al. 2015; King & Miller 2017; Experiment 6). Additionally, a separate network of frontoparietal regions in the T > NT response contrast emerged as the participant maintained an increasingly liberal criterion—a finding *not* previously observed at the group-level. The frontoparietal activations observed in Experiment 7 align quite well with previously defined networks based on resting-state functional connectivity analyses. Yeo and colleagues (2011) identified seven broad networks in which fMRI activity consistently co-fluctuated together within each network. As the participant maintained a more conservative decision criterion, the T > NT response contrast revealed greater activity in the frontoparietal, ventral attention, dorsal attention, and visual networks, which are networks associated with cognitive control and attentional processes (Yeo et al., 2011; Schaefer et al. 2017). This finding is driven both by the T > NT response contrast as the participant maintained a more conservative criterion *and* the reverse contrast of NT > T responses as the criterion becomes more liberal. Under a conservative criterion, nontarget responses are more common, whereas target responses are preponderant under a liberal criterion. When the participant made a response that could result in a critical error ("target" under a conservative criterion and "nontarget" when maintaining a liberal criterion), these attentional, cognitive control, and visual networks became more engaged. This finding is in line with a response bias account for the T > NT response contrast since it seems that prepotent responses needed to be inhibited, and stimuli more carefully scrutinized, in order to make a response that could result in a critical error.

In contrast to the cognitive control and attentional networks that became more active in the T > NT response contrast as the criterion becomes more conservative, the default mode network (DMN) became more active as the participant established a more liberal criterion.

This finding represents the T > NT response contrast under a liberal criterion and the reverse NT > T response contrast when a conservative criterion is maintained. In this case, the DMN becomes more active when the participant provided a prepotent response ("nontarget" under a conservative criterion and "target" when a liberal criterion is established). The DMN is described as the "default" state of the brain when a person is not engaged in an external task (Raichle et al., 2001). Although this is a novel finding only observed in the dense-sampling fMRI experiment, it reverberates the response bias account that prepotent responses require less attention and engagement compared to the opposite response, which must undergo extra scrutiny. While it is premature to draw group-level conclusions from a single participant, it is possible that assessments at the group-level are underpowered or need more than two criterion manipulations to capture this DMN activity associated with prepotent responses versus the rarer response type.

Although robust widespread frontoparietal networks were observed in the T > NT response contrast as a function of decision criteria, changes in discriminability only revealed sparse activations limited to regions within the parietal cortex. As discriminability increased, regions in the IPL, SPL, and Pc became more active which is consistent with previous fMRI findings showing that increases in memory strength elicit greater activity in these regions in various T > NT response contrasts (Wagner et al., 2005; Vilberg and Rugg, 2009; Gilmore et al., 2015; McDermott et al., 2017). However, the spatially sparse findings associated with increasing discriminability are very underwhelming relative to the robust frontoparietal networks observed as decision criteria become increasingly biased. Even when examining T > NT *item* contrasts, which removes the influence of response type, only spatially sparse activations emerged. This finding supports the conclusion from Experiment 6 that these fMRI contrasts are very insensitive to detecting brain activity associated with changes in discriminability. Although this dense-sampling experiment implemented many trials to try to circumvent this problem, it is possible that the data were still underpowered for detecting widespread changes in fMRI activity associated with differing levels of discriminability.

Experiment 8

Since Experiment 7 did not reveal robust fMRI activity in the T > NT response *or* item contrasts associated with changes in discriminability, Experiment 8 sought to better capture brain regions sensitive to varying levels of memory strength as well as replicate the fMRI findings associated with criterion placement in a different individual. Modifications to the Experiment 7 recognition memory paradigm included using visually rich scene stimuli as opposed to face stimuli, increasing the number of times images are shown during the study phase, doubling the number of sessions, and increasing the trial counts per session. These modifications intended to greatly increase the power for detecting brain areas associated with changes in discriminability to better dissociate activity related to decisional evidence versus the decision criterion.

Method

Participants

One healthy adult participant (female; age 29; right-handed) from UCSB completed the 31-session fMRI experiment and earned $20/hour plus monetary bonuses based on task performance. This participant was selected from a sample of fifty-five subjects (19 males; ages 18-29, *M* = 20, *SD* = 2.6) who completed an initial prescreen computer task and earned

$10/hour in addition to monetary bonuses. After 31 consecutive days of fMRI scanning, the participant experienced tinnitus at which point the experiment was terminated. The adverse event was reported to the UCSB and Army Research Office human subjects committees.

Procedure

The fMRI experiment consisted of 31 sessions conducted in the morning on consecutive days. Originally, the participant planned to conduct 32 sessions, but due to an adverse event after the 31st session the experiment was terminated early. The stimulus set for the recognition memory paradigm consisted of 20,480 unique scene images (640 stimuli per session) obtained from the SUN database (Xiao et al., 2010) and center-cropped to 330x330 pixels. During each session, the participant completed four cycles of a study block followed by four test mini-blocks for a total of 16 test mini-blocks (one per condition). The study block consisted of 80 unique stimuli either presented one, three, six, or nine times for a total of 380 stimuli presented in a random order. Initially, images were shown sequentially and continuously for 300 ms, but after the 12th session stimuli were presented for 600 ms to boost discriminability performance (see **Results**). Each study block lasted for about four minutes.

A test block scan followed each study block, in which four test mini-blocks appeared in a random order. Each mini-block included an instruction screen informing the participant of the payout structure for the upcoming test trials. The instruction screen appeared for 5.04 seconds (7 TRs) and listed the amount of money earned for correct responses (4 cents) and the amount of money lost for incorrect old and new responses (either 0, -1, or -8 cents). Afterwards a block of 40 trials appeared (20 old and 20 new) in a randomized order. Each test stimulus appeared for 2 seconds followed by a 160 ms presentation of a crosshair (3 TRs total). At the end of each test mini-block a feedback screen appeared for 5.04 seconds (7 TRs) to inform the participant of the amount of money earned on the previous test mini-block as well as the running total of money earned for the session. In between instruction, test stimuli, and feedback displays, zero to three jitter trials of a crosshair on a black background appeared for 720 ms (1 TR). A total of 186 jitter trials randomly appeared throughout the entire test block, which was preceded and ended by 7.2 seconds (10 TRs) of a crosshair presentation. Each test block lasted for approximately 9 minutes. The order of the 16 test conditions (one per mini-block) appeared randomly for each session with the exception that each of the four discriminability conditions appeared in each test block (in order to keep the length of the study phase the same throughout the experiment). The entire task took about 52 minutes to complete. **Figure 23** illustrates the Experiment 8 fMRI recognition memory paradigm.

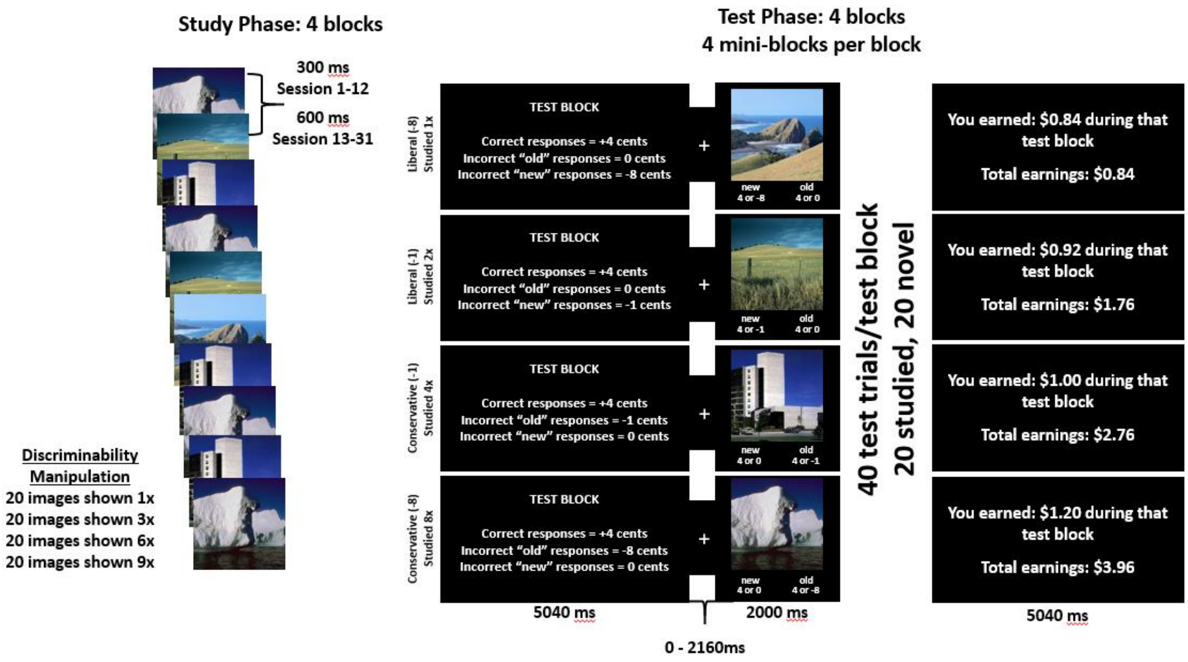


Figure 23: Experiment 8 recognition memory task that occurred during fMRI scanning across 31 sessions. The participant conducted four cycles of a ~4-minute study block, followed by a ~9-minute test block that included four test mini-blocks. During each session, the participant conducted 16 test mini-blocks (one per condition) in a randomized order with the exception that each test block needed to contain one mini-block per

discriminability condition in order to keep the length of the study phase consistent. The participant received feedback on the amount of money earned after each mini-block as well as the running total for the entire session. In sessions 1-12 study images appeared on screen for 300 ms each, but for sessions 13-31 images appeared for 600 ms each to boost discriminability performance.

Results

Behavioral findings

After the first 12 sessions, manipulations of discriminability proved less effective then intended as d_a ranged from 0.24 for images studied once to 1.00 for images studied nine times. Therefore, a small modification took place prior to session 13 in which images during the study phase appeared for 600 ms each, instead of 300 ms, which substantially improved discriminability performance. For sessions 13-31, maximum likelihood estimates of d_a increased for images studied once (0.51), three times (1.12), six times (1.43), and nine times (1.69). The participant successfully shifted decision criteria in each of the four discriminability conditions, which ranged from c_2 = -0.55 to c_2 = 2.48. However, in the weak liberal criterion condition, the participant maintained a conservative criterion on average (mean c_2 = 0.58) though this remained *relatively* more liberal than the weak conservative criterion (mean c_2 = 1.41). Table 10 shows the maximum likelihood estimates of c_2 and d_a values across all 16 conditions separately for when images appeared for 300 ms versus 600 ms during the study phase. Across the four discriminability manipulations, the participant shifted between the strong conservative and liberal conditions (*M* = 2.13) as well as the weak conservative and liberal conditions (*M* = 0.71). ROC curves derived from maximum likelihood estimates are presented in **Figure 24**.

Experiment 8 unequal-variance SDT measures					
	c_2 (300 ms study presentation in sessions 1-12)				**d_a**
	Strong Liberal	**Weak Liberal**	**Weak Conservative**	**Strong Conservative**	
Studied 1x	-0.28	0.67	2.19	2.48	0.24
Studied 3x	-0.46	0.51	1.72	2.11	0.64
Studied 6x	-0.55	0.29	1.49	1.86	0.69
Studied 9x	-0.39	0.36	1.43	1.80	1.00
	c_2 (600 ms study presentation in sessions 13-31)				**d_a**
	Strong Liberal	**Weak Liberal**	**Weak Conservative**	**Strong Conservative**	
Studied 1x	-0.19	0.87	1.39	1.82	0.51
Studied 3x	-0.48	0.67	1.09	1.66	1.12
Studied 6x	-0.22	0.64	0.94	1.67	1.43
Studied 9x	-0.50	0.59	0.99	1.56	1.69

Table 10: Experiment 8 maximum likelihood estimates for c_2 across the 16 conditions and d_a for each of the four discriminability conditions for when study images appeared for 300 ms in sessions 1-12 (top) and 600 ms in sessions 13-31 (bottom).

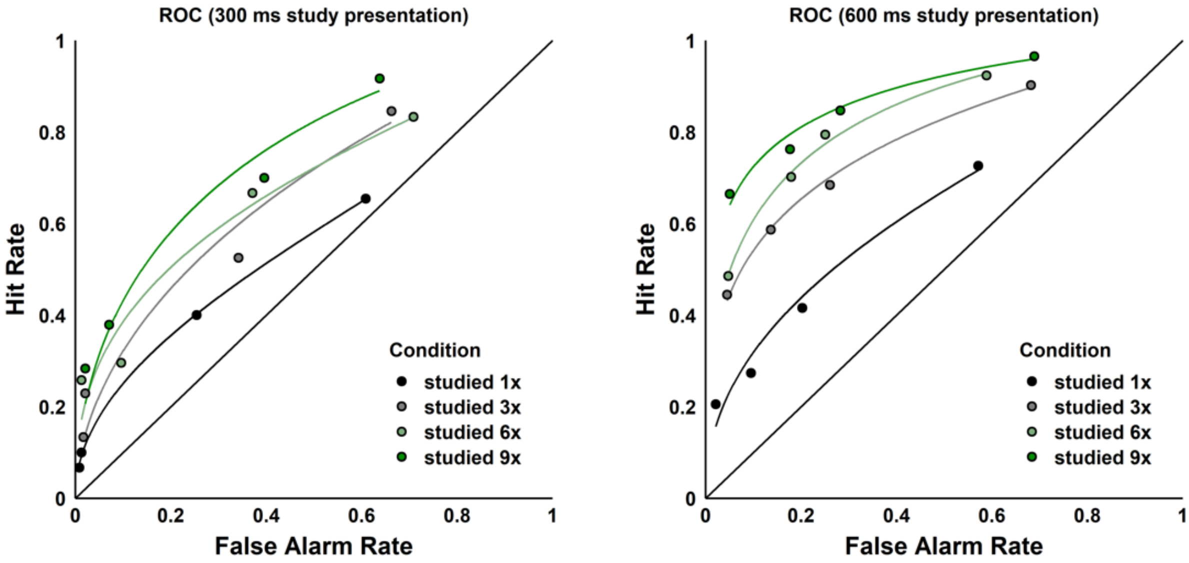


Figure 24: Experiment 8 ROC curves based on maximum likelihood estimates representing the hit and false alarm rates across criterion conditions from strong conservative (left) to strong liberal (right) across the four discriminability conditions. The left ROC figure represents sessions 1-12 in which study images appeared for 300 ms, whereas the right ROC figure represents session 13-31 in which study images appeared for 600 ms.

T > NT response and item contrasts

Although study images appeared for 300 ms in the first twelve sessions and 600 ms thereafter, all fMRI analyses reported here include results averaged from the 31 sessions together since no changes occurred in the test phase and the discriminability manipulation *did* work across all sessions, although to a lesser extent for sessions 1-12. Whole-brain GLM analyses computed separately for the 16 test conditions revealed widespread frontoparietal activity in the T > NT response contrast that largely varied as a function of the criterion conditions (**Figure 25**). Particularly striking is the reversal in activity in the T > NT response contrast between the *strong* conservative and liberal criterion conditions. When the participant maintained a strong conservative criterion, the T > NT response contrast revealed increased activity in regions of the frontoparietal, ventral attention, dorsal attention, and visual networks while revealing *decreased* activity in the DMN. When maintaining a strong liberal criterion, the T > NT response contrast revealed the opposite pattern of results with regions in the frontoparietal, ventral attention, dorsal attention, and visual networks showing *decreased* activity with increased activity in brain areas within the DMN. This again is consistent with the response bias account for T > NT response contrasts in that inhibiting prepotent responses engages cognitive control and attentional networks whereas providing a prepotent response requires less attentional resources. Interestingly, the weak liberal criterion condition (though the participant continued to maintain a slightly conservative criterion) showed the least robust effects regarding the spatial extent of frontoparietal activity in the T > NT response contrast. This suggests that the frontoparietal and DMN networks in this participant might become increasingly engaged as the criterion becomes more biased.

Whole-brain assessments of T > NT *item* contrasts within each of the 16 test conditions revealed little to no significant activity except for the strong conservative and liberal conditions, specifically when images were studied nine times (**Figure 26**). However, a similar trend occurred as in the T > NT *response* contrasts where images studied nine times under a conservative criterion elicited greater activity in the frontoparietal, ventral attention, dorsal attention, and visual networks, whereas regions within the DMN became more active under the strong liberal condition. While this again suggests that the criterion is largely driving the observed frontoparietal activity in the T > NT item contrast, there might be an interaction where greater discriminability heightens the activity of these criterion sensitive regions.

Covariate fMRI analyses revealed that varying levels of c_2 greatly affected T > NT response (**Figure 27**, top left) and item (**Figure 27**, bottom left) contrasts. Similar to findings in Experiment 7, an increasingly conservative criterion in the T > NT response contrast showed greater activity within widespread regions of the frontoparietal, ventral attention, dorsal attention, and visual networks, whereas the DMN became more active as the criterion became more liberal. Even the T > NT item contrast revealed this same pattern of results, though to a somewhat lesser spatial extent. As discriminability increases only a few regions within parietal and temporal cortex showed greater activity in the T > NT response contrast (**Figure 27**, top right) reverberating the findings of Experiment 6 and 7 that fMRI is seemingly insensitive for detecting activity related to changes in discriminability. Interestingly, the T > NT *item* contrast showed greater activity in frontoparietal regions including insula, angular gyrus and Pc as d_a increased regardless of criterion placement. It is possible that these regions are sensitive to the memory strength of items regardless of the response type in this participant, but if these are truly memory sensitive regions, then it is unclear why these same regions are not associated with increasing discriminability in the T > NT *response* contrast. Overall, these results replicate the findings of Experiment 7 in that specific frontoparietal networks appear to be consistently affected by changes in criterion placement. However, changes in discriminability do not show complimentary regional activations across the two dense-sampling participants. **Table 11** provides a complete list of local maxima fMRI activations associated with c_2 and d_a in both T > NT response and item contrasts.

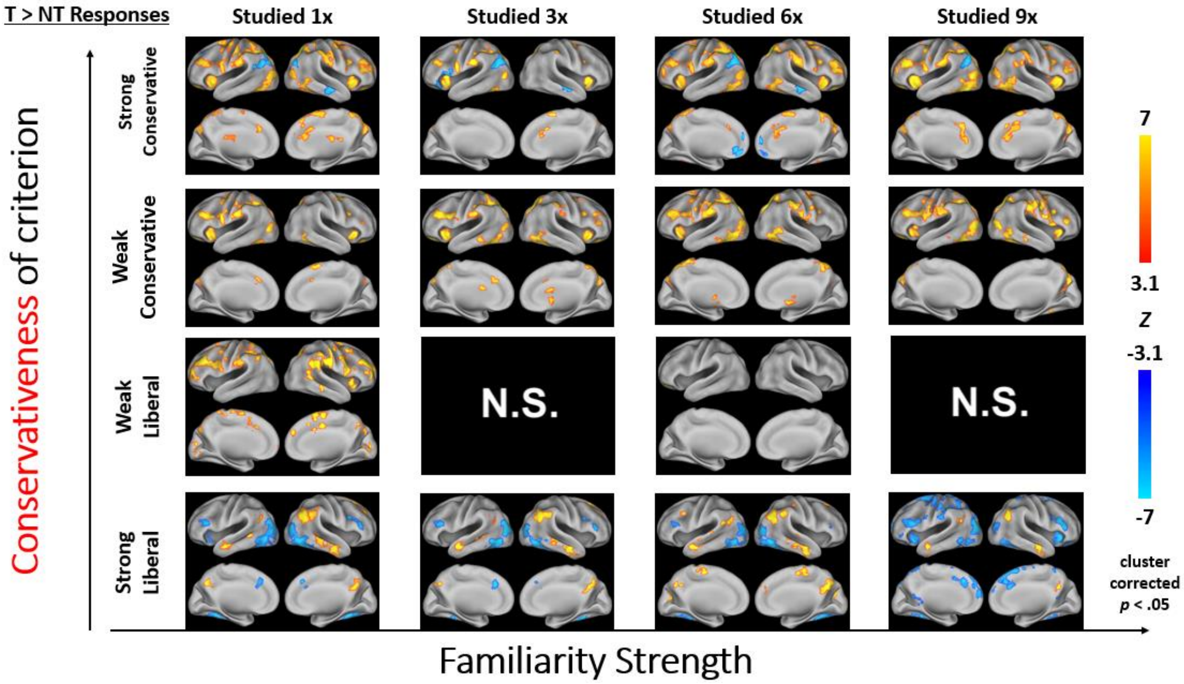


Figure 25: Experiment 8 whole-brain statistical Z-maps of T > NT *response* contrasts across the 16 conditions with thresholding at $Z > 3.1$ and cluster correction ($p < .05$). Values in orange and yellow represent brain areas with increased activity whereas values in blue represent *decreased* activity. Images containing “N.S.” represent conditions in which no significant activity occurred at the whole-brain level.

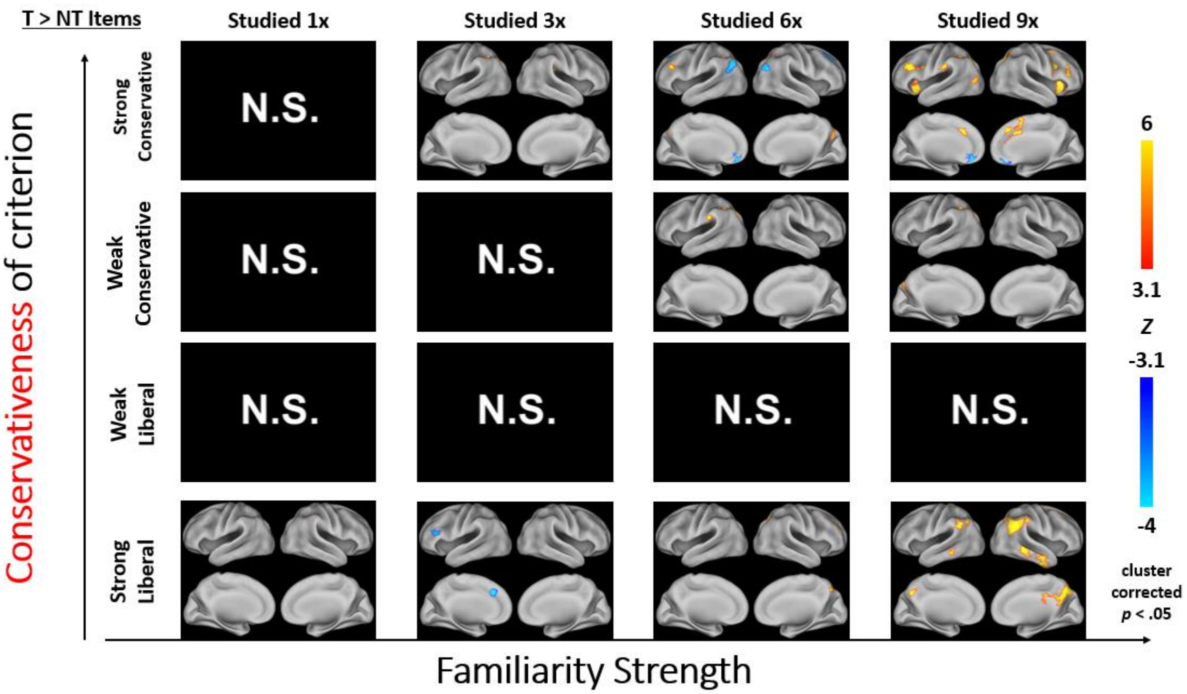


Figure 26: Experiment 8 whole-brain statistical *Z*-maps of T > NT *item* contrasts across the 16 conditions with thresholding at $Z > 3.1$ and cluster correction ($p < .05$). Values in orange and yellow represent brain areas with increased activity whereas values in blue represent *decreased* activity. Images containing "N.S." represent conditions in which no significant activity occurred at the whole-brain level.

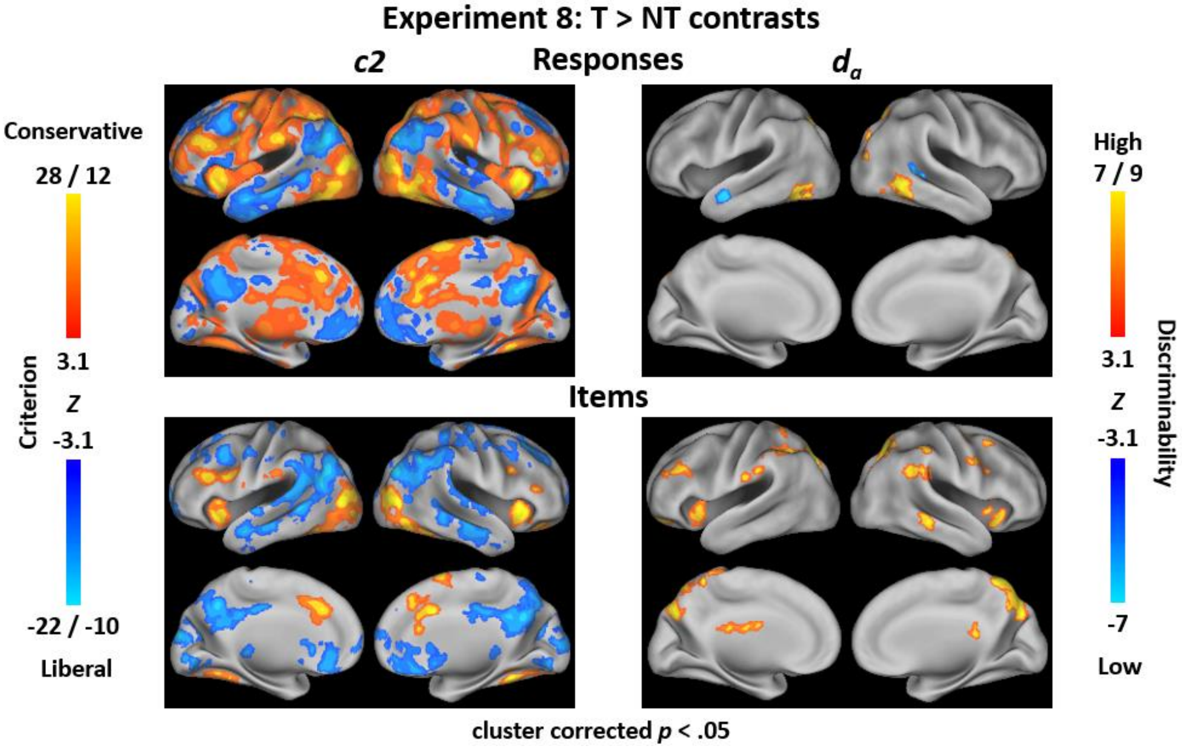


Figure 27: Experiment 8 whole-brain statistical *Z*-maps of T > NT response (top) and item (bottom) contrasts as a function of c_2 (left) and d_a (right) with thresholding at $Z > 3.1$ and cluster correction ($p < .05$). Values in orange and yellow represent brain areas that increase in activity as the criterion becomes more *conservative* (c_2 increases; left) or as discriminability *increases* (d_a increases; right). Values in blue represent brain areas that increase in activity as the criterion becomes more *liberal* (c_2 decreases; left) or as discriminability *decreases* (d_a decreases; right). Numbers on the color scales represent the minimum and maximum *Z*-value in each condition (when two values are separated by a slash, the first and second numbers represent the *Z*-value for the response and item contrasts, respectively). See also **Table 11**.

Experiment 8: fMRI local maxima covariate T > NT contrasts

T > NT response contrast as a function of c_2

Cluster	*Z*-value	X	Y	Z	Location	BA
1	27.70	-44	-78	10	Left Middle Occipital Gyrus	19
1	27.10	38	8	22	Right Inferior Frontal Gyrus	44
1	24.20	-54	-28	26	Left Supramarginal Gyrus	40
1	23.70	34	22	-6	Right Insula	13
1	23.70	-30	-46	34	Left Angular Gyrus	7
1	23.20	36	22	16	Right Inferior Frontal Gyrus	44
2	-9.52	-50	-50	-46	Left Cerebellum	37
2	-7.46	-50	-66	-38	Left Cerebellum	37
3	-6.51	6	-22	64	Right Supplementary Motor Area	6
3	-5.90	2	-22	72	Right Paracentral Lobule	6
3	-5.31	2	-16	52	Right Supplementary Motor Area	6
3	-4.88	-6	-18	58	Left Supplementary Motor Area	6
3	-4.49	-8	-22	64	Left Paracentral Lobule	4
3	-4.44	0	-34	50	Left Median Cingulate Gyrus	31
4	-7.93	40	-26	46	Right Postcentral Gyrus	1
4	-7.40	30	-26	44	Right Postcentral Gyrus	1
5	-11.90	-8	-50	-44	Left Cerebellum	37
5	-11.40	8	-50	-46	Right Cerebellum	37
5	-5.08	-2	-60	-58	Left Cerebellum	37
5	-4.42	4	-54	-70	Right Cerebellum	37
5	-4.09	-2	-54	-70	Left Cerebellum	37
6	-12.90	-46	6	8	Left Insula	44
6	-11.50	-46	30	-8	Left Inferior Frontal Gyrus	47
6	-9.77	-42	32	-14	Left Inferior Frontal Gyrus	47
6	-9.04	-52	18	6	Left Inferior Frontal Gyrus	45
6	-7.32	-32	20	-22	Left Inferior Frontal Gyrus	47
6	-5.13	-56	10	-4	Left Temporal Pole	22
7	-16.50	44	-54	28	Right Angular Gyrus	39
7	-16.40	44	-76	34	Right Middle Occipital Gyrus	39

7	-15.90	50	-56	38	Right Angular Gyrus	39
7	-14.30	54	-60	24	Right Angular Gyrus	39
7	-13.90	46	-64	50	Right Angular Gyrus	39
7	-13.00	54	-66	32	Right Angular Gyrus	39
8	-21.20	24	26	58	Right Frontal Superior Gyrus	8
8	-18.50	-14	34	38	Left Frontal Superior Gyrus	8
8	-17.70	20	24	44	Right Frontal Superior Gyrus	8
8	-16.40	22	16	66	Right Frontal Superior Gyrus	6
8	-16.10	-38	14	40	Left Middle Frontal Gyrus	8
8	-14.30	20	34	30	Right Frontal Superior Gyrus	8
9	-21.30	-40	-62	30	Left Angular Gyrus	39
9	-19.40	64	-18	-12	Right Middle Temporal Gyrus	21
9	-17.80	-38	-74	46	Left Inferior Parietal Lobule	39
9	-17.10	6	-56	32	Right Precuneus	31
9	-16.80	-48	-64	26	Left Angular Gyrus	39
9	-16.30	10	-50	26	Right Precuneus	23
T > NT response contrast as a function of d_a						
1	4.97	-20	-64	44	Left Superior Parietal Lobule	7
1	4.39	-16	-74	46	Left Superior Parietal Lobule	7
2	4.67	-48	-60	-10	Left Inferior Temporal Gyrus	37
2	4.02	-40	-68	-14	Left Fusiform Gyrus	19
2	3.73	-36	-72	-2	Left Middle Occipital Gyrus	19
2	3.50	-42	-70	-6	Left Inferior Occipital Gyrus	19
2	3.48	-50	-68	-12	Left Inferior Occipital Gyrus	37
3	6.37	34	-70	26	Right Middle Occipital Gyrus	39
3	5.36	32	-78	34	Right Middle Occipital Gyrus	39
3	4.48	26	-82	44	Right Superior Occipital Gyrus	7
3	3.72	36	-84	20	Right Middle Occipital Gyrus	19
3	3.60	42	-80	20	Right Middle Occipital Gyrus	19
4	5.69	20	-60	48	Right Superior Occipital Gyrus	7
4	4.26	18	-78	52	Right Superior Parietal Lobule	7
4	3.88	30	-50	40	Right Angular Gyrus	7
4	3.68	18	-66	68	Right Superior Parietal Lobule	7
4	3.42	14	-68	60	Right Superior Parietal Lobule	7
5	4.88	62	-52	-4	Right Middle Temporal Gyrus	37
5	4.85	58	-60	-8	Right Inferior Temporal Gyrus	37
5	4.62	54	-58	-2	Right Inferior Temporal Gyrus	37
5	4.13	52	-56	-18	Right Inferior Temporal Gyrus	37
5	4.01	60	-50	-18	Right Inferior Temporal Gyrus	37
5	3.99	54	-56	-22	Right Inferior Temporal Gyrus	37

6	-5.31	-54	-2	-20	Left Middle Temporal Gyrus	21
6	-3.27	-46	4	-28	Left Middle Temporal Gyrus	38
7	-6.98	46	-42	-2	Right Middle Temporal Gyrus	21
T > NT item contrast as a function of c_2						
1	6.59	24	34	-16	Right Superior Frontal Gyrus	11
1	5.71	16	44	-20	Right Superior Frontal Gyrus	11
1	5.59	30	40	-14	Right Middle Frontal Gyrus	47
1	3.90	26	38	-22	Right Middle Frontal Gyrus	11
1	3.66	24	32	-26	Right Superior Frontal Gyrus	11
2	6.21	8	10	66	Right Supplementary Motor Area	6
2	5.68	6	18	68	Right Supplementary Motor Area	6
2	4.46	10	0	56	Right Supplementary Motor Area	6
3	10.20	-30	18	-8	Left Insula	13
3	9.16	-30	14	2	Left Putamen	13
3	8.23	-40	14	-6	Left Insula	13
4	8.31	-30	-46	34	Left Angular Gyrus	7
4	8.05	-54	-28	26	Left Supramarginal Gyrus	40
4	6.95	-40	-46	36	Left Inferior Parietal Lobule	39
4	5.63	-64	-22	26	Left Supramarginal Gyrus	40
4	5.47	-50	-34	32	Left Supramarginal Gyrus	40
4	5.36	-20	-64	48	Left Superior Parietal Lobule	7
5	9.67	-4	22	36	Left Median Cingulate Gyrus	32
5	8.46	6	16	34	Right Median Cingulate Gyrus	32
5	7.97	10	26	20	Right Anterior Cingulate Gyrus	32
5	7.86	6	26	32	Right Median Cingulate Gyrus	8
5	5.10	6	34	44	Right Medial Superior Frontal Gyrus	8
6	9.42	-44	-2	22	Left Precentral Gyrus	6
6	8.57	-46	24	22	Left Inferior Frontal Gyrus	44
6	7.69	-48	32	28	Left Inferior Frontal Gyrus	9
6	7.62	-42	12	20	Left Inferior Frontal Gyrus	44
6	6.15	-36	34	24	Left Inferior Frontal Gyrus	9
6	6.07	-32	26	20	Left Inferior Frontal Gyrus	9
7	9.24	38	8	22	Right Inferior Frontal Gyrus	44
7	9.18	34	22	-6	Right Insula	13
7	9.00	38	26	14	Right Inferior Frontal Gyrus	45
7	8.83	36	18	-2	Right Insula	13
7	7.84	44	32	8	Right Inferior Frontal Gyrus	45
8	9.15	44	-74	8	Right Middle Occipital Gyrus	19
8	8.09	24	-46	-14	Right Fusiform Gyrus	37
8	7.89	28	-44	-14	Right Fusiform Gyrus	37

8	7.53	30	-80	10	Right Middle Occipital Gyrus	19
8	6.59	36	-50	-24	Right Cerebellum	37
8	6.34	48	-70	-10	Right Inferior Frontal Gyrus	37
9	11.40	-44	-78	8	Left Middle Occipital Gyrus	19
9	7.45	-32	-86	12	Left Middle Occipital Gyrus	19
9	7.09	-28	-68	24	Left Middle Occipital Gyrus	39
9	7.01	-46	-56	-16	Left Fusiform Gyrus	37
9	7.01	-34	-86	-2	Left Middle Occipital Gyrus	18
9	6.43	-26	-54	-12	Left Fusiform Gyrus	37
10	-5.62	-12	-86	-42	Left Cerebellum	18
11	-4.83	-50	-4	46	Left Precentral Gyrus	6
11	-4.35	-50	-16	28	Left Postcentral Gyrus	1
11	-4.11	-48	-20	38	Left Inferior Parietal Lobule	40
12	-6.89	-38	14	40	Left Middle Frontal Gyrus	8
12	-4.46	-48	16	38	Left Middle Frontal Gyrus	8
12	-4.43	-46	8	50	Left Precentral Gyrus	6
13	-5.53	-2	-28	4	Left Thalamus	50
13	-5.01	-10	-32	2	Left Thalamus	50
13	-4.24	0	-50	-10	Vermis	19
13	-4.16	0	-44	0	Vermis	36
13	-3.97	14	-32	4	Right Thalamus	50
14	-5.57	-22	-22	72	Left Precentral Gyrus	6
14	-5.25	-18	-20	50	Left Median Cingulate Gyrus	6
14	-5.22	-32	-26	62	Left Precentral Gyrus	4
14	-5.22	-14	-20	68	Left Paracentral Lobule	6
14	-5.03	-28	-32	54	Left Postcentral Gyrus	4
14	-4.95	-26	-26	58	Left Precentral Gyrus	4
15	-7.97	-22	62	4	Left Superior Frontal Gyrus	10
15	-7.12	-18	20	40	Left Superior Frontal Gyrus	8
15	-6.70	14	68	12	Right Medial Superior Frontal Gyrus	10
15	-6.26	-28	24	32	Left Superior Frontal Gyrus	9
15	-6.19	-14	34	38	Left Superior Frontal Gyrus	8
15	-6.10	0	36	-8	Right Medial Superior Frontal Gyrus	32
16	-9.62	44	-76	34	Right Middle Occipital Gyrus	39
16	-9.47	4	-54	32	Right Posterior Cingulate Gyrus	31
16	-9.29	-8	-92	18	Left Cuneus	18
16	-9.15	44	-52	26	Right Angular Gyrus	39
16	-8.92	28	24	56	Right Middle Frontal Gyrus	8
16	-8.86	34	26	50	Right Middle Frontal Gyrus	8
T > NT item contrast as a function of d_a						

1	5.21	-34	-54	-56	Left Cerebellum	20
2	8.01	-22	50	-14	Left Middle Frontal Gyrus	11
3	4.73	36	12	46	Right Middle Frontal Gyrus	8
3	4.68	40	-2	54	Right Middle Frontal Gyrus	6
3	4.02	28	-4	42	Right Precentral Gyrus	6
3	3.51	48	6	56	Right Middle Frontal Gyrus	6
4	4.98	-38	14	-4	Left Insula	13
4	4.86	-30	20	-8	Left Insula	13
4	3.79	-40	12	-16	Left Superior Temporal Gyrus	13
5	4.98	46	0	34	Right Precentral Gyrus	6
5	4.98	42	-2	32	Right Precentral Gyrus	6
5	4.82	52	2	36	Right Precentral Gyrus	6
5	4.13	56	8	42	Right Precentral Gyrus	6
5	3.80	36	10	18	Right Inferior Frontal Gyrus	44
5	3.77	40	8	22	Right Inferior Frontal Gyrus	44
6	6.11	0	-30	20	Right Posterior Cingulate Gyrus	23
6	5.20	4	-38	18	Right Posterior Cingulate Gyrus	30
6	4.15	-4	-6	22	Left Thalamus	50
6	4.12	-4	-16	20	Left Thalamus	50
6	3.69	6	-38	24	Right Posterior Cingulate Gyrus	23
6	3.20	-12	-2	20	Left Caudate	48
7	5.25	62	-32	-14	Right Inferior Temporal Gyrus	21
7	4.64	70	-34	-10	Right Middle Temporal Gyrus	21
7	4.26	48	-34	-12	Right Middle Temporal Gyrus	21
7	3.73	58	-40	-10	Right Inferior Temporal Gyrus	21
8	4.92	38	18	-10	Right Insula	13
8	4.89	30	16	-12	Right Insula	13
8	4.32	48	12	-8	Right Insula	13
8	4.29	32	22	-6	Right Insula	13
8	3.95	52	16	0	Right Inferior Frontal Gyrus	44
8	3.85	38	12	-2	Right Insula	13
9	4.98	-40	30	26	Left Inferior Frontal Gyrus	9
9	4.81	-32	28	20	Left Inferior Frontal Gyrus	9
9	3.94	-36	40	22	Left Middle Frontal Gyrus	10
9	3.90	-36	44	24	Left Middle Frontal Gyrus	10
9	3.83	-30	50	28	Left Middle Frontal Gyrus	10
10	8.53	-22	-64	28	Left Superior Occipital Gyrus	7
10	8.32	-20	-72	38	Left Superior Occipital Gyrus	7
10	8.18	-16	-74	56	Left Superior Parietal Lobule	7
10	8.05	-20	-64	42	Left Superior Parietal Lobule	7

10	7.74	-4	-56	62	Left Precuneus	7
10	7.59	18	-60	46	Right Precuneus	7

Table 11: Experiment 8 fMRI local maxima in the statistical *Z*-maps of T > NT response (top) and item (bottom) contrasts as a function of increasing c_2 or d_a (see also **Figure 27**). Negative *Z*-values represent areas that become more active in T > NT contrasts as c_2 or d_a decreases.

Discussion

Experiments 7 and 8 expand upon group-average fMRI findings of recognition memory T > NT response contrasts in that maintaining both conservative *and* liberal decision criteria engaged regions within the frontoparietal, ventral attention, dorsal attention, and visual networks when inhibiting prepotent responses. Previous studies have only observed this finding when participants maintain a conservative, but not a liberal criterion (Aminoff et al. 2015; King & Miller 2017), though Experiment 6 found some right frontal regions to be more active in the NT > T response contrast when participants maintained a liberal criterion. Findings from Experiment 8 suggest that more extreme biases may elicit frontoparietal activity to a greater extent, which is consistent with group-level findings from Aminoff and colleagues (2015) who found that the magnitude of frontoparietal activity in the H > CR contrast when comparing between conservative and liberal criterion conditions strongly depended on the extent of criterion shifting. One possible reason for why previous studies failed to observe a T > NT response contrast effect when participants maintained a liberal criterion could be that subjects on average may not have shifted to an extreme enough extent in the liberal conditions. Additionally, the dense-sampling experiments included many trials, which may have provided sufficient statistical power to detect such an effect.

Another novel finding from the dense-sampling experiments is that *providing* a prepotent response elicits DMN activity to a greater extent relative to *inhibiting* the prepotent response both when participants maintain a conservative and liberal criterion. DMN activity increases as people disengage from external tasks (Raichle et al., 2001), which is consistent with a response bias account for the T > NT response contrast since making prepotent responses should require less cognitive control and attention. DMN activity in Experiment 8, when the participant *provided* prepotent responses, appeared to be heightened when decisions biases were more extreme. It is possible that group-level fMRI findings do not include individuals who shift criteria extreme enough on average to observe DMN activity, or group-level experiments to date may just be underpowered to detect this finding.

A major goal of the dense-sampling experiments sought to better characterize regions sensitive to changes in memory strength in the T > NT response contrast, given that findings from Experiment 6 showed no differences in activity between low and moderate amounts of discriminability during recognition memory tests. Findings from Experiment 7 only showed a handful of parietal regions associated with increased levels of discriminability in the T > NT response contrast. This prompted modifications to the recognition memory paradigm in Experiment 8 to greatly increase the statistical power for detecting an effect related to changes in discriminability, if such an effect exists. Despite doubling the number of sessions and increasing the number of test trials, only spatially sparse regions within parietal and temporal cortices showed greater activity as d_a increased or decreased in the T > NT response contrast. The parietal regions observed in Experiment 8 did not overlap with those of Experiment 7, though the stimuli differed between the two experiments (faces in Experiment 7 vs. scenes in Experiment 8). The findings from Experiment 6-8 overwhelming suggest that fMRI is very insensitive for detecting activity related to changes in discriminability. Thus, T > NT response contrasts during recognition memory tests appear to be predominantly dependent on an individual’s decision criterion. This is not to suggest that memory strength plays *no* role in frontoparietal activity when making target versus nontarget responses, but rather other methods need to be implemented to observe neural effects related to discriminability, such as through patient studies. Theories of memory strength that draw conclusions from fMRI findings must consider an individual's criterion in order to dissociate neural activity related to decisional evidence versus the criterion.

Chapter III: Failed attempts to modulate decision criteria via neurostimulation

Findings from Experiments 6-8 and other fMRI studies of recognition memory (Aminoff et al., 2015; King & Miller 2017), overwhelming show that frontoparietal activity in T > NT response contrasts is dependent on an individual's decision criterion. However, fMRI findings are limited to correlations assessments of associating brain activity with behavioral performance. In order to confidently conclude that the observed frontoparietal activity is *necessary* for maintaining conservative versus liberal criteria requires an approach that causally manipulates brain activity and subsequent behavior. Despite the importance of establishing a decision criterion based on memory evidence, little is known about the causal neural mechanisms that underlie them (Gold & Shadlen, 2007; Ratcliff et al., 2016). Research in healthy individuals supports the notion that maintaining a conservative criterion during recognition memory requires engagement of the PFC. Aminoff and colleagues (2015) sought to manipulate criterion placement as participants performed recognition memory tests during fMRI scanning. An investigation of the H > CR contrast, yielded robust recruitment of widespread frontoparietal regions when participants maintained a conservative criterion—but not when maintaining a liberal criterion. These findings directly oppose theories that attribute increased BOLD activity in the H > CR contrast to differences in memory strength since hit trials on average carry more memory strength relative to correct rejection trials (Wheeler & Buckner, 2003; Kahn et al., 2004; Wagner et al., 2005; Vilberg & Rugg, 2009; Yu et al., 2012; Criss et al., 2013). However, the H > CR contrast also carries information about memory-based decisional processes because hit responses represent a *decision* that the memory strength of an item exceeds the established criterion whereas correctly rejected items do not carry enough memory strength to surpass the decision threshold (O'Conner et al., 2010; Jaeger et al., 2013; Miller & Dobbins, 2014). Through an individual differences analysis, Aminoff and colleagues (2015) revealed that participants who shifted decision criteria to a greater extent also showed greater frontoparietal activity in the H > CR contrast specifically when participants maintained a conservative decision criteria. No such relationship existed when comparing frontoparietal activity with individual differences in memory strength. This finding provides compelling evidence that the observed frontoparietal activity in the H > CR contrast is not only associated with the maintenance of a conservative criterion, but the magnitude of the frontoparietal activity correlates with the conservativeness of a decision criterion.

One potential explanation for the robust activity in the H > CR contrast when a conservative criterion is maintained is that suppressing a prepotent familiarity response may require cognitive control processes related to response inhibition (see Aminoff et al., 2015). Specifically, there is strong evidence indicating that the right inferior frontal gyrus (rIFG) is implicated in response inhibition (Wager et al, 2005; Chambers et al., 2009; Bari & Robbins, 2013) and may serve as a cognitive braking system (Aron et al., 2014, 2015). Other prefrontal areas, such as the dorsolateral prefrontal cortex (DLPFC), may also play a role in maintaining task goals to prepare for inhibiting a response (Jahfari et al., 2010; Swann et al., 2012). If maintaining a conservative criterion requires preparing for or executing response inhibition, then the rIFG and surrounding prefrontal areas provide promising sites for further investigation.

Functional MRI studies are of course limited in their ability to draw *causal* inferences between brain activity and behavior. However, the advent of neurostimulation techniques, such as repetitive transcranial magnetic stimulation (rTMS), offers a direct means of testing whether overt behavior can be altered by targeted cortical stimulation. Previous rTMS research demonstrated that offline continuous theta burst stimulation (cTBS) serves as an effective inhibitor of cortical excitability for up to 60 minutes in the hand area of the motor cortex (Huang, et al., 2005). Although it is unclear whether offline cTBS has equivalent inhibitory effects when applied to areas within the PFC (Grossheinrich et al., 2009), a handful of studies have successfully manipulated cognitive performance by applying offline cTBS to prefrontal regions. For example, Verbruggen and colleagues (2010) disrupted response inhibition and dual-task performance after applying cTBS to the rIFG. Georgiev and colleagues (2016) applied cTBS over the rDLPFC, which led to slower response times during a perceptual decision-making task. Additionally, Cho and colleagues (2010) reduced impulsivity in a delayed discounting task after applying cTBS to the rDLPFC. These studies provide evidence that offline cTBS can affect decision-making performance in a seemingly inhibitory manner.

Since cTBS appears to inhibit PFC excitability, an attempt to causally manipulate criterion placement occurred by applying cTBS to brain areas that Aminoff and colleagues (2015) identified as being associated with the magnitude of the conservativeness of a decision criterion, namely the rIFG, rMFG, and rDLPFC. The prediction was that cTBS to the rIFG, rMFG, and rDLPFC would inhibit the function of networks implicated in criterion placement without affecting recognition memory. More specifically, individuals were predicted to establish less conservative criteria when a conservative criterion is advantageous. In situations where a liberal criterion is advantageous, no changes in criterion placement were expected. This finding would suggest that the rIFG, rMFG, and rDLPFC play a crucial role in maintaining conservative decision criteria but are non-essential for maintaining liberal decision criteria during recognition memory. Importantly, this approach can provide more concrete evidence to support previous observations of increased frontoparietal activity during successful retrieval—but only when a conservative criterion is maintained (e.g. Aminoff et al., 2015)—and help explain why individuals with damaged and/or dysfunctional prefrontal cortices generally set liberal decision criteria relative to healthy controls.

Experiments 9 & 10: TMS of prefrontal cortex fails to modulate decision criteria

Method

Participants

Prior to the cTBS experiment, three hundred and fifty-two participants (126 males; ages 18-38, *M* = 20, *SD* = 2.5) conducted an initial prescreen task. Participants received an invitation to partake in the neuroimaging and neurostimulation phases of the experiment if they discriminated between old and new images, sufficiently shifted criteria between the conservative and liberal conditions, and met all eligibility requirements for MRI and TMS (see ***Procedure***). Twenty participants did not receive an invitation due to below chance discriminability performance. An additional one hundred and fifty participants did not receive an invitation because they did not adequately shift between conservative and liberal decision criteria. The one hundred and eighty-two eligible participants received an invitation to participate in the study on a rolling basis and enrollment occurred based on participants who replied quickest to the invitation.

Ultimately, a total of thirty-six participants (9 males; ages 18-26, *M* = 20, *SD* = 1.7) successfully completed all three cTBS sessions between Experiments 9 and 10. Experiment 9 consisted of 20 participants (5 males; ages 18-23, *M* = 20, *SD* = 1.6) with the exclusion of four additional participants due to computer malfunction (1), procedural error (1), or incomplete stimulation (2) during at least one of the three cTBS sessions. After observing a surprising trend (see **Results**) a follow-up experiment was conducted (Experiment 10) with 16 participants (4 males; ages 19-26, *M* = 21, *SD* = 1.8). Two additional participants withdrew from Experiment 10. Participants received $10/hour for performing the prescreen task and $20/hour for conducting the MRI and cTBS sessions.

Procedure

Experiments 9 and 10 consisted of a prescreen recognition memory task, an fMRI scanning session, and three cTBS sessions. The prescreen recognition memory task identified participants that discriminated between studied and unstudied face stimuli and adaptively shifted between conservative and liberal decision criteria. The task intentionally made discriminability difficult to ensure that the average memory strength of hit trials almost equals the average memory strength of correct rejection trials. This way the H > CR contrast subtracts out virtually all BOLD activity associated with memory strength while leaving behind activity associated with the decision criterion; hit trials represent correct decisions that the memory strength of items exceed the criterion while correct rejection trials serve as correct decisions that novel items lack the necessary amount of memory strength. However, above chance discriminability performance was required to ensure participants correctly conducted the recognition memory task.

In addition to performing above chance on the recognition memory task, participants also needed to adaptively shift their decision criteria. There are vast individual differences in the placement of a decision criterion during a recognition memory test (Aminoff et al., 2012; Kantner & Lindsay, 2012, 2014; Kantner et al., 2015; Frithsen et al., 2018; Layher et al., 2020) and it is necessary to identify when participants establish a *relatively* more conservative decision criterion. Additionally, Aminoff and colleagues (2015) found that individuals who failed to shift their decision criteria did not exhibit robust frontoparietal activity in the H > CR even in situations where maintaining a conservative criterion is advantageous. To test whether cTBS disrupts maintaining a *relatively* more conservative decision criteria and to ensure attainment of robust fMRI activation in the H > CR contrast for precise individualized cTBS targeting, individuals were only invited to participate in the study if they adaptively shifted criteria during the initial prescreen recognition memory tests. Once selected, participants performed recognition memory tests while maintaining a conservative decision criterion during fMRI scanning. The fMRI analyses provided subject-specific cTBS target sites based on each participant's peak voxel activity in the H > CR contrast within the rIFG (Experiments 9 and 10), rMFG (Experiment 9), and rDLPFC (Experiment 10). Finally, participants conducted recognition memory tests both before and after cTBS on three separate visits.

Recognition memory task

The recognition memory task followed the same procedure for the initial prescreen, fMRI, and cTBS phases of the experiment unless otherwise specified (**Figure 28**). During the study session, participants passively viewed a series of 100 novel face images displayed in the center of a computer screen with a black background. Each study image appeared for 300 ms followed by a 200 ms blank screen interstimulus interval. The study procedure intentionally induced low discrimination levels in order to make criterion shifting more advantageous. Every participant viewed a random series of face images which did not repeat across sessions.

After the study session, participants completed a test session that consisted of two test blocks in which participants made “old” (previously studied) or “new” (unstudied) recognition judgments. Prior to each test block, explicit instructions informed participants of the base rate probabilities of encountering an old test item. In the low probability (conservative criterion) condition, only 30% of the test items appeared during the study phase making it advantageous to respond “new” more often. In the high probability (liberal criterion) condition, 70% of test items appeared during the study phase making “old” responses more advantageous. Each block consisted of 100 test trials with 30 old images and 70 new images appearing in the conservative criterion condition and 70 old images and 30 new images appearing in the liberal criterion condition. Every test image appeared in the center of a computer screen surrounded by an orange or blue frame to help participants remember the test condition (conservative or liberal criterion). The images remained on screen until the participant made a response. During each trial, instructions appeared at the bottom of the screen to indicate whether the “0” or “1” keyboard button represented an “old” or “new” response. Each participant received a completely random assignment for the order of the conservative and liberal test blocks, the frame color assigned to each test condition, and the mapping of “old” and “new” response keys. In total, the recognition memory task spanned 10 to 15 minutes.

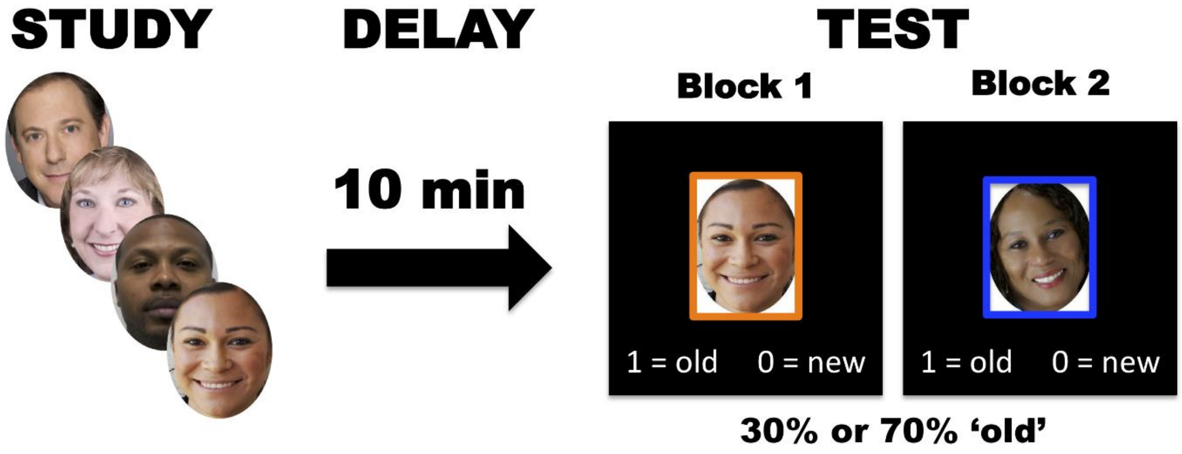


Figure 28: Experiments 9 and 10 recognition memory tasks. Participants studied 100 face images, followed by liberal and conservative test blocks with a base rate manipulation (either 30% or 70% of images, respectively, were “old”). A 10-minute delay ensued after completion of the study phase.

Deriving subject-specific cTBS targets

Participants selected from the prescreen initially conducted a modified version of the recognition memory task during fMRI scanning. In this version, participants studied 60 face images and performed two conservative testing blocks—both to precisely identify regions supporting conservative criterion placement, and because the H > CR contrast does not reveal robust activity when a liberal criterion is set (Aminoff et al., 2015). Each test block contained 100 test trials with 30 old and 70 new images that appeared for 3 seconds with random jitter to ensure separability of HRFs (interstimulus intervals of 0–6 s). Participants responded via a two-button response box held in the right hand.

A Siemens 3T PRISMA MRI scanner collected all imaging data using a 64 channel head and neck coil. An initial MPRAGE sequence acquired T_1-weighted anatomical images (208 slices; TE = 2.22 ms; TR = 2500ms; FoV = 241 mm^2; voxel size: 0.9 mm^3). A subsequent T_2*-weighted gradient recall echo (GRE) field map scan (48 oblique slices; TE 1 = 4.92 ms; TE 2 = 7.38 ms; FoV = 192 mm^2; voxel size: 3 mm^3) provided estimates of magnetic field inhomogeneities. Functional image acquisition employed a T_2*-weighted multi-band echo planar imaging (mbEPI) sequence sensitive to the BOLD contrast (48 oblique slices; TE = 35 ms; TR = 400 ms; FoV = 192 mm^2; voxel size: 3 mm^3; multiband factor = 8). Total scanning time lasted approximately 30 minutes.

All fMRI preprocessing and statistical analyses occurred using FSL, v5.0. Each functional scan underwent motion correction and realignment to the middle volume using FSL's MCFLIRT. FSL's FUGUE unwarped geometric deformations due to motion and field inhomogeneities. Temporal preprocessing of voxelwise timeseries included both high pass filtering (0.01 Hz) and prewhitening. The data underwent spatial smoothing using a 5 mm^3 FWHM Gaussian kernel. Coregistration of functional data to each individual's T_1-weighted anatomical image enabled cTBS target identification in subject space.

An event-related GLM identified within-subject activity related to successful retrieval. Each test block contained 4 regressors of interest: hits, correct rejections, misses, and false alarms. Nuisance regressors included trials with no old/new response in addition to head motion parameters derived from MCFLIRT realignment. FSL's default FLOBS provided model estimates to compute H > CR contrasts for each individual.

The H > CR contrast provided subject-specific cTBS target sites based on the peak voxel within each ROI. The Harvard-Oxford Cortical Structural atlas in FSL anatomically defined rIFG and rMFG ROIs, using probability maps with a threshold of 30% for the combined right pars opercularis and pars triangularis maps (rIFG) and the rMFG map. For Experiment 9, the location of the peak voxel within the H > CR contrast was encompassed by the large anatomical ROIs. In Experiment 10, the rDLPFC ROI was functionally defined from the Aminoff and colleagues (2015) group-level H > CR contrast (see **Results**), specifically for the conservative criterion condition of the recognition memory test for faces. FSL FLIRT registered each ROI to a participant's native brain space. To ensure replication of the fMRI findings from Aminoff and colleagues (2015), w a group-level mixed-effects analysis of variance was performed for the H > CR contrast. The resulting Z statistic maps underwent whole-brain voxelwise thresholding at $Z > 3.1$ and cluster correction ($p < .05$) using Gaussian random field theory, which determined statistical significance.

cTBS

Participants attended three cTBS sessions each separated by at least 48 hours to ensure that the effects of stimulation from one session did not carry over to another session. The location of the target site differed for each of the three cTBS sessions. In the first experiment participants received cTBS to the rIFG, rMFG, or occipital vertex (sham stimulation). The second experiment followed the same procedures except participants received cTBS to the rDLPFC instead of the rMFG (see **Results**). The order of stimulation over the three sessions occurred pseudo-randomly across participants to include all six possible order combinations. The cTBS stimulation intensity remained fixed at 35% of the maximum stimulator output because rIFG stimulation inadvertently contracts the temporalis muscle and the chosen intensity level minimized discomfort. During cTBS to the rMFG and rDLPFC, a researcher held the TMS coil handle at a 45° angle relative to the head's midline with the handle pointing posteriorly and to the right. The TMS coil handle pointed posteriorly while aligned parallel the head's midline during cTBS to the rIFG and occipital vertex. Participants unknowingly received sham stimulation that involved a slight tilting of the TMS coil away from the scalp to mitigate cTBS effects on the occipital vertex.

To precisely stimulate the functionally defined cTBS target sites, participants wore a headband with an infrared tracking device and earplugs to protect against hearing loss from the ambient TMS noise. A pointer tool registered the position of the tracking device in the participant's headband to the participant's T_1-weighted anatomical image using a Polaris infrared optical tracking system (Northern Digital Inc., Waterloo, ON, Canada) in conjunction with the Brainsight TMS navigation system, version 2.3.9 (Rogue Research, Montreal, QC, Canada). This allowed for real-time tracking of the position of the TMS coil relative to the cTBS target sites within each participant's brain. Researchers used the high-definition system to identify the location of the ROI for that session and applied cTBS on the scalp over the target site. A 70 mm figure of eight coil delivered cTBS in bursts of 50 Hz triplets at a rate of 5 Hz for 40 seconds (600 total pulses) using a Magstim Rapid^2 stimulator unit (Magstim Inc., Morrisville, NC).

During each of the three cTBS sessions, participants performed the recognition memory task twice, once before cTBS and again afterwards. Following the first study phase, a 10-minute delay ensued where participants sat in a chair while researchers provided information about the cTBS procedure for that session. Participants then completed the first test phase and immediately began the study phase of the second run. After the second study phase, researchers applied cTBS during another 10-minute delay period. Afterwards, participants performed the second recognition memory test phase (**Figure 29**).

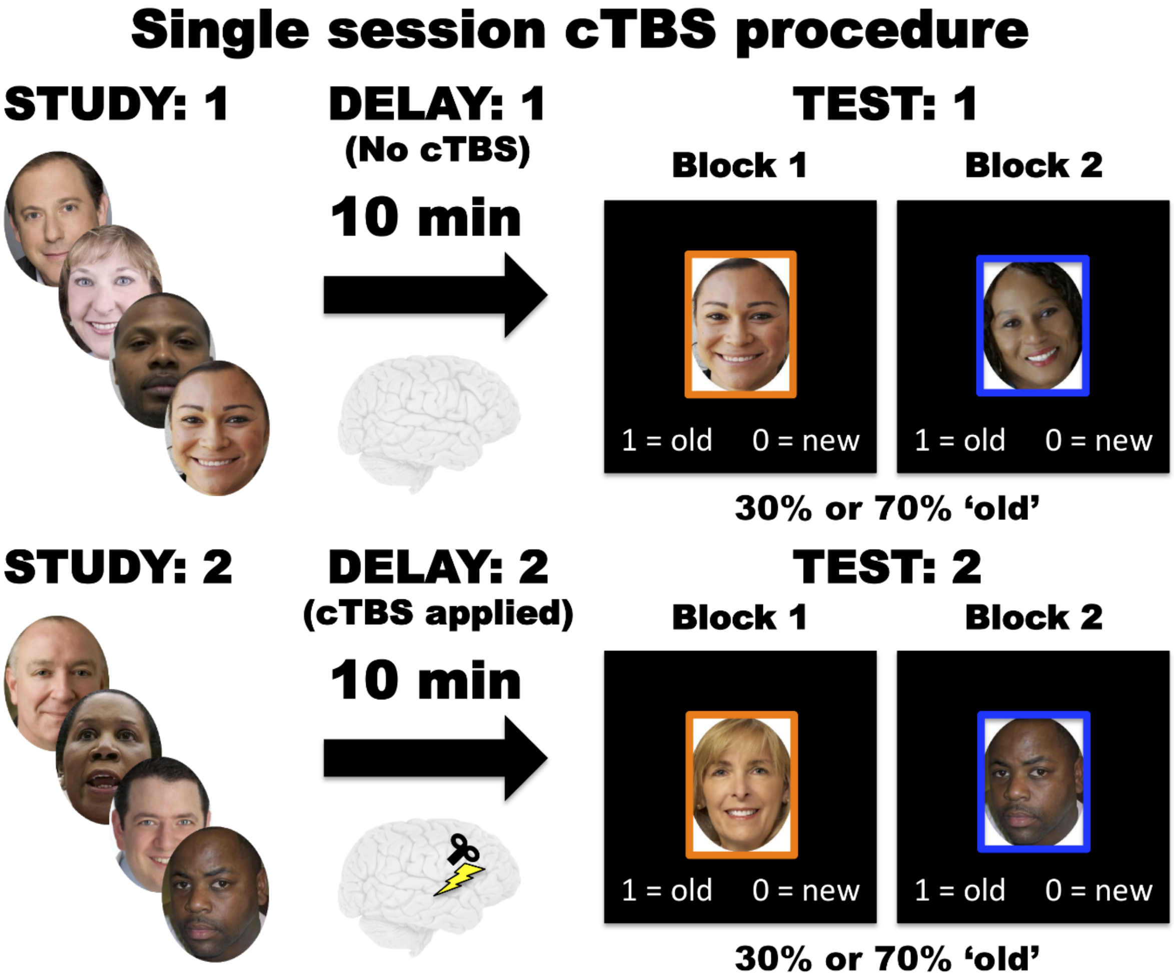


Figure 29: Experiments 9 and 10 cTBS session procedure. During each of the three cTBS sessions, participants initially conducted the recognition memory task without cTBS stimulation. Then participants performed the recognition memory task again with cTBS applied during the delay period.

Statistical analysis

The effects of cTBS stimulation on criterion placement and discriminability were tested using linear mixed models, which modeled mean differences in *d'* and c_n as functions of criterion condition (conservative vs. liberal [CON > LIB]), task time (post-cTBS vs. pre-cTBS [POST > PRE]), and cTBS target site (rIFG vs. sham [rIFG > Sham] and rMFG vs.

sham [rMFG > Sham]). Additionally, three-way interactions were modeled between criterion, time, and cTBS target contrasts, along with all marginal two-way interactions. Specification of a random effect on the model intercept accounted for baseline variation in c_n and d' values across subjects. Thus, the fixed effects models took the following form:

$$\hat{y} = b_0 + b_1(CON > LIB) + b_2(POST > PRE) + b_3(rIFG > Sham) + b_4(rMFG > Sham) + b_5(CON > LIB * POST > PRE) + b_6(CON > LIB * rIFG > Sham) + b_7(CON > LIB * rMFG > Sham) + b_8(POST > PRE * rIFG > Sham) + b_9(POST > PRE * rMFG > Sham) + b_{10}(CON > LIB * POST > PRE * \ rIFG > Sham) + b_{11}(CON > LIB * POST > PRE * rMFG > Sham) + \varepsilon.$$

The Experiment 10 models remained identical in form with the substitution of rDLPFC for rMFG.

Results

The successful retrieval effect

Group-level (*N* = 36) whole-brain fMRI analyses of the H > CR contrast yielded significant differential activity across widespread frontoparietal cortices (**Figure 30**). These results are consistent with the results of Aminoff and colleagues (2015) for participants who maintained a conservative decision criterion during recognition memory of faces (**Table 12**). The individualized cTBS target sites derived from subject-level fMRI analyses of the H > CR contrast are depicted in **Figure 31**.

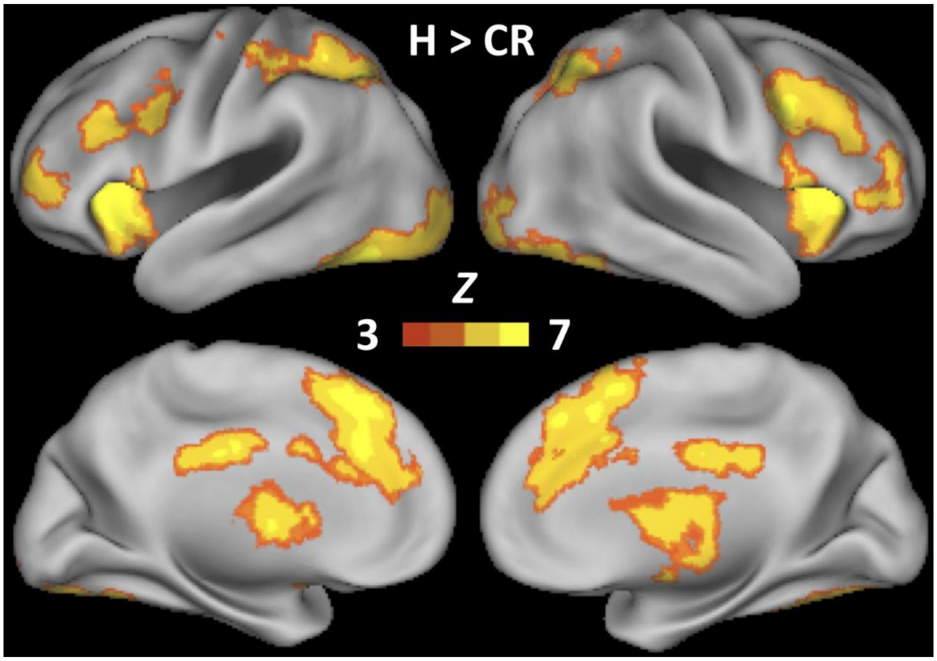


Figure 30: Experiments 9 and 10 combined whole-brain statistical *Z*-map of the H > CR contrast, estimated over two recognition memory tests requiring participants to maintain a conservative decision criterion ($N = 36$). Thresholding at $Z > 3.1$ and cluster correction ($p < .05$) determined significance (see also **Table 12**).

Experiments 9 and 10: fMRI local maxima covariate H > CR contrast						
Cluster	***Z*-value**	**X**	**Y**	**Z**	**Location**	**BA**
1	5.32	-40	-66	-20	Left Cerebellum	19
1	5.12	40	-56	-28	Right Cerebellum	37
1	5	36	-68	-16	Right Fusiform Gyrus	19
1	4.99	-36	-70	-20	Left Cerebellum	19
1	4.93	38	-54	-18	Right Fusiform Gyrus	37
1	4.92	34	-64	-24	Right Cerebellum	19
2	6.11	0	36	40	Left Medial Superior Frontal Gyrus	9
2	5.79	-4	26	32	Left Anterior Cingulate Cortex	24
2	5.69	0	30	46	Left Medial Superior Frontal Gyrus	8
2	5.33	-2	30	38	Left Medial Superior Frontal Gyrus	32
2	5.3	6	14	54	Right Supplementary Motor Area	6
2	5.27	-4	44	18	Left Anterior Cingulate Gyrus	32
3	6.3	32	20	6	Right Insula	48
3	5.32	30	30	2	Right Insula	47
3	5.14	42	10	28	Right Inferior Frontal Gyrus	44
3	4.96	44	12	-6	Right Insula	48
3	4.78	50	10	42	Right Precentral Gyrus	48
3	4.69	28	22	-10	Right Insula	47

4	5.4	-8	12	6	Left Caudate	25
4	5.02	-10	-2	20	Left Caudate	-
4	4.95	12	4	12	Right Caudate	-
4	4.89	-8	-2	8	Left Thalamus	-
4	4.69	10	14	12	Right Caudate	-
4	4.68	-10	4	14	Left Caudate	-
5	5.17	-28	-64	44	Left Inferior Parietal Lobule	7
5	5.09	-28	-52	42	Left Inferior Parietal Lobule	7
5	4.68	-36	-48	52	Left Inferior Parietal Lobule	40
5	4.52	-42	-40	52	Left Inferior Parietal Lobule	40
5	4.42	-42	-48	52	Left Inferior Parietal Lobule	40
5	4.39	-34	-56	52	Left Inferior Parietal Lobule	7
6	6.24	-34	16	4	Left Insula	48
6	5.85	-36	18	-2	Left Insula	47
6	5.84	-32	20	-2	Left Insula	47
6	5.56	-28	26	4	Left Insula	47
6	4.93	-38	12	-6	Left Insula	48
6	4.64	-48	16	-4	Left Inferior Frontal Gyrus	48
7	5.79	34	-62	52	Right Superior Parietal Lobule	7
7	5.27	32	-64	38	Right Middle Occipital Gyrus	7
7	3.6	30	-72	32	Right Middle Occipital Gyrus	19
7	3.3	42	-44	54	Right Inferior Parietal Lobule	40
7	3.27	46	-44	56	Right Inferior Parietal Lobule	40
8	4.94	-48	22	32	Left Inferior Frontal Gyrus	44
8	4.45	-50	8	36	Left Precentral Gyrus	44
8	4.33	-42	24	24	Left Inferior Frontal Gyrus	48
8	3.92	-42	2	32	Left Precentral Gyrus	6
8	3.88	-44	8	28	Left Inferior Frontal Gyrus	44
8	3.82	-44	4	28	Left Precentral Gyrus	44
9	4.98	2	-32	28	Left Posterior Cingulate Gyrus	23
9	4.82	-4	-24	30	Left Midcingulate Area	23
9	4.7	-2	-18	32	Left Midcingulate Area	23
9	4.63	4	-16	32	Right Midcingulate Area	23
10	4.52	-46	46	2	Left Inferior Frontal Gyrus	45
10	4.29	-42	52	6	Left Middle Frontal Gyrus	46
10	4.29	-38	54	10	Left Middle Frontal Gyrus	46
10	4.29	-36	52	6	Left Middle Frontal Gyrus	10
10	4.09	-36	50	2	Left Middle Frontal Gyrus	47
10	4.08	-36	46	0	Left Middle Frontal Gyrus	47

Table 12: Experiments 9 and 10 combined fMRI local maxima in the statistical *Z*-map of the H > CR contrast when participants (*N* = 36) maintained a conservative criterion (see also **Figure 30**).

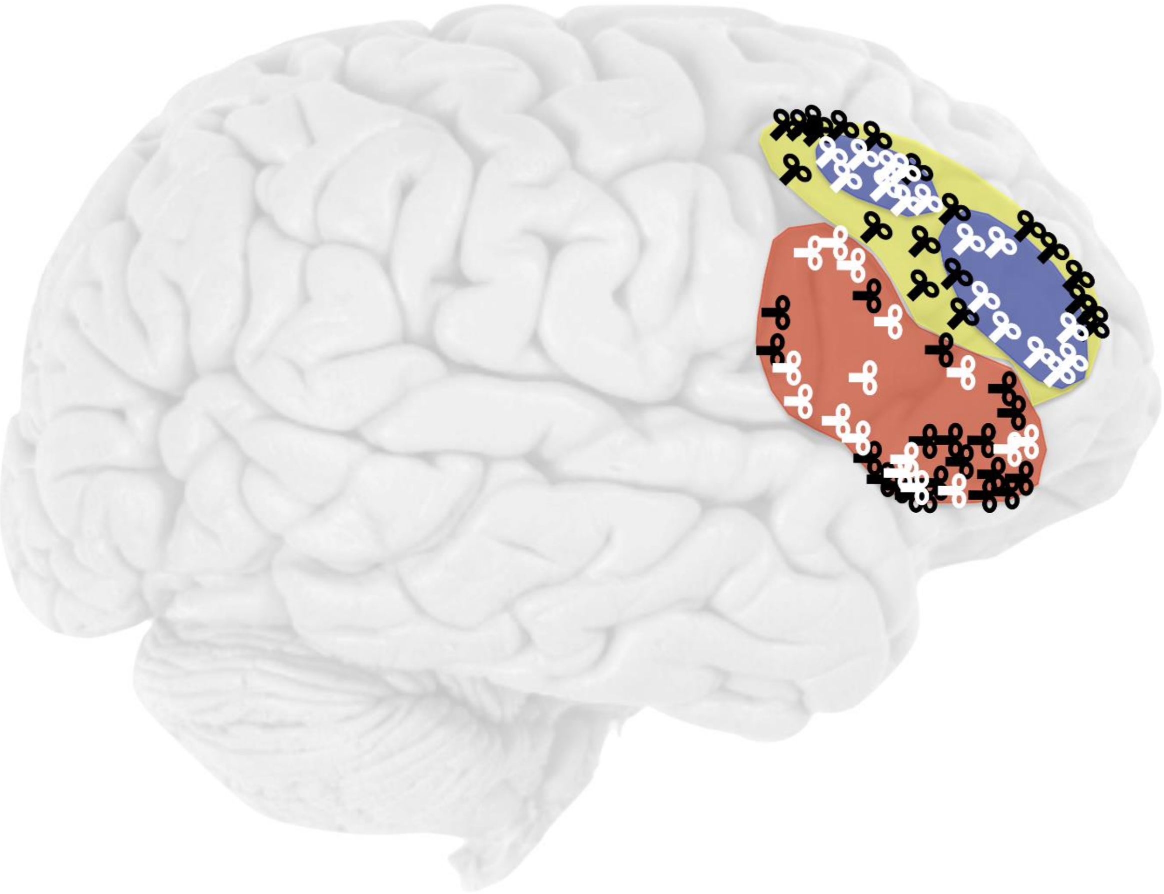


Figure 31: Location of cTBS in the rIFG (red), rMFG (yellow), and rDLPFC (blue). The coils represent the subject-specific target sites for the 20 participants in Experiment 9 (black) and the 16 participants in Experiment 10 (white).

cTBS effects on discriminability and criterion placement

Average behavioral performance during the pre-cTBS memory tests in Experiment 9 revealed participants successfully shifted their decision criteria in response to the conservative (c_n = 0.81, *SD* = 0.32) and liberal (c_n = 0.13, *SD* = 0.53) probability manipulation ($p < .001$, d = 1.84) (**Figure 32**; left). Experiment 10 revealed similar results in the conservative (c_n = 0.64, *SD* = 0.30) and liberal (c_n = 0.01, *SD* = 0.36) criterion conditions

($p < .001$, $d = 3.50$) (**Figure 32**; right). Although participants on average maintained a slightly conservative bias in the liberal condition, the important distinction is that participants shifted to a *relatively* more liberal criterion. Mean discriminability remained low in Experiment 9 ($d' = 0.36$, $SD = 0.37$) and Experiment 10 ($d' = 0.36$, $SD = 0.35$) for the pre-cTBS memory tests making it more strategic to shift decision criteria (**Figure 33**).

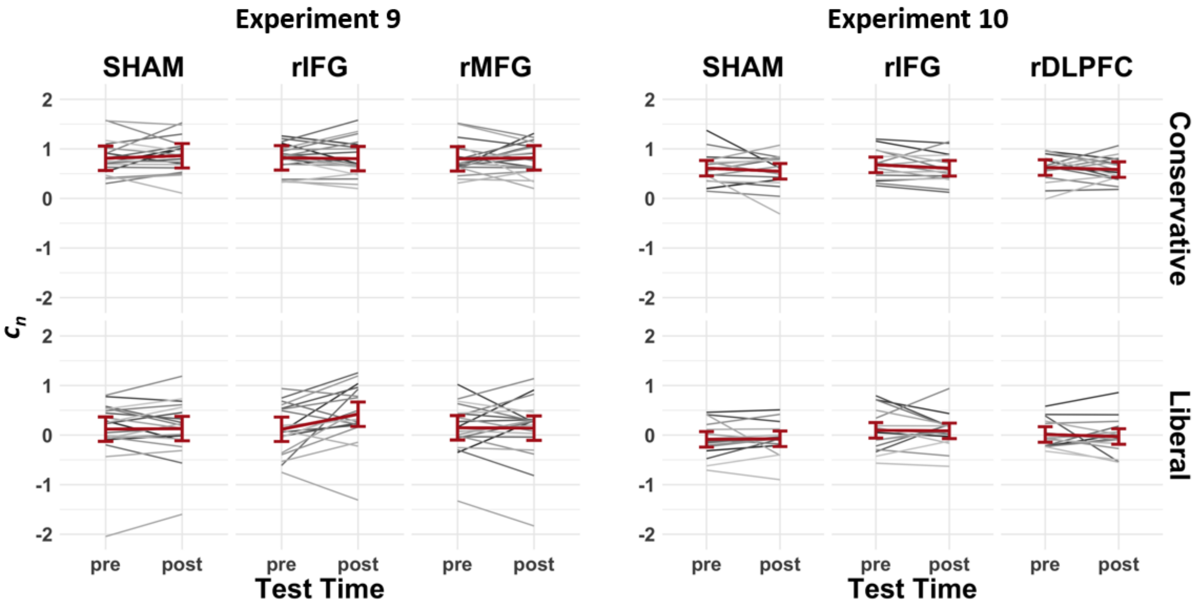


Figure 32: The pre-/post-cTBS c_n values for Experiment 9 (left) and 10 (right). Gray lines indicate individual subject performance and red lines represent group averages fitted with 95% confidence intervals.

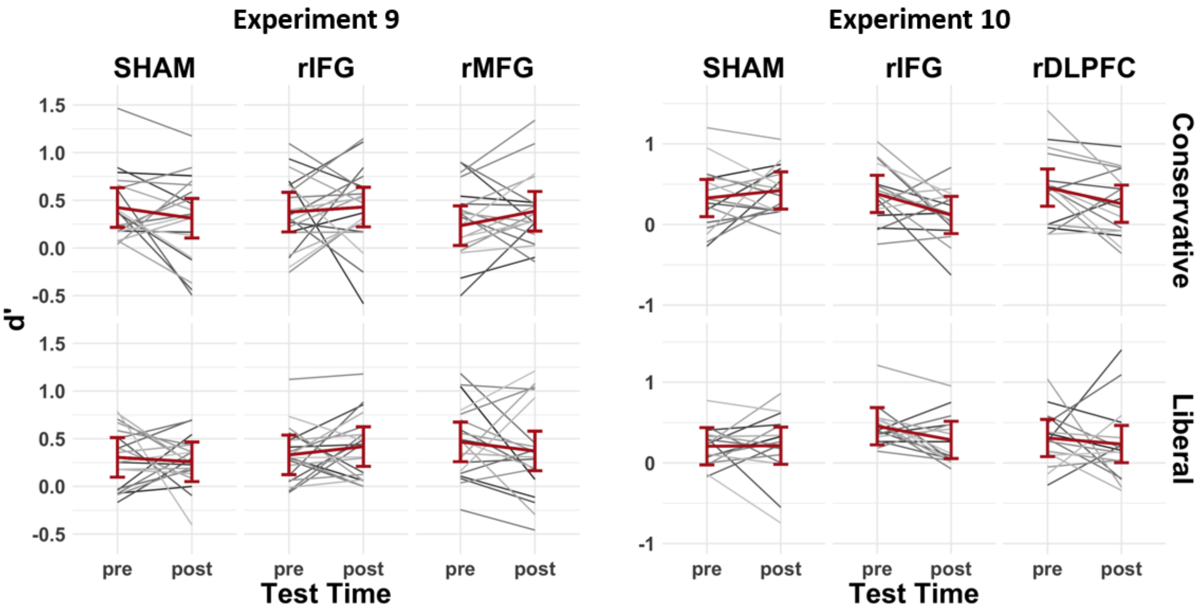


Figure 33: The pre-/post-cTBS *d'* values for Experiment 9 (left) and 10 (right). Gray lines indicate individual subject performance and red lines represent group averages fitted with 95% confidence intervals.

Experiment 9

As predicted, applying cTBS to regions previously associated with criterion placement did not affect *d'*. The criterion manipulation also did not affect discriminability nor did performing the task pre- versus post-cTBS. **Figure 34** (left) displays parameter estimates and mean discriminability across factor levels for *d'* fitted with 95% confidence intervals; **Table 13** (top) contains a summary of all model-level statistics.

With respect to c_n, a significant main effect of criterion condition was observed, such that participants set a more liberal decision criterion when target probability remained high (b = 0.64, 95CI = [0.54, 0.74], SE = 0.05, t = 12.53, d = 1.37). Contrary to expectation, however, cTBS failed to affect one's criterion placement. Rather than *decreasing* the conservativeness of decision criteria, a marginal trend towards *more stringent* decision criteria was observed following rIFG stimulation, as revealed by an interaction between criterion condition and cTBS to the rIFG, relative to sham (b = -0.10, 95CI = [-0.24, 0.04], *SE* = 0.07, *t* = -1.35, *d* = 0.21). A similar trend existed in the three-way interaction between the criterion manipulation, pre-/post-cTBS tests, and stimulation of the rIFG target (*b* = -0.12, 95CI [-0.26, 0.02], *SE* = 0.07, *t* = -1.63, *d* = 0.25). Summaries of model-level statistics and mean criterion placement are shown in **Figure 35** (left) and **Table 13** (top). This intriguing trend inspired further investigation to test whether the observed difference is truly a null result or merely an underpowered effect.

Experiment 10

Since cTBS to the rMFG proved completely ineffective at affecting decision criteria, the anatomically defined rMFG ROI was switched to a functionally defined rDLPFC ROI for subsequent data collection, in case the broad rMFG ROI encompassed brain areas unrelated to maintaining a conservative criterion. Analyses on an additional 16 participants also revealed no significant interactions between task time, cTBS target site (rIFG > sham, rDLPFC > sham), and criterion condition. Results from Experiment 10 suggest that the trending interaction of pre-/post-cTBS, criterion condition, and rIFG stimulation (relative to sham) in Experiment 9 is likely a true null result (*b* = 0.02, 95CI = [-0.07, 0.11], *SE* = 0.05, *t* = 0.49, *d* = 0.07). **Figure 35** (right) displays parameter estimates and mean criterion across factor levels for c_n fitted with 95% confidence intervals; **Table 13** (bottom) contains a summary of all model-level statistics.

Although no differences in discriminability was predicted, a significant interaction emerged between cTBS target site and criterion condition (**figure 34**, right; **table 14**, top). Relative to sham, cTBS to the rDLPFC improved *d'* performance—specifically in the liberal condition (*b* = -0.17, 95CI = [-0.30, 0.03], *SE* = 0.07, *t* = -2.44, *d* = 0.45). Although this is merely a two-way interaction (i.e. is agnostic to pre-/post-stimulation differences), it is nevertheless a moderately strong effect, and it raises the intriguing possibility that changing the target site from the rMFG to a more localized rDLPFC region directly affected a recognition memory network.

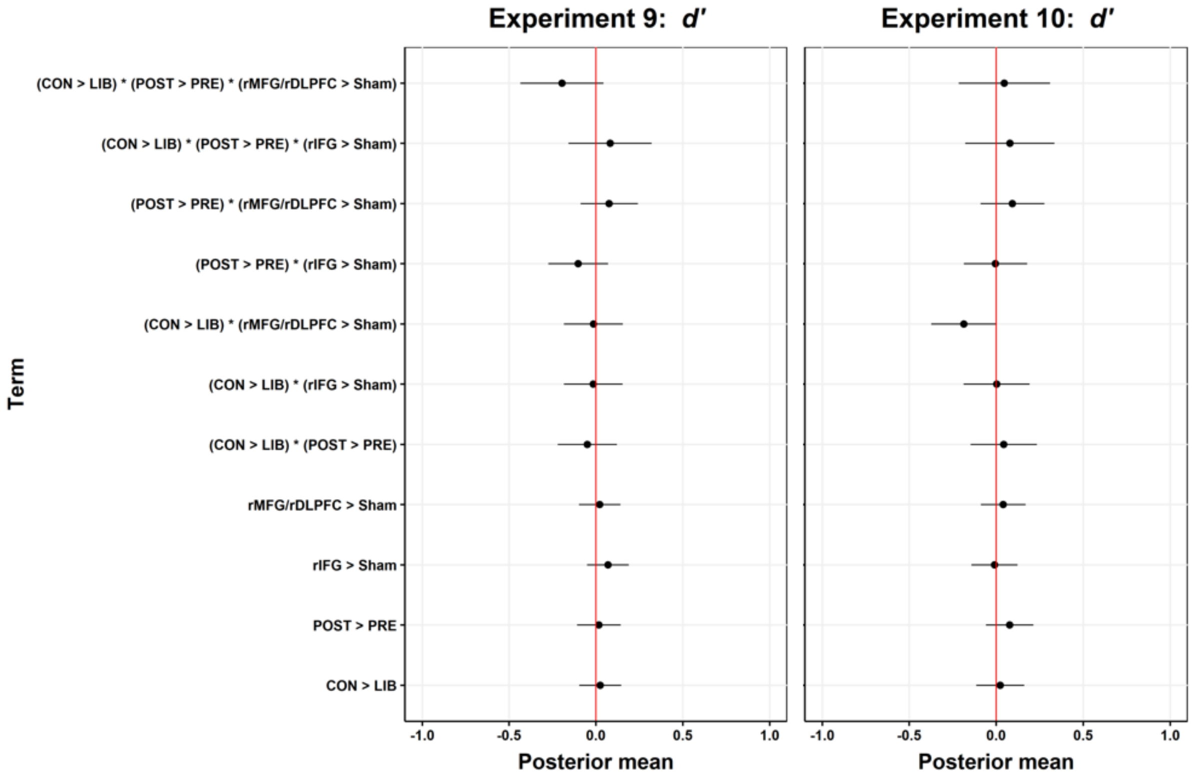


Figure 34: Experiments 9 (left) and 10 (right) posterior mean of parameter estimates across fixed effects for *d'* models, fitted with 95% confidence intervals. Estimates not intersecting zero are statistically significant (see also **Table 13**).

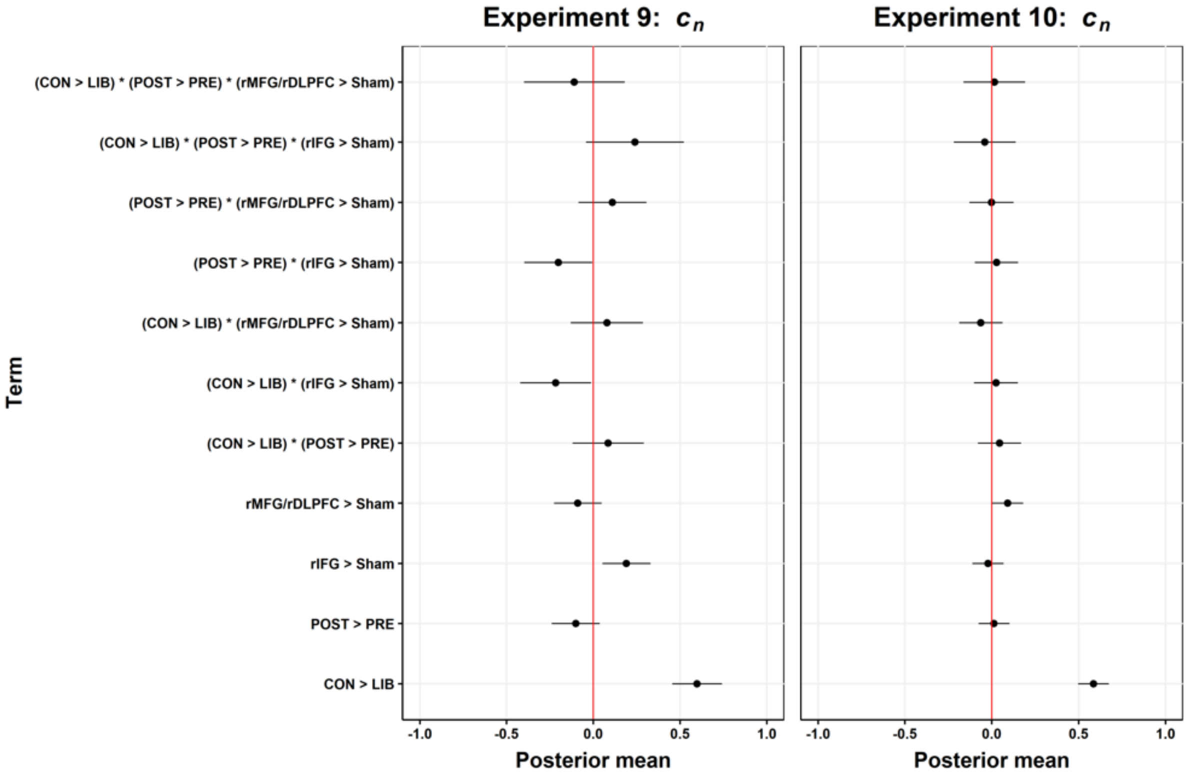


Figure 35: Experiments 9 (left) and 10 (right) posterior mean of parameter estimates across fixed effects for c_n models, fitted with 95% confidence intervals. Estimates not intersecting zero are statistically significant (see also **Table 14**).

Experiment 9 Model-Level Statistics: *d'*

Term	**Estimate (95CI)**	***SE***	***t***	**Effect size (*d*)**
(Intercept)	0.36 (0.26, 0.46)	0.05	6.95	0.94
CON > LIB	0.00 (-0.08, 0.08)	0.04	0.04	0.01
POST > PRE	0.02 (-0.05, 0.08)	0.03	0.51	0.04
rIFG > Sham	0.04 (-0.04, 0.13)	0.04	0.99	0.11
rMFG > Sham	-0.05 (-0.14, 0.03)	0.04	-1.18	0.13
(CON > LIB) * (POST > PRE)	0.02 (-0.06, 0.11)	0.04	0.57	0.06
(CON > LIB) * (rIFG > Sham)	0.03	0.06	0.44	0.07

	(-0.09, 0.15)			
(CON > LIB) * (rMFG > Sham)	-0.11 (-0.24, 0.02)	0.06	-1.84	0.29
(POST > PRE) * (rIFG > Sham)	0.01 (-0.07, 0.10)	0.04	0.26	0.03
(POST > PRE) * (rMFG > Sham)	0.06 (-0.02, 0.14)	0.04	1.40	0.16
(CON > LIB) * (POST > PRE) * (rIFG > Sham)	-0.04 (-0.16, 0.08)	0.06	-0.68	0.11
(CON > LIB) * (POST > PRE) * (rMFG > Sham)	0.10 (0.01, 0.21)	0.06	1.63	0.26
Random Effect: (Intercept | Subject)				
Subjects	20			
Intercept (*SD*)	0.19			
N	240			

Experiment 10 Model-Level Statistics: *d'*

Term	**Estimate (95CI)**	***SE***	***t***	**Effect size (*d*)**
(Intercept)	0.33 (0.23, 0.42)	0.05	6.49	0.90
CON > LIB	0.04 (-0.05, 0.14)	0.05	0.90	0.12
POST > PRE	-0.06 (-0.13, 0.01)	0.03	-1.81	0.17
rIFG > Sham	0.03 (-0.06, 0.13)	0.05	0.64	0.08
rDLPFC > Sham	-0.08 (-0.17, 0.02)	0.05	-1.64	0.22
(CON > LIB) * (POST > PRE)	-0.02 (-0.12, 0.08)	0.05	-0.45	0.06
(CON > LIB) * (rIFG > Sham)	0.04 (-0.09, 0.18)	0.07	0.64	0.12
(CON > LIB) * (rDLPFC > Sham)	-0.17 (-0.30, -0.03)	0.07	-2.44	0.45
(POST > PRE) * (rIFG > Sham)	-0.04 (-0.13, 0.06)	0.05	-0.81	0.11
(POST > PRE) * (rDLPFC > Sham)	-0.07 (-0.16, 0.03)	0.05	-1.46	0.19
	-0.04	0.07	-0.60	0.11

(CON > LIB) * (POST > PRE) * (rIFG > Sham)	(-0.17, 0.09)			
(CON > LIB) * (POST > PRE) * (rDLPFC > Sham)	-0.03 (-0.16, 0.11)	0.07	-0.36	0.07
Random Effect: (Intercept | Subject)				
Subjects	16			
Intercept (*SD*)	0.15			
N	192			

Table 13: Experiments 9 (top) and 10 (bottom) model-level statistics for *d’* values across the conservative (CON) and liberal (LIB) conditions both before (PRE) and after (POST) cTBS stimulation to the rIFG, rDLPFC, rMFG, or occipital vertex (Sham). See also **Figure 34**.

Experiment 9 Model-Level Statistics: c_n

Term	**Estimate (95CI)**	***SE***	***t***	**Effect size (*d*)**
(Intercept)	0.82 (0.69, 0.95)	0.07	12.41	1.76
CON > LIB	0.64 (0.54, 0.74)	0.05	12.53	1.37
POST > PRE	0.01 (-0.06, 0.08)	0.04	0.25	0.02
rIFG > Sham	-0.01 (-0.11, 0.09)	0.05	-0.15	0.02
rMFG > Sham	-0.01 (-0.11, 0.09)	0.05	-0.19	0.02
(CON > LIB) * (POST > PRE)	-0.04 (-0.14, 0.05)	0.05	-0.83	0.09
(CON > LIB) * (rIFG > Sham)	-0.10 (-0.24, 0.05)	0.07	-1.35	0.21
(CON > LIB) * (rMFG > Sham)	0.03 (-0.12, 0.17)	0.07	0.37	0.06
(POST > PRE) * (rIFG > Sham)	-0.02 (-0.12, 0.08)	0.05	-0.34	0.04
(POST > PRE) * (rMFG > Sham)	0.00 (-0.10, 0.10)	0.05	0.02	0.00
(CON > LIB) * (POST > PRE) * (rIFG > Sham)	-0.12 (-0.25, 0.01)	0.07	-1.63	0.25
(CON > LIB) * (POST > PRE) * (rMFG > Sham)	0.06 (-0.08, 0.19)	0.07	0.78	0.12

Random Effect: (Intercept | Subject)

Subjects	20
Intercept (*SD*)	0.25
N	240

Experiment 10 Model-Level Statistics: c_n

Term	Estimate (95CI)	*SE*	*t*	Effect size (*d*)
(Intercept)	0.61 (0.48, 0.74)	0.06	9.49	1.85
CON > LIB	0.61 (0.54, 0.67)	0.03	18.78	1.85
POST > PRE	-0.03 (-0.07, 0.02)	0.02	-1.25	0.09
rIFG > Sham	-0.01 (-0.07, 0.06)	0.03	-0.16	0.02
rDLPFC > Sham	0.03 (-0.03, 0.10)	0.03	1.07	0.11
(CON > LIB) * (POST > PRE)	-0.02 (-0.08, 0.04)	0.03	-0.68	0.07
(CON > LIB) * (rIFG > Sham)	0.00 (-0.08, 0.09)	0.05	0.08	0.01
(CON > LIB) * (rDLPFC > Sham)	-0.06 (-0.15, 0.03)	0.05	-1.22	0.17
(POST > PRE) * (rIFG > Sham)	0.01 (-0.06, 0.07)	0.03	0.26	0.03
(POST > PRE) * (rDLPFC > Sham)	-0.01 (-0.07, 0.06)	0.03	-0.19	0.02
(CON > LIB) * (POST > PRE) * (rIFG > Sham)	0.02 (-0.07, 0.11)	0.05	0.49	0.07
(CON > LIB) * (POST > PRE) * (rDLPFC > Sham)	-0.01 (-0.10, 0.08)	0.05	-0.16	0.02

Random Effect: (Intercept | Subject)

Subjects	16
Intercept (*SD*)	0.24
N	192

Table 14: Experiments 9 (top) and 10 (bottom) model-level statistics for c_n values across the conservative (CON) and liberal (LIB) conditions both before (PRE) and after (POST) cTBS stimulation to the rIFG, rDLPFC, rMFG, or occipital vertex (Sham). See also **Figure 35**.

Discussion

These experiments attempted to further illuminate neural mechanisms underlying the maintenance of a conservative decision criterion during recognition memory. Patients with damaged and/or dysfunction frontal lobes oftentimes establish overly liberal decision criteria when making recognition judgments (Biesbroek, et al., 2014; Deason et al., 2017). In healthy individuals, widespread fronto-parietal BOLD activity is present in the H > CR contrast of recognition memory tests when people maintain a conservative decision criterion, but not a liberal criterion (Aminoff et al., 2015; King & Miller 2017). These findings suggest that a conservative criterion may require an intact and functional PFC. Experiments 9 and 10 investigated whether regions involved in response inhibition, such as the rIFG (Aron et al., 2014), mediate a conservative criterion by suppressing a person's tendency to classify familiar items as old. The goal was to test whether cTBS to the rIFG, rMFG, and rDLPFC (regions where increased BOLD activity tracks with the conservativeness of a decision criterion, Aminoff et al., 2015) causes participants to establish less conservative decision criteria during recognition memory. Participants initially conducted a recognition memory test while maintaining a conservative decision criterion during fMRI scanning. Despite obtaining subject-specific target sites based on peak BOLD activity in the H > CR contrast and using high-definition TMS equipment, cTBS to the rIFG, rMFG, and rDLPFC did not significantly affect criterion placement.

There are several possible reasons why cTBS did not cause participants to establish less conservative decision criteria during recognition memory tests. First, the hypothesis that the rIFG, rMFG, and rDLPFC are *necessary* for maintaining a conservative decision criterion could simply be incorrect. Many studies investigating the neural substrates of criterion placement are correlational; thus, there might not be a direct causal relationship between the targeted regions and the maintenance of a conservative decision criterion. However, it would be difficult to reconcile this conclusion with results from studies showing that frontal lobe damage is commonly associated with more liberal decision criteria (but see Verfaellie, et al., 2004; Hwang et al., 2007). Another possibility is that the appropriate regions were not targeted within the rIFG, rMFG, and rDLPFC. Subject-specific target sites were obtained via the H > CR contrast from a relatively short fMRI scanning session, which may not have provided precise target sites of neural hubs that drive the maintenance of a conservative decision criterion. Future studies should consider obtaining subject-specific target sites via functional connectivity analyses from longer scanning sessions. There are also inherent technical difficulties with TMS that may explain the null findings, even *if* the correct brain regions were targeted. Sandrini and colleagues (2011) outline several technical considerations that affect the efficacy of TMS, including the type of stimulation protocol, intensity of stimulation, and the orientation of the coil handle.

Even if the most robust TMS protocol and coil positioning technique is employed, individual differences in anatomy and cortical excitability may cause wide variability in behavioral changes across participants. For instance, the efficacy of cTBS on inhibiting in the primary motor cortex is quite variable (Suppa et al., 2016) despite being one of the few brain regions that give a measurable output via motor evoked potentials. The frontal cortex is also highly interconnected, which may contribute to more inter-individual variability in the effects of cTBS. For instance, Lee and D’Esposito (2012) observed that individuals with greater functional connectivity between the left and right IFG tended to have less of a decrement in working memory performance following cTBS to the left IFG. The authors suggested that the right IFG might play a compensatory role that reduces the behavioral detriments caused by left IFG inhibition. It is possible that inhibiting a small region within the right PFC is easily compensated for since maintaining a conservative criterion may involve a widespread bilateral fronto-parietal network (Aminoff et al., 2015; King & Miller 2017). Lastly, the participants conducted the recognition memory task several times including twice before performing the first cTBS session. It is possible that performing the task multiple times allowed participants to develop efficient strategies that make it more difficult to disrupt behavior with cTBS. Due to the vast possibilities for null findings in a cTBS experiment, it is inappropriate to conclude that the rIFG, rMFG, and rDFLPC are *unnecessary* for maintaining a conservative decision criterion during recognition memory. Rather it's the cTBS technique that failed to affect decision criterion (whether or not the targeted regions are indeed implicated in maintaining a decision criterion) and should not be employed for future investigations of the neural mechanisms underlying decision criteria.

Surprisingly, a significant interaction occurred where discriminability improved, specifically when maintaining a liberal criterion following cTBS to the rDLPFC. This finding is difficult to interpret because changes in discriminability should be unaffected by criterion placement since they are behaviorally-independent processes (Macmillan & Creelman, 2005). Thus, an improvement in d' should be observed in *both* the conservative and liberal criterion conditions if cTBS indeed manipulated the discriminability of items at test. Due to the small sample size ($N = 16$) and unpredicted nature of the finding, future investigations are necessary to ensure this result is not a false positive. However, if this finding is upheld with future research, then it adds more complexity to the argument of whether the H > CR contrast is driven more strongly by memory strength or criterion setting.

Although this study failed to manipulate criterion placement with cTBS, other TMS methods might prove more successful. Future research studies should employ online TMS protocols to ensure target sites are being stimulated while participants perform recognition memory tasks that involve maintaining conservative and liberal decision criteria. An offline cTBS approach was implemented in hopes of conducting a future study where participants perform recognition tests during fMRI scanning following cTBS. However, the null results suggest that cTBS to the rIFG, rMFG, and rDLPFC are ineffective at manipulating criterion placement during a recognition memory test.

Conclusion

Offline cTBS to the rIFG, rMFG, and rDLPFC proved ineffective at altering decision criteria during recognition memory tests. This is not to suggest that these frontal regions are uninvolved in the maintenance of a conservative decision criterion or that TMS generally cannot affect criterion placement. However, it is not recommended to use offline cTBS to manipulate decision criteria during recognition memory. An unexpected finding of improved *d’* performance when maintaining a liberal criterion following cTBS to the rDLPFC could motivate future research investigating prefrontal neural networks involved in recognition memory.

Experiment 11: Diffuse tDCS across the insula fails to modulate decision criteria

Although widespread frontoparietal activity in recognition memory T > NT contrasts is greatly affected by criterion placement (Aminoff et al. 2015; King & Miller, 2017; Experiments 6-10), subject-specific TMS targeting of right PFC regions in Experiments 9 and 10 could not reliably alter criterion placement. Given the robust and widespread nature of frontoparietal activity within recognition memory T > NT contrasts, it is possible that modulating a focal portion of this large network is ineffective at altering decision-making behavior. Additionally, the *offline* cTBS protocol might be less effective than an online neurostimulation approach since an offline protocol requires long-lasting effects, which decay over time (Huang, et al., 2005). Transcranial direct current stimulation (tDCS) provides an alternative neurostimulation method that can affect neural activity across broad cortical regions and can be more easily administered as an online protocol. While TMS directly induces action potentials via strong magnetic pulses, tDCS alters the activation threshold of neurons via an electric current. Depending on the direction of the current, tDCS can make neurons more or less likely to fire, which mimics an excitatory or inhibitory effect, respectively (Reinhart et al., 2016). To attempt to more broadly stimulate the right PFC to alter decision criteria during recognition memory, an online diffuse tDCS experiment sought to alter activity in criterion-sensitive regions. Participants conducted recognition memory tests with criterion manipulations both before and during tDCS across three separate sessions. Each session differed based on the type of tDCS implemented, either anodal (increases neuronal excitability), cathodal (decreases neuronal excitability), or sham stimulation. It is predicted that anodal stimulation will induce a more conservative criterion (due to increased PFC excitability) whereas cathodal stimulation will make individuals maintain a more liberal criterion (due to PFC inhibition), relative to sham stimulation.

Method

Participants

Thirty participants (10 males; ages 18-33, *M* = 20, *SD* = 3.0) successfully completed all three sessions of the tDCS experiment. An additional four participants withdrew from the study before completing the three sessions and are excluded from all analyses. Participants earned $20/hour for participating.

Procedure

Unlike Experiments 9 and 10, the tDCS experiment did *not* include a prescreen task to exclude individuals who do not reliably shift decision criteria during recognition memory. Since diffuse tDCS does not require subject-specific targeting, participants do not need to undergo fMRI scanning and thus there is no concern that frontoparietal activity in T > NT response contrasts would be unobservable in individuals who do not shift criteria. Additionally, it is possible that decision-making strategies in individuals who do not shift criteria are more influenced by neurostimulation relative to individuals who regularly shift criteria, which may have contributed to the null findings in Experiments 9 and 10. Theoretically, manipulation of neural networks involved in criterion placement should affect performance regardless if participants strategically shift criteria or not.

Recognition memory task

Participants conducted two recognition memory tasks during each of the three sessions, once before tDCS and another during stimulation. The study phase consisted of a series of 100 face stimuli each displayed for 300 ms, followed by a 200 ms blank screen presentation. The quick study presentation time intended to induce low discriminability, making it more advantageous to shift criteria at test, and the entire study block lasted for less than a minute. After a two-minute delay, the participants conducted two recognition memory test blocks. Participants received explicit instructions about the likelihood of encountering an old image and were instructed to use that information to guide their decisions. In conservative test blocks, the likelihood of encountering an old face was 30%, whereas an old face appeared 70% of the time in the liberal condition. Each test block consisted of 100 images (30 old and 70 new, or vice versa) and stimuli remained on screen until the participant made a response. The order of the conservative and liberal test block following each study phase was pseudo-randomized across participants in which odd numbered participant always conducted the conservative test block first whereas even numbered participants completed the liberal test block first. The order of the test blocks remained fixed within each subject across sessions in order to keep the design as similar as possible when making behavioral comparisons across stimulation type.

tDCS

Prior to conducting the experiment, participants filled out a tDCS screening form to identify potential contraindications that may increase the likelihood of an adverse effect due to stimulation (e.g. history of seizures, neurological disorders, or metal/electrical implants). Any participant who indicated a potential contraindication could not participate in the experiment. Participants attended three tDCS sessions each separated by at least 48 hours to ensure that the effects of stimulation from one session did not carry over to another session. Neurostimulation occurred via a Soterix Medical High-Definition (HD) tDCS system. Stimulation parameters were designed to maximally target the right anterior insula, a criterion-sensitive region (Aminoff et al., 2015; King & Miller 2017; Experiments 6-10), which also requires stimulation of a large portion of right PFC in a relatively diffuse manner (**Figure 36**). The stimulation montage was derived from the Soterix Medical program "HD-Targets," which computes the best configuration and stimulation parameters to maximally stimulate an ROI on a standardized brain template. Participants wore an HD-cap that includes 68 holes where stimulating electrodes could potentially be placed. The configuration consisted of six-electrodes where current flowed from four electrodes (Fpz, FCz, C4, and F10) that provided 0.5 mA of current each, to two other electrodes that supplied 1.0 mA of current each (AF8 and F6), or vice versa. This created a total current flow of 2 mA, which is considered a safe, yet effective, amount of current to alter neuronal excitability (Bikson et al., 2016).

At the start of each session, participants were fitted with an HD-cap that held six plastic wells at locations AF8, C4, F6, F10, FCz, and Fpz. The positioning of the HD-cap involved centering location Fpz at 10% of the distance from the participant's nasion to inion. Researchers applied alcohol swabs to clean the scalp within the six wells and part the hair to expose the scalp, then filled each well with conducting gel. Afterwards, the tDCS electrodes were placed in the cap and swabbing occurred until the impedance of each electrode fell below 25 k (a measure of impedance termed "quality units" by Soterix Medical). Participants then conducted the first recognition memory task *without* stimulation. Prior to conducting the second recognition memory task, participants verbally completed a mood questionnaire to assess changes in mental state before and after stimulation.

After the participant completed the study phase of the second memory task, a two-minute break ensued in which the researcher began a 30 second ramp up of current until it reached 2 mA. Once the participant completed both self-paced test blocks, the researcher ramped the current down within 30 seconds. The participant then verbally completed the mood questionnaire a second time to ensure no major changes in mental state occurred as a result of the tDCS. During sham stimulation, the current ramped up to 2 mA within 30 seconds then slowly dropped to 0.1 mA before the test phase began and remained there for the duration of the test. The same procedure followed during sham stimulation sessions at the end of the test phase to mimic the tactile sensation felt as a result of the changing current. Pseudo-randomization determined the session order in which participants conducted the anodal, cathodal, and sham stimulations.

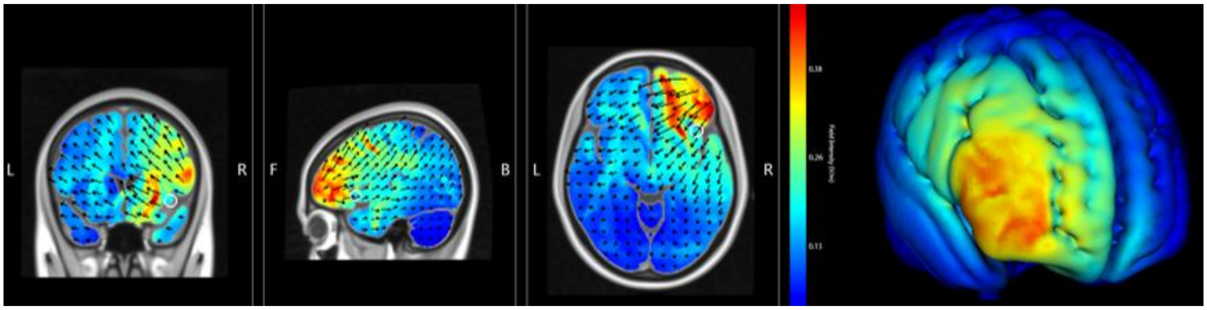


Figure 36: Experiment 11 current density map of the online diffuse tDCS protocol based on parameters that maximally target right anterior insula.

Statistical analysis

The effects of online tDCS on criterion placement and discriminability were tested using linear mixed models, which modeled mean differences in *d'* and c_n as functions of criterion condition (conservative vs. liberal [CON > LIB]), task time (tDCS stimulation vs. pre-tDCS [STIM > PRE]), and tDCS stimulation type (anodal vs. sham [A > Sham] and cathodal vs. sham [C > Sham]). Additionally, three-way interactions were modeled between criterion, time, and cTBS target contrasts, along with all marginal two-way interactions. Specification of a random effect on the model intercept accounted for baseline variation in c_n and *d*' values across subjects. The fixed effects models took the following form:

$$\hat{y} = b_0 + b_1(CON > LIB) + b_2(STIM > PRE) + b_3(A > SHAM) + b_4(C > SHAM) + b_5(CON > LIB * STIM > PRE) + b_6(CON > LIB * A > SHAM) + b_7(CON > LIB * C > SHAM) + b_8(STIM > PRE * A > SHAM) + b_9(STIM > PRE * C > SHAM) + b_{10}(CON > LIB * STIM > PRE * \ A > SHAM) + b_{11}(CON > LIB * STIM > PRE * C > SHAM) + \varepsilon.$$

Results

Average behavioral performance during the pre-tDCS memory tests revealed participants successfully shifted decision criteria in response to the conservative (c_n = 0.72, *SD* = 0.30) and liberal (c_n = -0.08, *SD* = 0.35) probability manipulations ($p < .001$, d = 2.00) (**Figure 37**; top). Mean discriminability remained low (*d'* = 0.26, *SD* = 0.33) making it more strategic to shift decision criteria (**Figure 37**; bottom).

As predicted, applying online tDCS to the right PFC during recognition memory tests did not affect *d'*. The criterion manipulation also did not affect discriminability nor did performing the task before or during tDCS. **Figure 38** (bottom) displays parameter estimates and mean discriminability across factor levels for *d'* fitted with 95% confidence intervals; **Table 15** (bottom) contains a summary of all model-level statistics. With respect to c_n, a significant main effect of criterion condition was observed, such that participants set a more liberal decision criterion when target probability remained high (b = 0.37, 95CI = [0.34, 0.41], *SE* = 0.02, t = 22.34, d = 0.99). Contrary to the hypotheses, online tDCS failed to affect criterion placement under any condition. Summaries of model-level statistics and mean criterion placement are shown in **Figure 38** (top) and **Table 15** (top).

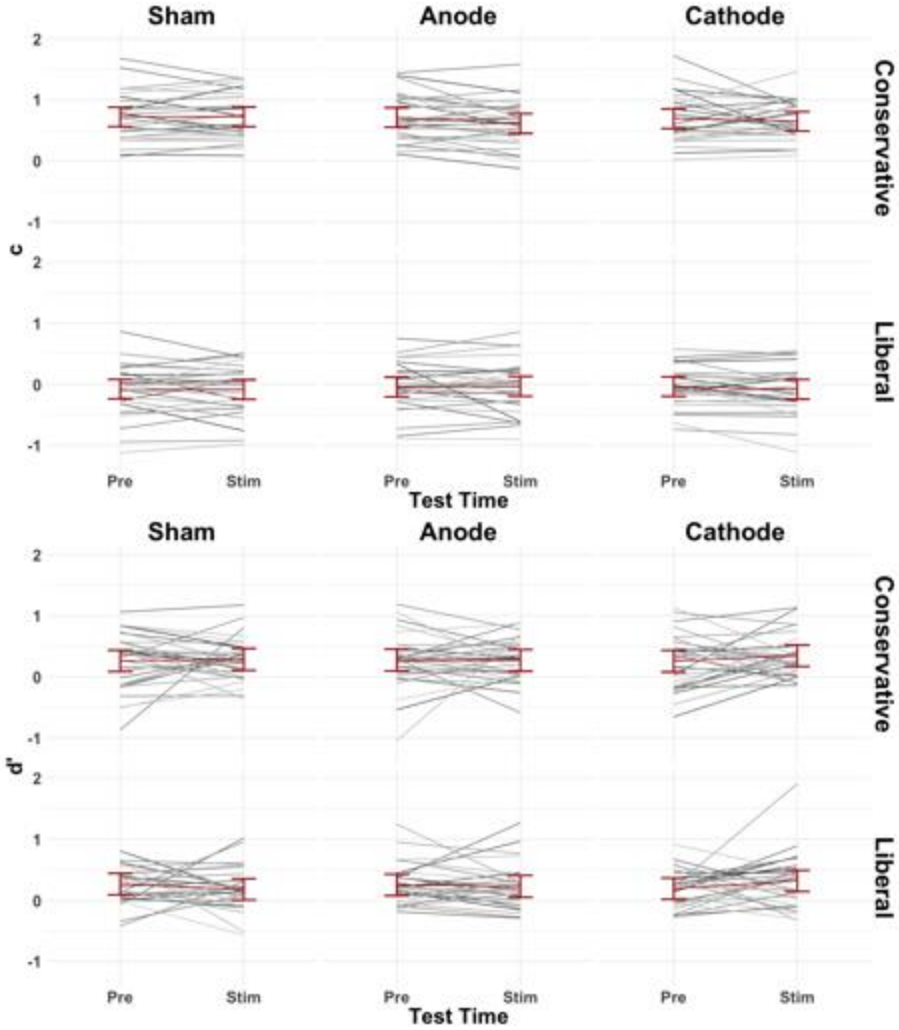


Figure 37: Experiment 11 c_n (top) and *d'* (bottom) values before and during tDCS. Gray lines indicate individual subject performance and red lines represent group averages fitted with 95% confidence intervals.

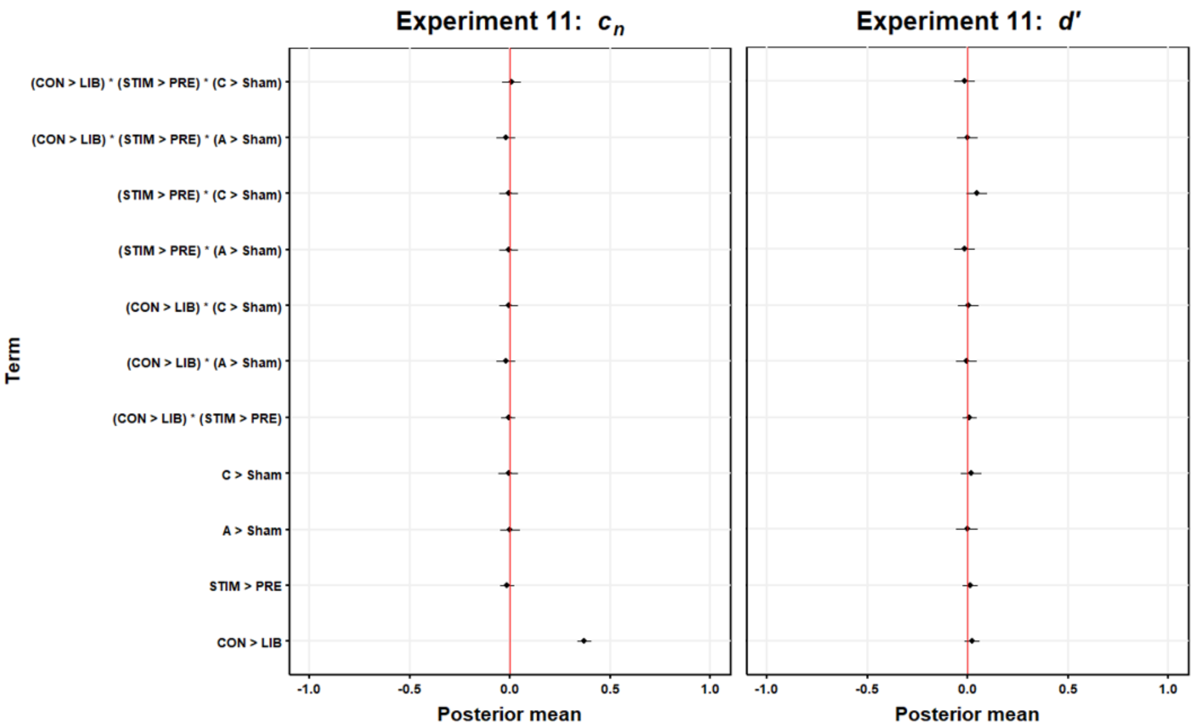


Figure 38: Experiments 11 posterior mean of parameter estimates across fixed effects models of c_n (top) and *d'* (bottom), fitted with 95% confidence intervals. Estimates not intersecting zero are statistically significant (see also **Table 15**).

Experiment 11 Model-Level Statistics: c_n				
Term	**Estimate (95 CI)**	***SE***	***t***	**Effect Size (*d*)**
Intercept	0.31 (0.23, 0.40)	0.04	7.68	0.83
CON > LIB	0.37 (0.34, 0.41)	0.02	22.34	0.99
STIM > PRE	-0.02 (-0.05, 0.02)	0.02	-0.92	0.04
Anodal > Sham	0.00 (-0.05, 0.05)	0.02	0.00	0.00
Cathodal > Sham	-0.01 (-0.05, 0.04)	0.02	-0.33	0.02
(CON > LIB) * (STIM > PRE)	-0.01 (-0.04, 0.02)	0.02	-0.50	0.02
(CON > LIB) * (Anodal > Sham)	-0.02 (-0.07, 0.03)	0.02	-0.96	0.06
(CON > LIB) * (Cathodal > Sham)	-0.01	0.02	-0.31	0.02

	(-0.05, 0.04)			
(STIM > PRE) * (Anodal > Sham)	-0.01 (-0.06, 0.04)	0.02	-0.30	0.02
(STIM > PRE) * (Cathodal > Sham)	-0.01 (-0.05, 0.04)	0.02	-0.28	0.02
(CON > LIB) * (STIM > PRE) * (Anodal > Sham)	-0.02 (-0.07, 0.03)	0.02	-0.84	0.05
(CON > LIB) * (STIM > PRE) * (Cathodal > Sham)	0.01 (-0.04, 0.05)	0.02	0.36	0.02

Random Effect: (Intercept | Subject)

Subjects	30
Intercept (*SD*)	0.20
N	360

Experiment 11 Model-Level Statistics: *d'*

Term	**Estimate (95 CI)**	***SE***	***t***	**Effect Size (*d*)**
Intercept	0.26 (0.19, 0.33)	0.03	8.00	0.69
CON > LIB	0.02 (-0.02, 0.06)	0.02	1.19	0.06
STIM > PRE	0.01 (-0.03, 0.05)	0.02	0.55	0.03
Anodal > Sham	-0.01 (-0.06, 0.05)	0.03	-0.19	0.01
Cathodal > Sham	0.02 (-0.03, 0.07)	0.03	0.66	0.05
(CON > LIB) * (STIM > PRE)	0.01 (-0.03, 0.04)	0.02	0.43	0.02
(CON > LIB) * (Anodal > Sham)	-0.01 (-0.06, 0.04)	0.03	-0.22	0.02
(CON > LIB) * (Cathodal > Sham)	0.00 (-0.05, 0.06)	0.03	0.08	0.01
(STIM > PRE) * (Anodal > Sham)	-0.02 (-0.07, 0.03)	0.03	-0.69	0.05
(STIM > PRE) * (Cathodal > Sham)	0.04 (-0.01, 0.10)	0.03	1.70	0.12
(CON > LIB) * (STIM > PRE) * (Anodal > Sham)	0.00	0.03	-0.15	0.01

	(-0.06, 0.05)			
(CON > LIB) * (STIM > PRE) * (Cathodal > Sham)	-0.02 (-0.07, 0.04)	0.03	-0.64	0.05
Random Effect: (Intercept | Subject)				
Subjects	30			
Intercept (*SD*)	0.15			
N	360			

Table 15: Experiments 11 model-level statistics for c_n (top) and *d'* (bottom) values across the conservative (CON) and liberal (LIB) conditions both before (PRE) and during (STIM) online tDCS with anodal, cathodal, or sham stimulation (see also **Figure 38**).

Discussion

Since focal TMS targeting of various right PFC regions in Experiments 9 and 10 failed to manipulate decision criteria, a tDCS paradigm was implemented to more broadly stimulate right PFC. Unfortunately, diffuse online tDCS failed to alter recognition memory performance for both criterion placement and discriminability. One possible explanation for this, is that the right PFC is simply not *necessary* for maintaining a conservative versus liberal criterion. However, this explanation is difficult to reconcile with the robust fMRI findings in Experiments 6-10 for T > NT response contrasts as well as other neuroimaging studies (Aminoff et al., 2015; King & Miller 2017) and studies that observe differences in criterion placement in various patient populations (Parkin et al., 1996; Schacter et al., 1996; Swick & Knight, 1999; Verfaellie, et al., 2004; Budson, el al., 2006; Moritz et al., 2008; Waring et al., 2008; Beth, et al., 2009; Callahan, et al., 2011; Biesbroek, et al., 2014; Deason et al., 2017). Another explanation is that tDCS does not sufficiently alter neuronal excitability to induce performance changes, even if the stimulated regions are indeed criterion-sensitive. Experiments 9-11 indicate that neurostimulation to the right PFC is ineffective at altering decision performance. Targeting other regions within the observed frontoparietal activity for T > NT response contrasts may prove more successful, though regions within the right PFC are consistently and robustly active in this contrast. Experiments 9-11 illustrate that manipulating decision criteria via neurostimulation is not an intuitive feat. If neurostimulation can indeed alter a decision criterion, a much more subject-specific montage is likely needed to account for individual differences in factors including structural and functional variability as well as brain excitability (Sandrini et al., 2011). Given the challenges of evoking behavioral changes in criterion placement during recognition memory via neural stimulation, it is recommended that future research focus on other techniques to identify a causal link between brain activity and criterion shifting, such as through patient studies.

Chapter IV: General Discussion

The overarching goal of this dissertation research sought to characterize the behavioral and neural mechanisms underlying the decision criterion during recognition memory with a focus on individual differences. Many theories revolving around the decision criterion during recognition memory derive from group-averaged findings (Ulehla, 1966; Parks 1966; Thomas & Legge, 1970; Kubovy, 1977; Hirshman, 1995; Maddox & Bohil, 2005; Benjamin et al., 2009; Lynn & Barret, 2014). However, systematic assessments at the individual level suggests that group averages obscure the true nature of criterion shifting. While there is immense between-subject variability in criterion shifting tendencies, the within subject stability and generalizability across situations, as well as the fact that these tendencies cannot be easily explained by other cognitive factors, make it a uniquely individualistic cognitive trait. Some regularly shift criteria to great extents across situations whereas others fail to shift entirely (Aminoff et al., 2012, 2015; Kantner et al., 2015; Frithsen, et al., 2018; Layher et al., 2018; Miller & Kantner 2019; Layher et al., 2020). Findings from Experiments 4 and 5 as well as Mickes and colleagues (2017) revealed that participants are *capable* of shifting criteria to greater extents, but some individuals appear unwilling to shift despite being explicitly aware of the advantages for doing so. Theories about the decision criterion *must* account for these stable individual differences and cannot simply rely on group-averaged results. While these findings provide important implications for the theoretical nature of criterion shifting and SDT models at a behavioral level, there are also important considerations when investigating neural mechanisms underlying the decision criterion.

A major challenge for investigating the neural mechanisms associated with various criterion placements is that the degree to which different people place a decision criterion during recognition memory tasks is highly variable across people (Kantner & Lindsay, 2012, 2014). Many neuroimaging studies assessing recognition memory fail to account for individual differences in how a decision criterion is placed (Wheeler & Buckner, 2003; Kahn et al., 2004; Wagner et al., 2005; Cabeza et al., 2008; Ciaramelli et al., 2008; Vilberg and Rugg, 2009; Gilmore et al., 2015; McDermott et al., 2017), making it difficult to distinguish between brain activity associated with mnemonic evidence versus decisional processes. Given the vast variability across participants, it is important to manipulate decision criteria and memory strength in *within*-subject research paradigms to better dissociate the underlying neural mechanisms of these processes. However, a major challenge of obtaining within-subject neural activity at distinct levels of decision criteria is that many individuals are unwilling to shift criteria, necessitating prescreening procedures to exclude non-shifting participants from neuroimaging studies. Findings from Experiments 6-8 indicate that regions within frontoparietal, ventral attention, dorsal attention, and visual networks become increasingly active in the T > NT response contrast during recognition memory tests as a decision criterion becomes more extremely biased. Additionally, Experiments 7 and 8 revealed that the DMN becomes more engaged when participants provide a prepotent response versus inhibiting the preponderant response type. These findings support a response bias account for explaining widespread frontoparietal activity in T > NT response contrasts during recognition memory, since greater cognitive control and attentional resources are needed to inhibit versus provide prepotent responses. While the placement of a decision criteria robustly affects frontoparietal activity when making target versus nontarget responses, changes in discriminability show little to no differences across these networks. Experiments 6-8 strikingly revealed that fMRI measures are quite insensitive for identifying activity related to changes in memory strength, when the decision criterion is controlled for. Future neuroimaging investigations of recognition memory *must* account for the decision criterion in order to accurately assess neural mechanisms associated with memory versus decisional processes. Although the fMRI findings in Experiments 6-8 robustly indicate that shifts in decision criteria are associated with widespread changes in frontoparietal activity, it is necessary to provide a *causal* link between brain activity and the placement of a criterion to appropriately develop a neural model of the decision criterion.

Neurostimulation provides a means to modulate brain activity and potentially modify behavior directly. Initial attempts to focally stimulate regions within the right PFC via offline rTMS protocols proved unsuccessful at modulating decision criteria during recognition memory tests. Experiment 11 implemented an online tDCS protocol to more broadly stimulate the right PFC but attempts to modulate decision criteria again proved futile. While it is possible that the right PFC does not play a critical role in maintaining a decision criterion, there are a plethora of logistical factors that can lead to null results in neurostimulation experiments, such as individual differences in anatomy and brain excitability (Sandrini et al., 2011). Experiments 9-11 revealed that there is not an intuitive way for modulating decision criteria through neurostimulation. While it is advised to investigate causal relationships between the placement of a criterion and brain activity through other means, anyone daring enough to implement neurostimulation methods in future investigations of decision criteria should carefully control for the many factors that could lead to a null finding while also implementing a way to assess the efficacy of stimulation at the individual-level. Additionally, new stimulation protocols are likely needed to more effectively stimulate prefrontal regions in a way that can alter decision-making behavior. Perhaps criterion shifting tendencies are so ingrained in people that not even brain stimulation can disrupt the stability of these individual decision strategies.

References

Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Tipper, C., Johnson, A., Grafton, S. G., & Miller, M. B. (2012). Individual differences in shifting decision criterion: A recognition memory study. *Memory & Cognition, 40*(7), 1016-1030. https://doi.org/10.3758/s13421-012-0204-6

Aminoff, E., Freeman, S., Clewett, D., Tipper, C., Frithsen, A., Johnson, A., Grafton, S.G., & Miller, M. (2015). Maintaining a cautious state of mind during a recognition test: A large-scale fMRI study. *Neuropsychologia*, *67*, 132-147. https://doi.org/10.1016/j.neuropsychologia.2014.12.011

Aron, A. R., Cai, W., Badre, D., & Robbins, T. W. (2015). Evidence Supports Specific Braking Function for Inferior PFC. *Trends in Cognitive Sciences*, *19*(12), 711–712. https://doi.org/10.1016/j.tics.2015.09.001

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The Intrinsic Memorability of Face Photographs. *Journal of Experimental Psychology: General, 142,*1323-1334. https://doi.org/10.1037/a0033872

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*, 81–99. http://dx.doi.org/10.1037/h0029531

Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: Behavioral and neural basis of response control. *Progress in Neurobiology*, *108*, 44–79. https://doi.org/10.1016/j.pneurobio.2013.06.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review*, *116*(1), 84-115. https://doi.org/10.1037/a0014351

Benjamini, Y. & Hochberg, Y. (1995). Controlling The False Discovery Rate: A Practical and Powerful Approach To Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Beth, E. H., Budson, A. E., Waring, J. D., & Ally, B. A. (2009). Response Bias for Picture Recognition in Patients With Alzheimer Disease. *Cognitive and Behavioral Neurology*, *22*(4), 229–235. https://doi.org/10.1097/wnn.0b013e3181b7f3b1

Biesbroek, J. M., van Zandvoort, M. J., Kappelle, L. J., Schoo, L., Kuijf, H. J., Velthuis, B. K., Biessels, G. J., & Postma, A. (2014). Distinct anatomical correlates of discriminability and criterion setting in verbal recognition memory revealed by lesion-symptom mapping. *Human Brain Mapping*, *36*(4), 1292–1303. https://doi.org/10.1002/hbm.22702

Bikson, M., Grossman, P., Thomas, C., Zannou, A. L., Jiang, J., Adnan, T., Mourdoukoutas, A. P., Kronberg, G., Truong, D., Boggio, P., Brunoni, A. R., Charvet, L., Fregni, F., Fritsch, B., Gillick, B., Hamilton, R. H., Hampstead, B. M., Jankord, R., Kirton, A., . . . Woods, A. J. (2016). Safety of Transcranial Direct Current Stimulation: Evidence Based Update 2016. *Brain Stimulation*, *9*(5), 641–661. https://doi.org/10.1016/j.brs.2016.06.004

Bossong, M. G., Jager, G., van Hell, H. H., Zuurman, L., Jansma, J. M., Mehta, M. A., van

Gerven, J. M. A., Kahn, R. S., & Ramsey, N. F. (2012). Effects of Δ9-Tetrahydrocannabinol Administration on Human Encoding and Recall Memory Function: A Pharmacological fMRI Study. *Journal of Cognitive Neuroscience*, *24*(3), 588–599. https://doi.org/10.1162/jocn_a_00156

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436. https://doi.org/10.1163/156856897x00357

Budson, A. E., Wolk, D. A., Chong, H., & Waring, J. D. (2006). Episodic memory in Alzheimer's disease: Separating response bias from discrimination. *Neuropsychologia*, *44*(12), 2222–2232. https://doi.org/10.1016/j.neuropsychologia.2006.05.024

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience*, *9*(8), 613–625. https://doi.org/10.1038/nrn2459

Callahan, B. L., Ueda, K., Sakata, D., Plamondon, A., & Murai, T. (2011). Liberal bias mediates emotion recognition deficits in frontal traumatic brain injury. *Brain and Cognition, 77*(3), 412–418. https://doi.org/10.1016/j.bandc.2011.08.017

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*(5), 319-333. https://doi.org/10.1037//0022-3514.67.2.319

Chambers, C. D., Garavan, H., & Bellgrove, M. A. (2009). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neuroscience & Biobehavioral Reviews*, *33*(5), 631–646. https://doi.org/10.1016/j.neubiorev.2008.08.016

Cho, S. S., Ko, J. H., Pellecchia, G., van Eimeren, T., Cilia, R., & Strafella, A. P. (2010). Continuous theta burst stimulation of right dorsolateral prefrontal cortex induces changes in impulsivity level. *Brain Stimulation*, *3*(3), 170–176. https://doi.org/10.1016/j.brs.2009.10.002

Ciaramelli, E., Burianová, H., Vallesi, A., Cabeza, R., & Moscovitch, M. (2020). Functional Interplay Between Posterior Parietal Cortex and Hippocampus During Detection of Memory Targets and Non-targets. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.563768

Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, *46*(7), 1828–1851. https://doi.org/10.1016/j.neuropsychologia.2008.03.022

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language*, *55*(4), 461-478. https://doi.org/10.1016/j.jml.2006.08.003

Criss, A. H. (2009). The distribution of subjective memory strength: List strength and response bias. *Cognitive Psychology*, *59*, 297-319. https://doi.org/10.1016/j.cogpsych.2009.07.003

Criss, A. H., Wheeler, M. E., & McClelland, J. L. (2013). A Differentiation Account of Recognition Memory: Evidence from fMRI. *Journal of Cognitive Neuroscience*, *25*(3), 421–435. https://doi.org/10.1162/jocn_a_00292

Deason, R. G., Tat, M. J., Flannery, S., Mithal, P. S., Hussey, E. P., Crehan, E. T., Ally, B. A., & Budson, A. E. (2017). Response bias and response monitoring: Evidence from healthy older adults and patients with mild Alzheimer's disease. *Brain and Cognition*, *119*, 17–24. https://doi.org/10.1016/j.bandc.2017.09.002

Dennis, N. A., Bowman, C. R., & Turney, I. C. (2015). Functional Neuroimaging of False Memories. In D. R. Addis, M. Barense, & A. Duarte (Eds.), *The Wiley Handbook on the Cognitive Neuroscience of Memory* (1st ed., pp. 150–171). John Wiley & Sons, Ltd.

Doss, M. K., Weafer, J., Gallo, D. A., & de Wit, H. (2018). Δ9-Tetrahydrocannabinol at Retrieval Drives False Recollection of Neutral and Emotional Memories. *Biological Psychiatry*, *84*(10), 743–750. https://doi.org/10.1016/j.biopsych.2018.04.020

Dube, C., & Rotello, C. M. (2012). Binary ROCs in Perception and Recognition Memory Are Curved. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 130-151. https://doi.org/10.1037/a0024957

Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Tech. Note No. AFCRC-TN-58–51, AO-152650). Bloomington, IN: Indiana University Hearing and Communication Laboratory.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 1-9. https://doi.org/10.3389/fnhum.2014.00443

Franks, B. A., & Hicks, J. L. (2016). The reliability of criterion shifting in recognition memory is task dependent. *Memory & Cognition*, *44*(8), 1215–1227. https://doi.org/10.3758/s13421-016-0633-8

Frithsen, A., Kantner, J., Lopez, B. A., & Miller, M. B. (2018). Cross-task and cross-manipulation stability in shifting the decision criterion. *Memory*, *26*(5), 653-663. https://doi.org/10.1080/09658211.2017.1393090

Georgiev, D., Rocchi, L., Tocco, P., Speekenbrink, M., Rothwell, J. C., & Jahanshahi, M. (2016). Continuous Theta Burst Stimulation Over the Dorsolateral Prefrontal Cortex and the Pre-SMA Alter Drift Rate and Response Thresholds Respectively During Perceptual Decision-Making. *Brain Stimulation*, *9*(4), 601–608. https://doi.org/10.1016/j.brs.2016.04.004

Gilmore, A. W., Nelson, S. M., & McDermott, K. B. (2015). A parietal memory network revealed by multiple MRI methods. *Trends in Cognitive Sciences*, *19*(9), 534–543. https://doi.org/10.1016/j.tics.2015.07.004

Glanzer, M. & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8-20. https://doi.org/10.3758/BF03198438

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

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

Grossheinrich, N., Rau, A., Pogarell, O., Hennig-Fast, K., Reinl, M., Karch, S., Dieler, A., Leicht, G., Mulert, C., Sterr, A., & Padberg, F. (2009). Theta Burst Stimulation of the Prefrontal Cortex: Safety and Impact on Cognition, Mood, and Resting Electroencephalogram. *Biological Psychiatry*, *65*(9), 778–784. https://doi.org/10.1016/j.biopsych.2008.10.029

Gur, D., Sumkin, J. H., Hardesty, L. A., Clearfield, R. J., Cohen, C. S., Ganott, M. A., Hakim, C. M., Harris, K. M., Poller, W. R., Shah, R., Wallace, L. P., & Rockette, H. E. (2004). Recall and detection rates in screening mammography: a review of clinical experience - implications for practice guidelines, *Cancer, 100*(8), 1590–1594. https://doi.org/10.1002/cncr.20053

Han, S., & Dobbins, I. G. (2008). Examining recognition criterion rigidity during testing using a biased-feedback technique: Evidence for adaptive criterion learning. *Memory & Cognition*, *36*(4), 703-715. https://doi.org/10.3758/MC.36.4.703

Han, S., & Dobbins, I. G. (2009). Regulating recognition decisions through incremental reinforcement learning. *Psychonomic Bulletin & Review*, *16*(3), 469-474. https://doi.org/10.3758/PBR.16.3.469

Healy, A. F., & Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of Performance and cutoff location in recognition memory. *Memory & Cognition*, *6*(5), 544-553. https://doi.org/10.3758/BF03198243

Hicks, J. L., & Starns, J. J. (2014). Strength cues and blocking at test promote reliable within-list criterion shifts in recognition memory. *Memory & Cognition*, *42*(8), 742-754. https://doi.org/10.3758/s13421-014-0397-y

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 302-313. http://dx.doi.org/10.1037//0278-7393.21.2.302

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta Burst Stimulation of the Human Motor Cortex. *Neuron*, *45*(2), 201–206. https://doi.org/10.1016/j.neuron.2004.12.033

Hwang, D. Y., Gallo, D. A., Ally, B. A., Black, P. M., Schacter, D. L., & Budson, A. E. (2007). Diagnostic retrieval monitoring in patients with frontal lobe lesions: Further exploration of the distinctiveness heuristic. *Neuropsychologia*, *45*(11), 2543–2552. https://doi.org/10.1016/j.neuropsychologia.2007.03.018

Jaeger, A., Konkel, A., & Dobbins, I. G. (2013). Unexpected novelty and familiarity orienting responses in lateral parietal cortex during recognition judgment. *Neuropsychologia*, *51*(6), 1061–1076. https://doi.org/10.1016/j.neuropsychologia.2013.02.018

Jahfari, S., Stinear, C. M., Claffey, M., Verbruggen, F., & Aron, A. R. (2010). Responding with Restraint: What Are the Neurocognitive Mechanisms? *Journal of Cognitive Neuroscience*, *22*(7), 1479–1492. https://doi.org/10.1162/jocn.2009.21307

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Kahn, I., Davachi, L., & Wagner A. (2004). Functional-Neuroanatomic Correlates of Recollection: Implications for Models of Recognition Memory. *Journal of Neuroscience*, *24*(17), 4172–4180. https://doi.org/10.1523/jneurosci.0624-04.2004

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working Memory, Attention Control, and the *N*-Back task: A Question of Construct Validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615-622. https://doi.org/10.1037/0278-7393.33.3.615

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition, 40*(8), 1163-1177. https://doi.org/10.3758/s13421-012-0226-0

Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review*, *21*(5), 1272-1280. https://doi.org/10.3758/s13423-014-0608-3

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

Kim, H. (2020). An integrative model of network activity during episodic memory retrieval and a meta-analysis of fMRI studies on source memory retrieval. *Brain Research*, *1747*, 147049. https://doi.org/10.1016/j.brainres.2020.147049

King, D. R., & Miller, M. B. (2017). Influence of response bias and internal/external source on lateral posterior parietal successful retrieval activity. *Cortex*, *91*, 126–141. https://doi.org/10.1016/j.cortex.2017.04.002

Koen, J. D., Barrett, F. S., Harlow, I. M., & Yonelinas, A. P. (2016). The ROC Toolbox: A toolbox for analyzing receiver-operating characteristics derived from confidence ratings. *Behavior Research Methods*, *49*(4), 1399–1406. https://doi.org/10.3758/s13428-016-0796-z

Koen, J. D., & Yonelinas, A. P. (2011). From humans to rats and back again: Bridging the divide between human and animal studies of recognition memory with receiver operating characteristics. *Learning & Memory, 18*, 519-522. https://doi.org/10.1101/lm.2214511

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness*, Cambridge University Press, 289-325. https://doi.org/10.1017/CBO9780511816789.012

Kubovy, M. (1977). Response availability and the apparent spontaneity of numerical choices. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(2), 359-364. http://dx.doi.org/10.1037/0096-1523.3.2.359

Layher, E., Dixit, A., & Miller, M. B. (2020). Who gives a criterion shift? A uniquely individualistic cognitive trait. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2075–2105. https://doi.org/10.1037/xlm0000951

Layher, E., & Miller, M. (2019, July 1). Who gives a criterion shift? A uniquely individualistic cognitive trait. Retrieved from osf.io/4k2hb

Layher, E., Santander, T., Volz, L. J., & Miller, M. B. (2018). Failure to Affect Decision Criteria During Recognition Memory With Continuous Theta Burst Stimulation. *Frontiers in Neuroscience, 12*(705). https://doi.org/10.3389/fnins.2018.00705

Lee, T. G., & D'Esposito, M. (2012). The Dynamic Nature of Top-Down Signals Originating from Prefrontal Cortex: A Combined fMRI-TMS Study. *Journal of Neuroscience*, *32*(44), 15458–15466. https://doi.org/10.1523/jneurosci.0627-12.2012

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8*(2), 75-84. https://doi.org/10.1037/1076-898x.8.2.75

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological science*, *25*(9), 1663-1673. https://doi.org/10.1177/0956797614541991

Macmillan, N. A., and Creelman C. D. (2005). *Detection Theory, A User's Guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Macmillan, N. A., and Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*, 185-199. http://dx.doi.org/10.1037/0033-2909.98.1.185

Maddox, W. T., & Bohil, C. J. (2005). Optimal classifier feedback improves cost-benefit but not base-rate decision criterion learning in perceptual categorization. *Memory & Cognition*, *33*(2), 303-319. http://dx.doi.org/10.3758/BF03195319

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*(4), 335–384. https://doi.org/10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Annis, J. (2012). On the Relationship Between Memory and Perception: Sequential Dependencies in Recognition Memory Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *141*(2), 233-259. https://doi.org/10.1037/a0025277

McDermott, K. B., Gilmore, A. W., Nelson, S. M., Watson, J. M., & Ojemann, J. G. (2017). The parietal memory network activates similarly for true and associative false recognition elicited via the DRM procedure. *Cortex*, *87*, 96–107. https://doi.org/10.1016/j.cortex.2016.09.008

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review*, *14*(5), 858-865. https://doi.org/10.3758/bf03194112

Mickes, L., Hwe, V., Wais, P. E., & Wixted, J. T. (2011). Strong Memories Are Hard to Scale. *Journal of Experimental Psychology: General*, *140*(2), 239-257. https://doi.org/10.1037/a0023007

Mickes, L., Seale-Carlisle, T. M., Wetmore, S. A., Gronlund, S. D., Clark, S. E., Carlson, C. A., Goodsell, D. W., Weatherford, D., & Wixted, J. T. (2017). ROCs in Eyewitness Identification: Instructions versus Confidence Ratings. *Applied Cognitive Psychology*, *31*, 467-477. https://doi.org/10.1002/acp.3344

Miller, M. B., & Dobbins, I. G. (2014). Memory as decision-making. In M. S. Gazzaniga &

G. R. Mangun (Eds.), *The Cognitive Neurosciences* (5th ed., pp. 551–563). Cambridge, MA: MIT Press.

Miller, M. B., & Kantner, J. (2019). Not All People Are Cut for Strategic Criterion Shifting. *Current Direction in Psychological Science*, 1-7. https://doi.org/10.1177/0963721419872747

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*, 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2018, May 19). http://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf

Moritz, S., Woodward, T., Jelinek, L., & Klinge, R. (2008). Memory and metamemory in schizophrenia: a liberal acceptance account of psychosis. *Psychological Medicine*, *38*(6), 825–832. https://doi.org/10.1017/s0033291707002553

O'Connor, A. R., Han, S., & Dobbins, I. G. (2010). The Inferior Parietal Lobule and Recognition Memory: Expectancy Violation or Successful Retrieval? *Journal of Neuroscience*, *30*(8), 2924–2934. https://doi.org/10.1523/jneurosci.4225-09.2010

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rhinehart, & Winston.

Parkin, A. J., Bindschaedler, C., Harsent, L., & Metzler, C. (1996). Pathological False Alarm Rates Following Damage to the Left Frontal Cortex. *Brain and Cognition*, *32*(1), 14–27. https://doi.org/10.1006/brcg.1996.0055

Parks, T. E. (1966). Signal-detectability theory of recognition-memory performance. *Psychological Review, 73*(1), 44-58. http://dx.doi.org/10.1037/h0022662

Poldrack, R. A. (2017). Precision Neuroscience: Dense Sampling of Individual Brains. *Neuron*, *95*(4), 727–729. https://doi.org/10.1016/j.neuron.2017.08.002

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682. https://doi.org/10.1073/pnas.98.2.676

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences, 20*(4), 260-281. https://doi.org/10.1016/j.tics.2016.01.007

Reinhart, R. M. G., Cosman, J. D., Fukuda, K., & Woodman, G. F. (2016). Using transcranial direct-current stimulation (tDCS) to understand cognitive processing. *Attention, Perception, & Psychophysics*, *79*(1), 3–23. https://doi.org/10.3758/s13414-016-1224-2

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*(2), 305-320. https://doi.org/10.1037/0278-7393.33.2.305

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

Rogers, R. D., & Monsell, S. (1995). Costs of a Predictable Switch Between Simple Cognitive Tasks. *Journal of Experimental Psychology: General, 124*(2), 207-231. https://doi.org/10.1037//0096-3445.124.2.207

Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review*, *12*(5), 865-873. https://doi.org/10.1037/0033-295X.111.3.588

Ryan, R. M. (1982). Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory. *Journal of Personality and Social Psychology, 43*(3), 450-461. https://doi.org/10.1037/0022-3514.43.3.450

Sandrini, M., Umiltà, C., & Rusconi, E. (2011). The use of transcranial magnetic stimulation in cognitive neuroscience: A new synthesis of methodological issues. *Neuroscience & Biobehavioral Reviews*, *35*(3), 516–536. https://doi.org/10.1016/j.neubiorev.2010.06.005

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2017). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Schacter, D. L., Curran, T., Galluccio, L., Milberg, W. P., & Bates, J. F. (1996). False recognition and the right frontal lobe: A case study. *Neuropsychologia*, *34*(8), 793–808. https://doi.org/10.1016/0028-3932(95)00165-4

Selmeczy, D., & Dobbins, I. G. (2013). Metacognitive awareness and adaptive recognition biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 678-690. https://doi.org/10.1037/a0029469

Shiffrin, R. M., & Steyvers, M. (1997). A model of recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145-166. https://doi.org/10.3758/BF03209391

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures.

Behavior Research Methods, Instruments, & Computers, *31*, 137-149. https://doi.org/10.3758/bf03207704

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. https://doi.org/10.3758/s13421-014-0433-y

Starns, J. J., White, C. N., & Ratcliff, R. (2010). A direct test of the differentiation mechanism: REM, BCDMEM, and the strength-based mirror effect in recognition memory. *Journal of Memory and Language*, *63*(1), 18-34. https://doi.org/10.1016/j.jml.2010.03.004

Stretch, V., & Wixted, J. T. (1998a). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1379–1396. https://doi.org/10.1037/0278-7393.24.6.1379

Stretch, V., & Wixted, J. T. (1998b). Decision Rules for Recognition Memory Confidence Judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1397-1410. https://doi.org/10.1037/0278-7393.24.6.1397

Suppa, A., Huang, Y. Z., Funke, K., Ridding, M., Cheeran, B., di Lazzaro, V., Ziemann, U., & Rothwell, J. (2016). Ten Years of Theta Burst Stimulation in Humans: Established Knowledge, Unknowns and Prospects. *Brain Stimulation*, *9*(3), 323–335. https://doi.org/10.1016/j.brs.2016.01.006

Swann, N. C., Tandon, N., Pieters, T. A., & Aron, A. R. (2012). Intracranial Electroencephalography Reveals Different Temporal Profiles for Dorsal- and Ventro-lateral Prefrontal Cortex in Preparing to Stop Action. *Cerebral Cortex, 23*(10), 2479–2488. https://doi.org/10.1093/cercor/bhs245

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1955). *The Evidence for a Decision-Making Theory of Visual Detection*. Technical Report No. 40, Electronic Defense Group, University of Michigan.

Swick, D., & Knight, R. T. (1999). Contributions of prefrontal cortex to recognition memory: Electrophysiological and behavioral evidence. *Neuropsychology*, *13*(2), 155–170. https://doi.org/10.1037/0894-4105.13.2.155

Tanner, W. P., Swets, J. A., & Green, D. M. (1956). *Some General Properties of The Hearing Mechanism*. Technical Report No. 30, Electronic Defense Group, University of Michigan.

Thomas, E. A., & Legge, D. (1970). Probability matching as a basis for detection and recognition decisions. *Psychological Review, 77*(1), 65-72. https://doi.org/10.1037/h0028579

Ulehla, Z. J. (1966). Optimality of perceptual decision criteria. *Journal of Experimental Psychology*, *71*(4), 564-569. https://doi.org/10.1037/h0023007

Verbruggen, F., Aron, A. R., Stevens, M. A., & Chambers, C. D. (2010). Theta burst stimulation dissociates attention and action updating in human inferior frontal cortex. *Proceedings of the National Academy of Sciences*, *107*(31), 13966–13971. https://doi.org/10.1073/pnas.1001957107

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition, 35*(2), 254-262. https://doi.org/10.3758/bf03193446

Verfaellie, M., Rapcsak, S. Z., Keane, M. M., & Alexander, M. P. (2004). Elevated False Recognition in Patients With Frontal Lobe Damage Is Neither a General Nor a Unitary Phenomenon. *Neuropsychology*, *18*(1), 94–103. https://doi.org/10.1037/0894-4105.18.1.94

Vilberg, K. L., & Rugg, M. D. (2009). An investigation of the effects of relative probability of old and new test items on the neural correlates of successful and unsuccessful source memory. *NeuroImage*, *45*(2), 562–571. https://doi.org/10.1016/j.neuroimage.2008.12.020

Waldmann, M. R., & Göttert, R. (1989). Response Bias in Below-Chance Performance: Computation of the Parametric Measure ß. *Psychological Bulletin*, *106*(2), 338-340. https://doi.org/10.1037/0033-2909.106.2.338

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445–453. https://doi.org/10.1016/j.tics.2005.07.001

Waring, J. D., Chong, H., Wolk, D. A., & Budson, A. E. (2008). Preserved metamemorial ability in patients with mild Alzheimer's disease: Shifting response bias. *Brain and Cognition*, *66*(1), 32–39. https://doi.org/10.1016/j.bandc.2007.05.002

Watson, D., & Clark L. A. (1994). The PANAS-X: Manual for the Positive and Negative Affect Schedule – Expanded Form. *Iowa City: University of Iowa*. https://doi.org/10.17077/48vt-m4t2

Wessel, J. R. (2017). Prepotent motor activity and inhibitory control demands in different variants of the go/no-go paradigm. *Psychophysiology, 55*(3), 1-14. https://doi.org/10.1111/psyp.12871

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020-2045. https://doi.org/10.1037/xge0000014

Wheeler, M. E., & Buckner, R. L. (2003). Functional Dissociation among Components of Remembering: Control, Perceived Oldness, and Content. *The Journal of Neuroscience*, *23*(9), 3869–3880. https://doi.org/10.1523/jneurosci.23-09-03869.2003

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, 44, 289-300. https://doi.org/10.1068/p7908

Wixted, J. T., & Gaitan, S. C. (2002). Cognitive theories as reinforcement history surrogates: The case of likelihood ratio models of human recognition memory. *Animal Learning & Behavior*, *30*(4), 289-305. https://doi.org/10.3758/bf03195955

Xiao, J., Hays, J., Ehinger, K., Oliva, A., & Torralba A. (2010). SUN Database: Large-scale Scene Recognition from Abbey to Zoo. *IEEE Conference on Computer Vision and Pattern Recognition*.

Xu, Z., Adam, K. C. S., Fang, X., & Vogel, E. K. (2018). The Reliability and Stability of Visual Working Memory Capacity. *Behavior Research Methods*, *50*(2), 576–588. https://doi.org/10.3758/s13428-017-0886-6

Yankaskas, B. C., Cleveland, R. J., Schell, M. J., & Kozar, R. (2001). Association of Recall Rates with Sensitivity and Positive Predictive Values of Screening Mammography. *American Journal of Roentgenology*, *177*(3), 543-549. https://doi.org/10.2214/ajr.177.3.1770543

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Yonelinas, A. P., & Parks, C. M. (2007). Receiver Operating Characteristics (ROCs) in Recognition Memory: A Review. *Psychological Bulletin*, *133*(5), 800-832. https://doi.org/10.1037/0033-2909.133.5.800

Yu, S. S., Johnson, J. D., & Rugg, M. D. (2012). Dissociation of recollection-related neural activity in ventral lateral parietal cortex. *Cognitive Neuroscience*, *3*(3–4), 142–149. https://doi.org/10.1080/17588928.2012.669363

Appendix

Intrinsic Motivation Inventory: Effort/Importance subscale (Ryan, 1982)

Please indicate how true this statement is for you

1	2	3	4	5	6	7
not at all true			somewhat true			very true

1. I put a lot of effort into this.
2. I did NOT try very hard to do well at this task.*
3. I tried very hard on this task.
4. It was important to me to do well at this task.
5. I did NOT put much energy into this.*

BIS/BAS: fun-seeking subscale (Carver & White, 1994)

How true or false is this statement for you?

1	2	3	4
very true	somewhat true	somewhat false	very false

1. I'm always willing to try something new if I think it will be fun.
2. I will often do things for no other reason than that they might be fun.
3. I often act on the spur of the moment.
4. I crave excitement and new sensations.

PANAS-X: Negative Affect (Watson & Clark, 1999)

Indicate to what extent you have felt this way during the past few weeks

1	2	3	4	5
slightly	a little	moderately	quite a bit	extremely

1. afraid
2. scared
3. jittery
4. nervous
5. irritable
6. hostile
7. guilty
8. ashamed
9. upset
10. distressed

VVQ (modified): Verbalizer score (Richardson, 1977; Paivio 1971)

Indicate how much you agree or disagree with this statement

1 2 3 4 5 6

7

strongly disagree neither agree or disagree

strongly agree

1. I enjoy doing work that requires the use of words.
2. I enjoy learning new words.
3. I can easily think of synonyms for words.
4. I read rather slowly.*

5. I prefer to read instructions about how to do something rather than have someone show me.
6. I have better than average fluency in using words.
7. I spend very little time attempting to increase my vocabulary.*

All questionnaires are scored by summing the numeric value (or reversely coded value) assigned to each item.

*Reverse code items