Title

Exploring Criterion Shifting with Ecologically Valid Contexts: An Investigation of Recollection Processes and the Impact of Social External Factors on Decision-Making Strategies for Memory-Based Tasks

Author

Durdle, Courtney Alexandra Susan

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

Exploring Criterion Shifting with Ecologically Valid Contexts:

An Investigation of Recollection Processes and the Impact of Social External Factors on Decision-Making Strategies for Memory-Based Tasks

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychological and Brain Sciences

by

Courtney A. S. Durdle

Committee in charge:

Professor Michael B. Miller, Chair

Professor Michael Gazzaniga

Professor Richard Mayer

Professor Jonathan Schooler

December 2024

The dissertation of Courtney A. S. Durdle is approved.

Michael Gazzaniga

Richard Mayer

Jonathan Schooler

Michael B. Miller, Committee Chair

August 2024

Exploring Criterion Shifting with Ecologically Valid Contexts:

An Investigation of Recollection Processes and the Impact of Social External Factors on

Decision-Making Strategies for Memory-Based Tasks

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to the following people, whose advice, support, and encouragement have been indispensable throughout the past six years.

Firstly, I want to thank my advisor, Dr. Mike Miller, for the opportunity to work on this project and for encouraging me to think more creatively and push beyond conventional boundaries. Though the journey was challenging at times, your guidance on fostering novel methodologies shaped my approach to research. I appreciate the freedom you gave me to explore, which ultimately strengthened my resilience as a scientist.

I extend my heartfelt thanks to the entire Miller Lab team: Dr. Evan Layher, Dr. Tyler Santander, Sara Leslie, Selin Bekir, Luna Li, Henri Skinner, and Christina Boardman. Your collective expertise, thoughtful feedback, and collaborative spirit have greatly contributed to my development as a researcher. A special thank you goes to the many research assistants who have been an integral part of my work over the past six years: Jessica Simonson, Kaitlyn Deen, Shiva Areff, Aarushi Agrawal, Chinmayee Balachandra, Ana Bobrycki, Jasmine Chuey, Galia Broussi Raich, Angelo "Jello" Joseph Bulahan, Manasvi Chandra, Ziyuan Chen, Rosalyn Collins, Luke Hamilton, Samridi Iyer, Eleanor Khan, Maliha Khan, Hannah Kim, Grace Liu, Matejas Mackin, Leighanne Masri, Anoushka Menon, Natasha Pansare, Alex Reed Aparicio, Claudine Ushana, Alissa Wong, Albert Yang, Jamie Yoo, and Tongxi Yu. Your contributions have been invaluable, and this dissertation would not have been possible without your support.

I would like to especially highlight Dr. Evan Layher for being an exceptional mentor and friend. Evan, your generosity, time, and patience have been a cornerstone of my graduate experience. From our first meeting—when you graciously spent a day answering my questions before I had even applied to the program—you have consistently gone above and beyond. Your mentorship has been a defining aspect of my journey, and I am profoundly grateful for your guidance, feedback, and support. I am a far better researcher, and person, because of you.

I also want to extend my deepest thanks to my committee members: Dr. Mike Gazzaniga, Dr. Rich Mayer, and Dr. Jonathan Schooler. Your expertise, critical feedback, and guidance throughout my project were invaluable. Rich and Mike, the classes I took with you helped sharpen my thinking and encouraged me to critically evaluate the field of psychology. Jonathan, your feedback on my second-year paper was instrumental in shaping parts of this dissertation. I am grateful to each of you for your thoughtful advice during my defense and for the lasting impact you have had on my academic growth.

To my amazing cohort and members of the UCSB community—Alyssa Lawson, Sara Leslie, Jordan Garrett, Shuying Yu, Paige Harris, Viki Papadakis, Angela Morales, Will Redman, Sudhanshu Srivastava, Liz Rizor, Vik Babenko, Nicole Alea Albada, Greg Ashby, R "Baker" Baker, Dylan Benkley, Henry Biedron, Wade Bloomquist, Kyle Brown, Tom Bullock, Stevie Bullock, Fernando Cano, Tikal Catena, Carly Chak, Jake Chak, Nancy Collins, Ashley Coventry, Jamiee Cook, Neil Dundon, Lexxus Edison, Barry Geisbrecht, Scott Grafton, Hannah Grotzinger, Sebaah Hamad, Justin Haiman, Adeel Hussaini, Leslie Huang, Skirmantas Janusonis, Regina Lapate, Mengsi Li, Katherine Lowe, Mario Medoza, Chris McFerron, Bedlam Oak, Elle Murata, Josh Ortega, Sophie Peterson, Anita Raychawdhuri, Justin Rogers, Rammy Salem, Gabrielle Salo, John Saunders, Sam Scudder, Shivang Shelat, Joanne Stasiak, Caitlin Taylor, Daniel Thayer, Lexie Topete, Jaime Vandeveer, Lucas Vincenzi, Em Waters, Jingyi Wang, Vanessa Woods, and Hongbo Yu—I am so grateful to have shared this journey with all of you. Your kindness and encouragement helped me through the challenges of grad school, and I really appreciate the time we spent together. Whether through brief conversations

or shared experiences, your presence made a positive difference. I feel incredibly fortunate to have been part of such a supportive community, and I know each of you will continue to achieve great things!

I am profoundly grateful to the McNair Scholars Program, which has been one of the most meaningful parts of my time at UCSB. First, I want to express my deep appreciation to Dr. Veronica Fematt, who was my direct supervisor and became an incredible mentor and friend. Veronica, working with you was both a privilege and a joy. Your tough-love approach, paired with your dedication to student success, made you an inspiring leader, and our dynamic as a team was something I truly cherished. From the insightful seminars you led to the memorable UCLA McNair Conference trip, I learned so much from you—both professionally and personally. You gave me the freedom to explore my own teaching style, which I am incredibly grateful for. Your transition to Hispanic Serving Institution Director is a testament to your incredible skill and passion for making a difference.

I would also like to thank Dr. Miros Chávez-García, the Faculty Director, whose dedication to the McNair students is truly inspiring. Miros, your hard work and tireless commitment to creating opportunities for the scholars left a lasting impression on me. Your support has meant a great deal, especially when writing letters of recommendation and serving as a reference. I am incredibly grateful for everything you have done.

My gratitude also extends to the incredible graduate mentors, Jason Fly and Mary Michael. Jason, your calm and thoughtful presence brought so much warmth to the program. Your kindness, patience, and willingness to share your experiences made a huge impact on both the students and me. Mary, my “zesty bestie,” your lively, energetic personality always brought laughter and fun to our work together. I respect and admire both of you for your

dedication to the students. Working with you both has been a highlight, and I truly appreciate the strong bond we formed.

I would also like to recognize Angelica Hernandez, our Program Coordinator. Angelica, your kindness and adventurous spirit have always been a bright spot in the program. I admire your courage and how willing you are to embrace new experiences. Your contributions to McNair have been so valuable, and I've loved getting to know you. Finally, I want to give my deepest thanks to the McNair students: Jada Alexander, Nicholas Alexander, Natalia Almanza, Fátima Aguilar, Paulos Berhane, Andre Braga, Mayra Canseco, Julissa Cesareo, Jaden Chau, Chiara Cimarusti, Ezequiel Contreras Martinez, Rocio Curiel, Maria Donan, Naka Elelleh, Martha Garcia, Andrea Gomez, Fatima Gonzalez, Karen Gonzalez-Salazar, Arsalan Hashmi, Genesis Hernandez, Angeline Higby, Citlali Ibarra, Evelyn Isais, Jasmine Labastida, Katherine Larina, Collin Lejano, Erin Lewis, Jacky Lopez, Laila Lore, Monsi Lupian, Ixel Medrano, Brandon Nguyen, Rio Ondo, Zeena Omer, Michael Osio, Madeleine Pacheco, Eliza Pasha, Sherry Perez, Bree Pernas, Alex Reed Aparicio, Destiny Rivera, Lizette Rivera, Nataly Rincon Garcia, Natalia Rodriguez Martinez, Melissa Romero, Isa Ruacho, Zacharias Seitz, Gil Marc Sia, Carla Suarez Soto, Mackenzie Taradalsky, Lucero Torres Ojeda, Angie Torrico, Marian Walker, Britney Walton, Abby Welch, Rick Zheng. Your dedication, curiosity, and enthusiasm were a constant source of inspiration during my time as the Writing Specialist. Working with each of you has been one of the most rewarding experiences of my career, and I am incredibly proud of you. Your passion for learning and commitment to your research projects have left a lasting impression on me, both as a mentor and as a scientist.

To my fellow Brain Campers at the 2022 Kavli Summer Institute in Cognitive Neuroscience, being part of this program with you was one of the most memorable experiences of my time at UCSB. The friendships we built and the conversations we had were incredibly rewarding, both personally and professionally. It was a privilege to meet such a passionate group of researchers, and I am excited to see the groundbreaking work you will accomplish in the future!

I would like to express my deepest gratitude to my UC Davis family, whose unwavering support and encouragement have profoundly shaped my academic journey. Dr. Tony Simon and Dr. Kathy Angkustsiri, working with you at the MIND Institute provided countless opportunities to expand my skills. From hands-on training with EEG and fMRI to conducting clinical interviews and engaging with physiology equipment, the experiences you offered allowed me to develop as a researcher. Your mentorship and trust in my abilities empowered me to embrace new challenges and explore diverse research methodologies. I am also extremely thankful to Dr. Abbie Popa and Dr. Kathryn McCabe for your invaluable guidance during a pivotal phase of my academic career. Your support throughout the grad school application process, as well as your insights into the often unspoken aspects of academia, were critical to my success. Dr. Gail Goodman and Dr. Debbie Goldfarb, your groundbreaking work in law and psychology inspired my decision to pursue this path, and I continue to look up to both of you as role models. I greatly appreciate your ongoing mentorship and advice. Lastly, I want to extend my heartfelt thanks to Dr. Simona Ghetti and Dr. Christine Coughlin, whose early mentorship introduced me to the world of research and set me on the path toward academia. The sense of community and mentorship I experienced at UC Davis will forever hold a special place in my heart.

To my beloved pups, Cali and Uni, your unconditional love and infectious joy have been the brightest part of my days. Your energy, playfulness, and ability to make me smile have carried me through the toughest times. During the pandemic, your enthusiasm gave me purpose, motivating me to find peace and joy in our daily walks to the beach. You’ve both taught me the beauty of living in the moment, and your snuggles and playful spirit have been a constant reminder of the happiness found in life’s simple moments. I will always be grateful for the light you’ve brought into my life.

To my best friend, Sarah Wagner, your friendship has been a lifeline through every stage of this journey. Throughout the ups and downs, especially during the long days of my qualifying exams, you were always there to help me stay focused and positive. The daily Zoom sessions and your constant encouragement gave me the strength to keep pushing forward. I’m endlessly thankful for your belief in me and your unwavering support over the years. To Annie Lo, your steady support and kindness have been a true source of comfort. The bond we share means so much to me, and I’m incredibly lucky to have both you and Sarah in my life. I love you both so much, and I will always be here for you, no matter what.

To my partner and best friend, Ian Durdle, your presence in my life has been the greatest gift, especially during these last six years. From the moment you proposed at Butterfly Beach, surrounded by family and friends, to our beautiful wedding in Huntington Beach, you have brought so much thoughtfulness, love, and joy into every chapter of this journey. That proposal, during my first week at UCSB, was a reflection of the care and intention you bring to everything we do. Over our 14 years together, you have always been there to support me, taking on the responsibilities of the household, caring for our pups, and ensuring I had the

space to focus on my work. You have done all of this with such generosity and patience, and I am so deeply thankful for all you have done to lighten my load.

What stands out most to me is the way you have made this time so much more fun. Together, we've created memories that I'll carry with me always—whether organizing trivia nights for the lab, hosting Family Feud for my research assistants, or planning our Durdle Disney trips. Your energy and enthusiasm for making these moments special is something I admire and cherish. You have made our home a place where laughter and connection thrive, and where we have always found time to celebrate the people around us.

One of the most meaningful aspects of our relationship is the partnership we have built. You took my last name, a gesture that symbolizes the team we have always been. We have shared the responsibilities, the joys, and the challenges together, and your unwavering support has meant the world to me. I have relied on your calm presence to keep me steady and your humor to remind me that even in the busiest of times, there is room for joy.

I am endlessly grateful for your friendship, love, and for the way we have navigated this journey as a team. You have helped me find balance in both work and life, and you have made this experience so much more than I could have imagined. I feel incredibly lucky to walk through life with you, and I look forward to all the adventures that await us. I love you so much, and I am so proud of everything we have built together.

VITA OF COURTNEY A. S. DURDLE

AUGUST 2024

Education

University of California, Santa Barbara (UCSB) ***August 2024***

Doctor of Philosophy (Ph.D.), Cognition, Perception, & Cognitive Neuroscience
Advanced to Candidacy: Fall 2020
Advisor: Dr. Michael B. Miller
Committee: Dr. Michael Gazzaniga, Dr. Rich Mayer, & Dr. Jonathan Schooler
Department: Psychological & Brain Sciences

University of California, Davis (UC Davis) ***June 2014***

Bachelor of Science, Human Development
Bachelor of Arts, Psychology
Minor in Education

Grants, Awards, and Honors

UCSB McNair Scholars Program Graduate Student Fellowship – 2022-24 ($65,000)
UCSB Psychological & Brain Sciences Department Nominee Grad Division STEM Mentoring Award – 2023
UCSB Graduate Student Association Conference Travel Grant – 2023 & 2024 ($1000)
UCSB Academic Senate Doctoral Travel Grant – 2022 & 2024 ($1800)
Womxn in Science & Engineering (WiSE) Decolvenaere Research Accelerator Award Finalist – 2022 & 2023
WiSE Karl Storz Research Accelerator Award Finalist – 2022 & 2023
National Science Foundation Graduate Research Fellow – 2020-22 & 2023-24 ($145,000)
UCSB Associated Students Community Foundation Grant to fund P.U.G.S. – 2021 ($5,400)
Volunteer of the Month at WEAVE Inc., Sacramento, CA – August 2016
UC Davis MIND Institute Make the Difference Award for Mentorship – 2015
UC Davis Undergraduate Dean’s List – 2012
UC Davis University Grant – 2010 to 2014 ($28,840)
Kelly J. Kolozsi Memorial Scholarship – 2010 & 2012 ($4,500)
Pell Grant – 2010 to 2011 ($7,400)
UC Davis Summer UC Grant – 2011 ($4,400)
CAL Grant A – 2010 ($13,266)
UC Davis Entering Undergraduate Scholarship – 2010 ($2,727)
Federal Academic Competitiveness Grant – 2010 ($750)
Pearl Moulden Community Activities Committee Scholarship – 2010 ($300)
Silicon Valley Scholarship – 2010 ($1,000)
Curry Award for Girls and Young Women – 2010 ($1,000)
Diane and Howard Crittenden Grant – 2010 ($5,000)

Publications

2024

Ge R., Ching C.R.K., Bassett A.S., Kushan L., Antshel K.M., van Amelsvoort T., Bakker G., Butcher N.J., Campbell L.E., Chow E.W.C., Craig M., Crossley N.A., Cunningham A., Daly E., Doherty J.L., **Durdle C.A.**, Emanuel B.S., Fiksinski A., Forsyth J.K., Fremont W., Goodrich-Hunsaker N.J., Gudbrandsen M., Gur R.E., Jalbrzikowski M., Kates W.R., Lin A., Linden D.E.J., McCabe K.L., McDonald-McGinn D., Moss H., Murphy D.G., Murphy K.C., Owen M.J., Villalon-Reina J.E., Repetto G.M., Roalf D.R., Ruparel K, Schmitt JE, Schuite-Koops S, Angkustsiri K, Sun D, Vajdi A, van den Bree M, Vorstman J, Thompson PM, Vila-Rodriguez F., & Bearden C.E. (2024). **Source-based morphometry reveals structural brain pattern abnormalities in 22q11.2 deletion syndrome**. *Human Brain Mapping*, *45*(1).

2023

Simonson, J.M., **Durdle, C.A.,** Chen, Z., Pansare, N., & Miller, M.B. (2023). **The Effects of Hormonal Contraception on Auditory Emotional Memory**. *Psi Chi Journal of Psychological Research*, *28*(4).

Agrawal, A., **Durdle, C.A.,** Pansare, N., & Miller, M.B. (2023). **The Effects of Oral Hormonal Contraceptives on the Ease of Recall of an Emotional Autobiographical Memory.** *Undergraduate Research & Creative Activity Journal.*

2022

Bobrycki, A.M., **Durdle, C.A.,** & Miller, M.B. (2022). **Increased Negative Affect Decreases Donations in Adults with a Fearful-Avoidant Attachment Style.** *Undergraduate Research & Creative Activity Journal.*

2021

Sønderby, I.E., Ching, C.R., Thomopoulos, S.I., Meer, D., Sun, D., Villalon-Reina, J.E., Agartz, I., Amunts, K., Arango, C., Armstrong, N.J., Ayesa-Arriola, R., Bakker, G., Bassett, A.S., Boomsma, D.I., Bülow, R., Butcher, N.J., Calhoun, V.D., Caspers, S., Chow, E.W., Cichon, S., Ciufolini, S., Craig, M.C., Crespo-Facorro, B., Cunningham, A.C., Dale, A.M., Dazzan, P., de Zubicaray, G.I., Srdjan, D., Doherty, J.L., Donohoe, G., Draganski, B., **Durdle, C.A.,** Ehrlich, S., Emanuel, B.S., Espeseth, T., Fisher, S.E., Ge, T., Glahn, D.C., Grabe, H.J., Gur, R.E., Gutman, B.A., Haavik, J., Håberg, A.K., Hansen, L.A., Hashimoto, R., Hibar, D.P., Holmes, A.J., Hottenga, J., Pol, H.E.H., Jalbrzikowski, M., Knowles, E.E.M., Kushan, L., Linden, D.E.J., Liu, J., Lundervold, A.J., Martin-Brevet, S., Martínez, K., Mather, K.A., Mathias, S.R., McDonald-McGinn, D.M., McRae, A.F., Medland, S.E., Moberget, T., Modenato, C., Sánchez, J.M., Moreau, C.A., Mühleisen, T.W., Paus, T., Pausova, Z., Prieto, C., Ragothaman, A., Reinbold, C.S., Marques, T.R., Repetto, G.M., Reymond, A., Roalf, D.R., Rodriguez-Herreros, B., Rucker, J.J., Sachdev, P.S., Schmitt, J.E., Schofield, P.R., Silva, A.I., Stefansson, H., Stein, D.J., Tamnes, C.K., Tordesillas-Gutiérrez, D., Ulfarsson, M.O., Vajdi, A., van 't Ent, D., van den Bree, M.B.M., Vassos, E., Vázquez-

Bourgon, J., Vila-Rodriguez, F., Walters, G.B., Wen, W., Westlye, L.T., Wittfeld, K., Zackai, E.H., Stefánsson, K., Jacquemont, S., Thompson, P.M., Bearden, C.E., & Andreassen, O.A. (2021). **Effects of copy number variations on brain structure and risk for psychiatric illness: Large-scale studies from the Enigma working groups on CNVs.** *Human Brain Mapping*, *43*(1), 300–328.

2020

Ching, C. R. K., Gutman, B. A., Sun, D., Villalon Reina, J., Ragothaman, A., Isaev, D., Zavaliangos-Petropulu, A., Lin, A., Jonas, R. K., Kushan, L., Pacheco-Hansen, L., Vajdi, A., Forsyth, J. K., Jalbrzikowski, M., Bakker, G., van Amelsvoort, T., Antshel, K. M., Fremont, W., Kates, W. R., Campbell, L.E., McCabe, K.L., Craig, M.C., Daly, E., Gudbrandsen, M., Murphy, C.M., Murphy, D.G., Murphy, K.C., Fiksinski, A., Koops, S., Vorstman, J., Crowley, T.B., Emanuel, B.S., Gur, R.E., McDonald-McGinn, D.M., Roalf, D.R., Ruparel, K., Schmitt, J.E., Zackai, E.H., **Durdle, C.A.,** Goodrich-Hunsaker, N.J., Simon, T.J., Bassett, A.S., Butcher, N.J., Chow, E.W.C., Vila-Rodriguez, F., Cunningham, A., Doherty, J., Linden, D.E., Moss, H., Owen, M.J., van den Bree, M., Crossley, N.A., Repetto, G.M., Thompson, P.M., & Bearden, C.E. (2020). **Mapping subcortical brain alterations in 22q11.2 deletion syndrome: Effects of deletion size and convergence with idiopathic neuropsychiatric illness.** *American Journal of Psychiatry*, *177*(7), 589–600.

2019

McCabe, K. L., Popa, A. M., **Durdle, C.**, Amato, M., Cabaral, M. H., Cruz, J., Wong, L., Harvey, D., Tartaglia, N., & Simon, T. J. (2019). **Quantifying the resolution of spatial and temporal representation in children with 22q11. 2 deletion syndrome.** *Journal of Neurodevelopmental Disorders*, *11*(1), 40.

Balachandra, C., **Durdle, C.A.**, & Miller, M.B. (2019). **Effects of Stress on Cognition and Performance (ESCaPe)**. *Undergraduate Research & Creative Activity Journal,* 1, 23-33.

Villalón-Reina, J. E., Martínez, K., Qu, X., Ching, C. R., Nir, T. M., Kothapalli, D., Corbin, C., Sun, D., Lin, A., Forsyth, J.K., Amelsvoort, T., Bakker, G., Kates, W.R., Antshel, K.M., Fremont, W., Campbell, L.E., McCabe, K.L., Daly, E., Gudbrandsen, M., Murphy, C.M., Murphy, D., Craig, M., Emmanuel, B., McDonald-McGinn, D.M., Vorstman, J.A.S., Fiksinski, A., Koops, S., Ruparel, K., Roalf, D.R., Gur, R.E., Schmitt, J.E., Simon, T.J., Goodrich-Hunsaker, N.J., **Durdle, C.A.**, Doherty, J.L., Cunningham, A., van den Bree, M., Linden, D.E.J., Owen, M.J., Moss, H., Kelly, S., Donohoe, G., Murphy, K.C., Arango, C., Jahanshad, N., Thompson, P.M., & Bearden, C.E. (2019). **Altered white matter microstructure in 22q11. 2 deletion syndrome: a multisite diffusion tensor imaging study.** *Molecular psychiatry*, *25*(11), 2818-2831.

2018

Villalon-Reina, J. E., Ching, C. R., Kothapalli, D., Sun, D., Nir, T., Lin, A., Forsyth, J.K., Kushan, L., Vajdi, A., Jalbrzikowski, M., Hansen, L., Jonas, R.K., Amelsvoort, T.v., Bakker, G., Kates, W.R., Antshel, K.M., Fremont, W., Campbell, L.E., McCabe, K.L., Daly, E., Gudbrandsen, M., Murphy, C., Murphy, D., Craig, M., Emanuel, B., McDonald-McGinn, D., Ruparel, K., Roalf, D., Gur, R.E., Schmitt, J.E., Simon, T.J., Goodrich-Hunsaker, N.J., **Durdle, C.A.,** Doherty, J., Cunningham, A.C., van den Bree, M., Linden, D.E.J., Owen, M., Moss, H., Jahanshad, N., Bearden, C.E., Thompson, P.M. (2018). **Alternative diffusion anisotropy measures for the investigation of white matter alterations in 22q11. 2 deletion syndrome.** In *Proceedings of the 14th International Symposium on Medical Information Processing and Analysis* (Vol. 10975, pp. 256-269). SPIE.

Sun, D., Ching, C.R.K., Lin, A., Forsyth, J., Kushan, L., Vajdi, A., Jalbrzikowski, M., Hansen, L., Villalon-Reina, J.E., Qu, X., Jonas, R.K., van Amelsvoort, T., Bakker, G., Kates, W.R., Antshel, K.M., Fremont, W., Campbell, L.E., McCabe, K.L., Daly, E., Gudbrandsen, M., Murphy, C.M., Murphy, D., Craig, M., Vorstman, J.A.S., Fiksinski, A., Koops, S., Ruparel, K., Roalf, D.R., Gur, R.E., Schmitt, J.E., Simon, T.J., Goodrich-Hunsaker, N.J., **Durdle, C.A.**, Bassett, A.S., Chow, E.W.C, Butcher, N., Vila-Rodriguez, F., Doherty, J., Cunningham, A., van den Bree, M., Linden, D.E.J., Owen, M.J., McDonald-McGinn, D., Emanuel, B., van Erp, T.G.M., Turner, J.A., Thompson, P.M., Bearden, C.E. (2018). **Large-Scale Mapping of Cortical Alterations in 22q11.2 Deletion Syndrome: Convergence with Idiopathic Psychosis and Effects of Deletion Size**. *Molecular Psychiatry, 25*(8), 1822-1834.

Zhan, L., Jenkins, L., Zhang, A., Conte, G., Forbes, A., Harvey, D., Angkustsiri, K., Hunsaker, N., **Durdle, C.A.**, Lee, A., Schumann, C., Carmichael, O., Kalish, K., Leow, A., & Simon, T.J. (2018). **Baseline connectome modular abnormalities in the childhood phase of a longitudinal study on individuals with chromosome 22q11.2 deletion syndrome**. *Human Brain Mapping*, *39*(1), 232-248.

2017

Popa, A., **Durdle, C.A**., Morgan, H., Shapiro, H., Niendam, T., Carter, C.S., Luck, S.J., & Simon, T.J. (2017). **Highly Psychosis-Prone Adolescents Show Increased Capture by Distractor Stimuli and More Effort to Inhibit Emotional Stimuli Than Typically Developing Controls.** *Schizophrenia Bulletin*, *43*(Suppl 1), S27.

2016

Simon, T.J., **Durdle, C.**, Garner, J., Popa, A., Hunsaker, N., Hunsaker, M.R., Schumann, C., Lee, A., Kalish, K., Carmichael, O.T., Coronado, R., & Harvey, D. (2016). **Three distinct brain structure patterns as potential biomarkers for behavioral and psychiatric outcomes in chromosome 22q11. 2 deletion syndrome**. *Biological Psychiatry. 79*(9), pp. 45S-45S.

Poster Presentations

Durdle, C.A., Bobrycki, A., Raich Broussi, G., Bulahan, A.J., Chen, Z., Chuey, J., Deen, K.P., Kim, H., Panasare, N., Simonson, J.M., Yu, T., & Miller, M.B. (2023, April). Bribing the Witness? Investigating the Effect of Monetary Incentives as Criterion Manipulations for Freely Recalled Episodic Events. Presented at the 2023 International Conference on Learning & Memory in Huntington Beach, CA.

Durdle, C.A., Leslie, S., Layher, E., Deen, K.P., Simonson, J.M., Eckstein, M., & Miller, M.B. (2023, March). Resting-State Functional Connectivity as a Predictor for Shifting Decision Criterion. Presented at the 30th Cognitive Neuroscience Society in San Francisco, CA.

Walton, B., **Durdle, C.A.**, Whited, B., & Bennahum, N. (2022, November) Examining Key Factors that Help Facilitate Difficult Dialogues on Race in Pedagogical Spaces. Presented at the 2022 Southern California Conferences for Undergraduate Research in Malibu, CA.

Rivera, L., **Durdle, C.A.**, Frizzell, R., & Baylis, K. (2022, August) Criminal Justice & Public Opinion. Presented at the Summer 2022 Center for Science & Engineering Partnership Conference in Santa Barbara, CA.

Durdle, C.A., Areff, S., Bobrycki, A., Chen, Z., Pansare, N., Simonson, J.M., Yoo, J., & Miller, M.B. (2022, May) How do we Criterion Shift Free Recall?: The Difficulty of Systematically Manipulating Criterion for Freely Recalled Episodic Events. Presented at the 2022 Association of Psychological Science (APS) Conference in Chicago, IL.

Simonson, J.M., **Durdle, C.A.,** Chen, Z., Pansare, N., & Miller, M.B. (2022, May) The Effects of Hormonal Contraception on Auditory Emotional Memory. Submitted to be presented at the 2022 Organization for the Study of Sex Differences (OSSD) Conference in Los Angeles, CA.

Agrawal, A., **Durdle, C.A.,** Pansare, N., & Miller, M.B. (2022, May) Effects of Oral Hormonal Contraceptives on the Ease of Recall of an Emotional Autobiographical Memory. Presented at the 2022 Undergraduate Research & Creative Activity (URCA) Conference in Santa Barbara, CA.

Bobrycki, A., **Durdle, C.A.**, & Miller, M.B. (2021, May) From Mom to University: Attachment Style and Autobiographical Recall Do Not Influence Students' Decisions. Presented at the 2021 Association of Psychological Science (APS) Virtual Conference and the 2021 UCLA Psychology Undergraduate Research Conference online.

Layher, E., Abbey, C. K., **Durdle, C.A.,** Leslie S., Santander, T., & Miller, M.B. (2020, November). Simultaneous ROC modeling of prevalence and difficulty across recognition memory tests; the best model is individual specific. Presented at the 61st Psychonomic Society: Virtual Conference.

Layher, E., **Durdle, C.A.,** Leslie, S., Santander, T. & Miller, M.B. (2020, May). Dissociating fMRI activity related to familiarity strength vs. decision criteria during recognition memory. Presented at the 27th Cognitive Neuroscience Society: Virtual Conference.

Balachandra, C., **Durdle, C.A.**, & Miller, M.B. (2019, May). Effects of Stress on Cognition and Performance (ESCaPe). Presented at the 2019 Undergraduate Research & Creative Activity (URCA) Conference in Los Angeles, CA.

Layher, E., **Durdle, C.A.**, Leslie, S., & Miller, M.B. (2019, March). Does the fMRI scanning environment weaken criterion shift stability? Presented at the 26th Cognitive Neuroscience Society in San Francisco, CA.

Durdle, C.A., Morgan, H.L., Popa, A.M., Bassal, A.M., Ferrer, E., Luck, S.J., Bolden, K.A., Niendam, T.A., Carter, C.S., and Simon, T.J. (2018, August) Youth with 22q11.2 Deletion Syndrome are Impaired by Emotional Stimuli in Executive Functioning Tasks. Presented at the 21st Society for the Study of Behavioural Phenotypes (SSBP) International Research Symposium in Melbourne, Australia.

Durdle, C.A., Goldfarb, D., Qin, J., Gonzalez, A., Eisen, M., & Goodman, G.S. (2018, August). From Child Eyewitness to Adult Memory: Identification Responses After 20 Years. Presented at the Bay Area Memory Meeting in Davis, CA.

Durdle, C., McCabe, K.L., Popa, A., Cabaral, M., Wong, L., Harvey, D., & Simon, T. J. (2017, March). Bigger is Better! Spatiotemporal Hypergranularity Theory Explored: Quantifying Spatial and Temporal Awareness in Children with Chromosome 22q11.2 Deletion Syndrome. Presented at the 50th Annual Gatlinburg Conference in San Antonio, TX.

Mayo, D., Niendam, T., **Durdle, C.**, Morgan, H., Carter, C., & Simon, T.J. (2017, March). Are you my friend? Bullying and its association to clinical symptoms among youth with 22q11.2 Chromosome Deletion Syndrome. Presented at the 16th International Congress on Schizophrenia Research Conference in San Diego, CA.

Durdle, C., Popa, A., Harvey, D., & Simon, T. J. (2016, July). Comparing Space and Time: An Exploration of the Spatiotemporal Hypergranularity Theory in Children with Chromosome 22q11.2 Deletion Syndrome. Presented at the 10th Biennial International

22q11.2 Conference in Sirmione, Italy & UC Davis Center for Neuroscience Public Seminar in Davis, CA.

Popa, A., Shapiro, H., Harvey, D., Amato, M., Garner, J., Cruz, J., Cung, N., Reyes, D., **Durdle, C.**, Morgan, H., & Simon, T.J. (2016, July). Children with 22q11.2 Deletion Syndrome Show Lower Spatial and Temporal Acuity than Typically Developing Children in Continuously Varying Tasks. Presented at the 10th Biennial International 22q11.2 Conference in Sirmione, Italy.

Garner, J., Morgan, H., Quintero, A., Popa, A., Harvey, D., Reyes, D., Cruz, J., **Durdle, C.**, Amato, M., Chang, A., & Simon, T.J. (2016, July) Children with Chromosome 22q11.2 Deletion Syndrome Show Common Visuospatial Impairments in Two Analogous Line Bisection Tasks. Presented at the 10th Biennial International 22q11.2 Conference in Sirmione, Italy.

Amato, M., Scott, J., Schuuman, C., Carmichael, O., Hunsaker, N., Garner, J., Popa, A., **Durdle, C.**, Morgan, H., Harvey, D., & Simon, T.J. (2016, July) Anatomical Differences in the Hippocampus may explain Cognitive Ability during Spatiotemporal Tasks in those with 22q11.2 Deletion Syndrome. Presented at the 10th Biennial International 22q11.2 Conference in Sirmione, Italy.

Simon, T.J., Aronson, T., Popa, A., Garner, J., **Durdle, C.**, Hollinger, El, Hunicke, R., Middleton, M. (2015, May). Translating Basic and Clinical Neuroscience into a Spatiotemporal Cognition Therapeutic Video Game. Presented at the Entertainment Software & Cognitive Neurotherapeutics Society (ESCoNS) Conference in San Francisco, CA.

Research Experience

UC Santa Barbara Department of Psychological & Brain Sciences
September 2018 to August 2024
Miller Lab, Santa Barbara, CA
Principal Investigator: Michael Miller, PhD
Graduate Student Researcher

UC Davis MIND Institute
August 2014 to June 2018
22q11.2 Deletion Syndrome (22q) Research Center and Clinic, Sacramento, CA
Director: Tony J. Simon, PhD
Staff Research Associate
August 2017 – June 2018

- Analyzing longitudinal resting-state fMRI connectivity data and psychosis risk of adolescents with and without 22q.
- Collaboration to develop scripts in MatLab and Bash with Dr. Vinod Menon at Stanford University.
- Operate MRI 3T Siemens scanner and ERP Brain Products equipment with adolescent participants.

Junior Specialist (Funded Research Assistant) and Lab Manager
August 2014 – July 2017

- Responsible for daily lab operations.
- Coordinate use of lab resources for outside researchers, such as Eye Tracking equipment and testing rooms.
- Oversee Institutional Review Board (IRB) documents for six total studies.
- Direct, train, and supervise Undergraduate Volunteer Interns and incoming staff.
- Organize public events (i.e., 22q At the Zoo, MIND Institute Participant Appreciation Day, etc.).
- Recruit and enroll participants for four studies, including those with 22q11.2DS, Trisomy X, and Klinefelter Syndrome.
- Administer research clinical assessments, computer, and paper based tasks for six total studies.

- **Cognitive Affective Risk and Protection for Psychosis Study (PI: Tony J. Simon, PhD):** Investigate psychosis proneness in adolescents ages 12- to 18-years-old with chromosome 22q11.2 deletion syndrome (22q11.2DS) through clinical assessments, cognitive testing, ERP, MRI, and physiological measures.
  - Operate MRI 3T Siemens scanner and ERP Brain Products equipment with adolescent participants, many of whom have anxiety diagnoses.
  - Compose Institutional Review Board (IRB) documents for initial and subsequent reviews.
  - Design and administer research tasks using E-prime, including a visually cued card sort that measures interactions of emotion and task switching.
  - Create data analysis processing scripts in R software such as ANOVA and Correlation Coefficients.

- **Parenting and Adaptive Functioning in Children Study (PI: Kathleen Angkustsiri, MD):** Explore how parent-child relationships influence emotion development and independent living skills in children ages 4- to 11-years-old both with and without 22q11.2 Deletion Syndrome.
  - Recruit and administer testing to participants of 60 child-parent dyads with and without 22q11.2DS.
  - Collect anxiety symptom, physiological, and behavioral data from child parent interactions.
  - Organize clinicians' schedules to meet with families to provide additional testing and feedback.
  - Analyze interrater reliability of Undergraduate Coding team using SPSS.

- **22q11.2DS Brain and Behavior Consortium:** Compile DNA and medical data from previous participants to contribute to a fourteen-university repository to investigate the prevalence of schizophrenia in children and adults with 22q11.2DS.
  - Recruit previous participants to share their DNA samples with the consortium and coordinate staff for shipment.
  - Manage RedCap database to enter behavioral, medical, and cognitive data points for 68 participants, as well as create comprehensive clinical case summaries.

UC Davis Department of Psychology
July 2015 to September 2018
Developmental Research Center, Davis, CA
Principal Investigator: Gail Goodman, PhD
Research Assistant

- **Child Maltreatment and Long Term Memory Study:** Focus on individuals who experienced trauma or suspected abuse during the 1990's and their memories of these events.
  - Interview approximately 25 participants over the phone for either one of two sessions, the second session using Cognitive Interview or Federal Law Enforcement Training methods regarding childhood maltreatment.
  - Search for previous participant contact information by using Spokeo, LexisNexis, and social media sources.
  - Analyze and organize data in SPSS and Microsoft Excel.

UC Davis MIND Institute
April 2014 to August 2014
Social Attention Virtual Reality Laboratory, Davis, CA
Principal Investigator: Peter Mundy, PhD
Research Assistant

- **Virtual Reality Applications for Attention and Learning in Children with Autism and ADHD:** Worked on the second time point in a longitudinal study focusing on social-cognitive and academic difficulties in children with Autism, ADHD, and dual diagnoses, as well as typically developing controls.
  - Administered child and parent assessments, including Theory of Mind computer activities.
  - Performed general lab organization tasks such as participant packets and data entry in Microsoft excel.
  - Reached out to families to maintain contact.

UC Davis Center for Mind and Brain
June 2013 to August 2014
Memory and Development Lab, Davis, CA
Principal Investigator: Simona Ghetti, PhD
Research Assistant

- **Development of Prospection about Future Events study:** Investigated the relationship between episodic thinking and future prospection in children ages 5-, 7-, 9-, and 11-years-old, as well as college age adults.
  - Administered experimental tasks and interviewed participants for episodic memories.
  - Transcribed episodic memory interviews.
  - Coded transcriptions based on experimenter prompting verbiage and other rules.
  - Tagged white and gray matter abnormalities using FreeSurfer brain tracing software.

Volunteer Experience

University of California, Santa Barbara, Santa Barbara, CA
March 2022 to June 2023
College of Letters & Sciences Faculty Executive Committee
Graduate Student Representative
The Faculty Executive Committee (FEC) is responsible for coordinating the academic affairs of the university and overseeing the welfare of its students. The FEC regularly opines on curricular issues, authorizes new courses, course revisions, and minor changes to degree programs (majors and minors), and reviews significant changes to degree programs.

University of California, Santa Barbara, Santa Barbara, CA
October 2021 to February 2024
Department of Psychological & Brain Sciences Graduate Admissions Committee
Graduate Student Representative
Participated in outreach efforts to prospective Ph.D. candidates; advised on the allocation of recruitment resources and recruitment slots to faculty; assisted in planning for the recruitment weekend for the department.

University of California, Santa Barbara, Santa Barbara, CA
May 2021 to September 2022
Queer & Trans Graduate Student Union (QTGSU)
Treasurer
The purpose of the QTGSU organization is to provide a safe space for queer and trans identified UCSB graduate students to socialize, organize, and seek support. Our mission is to advocate for the rights and well-being of LGBTQQIA+ identified students and will provide mentorship to undergraduate students. Furthermore, the QTGSU will work to foster ties with the LBGTQQIA+ community in the greater Santa Barbara area. Assisted the President and Vice President in the procurement of funding for all projects, programs, and activities and handled all financial affairs and budgeting of the organization.

WEAVE Inc., Sacramento, CA
July 2015 to March 2017
Support Line Volunteer & Sexual Assault Response Team (SART) Advocate
Completed 70 hours of California Peer Counselor training to prepare for working with survivors of domestic violence and sexual assault; provide emotional and informational support over the phone to residents of Sacramento County; connect survivors (of sexual assault, domestic violence, and other trauma) to advocate response teams; accompany survivors of sexual assault to evidentiary exams and inform them of available resources provided by the State and WEAVE.

Teaching Experience

University of California, Santa Barbara, Santa Barbara, CA
June 2022 to Present
Institute for Social, Behavioral and Economic Research (ISBER)
McNair Scholar and Edison STEM Graduate Student Mentor & Writing Specialist
As a Graduate Student Mentor, I prepared qualified undergraduates for entrance to graduate programs in all fields of STEM studies. The goals of the McNair scholars and Edison STEM programs are to increase the number of first-generation, low-income, and/or historically underrepresented students in doctoral programs to ultimately help diversify the faculty in colleges and universities across the country. I led courses, seminars, and workshops on topics related to graduate school preparation, and gave students individualized guidance to help them complete a research project with the goal of presenting at local, regional, and national conferences.

University of California, Santa Barbara, Santa Barbara, CA
Founder and Director, Preparing Undergraduates for Graduate School (P.U.G.S.)
P.U.G.S. is a free program I created for UCSB affiliates with the goal of helping students navigate the graduate school application process and prepare to take the GRE. Specifically, PUGS is designed to support and advocate for underrepresented students and promote a more diverse and inclusive academic community.

- **Summer Graduate School Workshops**
  July to November 2022
    - Organized a series of in-person and virtual workshops to aid underrepresented students with the process of applying to graduate school.
    - Topics included graduate school application timelines, information on how to find programs, developing school lists, writing C.V.s, how to network, learning to use social media in academia, statements of purpose, personal history statements, diversity statements, requesting letters of recommendation, and how to find funding for graduate school.
    - Mediated student presentations regarding their work and subsequent feedback from me and their peers.
    - Met with students individually to give more personal feedback and guidance.

- **Ten-day Summer Graduate Readiness Exam (GRE) Bootcamp**
  June 21 to July 6, 2021
    - Organized a Graduate Readiness Exam (GRE) 10-day Bootcamp for over 170 students via Zoom.
    - Awarded an Associated Students Community Foundation grant to fund instructors for the Bootcamp ($5400)
    - Collaborated with other UCSB entities such as the Campus Learning Assistance Services (CLAS), Career Services, Graduate Division, the McNair Scholars Program, the Undergraduate Research & Creative Activities (URCA) program, and Opening New Doors to Accelerate Success (ONDAS) to gather more resources and give presentations to students to prepare them with the graduate school application process.

- **Winter 2019 GRE Bootcamp**
  December 2018 to March 2019
    - Organized an in-person GRE Bootcamp for over 70 students.
    - Connected students with other UCSB entities such as the Education Opportunity Program (EOP) and ONDAS to help them secure funding and other resources for graduate school applications and taking the GRE.
    - Created a resource bank of GRE and Graduate Application materials for students in the program and for those who were unable to attend the Bootcamp.
    - Tutored small groups and individual students on GRE strategies and graduate school applications.

University of California, Santa Barbara, Santa Barbara, CA
Summer & Fall of 2020, 2021, 2022, 2023
Graduate Division
National Science Foundation (NSF) Graduate Research Fellowship Writing Consultant

- Met with individual students to review and provide feedback on their NSF GRFP applications, specifically the student's personal statement and research project proposal.

University of California, Santa Barbara, Santa Barbara, CA
Teaching Assistant, Psychology 120L: Advanced Research Method
Fall 2019 & Spring 2020
Teaching Assistant, Psychology 10A: Research Methods
Fall 2018 & Winter 2019

- Led classes of 5-30 undergraduate students in lessons designed by the instructor.
- Graded written homework assignments, in class activities, and a final research paper.
- Tutored small groups and individual students.

UC Santa Barbara Department of Psychological & Brain Sciences
September 2018 to June 2023
Miller Lab, Santa Barbara, CA
Graduate Student Mentor for Undergraduate Honors Thesis Students

Jasmine Chuey, 2022/2023,
- "The Effects of Social Pressure on Criterion Shifting."

Aarushi Agrawal, 2021/2022,
- "Effects of Oral Hormonal Contraceptives on the Ease of Recall of an Emotional Autobiographical Memory."

Jessica Simonson, 2021,
- "The Effects of Hormonal Contraception on Auditory Emotional Memory."

Ana Bobrycki, 2020/2021,
- "From mom to university: attachment style and autobiographical recall do not influence students' decisions."

Chinmayee Balachandra, 2018/2019,
- "Effects of stress on cognition and performance."

Workshops & Training

4T|Phys BIOPAC Human Physiology Conference (2023, July) held at UC Santa Barbara by Frazer Findlay, CEO, BIOPAC

Kavli Summer Institutes in Cognitive Neuroscience (2022, June) held at UC Santa Barbara by George (Ron) Mangun, Ph.D., Barry Giesbrecht, Ph.D., Michael Miller, Ph.D., & Michael Gazzaniga, Ph.D.

Weil Cornell Summer Institute on Psychology and the Law (2016, July) held at the Weil Cornell Medical College in New York, New York by BJ Casey, Ph.D.

ERP Bootcamp (2015, July) at the UC Davis Center for Mind and Brain by Steven J. Luck, Ph.D. & Emily Kappenman, Ph.D.

Administration Workshop for the Structured Clinical Interview for DSM-IV & Structured Interview for Prodromal Symptoms (2015, July) held at the UC Davis Imaging Research Center by Tara Niendam, Ph.D. & Tyler Lesh, Ph.D.

Cognitive Interview Training (2015, August) at UC Davis by Gail Goodman, Ph.D., Deborah Goldfarb, J.D., Ph.D., & Ronald Fisher, Ph.D.

Assessments trained by Clinical Staff:

- Wechsler Abbreviated Scale of Intelligence I & II (WASI-I; WASI-II) – Total approximately 100 hours.
- Structured Clinical Interview for DSM-V – Total approximately 50 hours.
- Anxiety Disorders Interview Schedule – Parent (ADIS) – Total approximately 70 hours.
- Diagnostic Interview Schedule for Children (DISC-IV) – Total approximately 20 hours.

Additional Skills

<u>Software and Programs:</u> Matlab, R, Python (Code School), PsychoPy, OpenSesame, Qualtrics, Airtable, REDCap, PyCorder, Brain Products ERP System, Microsoft Access, DataGraph, E-Prime, SPSS.
<u>Physiology Recording Software:</u> Nexus and BIOPAC.
<u>Scoring Software:</u> ABAS, BASC.

ABSTRACT

Exploring Criterion Shifting with Ecologically Valid Contexts:

An Investigation of Recollection Processes and the Impact of Social External Factors on

Decision-Making Strategies for Memory-Based Tasks

by

Courtney A. S. Durdle

The ability to accurately recall rich episodic memories and the decision-making strategies used in memory-based tasks are critical in many real-world contexts, such as legal testimony, healthcare, and education. Criterion shifting is a decision-making strategy that adjusts decision thresholds based on individual task demands and has been found to be a uniquely stable cognitive trait. The goal of this research was to systematically explore criterion shifting behavior in ecologically valid scenarios, focusing on recollection processes and the influence of social external factors.

The research begins with an ecologically valid scenario in Experiment 1 (Chapter 2), where criterion manipulations were introduced during an interview to examine how participants adjusted their recall strategies after a stressful autobiographical episodic event. The findings revealed that conservative instructions led to stricter decision criteria and more selective recall, whereas liberal instructions did not significantly lower the already liberal baseline criterion. Given these results, the subsequent experiments aimed to systematically examine how criterion shifting functions in both controlled laboratory settings and more

ecologically valid scenarios, providing insights into how individuals adapt their decision-making strategies across different contexts.

Experiments 2 and 3 (Chapter 3), in a controlled laboratory setting, isolate the components of social context and criterion shifting. Experiment 2 investigates how competition and public rankings influence criterion shifting during recognition memory tasks, while Experiment 3 considers how collective outcomes influence performance in a group-based task. Neither experiment found significant impacts from social pressure on criterion shifting tendencies, supporting the view that such tendencies are stable cognitive traits. However, Experiment 3 revealed that collective outcomes can enhance discriminability when tasks are relatively easy, suggesting that social influences may improve memory performance under favorable conditions.

Experiment 4 (Chapter 4) focused on understanding recollection processes through a within-subject design that systematically increased reliance on recollection across image recognition and word cue tasks. This study tested predictions from the Dual Process Signal Detection (DPSD) model, which views recollection as a thresholded process, and the Unequal Variance Signal Detection (UVSD) model, which treats recollection as a continuous process with Gaussian memory strength distributions. The results revealed consistent criterion shifting across both tasks, indicating that participants employ a generalized decision-making strategy that adapts to varying cognitive demands. The linearity of the z-transformed ROC curves further supports the UVSD model's predictions, which challenges the DPSD model's assumption that recollection is a discrete process.

Overall, this dissertation contributes to a broader understanding of how criterion shifting operates in diverse contexts, providing insights into the stability of this cognitive trait across different memory tasks and the extent to which social factors and recollection processes influence decision-making. The findings underscore the need for future research to further explore criterion shifting with ecological validity in mind.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

"This research was not carried out by uneducated or incompetent people; some of the best minds in psychology have worked and are presently working in the area of memory. Why, then, have they not turned their attention to practical problems and natural settings?" This quote by Ulric Neisser, the "father of cognitive psychology," demonstrates his zealous advocacy for ecological validity in psychological experiments (Neisser, 1982). Neisser gained recognition for his focus on memory within real-world settings, notably through his well-known case study of John Dean, President Richard M. Nixon's former White House Counsel, who testified to the Senate Watergate Committee regarding his role in the Watergate scandal. Neisser compared Dean's testimony to tape recordings from the Oval Office and used the many discrepancies he found to shed light on how malleable memory can be. Neisser noticed that Dean reported events more accurately if the same type of event happened more than once. However, the unique details of particular events were often wrongly reported as occurring during a different but related event. This phenomenon led Neisser to coin the term "repisodic" memory to describe a generic memory resulting from a mixture of repeated memories (Neisser, 1981). While Neisser's research shows how Dean distorted his own memory, it fails to touch on how the persistent questioning of the Senate Watergate Committee could have influenced Dean's memory.

In June 1973, John Dean was doggedly questioned by Republican Senator Howard Baker to determine precisely what happened during the meetings Dean had with President Nixon. At the beginning of his testimony, Dean relayed information about what occurred in President Nixon's office as if he were a tape recorder, describing "exact" details about various events and providing the "exact" words and phrases people said in conversation. In one instance, Dean used the phrase “I recall the President very clearly saying to me…” in responding to a question about a conversation he had. Senator Baker was determined to get as accurate information as possible from Dean, asking him repeatedly to be more specific with his reporting. By pressing Dean to be more precise, Senator Baker was essentially prompting Dean to be more cautious and strict with how he recalled information about what he had experienced. Dean reacted to Senator Baker’s more intense questioning by providing more conservative responses, such as stating that he was making inferences or could not remember exactly what happened. Testimony, or relaying any memory to another person, involves making a decision about what information to report. Each individual uses a unique threshold, or *criterion*, to determine the amount of information they need to feel comfortable sharing a given memory. Dean’s testimony is a prime example of how the criterion shifting process can be manipulated to change how people recall events they experienced.

In many scenarios, such as eyewitness accounts, the context can be dynamic, and the consequences of a decision error can be severe, which makes having the cognitive flexibility to optimize decisions very beneficial. This ability to adapt to the most optimal decision strategy is known as *criterion shifting*. Extensive research has shown that people performing memory recognition tasks can be easily influenced to shift their decisions to be more liberal or conservative (depending on what is most optimal) when reporting whether they have seen something before. What deserves further study is the extent to which a witness freely recalling their experience in an ecologically valid setting can optimize their memory-based decision-making in the same way.

1.1 Episodic Memory and Free Recall

The term “memory” has frequently been described as any state that results from the linear process of encoding, storage, and retrieval. However, this perspective, titled the “received view,” is too broad and can apply to almost all human executive functioning. More updated definitions of memory describe it as an autonoetic experience, where the emphasis is on reliving an experienced event from the individual’s point of view (Tulving, 1984). This sense of self and spatial and temporal information from the remembered event are the three necessary components of episodic memory (Tulving, 2002). This experience of temporality (reacquainting oneself with one’s own *past* experiences) differs from simply knowing facts about the past, such as semantic information (i.e., referencing past information in the present). Thus, reliving an episodic memory should be considered the true definition of “memory,” not the more all-encompassing declarative definitions of memory—such as procedural or semantic—often used in today’s literature (Klein, 2015).

Recognition memory and cued recall are different methods by which to probe memory. Early work in the memory literature by Hollingworth (1913) compared recognition memory to recall and found that the primary difference is that, for recall, the individual is provided the context and then must retrieve the target information, while in recognition, it is the opposite (the individual is provided the target and then must retrieve the context). Free recall tasks allow participants to recall items in any order and with as much or as little detail as they choose. This method of studying memory tracks back to Ebbinghaus’s time when subjects would be asked to recall word lists (Ebbinghaus, 1913). In the early 1970s, scientists such as Kintsch (1970) and Anderson and Bower (1972) theorized that recall, in general, involves two stages: first, a person generates knowledge about something, then later, that person goes through a recognition process about that knowledge. This is known as the generate-and-recognize model. In a typical study under this model, it is believed that when recalling an item, a subject will first begin a search process and bring items to mind, then go through a recognition process where they will determine if the item is from the original task. In other words, generation involves recalling details or information, while recognition is a decision process that excludes generated details or information from being reported.

Jacoby and Hollingshead (1990) later revised the generate-and-recognize model to specify that people may bypass the recognition assessment if they can generate what they recall relatively easily, quickly, or fluently. Task design could play a significant role in the memory-based decision process, not only for the accuracy and number of items or details but for the intentional effort given to the task. It is important to point out that this revised generate-and-recognize model is based on a task that allows participants to rely on perceptually driven cues (e.g., word stem test cues) and thus allows participants to base their decisions on their feelings of ease, speed, and fluency. Most free recall tests involve less perceptual memory-based evidence and more conceptual information (Roediger et al., 1989). These two cognitive processes rely on different variables to trigger memory-based behaviors. The perceptual approach is cued by the appearance of the recalled item, while the conceptual approach concentrates on the meaning of the item. In their study, Guynn and McDaniel (1999) updated the design of Jacoby and Hollingshead's (1990) task by creating semantic category label cues in order to be more conceptually than perceptually driven (e.g., perceptual: kan_ for kangaroo vs. conceptual: animal category for kangaroo). They found that conceptually-based free recall tasks sometimes showed results in line with the revised model's specification of bypassing the recognize phase and directly retrieving the information. Under other circumstances, the participants would rely on the more reflective generate-and-recognize strategy. Ultimately, with these conflicting results, they concluded that free recall involves a variety of retrieval strategies that can be employed depending on the requirements of the recall task. While these studies focus on the original and revised generate-and-recognize models, there may be other means or methods that a participant may utilize to make memory-based decisions.

Signal Detection Theory (SDT) is one such prominent theoretical framework for understanding how people make decisions. This theory has been applied to decision-making in many contexts, including medicine (to evaluate competing diagnostic tests), machine learning (to evaluate competing pattern-recognition algorithms), and weather forecasting (to test competing weather prediction models) (Wixted, 2020). Researchers have also applied SDT to the study of memory, specifically to people's decisions regarding whether they remember an event. SDT is especially applicable to recognition memory research, as the participant is attempting to decipher whether a stimulus is "old" (acting as the target/signal) or "new" (acting as the lure/noise). How well a participant distinguishes between the signal and noise of the stimuli is known as the strength of discriminability, or *d'*. In their studies designed to test memory and decision-making under a SDT framework, researchers have judged participant responses to recognition memory tasks by categorizing them as hits, misses, correct rejections, or false alarms. If a participant correctly identified a piece of studied or previously seen information as "old" (i.e., reporting that they remembered it), that response would be considered a "hit," and if they incorrectly identified that information as "new," then the response would be deemed a "miss." If a participant correctly identified new information (also known as a lure) as "new," the response would be categorized as a "correct rejection," while if they incorrectly designated it as remembered information, then this response would be regarded as a "false alarm."

SDT provides a possible explanation for the decision-making process that people go through when freely recalling a rich episodic memory. In their 2008 study, Wright and colleagues examined how memory recall is affected by varied response criteria. First, participants were shown four complex still images and then given lenient or strict recall instructions. These instructions asked participants not to worry about mistakes or guessing (the liberal/lenient criterion) or to report items if they were sure their response would be accurate (the conservative/strict criterion). They found that those with liberal/lenient instructions recalled more accurate details than those with conservative/strict instructions but also had higher rates of reporting inaccurate information. The conservative/strict instructions were found to reduce the amount of inaccurate information reported, and the number of accurate details recalled.

Some scientists, such as Roediger and Payne (1985), argue that recall accuracy is independent of criterion thresholds. In their experiment, Roediger and Payne (1985) wished to determine if the accuracy of recalled items would increase if participants relaxed their criterion threshold. They tested this using three different levels of criterion thresholds based on recall type: (1) uninhibited (guessing allowed - liberal/relaxed threshold), (2) free recall (no guessing allowed - middle/mixed threshold), and (3) forced (having to write a certain number of words - conservative/strict threshold). Their results showed that having a more conservative criterion threshold did not mean that subjects had lower rates of recall. In fact, there was no significant difference in accuracy between any of the conditions. More generally, they held that there is no criterion effect in any recall task. However, it is possible that Roediger and Payne did not sufficiently motivate their participants to shift their criteria. Influencing participants to act more conservatively or liberally in their memory reports without more explicit direction likely reduces the extent to which a person may shift, while greater criterion shifting is likely to be shown with increased motivation. In Rhodes and Jacoby's (2007) study on criterion shifting in recognition memory, they ensured that participants knew the probability of the correct answers they would use, resulting in the criterion shift behavior. When the researchers were not explicit in pointing out the probability changes to participants, the subjects did not show a criterion shift.

Miller and Wolford (1999) argue that the recall of the critical lure in a word list recall task is based on a criterion shift. They proposed that participants who showed high false alarm rates for critical lures likely had a more liberal criterion threshold for items related to the word lists' theme. In their 2011 paper, Miller & Wolford followed up on their previous research, specifically arguing that a participant's criterion will be more liberal for any item related to the gist of the list items (including the critical lure). They found that participants believed that if they relied on the memory gist for the list of words, knowledge of the gist would help them decide if they had seen an item before. This belief resulted in high memory errors. If participants were warned not to rely on the gist memory, they showed significantly fewer memory errors. This study provides evidence for the idea that participants are able to control their responses strategically and have the potential to mitigate false alarms. If the experimental task is designed to provide participants with explicit information on optimizing performance, then they would not necessarily rely on the gist of a memory to make a decision. Instead, they would be forced to acknowledge a decision strategy that likely depends on generated recollections.

1.2 Criterion Shifting

According to SDT, decisions based on memory can be especially difficult because memories can be vague and ambiguous (i.e., the strength of discriminability is low), so people will set a standard of evidence, called a decision criterion threshold, on a spectrum of memory strength (Hirshman, 1995; Macmillan & Creelman, 2005). Criterion shifting is the ability to adaptively shift the placement of this decision threshold. In some situations, the need to rely on criterion shifting to make a decision is not entirely necessary due to high discriminability strength (*d'*), which is the increased ability to discriminate between signal and noise. In other words, as the strength of discriminability increases, the need for criterion shifting decreases (Macmillan & Creelman, 2005). Under other decision-making circumstances, where the strength of discriminability is low, an individual may rely on relatively weaker memory evidence (i.e., relying on a more lax or liberal criterion), while others would only accept strong memory evidence (i.e., acting more conservatively) (Miller & Kantner, 2019). It has been established from recognition memory studies that criterion placement and shifting are considered to be two independent behaviors (Kantner & Lindsay, 2012; Layher et al., 2020). SDT defines strategic criterion shifting as an intentional decision strategy that requires a person to be more aware of the relative rewards or consequences of different response types as situations change (Macmillan & Creelman, 2005). Shifting decision criteria involves proactive control, which is when a person maintains a particular goal to optimize memory-based decisions. Explicit information is provided to the individual so they may form this decision strategy that will provide them with the greatest payoff. Criterion shifting is a particularly important strategy to maximize potential benefits when there is uncertainty in the detected signal.

Much of the previous research and theories developed around criterion shifting behaviors have relied on group-averaged analysis, which has only highlighted the between-subject variability (Ulehla, 1966; Parks, 1966; Thomas & Legge, 1970; Kubovy, 1977; Hirshman, 1995; Maddox & Bohil, 2005; Benjamin et al., 2009; Lynn & Barret, 2014; Layher et al., 2020). More recent works have systematically evaluated these individual differences as well as explored the within-subject stability of criterion shifting habits. Results from these studies have provided support for the idea that this behavior is a uniquely individualistic and consistent cognitive trait. While some people routinely shift their criteria to extreme degrees across situations, others do not tend to shift their criteria at all (Aminoff et al., 2012, 2015; Kantner et al., 2015; Frithsen et al., 2018; Layher et al., 2018; Miller & Kantner, 2019; Layher et al., 2020; Layher, Santander et al., 2023). Many studies have attempted to find predictors for criterion shifting tendencies and have found very few significant correlations (Aminoff et al., 2012, 2015; Kantner et al., 2015; Kantner & Lindsay, 2012, 2014; Frithsen et al., 2018; Layher et al., 2018; Miller & Kantner, 2019; Layher et al., 2020). Individual strategic criterion shifting tendencies have been shown to be largely consistent across different task stimuli (i.e., word vs. facial recognition; Aminoff et al., 2012; Kantner et al., 2015; Frithsen et al., 2018), stable across criterion bias manipulations (i.e., base rates of targets and nontargets shown vs. monetary incentivization; Kantner et al., 2015; Rhodes & Jacoby, 2007; Frithsen et al., 2018; Layher et al., 2020), and constant in various decision domains such as recognition memory, visual detection, or visual discrimination judgments (Frithsen et al., 2018; Layher et al., 2020).

These individual differences in strategic criterion shifting do not necessarily indicate that some people are incapable of criterion shifting. On the contrary, certain individuals appear *unwilling* to shift their criteria (Miller & Kantner, 2019). For example, Layher and colleagues (2020) conducted a recognition memory experiment in which participants simply rated their confidence with low, medium, or high confidence on each test trial. Participants received instructions in a separate memory test to alter their criteria based on their confidence levels. In the conservative criterion condition, participants were instructed only to respond "old" when they had high confidence an image was old; otherwise, they were to respond "new." The liberal criterion condition had the opposite instructions, in which participants should only respond "new" when they had high confidence that an image was new. For most subjects, the extent of criterion shifting proved to be much smaller than the extent to which they established criteria for high confidence in the confidence rating tasks. This demonstrates that participants are *capable* of shifting to more extreme extents if they simply use the same high confidence thresholds in the confidence ratings task as they do for the criterion shifting task, but individuals appear unwilling to do so. While some individuals shift well by optimizing their decisional outcomes, others are considered poor shifters because they rely too heavily on their sense of familiarity, which typically leads to suboptimal outcomes.

When participants were given more general recognition judgment instructions, such as to base their decisions on high confidence only (or 100% confidence), they shifted to a lower degree compared to when the confidence measurement was integrated into the decision (a multipoint scale with varying degrees of confidence for an "old" and "new" choice; Mickes et al., 2017; Layher et al., 2020). Researchers have suggested that strategic criterion shifting (adapting one's standard of evidence to the specific demands of the situation) may result from a participant's willingness to employ an explicit strategy and optimize their outcome (Green & Swets, 1966; Wixted & Stretch, 2000; Aminoff et al., 2012; Kantner et al., 2015; Miller & Kantner, 2019; Layher et al., 2020; Layher, Santander et al., 2023). When making a memory-based decision, a higher willingness to consider criterion shift strategies instead of relying on weak memory evidence may explain why some people are more apt at shifting than others. The literature holds that people do not shift when they are more focused on the memory discrimination component of a task than on placing their criterion threshold (Wickelgren, 1967; Macmillan & Creelman, 2005; Aminoff et al., 2012).

Past research has evaluated the sensitivity needed from a criterion shifting manipulation in order to affect participant decisions. Manipulations can be implemented explicitly (i.e., directly communicating the circumstances surrounding a decision; Banks, 1970; Healy & Kubovy, 1978; Rotello et al., 2005; Aminoff et al., 2012) or implicitly (i.e., providing feedback for reinforcement learning; Wixted & Gaitan, 2002; Han & Dobbins, 2008, 2009). When using implicit means to influence criterion shifting behavior, previous studies have utilized trial-by-trial feedback to either encourage more liberal or conservative decisions during recognition recall (Han & Dobbins, 2008). This feedback may be true or false depending on how liberal or conservative the experimenter wants the participant to be. Studies that employ reinforcement learning methods with feedback may improve criterion shifting behavior in “poor shifters” by helping them “calibrate” their reliance on feelings of familiarity and counteract its potential power on the decision-making process. Because of this feedback, participants are more likely to “think twice” before basing their decision solely on their level of familiarity. However, with false feedback designs, this implicit method can take hundreds of trials before a participant demonstrates a shift in criterion placement. Explicit instructions, on the other hand, take no training or sequence of trials for the manipulation to occur. In summary, having implicit or explicit instructions will highly influence how quickly a participant is manipulated. Due to the fact that the convenience of explicit instructions is highly valued, proactive block designs will be used to explore the facets of criterion shifting under more real-world circumstances and with free recall.

1.3 Criterion Manipulations and False Memories

Decades of research have shown that memories are malleable. However, it is challenging to determine the extent of how malleable memory truly is and how easily it can be distorted. One of the primary reasons why researching this topic is difficult is that there are so many potential confounds surrounding these possible distortions to account for, such as how the memory is framed during reporting or whether misleading information is introduced. These questions are particularly prevalent in the study of autobiographical memory.

Many research studies have focused on paradigms to measure the rate of recall, but these studies can have an alternative interpretation. Their designs can be seen as manipulations of criterion thresholds that affect recall, especially studies done using hypnosis (Dywan & Bowers, 1983; Whitehouse et al., 1988; Klatzky & Erdelyi, 1985). Researchers have found that participants under hypnosis recalled more items and classified them as memories with higher confidence ratings (for both correct and false memories; Dywan, 1988). This frequency of recall and greater confidence may be because hypnosis alters a participant’s retrieval experience. Many subjects experience an “illusion of familiarity” during hypnotic retrieval, meaning the level of perceptual fluency and vividness is on par with remembering normally (Dywan, 1995). Whitehouse, Dinges, Orne, and Orne (1988) substantiated this phenomenon in their study, where they had participants view videos and then split them into control and hypnosis groups to recall details from the video. They found that accuracy was no different for either condition, but those under hypnosis tended to report higher confidence for guessed responses. Signal detection theory provides a possible explanation for this, suggesting that many of these distortions can be attributed to people's decision-making process when recalling a memory. Here, the confidence for guessed items may be due to hypnosis, making a person more likely to accept vague details or inferences as true, which would lower or relax the criterion thresholds participants use to report a memory.

Signal detection theory may also provide an alternative explanation for the changes in participants' free recall reporting after being exposed to misleading post-event information. There is an ongoing debate about the "misinformation effect," a phenomenon that Loftus, Miller, and Burns (1978) found in their study of how event information supplied after the event impacts a person's memory of that event. Loftus and her colleagues found that misleading information provided to participants post-event, such as altering a detail in the memory or implanting new information, could impair what they recalled. They determined that this misinformation could result in witnesses developing very richly detailed false memories, a finding that has direct ties to the efficacy of eyewitness testimony. This research led Loftus to develop the alteration hypothesis that post-event information irreparably changes, or even replaces, an original memory (Loftus, Miller, & Burns, 1978; Loftus, 2005). However, some researchers believe Loftus's findings are too broad and that the type of post-event information matters as to whether that information impacts the original memory trace (McCloskey & Zaragoza, 1985).

McCloskey and Zaragoza (1985) argue against the alteration hypothesis and point out major methodological flaws in Loftus's work on this topic, such as having subjects make a forced choice decision between the original and the misleading information. McCloskey and

Zaragoza redesigned the study to give participants the choice between the original and unencountered information, leaving out the misleading information in the forced choice task. The results showed that participants would more often than not choose the original piece of information, thus showing that the original memory trace is still available even though the participant was exposed to post-event information. Their study, along with others (Dodd & Bradshaw, 1980), showed that a post-event effect on memory depends on the types of test, original information, and source of misleading information. They further state that misleading information affects participants by biasing them toward the new information, especially if the original memory was weak. They posit that misleading information does not necessarily change a participant's memory but provides more evidence for a participant to report that false item as a memory. This supports the theory that while memories do not change, decision criterion thresholds do shift.

While Loftus and McCloskey, in this debate, do not explicitly investigate the impact of criterion thresholds on memory, it is possible that the effects of misinformation they witnessed came from a change in the participant's criterion. Providing a participant with misleading post-event information is similar to lowering a participant's inhibitions under hypnosis - in both scenarios, the experimenter is giving the participant a reason to change their criterion threshold. Post-event information is functionally equivalent to any other information that a person typically draws on when deciding to report on what they recall. This information may be used as new evidence that moves a participant towards or away from their already established criterion threshold.

Another primary method of looking at false memory effects is the Deese, Roediger, and McDermott (DRM) paradigm. Under this method, researchers instruct subjects to memorize a list of words. This list of words has a high association with a single target word that is not present in the list. This single word is known as a critical lure. For example, if the target word is "sleep," the list will contain words that relate to the word sleep, such as "bed" or "tired," but not the actual word "sleep." It has been found that the critical lures are recalled above chance during the free recall portion of the task, thus showing evidence of false memory implantation (Roediger & McDermott, 1995). On the other hand, Miller and Wolford (1999) argue that the recall of the critical lure is based on a criterion shift instead. As previously discussed, Miller & Wolford showed in their 2011 follow-up study that a participant's criterion was more liberal for any item related to the gist of the list items, which may include the critical lure. However, when participants were warned not to rely on the gist memory, participants showed significantly fewer memory errors, providing evidence that participants can strategically control their responses and potentially suppress a false memory effect.

Collectively, these studies indicate that criterion shifting exists in free recall tasks and can help explain the errors many researchers have found in their studies on human memory. The next step in this line of research should be to utilize free recall in a task that tests autobiographical memory, the type of memory most members of the public consider "memory" in general, and the type of memory constantly scrutinized in our legal system.

1.4 Taking an Ecologically Valid Approach to Criterion Shifting and Free Recall

John Dean's testimony during the Watergate scandal highlights the real-world implications of society's reliance on free recall. This reliance can be readily seen in the institutions that impact everyday lives—the legal system, the healthcare system, and many academic fields, to name a few—and illustrates how important a better understanding of decision-making processes in the context of the free recall of rich episodic memories can be. Building off of the previous criterion shifting literature, especially the studies that establish this behavior as a uniquely individualized cognitive trait, one aim of this dissertation is to explore the application of criterion shifting to new contexts. More specifically, the goal is to further the understanding of criterion shifting for ecologically valid situations, which means considering recollection processes and whether social external factors influence decision strategies.

The first experiment, discussed in Chapter 2, intended to create an ecologically valid scenario (i.e., one with stressful social pressure) to examine how criterion manipulations impact the free recall of a rich autobiographical event. Observing the results of that study then led to a decision to break out the components—social pressure and recollection processes—to examine them in a more controlled setting. In Chapter 3, Experiments 2 and 3 adopted a laboratory approach to investigate how different social contexts (competition and public rankings in Experiment 2, collective outcomes in Experiment 3) may affect criterion shifting during recognition memory tasks. Chapter 4 focused on exploring theories of recollection processes through a within-subject experiment that systematically increased the reliance on recollection while ensuring that the shifting conditions were easily identifiable for participants. This design was created to compel participants to identify and utilize a decision strategy so that the findings would lend support to either the Dual Process Signal Detection (DPSD) model or the Unequal Variance Signal Detection (UVSD) model. These models have been highly debated, with the former outlining that recollection is a thresholded process while the latter promotes a continuous process. Finally, Chapter 5 will discuss the overall findings from these experiments and lay out future directions.

Chapter 2: Experiment 1 - Free Recall of a Stressful Real-World Event

2.1 Introduction

In this study, participants experienced a richly detailed event that generated a stress level analogous to what an eyewitness may encounter in real life. Most previous research on the relationship between stress and memory has used stressors unrelated to the material participants were recalling (Wolf, 2019). For example, Smeets, Otgaar, Candel, and Wolf (2008) would have participants complete the DRM paradigm, followed by the cold pressor stress (CPS) task before the recall phase. The CPS task is a low-risk and reliable method to induce stress in participants by having them dip one of their elbows into very cold water for about three minutes. While these studies provide insights into the connection between stress and memory, they offer a limited understanding of the recall of stressful events (like eyewitness testimony). Every moment of our lives is rich with a myriad of details – objects, people, sights, sounds, smells, tastes, and feelings surround us at any given second. Recalling these rich details requires activating complex cognitive processes, such as recall and decision-making, and setting appropriate decision criteria before making any memory-related judgments. Researchers have debated the malleability of memory, with some positing that distortions will change the memory itself and others believing these distortions affect decision thresholds, not the memory itself. Previous research on decision-making thresholds has focused on recognition memory tasks or simple episodic memory events. This study explores whether criterion manipulations affect the recall of live, ecologically valid events.

2.1.1 - Trier Social Stress Test

Understanding the effects of criterion threshold manipulations during the free recall of live events has the potential to improve how eyewitness testimony is collected and presented in the legal system. This is why the present study utilizes the Trier Social Stress Test (TSST), which integrates the experience and the stress as part of the event to be recalled. In 1993, Kirschbaum, Pirke, and Hellhammer developed the TSST to induce moderate psychological stress that can be measured physiologically in a laboratory setting. This protocol has been utilized for over two decades to reliably induce stress in participants (Kudielka, Hellhammer, & Kirschbaum, 2007). The TSST consists of a ten-minute anticipation period followed by ten minutes of testing. The testing procedure entails delivering a five-minute free speech about the participants' dream job and their qualifications. During the following five minutes participants completed an arithmetic task where they had to count down from 2023 in increments of 17.

Wolf (2019), in his study on long-term memory, found that participants who experience a stressful event are more likely to remember the central aspects of it due to the positive effects of stress on memory consolidation. Participants in Wolf's study completed the TSST while several common office supplies were scattered around the judge's table, some of which the judges used during the task. These items were considered "central" because they were directly related to the judges, i.e., the stressor. Wolf, as well as those who replicated his study, found that subjects recalled the central items more than any other item type, suggesting that items associated with the primary source of stress are more memorable (Herten, Otto, & Wolf, 2017; Wiemers, Sauvage, & Wolf, 2014). This study used the TSST protocol because it provides a reliable, standardized, and, most importantly, a richly detailed autobiographical event for participants to recall, all while inducing a type of stress (public speaking) that people encounter in their everyday lives. Additionally, having participants speak to a panel of confederates (acting as judges) allowed this study to incorporate measures similar to eyewitness testimony research, and filling the room with objects (both "central" and background) allowed participants to be tested with recognition memory tasks (similar to Wolf's 2019 study).

2.1.2 - Cognitive Interview Method

Ed Geiselman and Ron Fisher (1985) developed the cognitive interview method to enhance witness interviews, and it has since been established as an effective approach for gathering detailed information from eyewitnesses. This information-gathering technique has been tested thoroughly in over 100 laboratory experiments, with volunteers acting as witnesses to either a live, innocuous event or a videotape of a simulated crime (Geiselman, Fisher, MacKinnon, & Holland, 1985). These studies have involved a diverse array of witnesses, including adults, children, people with learning disabilities, and the elderly. Some studies have found that interviews completed soon after the event can improve recall accuracy, while delays in recall and biased questioning procedures have been shown to cause errors. The cognitive interview has been successfully implemented by a wide variety of interviewers, from trained and experienced police officers to students with minimal interviewing experience (Memon, Meissner, & Fraser, 2010).

The cognitive interview is designed to have witnesses freely recall the event and then have investigators ask open-ended questions. The four most important techniques for conducting a successful cognitive interview are: emphasizing context, encouraging participants to report everything, asking participants to consider a variety of perspectives, and varying temporal order. First, successful interviewers will emphasize the physical and personal context of the event and have the witness mentally reconstruct these components. The second technique involves encouraging participants to report all the details of their recalled event, including partial or incomplete ones. Thirdly, interviewers should have participants consider other witnesses' perspectives and recall events from another person's point of view. The final technique has participants recall information from differing time points (whether from beginning to end, backward, or at other salient points in the event). A good cognitive interviewer should meet the individual needs of the witness by making the interview experience witness-centered and giving the person control of the narrative.

Including a cognitive interview helps align the proposed study with a typical eyewitness experience. The process begins by transferring control to the participant acting as an eyewitness, putting them at ease. This is a key component because it reduces the unintentional bias some individuals feel with perceived power and authority figures, such as law enforcement personnel. For example, the interviewer should demonstrate patience by allowing time for long pauses while the participant gathers their thoughts (Lacy & Stark, 2013). The cognitive interview method has been shown to lessen the effects of misleading questions (Geiselman, Fisher, Cohen, Holland & Surtes, 1986; Memon et al. 2010; Milne & Bull, 2003) and avoid contributing to the creation of false memories (Sharman & Powell, 2013). Geiselman et al. (1986) found that the likely reasons for this are a greater reliance on open-ended questions and the use of context reinstatement, where the interviewer directs the witness back to the original memory record. Geiselman and Fisher (1992) then enhanced the cognitive interview by adding components such as building rapport, instructing witnesses not to guess, and encouraging focused memory techniques like having the witness concentrate on mental images to guide recall. These additions resulted in a greater number of details being recalled with higher accuracy compared to previous cognitive interview protocols or other types of interviews.

2.1.3 - Present Study

This study's primary motivation is to explore people's decisions when recalling a rich autobiographical event. While some researchers argue that any manipulation would change the memory itself, this work aims to demonstrate how it is actually the decision criteria of a memory that will be affected by criterion shift manipulations. The analysis aimed to compare the number of details reported between the first and second interview sessions while also comparing the conditions (conservative vs liberal) to see whether results from this novel free recall task aligned with prior criterion shifting research based on recognition memory tasks (e.g., participants in the liberal manipulation group reporting more details overall, accurate or not, than those in the conservative manipulation group). The hypothesis of this study was that the cognitive interview manipulations would shift the participants' criterion for their freely recalled experience to be more conservative or liberal, depending on the conditions they were randomly assigned.

This study utilized a significantly tested social stressor as a controlled live event to determine an individual's ability to accurately recall details of said event. Additionally, this study used the well-established cognitive interview method to determine the baseline amount of information a participant would provide. The cognitive interview method has been a reliable way for interviewers to gain the highest volume of and most accurate details from witnesses while simultaneously lessening the likelihood of possible distortions. Most importantly, the study aims to expand our knowledge of episodic memory in an ecologically valid manner.

2.2 Method

2.2.1 - Participant Recruitment

Participants for Experiment 1 enrolled through the University of California, Santa Barbara (UCSB) research participation website. The experiments were approved by the UCSB Human Subjects Committee Institutional Review Board (IRB). All participants gave signed informed consent.

A total of 62 subjects were recruited to participate in this study, split evenly into two groups. However, only 56 participants' data were ultimately used (n = 56 [44 female], M = 19.4 years, range = 18-24 years, SD = 1.43), as four participants did not finish the task and two encountered technical difficulties that prevented them from completing the task. Study participants self-identified their race on a demographic questionnaire: 3 participants (5%) identified as Black, 8 participants (14%) identified as Asian, 16 participants (29%) identified as Latine, 3 participants (5%) identified as Middle Eastern, 18 participants (32%) identified as White, and 8 participants (15%) identified as multiracial, which included 5 (9%) identifying as Asian and White, 2 (4%) identifying as Latine and White, and 1 (2%) identifying as Middle Eastern and White.

2.2.2 - Procedure

Each participant made one visit split into two segments. The first was the encoding segment, where subjects filled out questionnaires and then underwent the TSST. The second segment, the recall phase, began with participants completing an unrelated recognition memory-based criterion shift task on a computer for at least seventeen minutes before sitting for their first cognitive interview. This time between the TSST and the interview was deliberately included to give enough time for participants to process their experience into long-term memory and simulate the amount of time in a real-world scenario that would elapse before someone would be interviewed by a first responder. The first cognitive interview established a baseline report for each participant. After the first interview, participants completed written questionnaires for roughly five minutes while the interviewer exited the room. Upon returning, the interviewer would inform participants that they needed to do another interview and provide them with instructions that contained either a conservative or liberal criterion manipulation. The different instructions were as follows:

> Liberal: "I just had our behavioral analyst review my notes of what you reported happened during the speech task and they said that you had missed a lot of things that happened. I was not there, so I do not know what did or did not happen. I'm going to have you report what happened again but this time, please report all information – even little details that you think may not be important. Guessing is ok, just let me know if you are."
>
> Conservative: "I just had our behavioral analyst review my notes of what you reported happened during the speech task and they said that you had reported quite a few things that actually did not happen or were not present. I was not there, so I do not know what did or did not happen. I'm going to have you report what happened again but this time, please only report information if you are very sure about it, do not guess."

The second interview was designed to evaluate the participants' episodic memory accuracy and durability for what they reported. Both interviews began with a complete free recall (no cues), followed by open-ended questions. This has been shown to be the most efficient method to encourage comprehensive reporting of details by participants. Additionally, the people involved and set up of the testing environment were the exact same for every participant to reduce any variance caused by external factors and possible biases during the recall phase.

The questionnaires participants completed over the course of the task were meant to investigate any potential individual differences. For example, the Big 5 Inventory was used to evaluate if a personality trait could help to predict the amount of reported information (especially after criterion manipulation) (John, Donahue, & Kentle, 2010). The State-Trait Anxiety Inventory, which measures the current and general feelings of anxiety levels, was administered before and after the TSST as a manipulation check to evaluate its effectiveness, as well as before and after the second interview to investigate if the autobiographical recall stimulated any anxiety (Spielberger, 2010).

2.2.3 - Qualitative Coding Protocol

This experiment employed an open coding methodology to identify themes in the data. This process involved two undergraduate research assistants conducting an initial review of the baseline and criterion manipulation interview transcripts to identify potential coding category themes. The identified codes were then organized into broader categories. This process helped create categories that encapsulated related concepts, structuring the data into coherent groups. Detailed definitions were written for each category to ensure consistency in the application of the codes. The final list of categories and their descriptions, drawn from our codebook, included both “Gist” and “Detail” codes. “Gist” codes captured broader, more conceptual themes, while “Detail” codes focused on specific details mentioned by participants. Additionally, the coding scheme included categories for reported “Correct Items” and “False Alarms.” In total, there were 53 categories: 12 “Gist” categories for correct information, 0 “Gist” categories for false alarm information, 31 “Detail” categories for correct information, and 10 “Detail” categories for false alarm information. Research assistants were asked to review the baseline and criterion manipulation interviews to compare themes between the time points and the two manipulations. No additional themes were identified in the criterion manipulations that would fit within the “Correct Items” or “False Alarms” categories. See Tables 1, 2, and 3 for the complete codebook separated by Correct Gist items, Correct Detail items, and False Alarm items.

Table 1
Qualitative Codebook: Correct Gist Items

Category Name	Detail / Gist	Definition	Baseline Frequency	Post-Manipulation Frequency	Liberal Manipulation Frequency	Conservative Manipulation Frequency
Dream job speech	Gist	The participant mentioned that they gave a speech about their dream job.	53	43	26	17
Prepare speech	Gist	The participant mentioned that they prepared a speech.	50	44	25	19
Experimenter knocked on door	Gist	The participant mentioned that the experimenter knocked on the door to signal for the judges to come in.	7	13	9	4
Mentions that they had to stand	Gist	The participant mentioned that they had to stand.	44	36	20	16
Give a speech	Gist	The participant mentioned that they presented their speech, were interviewed by the judges, or had an interview.	55	30	26	4
Math task	Gist	The participant mentioned that they had to do a math task or do calculations.	54	51	26	25
Restart math task	Gist	The participant mentioned that they had to restart the math task if they subtracted incorrectly.	39	29	15	14
Not allowed to stop talking until 5 minutes were over	Gist	The participant mentioned that they were not allowed to stop talking during their speech until the five-minute timer had expired. This gist was counted if the participant mentioned they were running out of things to say or that they were trying to fill the time, which insinuates they knew they had to talk for the full five minutes.	30	17	7	10
3 judges	Gist	The participant mentioned that there were three judges.	51	44	21	23
Judge set a timer	Gist	The participant mentioned that the judges set a timer or had a timer. This gist was included even if the judges' use of a timer was only insinuated as opposed to explicitly stated.	27	41	22	19
Filled out a questionnaire	Gist	The participant mentioned that they were asked or told they had to complete a questionnaire or survey	16	15	7	8
Camera person / initial experimenter present	Gist	The participant mentioned that a person was operating the camera or standing behind the camera. This gist was included even if the participant did not say "camera person" specifically, but instead described the experimenter operating the camera (e.g. "the experimenter who brought me in" and “the blonde girl").	31	22	13	9

Table 2
Qualitative Codebook: Correct Detail Items

Category Name	Detail / Gist	Definition	Baseline Frequency	Post-Manipulation Frequency	Liberal Manipulation Frequency	Conservative Manipulation Frequency
Preparation time was 5 minutes	Detail	The participant mentioned that they had five minutes to brainstorm or that the experimenter set a timer for five minutes.	44	44	23	21
Were given a pen and paper	Detail	The participant mentioned that they were given a pen and paper for brainstorming.	34	34	19	15
Pen provided was broken	Detail	The participant mentioned that the pen they were originally given was broken.	6	17	10	7
First pen provided was Blue	Detail	The participant mentioned that the pen that they were originally provided was blue.	0	4	2	2
The new pen provided is grey with black ink	Detail	The participant mentioned that the new pen was grey or had black ink. This detail was not counted if the participant said the pen itself was black.	0	4	2	2
Three knocks on door	Detail	The participant mentioned that the experimenter knocked on the door exactly three times.	1	5	1	4
Brainstorm paper was taken away	Detail	The participant mentioned that the paper they used for brainstorming their speech was taken away or was otherwise unavailable when they gave their speech.	23	28	13	15
Stand on the X	Detail	The participant mentioned that they had to stand on an "X" marked on the floor.	39	35	21	14
X made of duct tape	Detail	The participant mentioned that the "X" on the floor was made out of duct tape.	9	8	6	2
Speech was 5 minutes (detail)	Detail	The participant mentioned that their speech was timed to be no more than five minutes or that a five-minute timer was set.	45	42	21	21
Math task was 5 minutes	Detail	The participant mentioned that the math task was timed to be no more than five minutes or that a five-minute timer was set.	38	27	15	12
Subtraction in increments of 17	Detail	The participant mentioned that they had to subtract in increments of seventeen during the math task.	45	39	19	20
Math task started from 2023 (detail)	Detail	The participant mentioned that they had to start the math task by subtracting from 2023.	45	41	18	23
Math task was mental math	Detail	The participant mentioned that they had to do mental math, do the subtraction calculations in their head, or had to do the math calculations without using a pen and paper or calculator.	16	9	7	2

Table 2 (continued)
Qualitative Codebook: Correct Detail Items

Category Name	Detail / Gist	Definition	Baseline Frequency	Post-Manipulation Frequency	Liberal Manipulation Frequency	Conservative Manipulation Frequency
Chair present	Detail	The participant mentioned that there was a chair present or that they sat in a chair when they were brainstorming.	48	37	20	17
Yellow chair	Detail	The participant mentioned that the chair was yellow.	22	17	10	7
Table present	Detail	The participant mentioned the table that they sat at while they were brainstorming.	47	35	18	17
Brainstorm table is square	Detail	The participant mentioned that the table they sat at while they were brainstorming was square.	2	1	0	1
Camera present	Detail	The participant mentioned there was a camera. This detail was not counted if the participant only said they were being filmed or otherwise failed to explicitly state that there was a camera.	42	33	20	13
Judges had straight faces	Detail	The participant mentioned that the judges had straight faces, had "poker faces," were not responsive, did not smile, or otherwise did not show any emotion.	45	27	18	9
Judges wore white lab coats	Detail	The participant mentioned that the three judges were all wearing white lab coats.	22	17	12	5
Judges sitting	Detail	The participant mentioned that the judges were sitting during the task.	30	33	18	15
Judge was writing notes	Detail	The participant mentioned that one of the judges was taking notes. This detail was counted if the participant mentioned that "the judges" were taking notes.	7	9	9	0
Judge coughed/sneezed	Detail	The participant mentioned that a judge coughed or sneezed.	13	12	9	3
Timer mentioned	Detail	The participant mentioned that a timer was present in the task.	50	47	26	21
Plant present	Detail	The participant mentioned that there was a plant.	11	9	8	1
Plant is fake	Detail	The participant mentioned that the plant was fake or that they did not think it was real.	0	0	0	0
Curtain present	Detail	The participant mentioned a curtain.	42	26	15	11
Camera person standing at a table with the computer	Detail	The participant mentioned that the camera person was standing at the podium, next to the table with the computer, or generally by the computer.	18	24	14	10
Phrenology head present	Detail	The participant mentioned that there was a small sculpture of a head present. This detail was included even if the participant did not specifically describe the sculpture as a "phrenology head."	6	6	5	1
Turtle present inside plant	Detail	The participant mentioned that the animal inside the plant was a turtle.	2	2	1	1

Table 3
Qualitative Codebook: False Alarm Items

Category Name	Detail / Gist	Definition	Baseline Frequency	Post-Manipulation Frequency	Liberal Manipulation Frequency	Conservative Manipulation Frequency
Different time other than 5 min to prepare speech	Detail	The participant explicitly stated an amount of time other than five minutes when referring to how much time they had to brainstorm for their speech.	0	0	0	0
Different time other than 5 min to give a speech	Detail	The participant explicitly stated an amount of time other than five minutes when referring to how much time they had to give their speech.	0	0	0	0
Wrong number of judges	Detail	The participant explicitly stated a number of judges present besides three.	1	0	0	0
Different color pen	Detail	The participant explicitly stated that the pen they were originally given was not blue or that the replacement pen they were given was black.	0	0	0	0
Wrong writing tool	Detail	The participant explicitly stated that they were given something other than a pen to use for brainstorming (such as a pencil).	1	1	0	1
Real plant	Detail	The participant explicitly stated that the plant was real.	0	0	0	0
Wrong animal inside plant	Detail	The participant explicitly stated that an animal other than a turtle was in the plant.	1	1	1	0
Math task did not start at 2023	Detail	The participant explicitly stated that they had to start the math task by subtracting from a number other than 2023.	1	1	1	0
Different time other than 5 min for math task	Detail	The participant explicitly stated an amount of time other than five minutes when referring to how much time they had to do the math task.	0	1	1	0
Chair color was wrong	Detail	The participant explicitly stated that the color of the chair was a color other than yellow.	0	1	1	0

Three new research assistants were then trained on the category system and asked to independently apply the codes to the interview transcripts. Qualitative coders met weekly to discuss the coding process, working through ten interviews during each session. This process enabled the coders to reconcile any disagreements that emerged and have a discussion to achieve consensus. Coders first coded the baseline interviews and then coded the post-criterion shift manipulation interviews. Each coder rated all 56 subjects across 53 categories, resulting in 2968 ratings per coder for the baseline interview and 2968 ratings for the criterion manipulation interview.

2.2.4 - Qualitative Coding Inter-rater Reliability Method of Analysis

Krippendorff's Alpha was selected as the primary measure for assessing inter-rater reliability (IRR) due to its robustness and flexibility to accommodate various data types, handle missing data, and suitability for any number of coders (Krippendorff, 2011). Krippendorff's Alpha compares observed agreement among coders to the agreement expected by chance, providing a comprehensive measure of coding consistency. To assess the consistency and reliability of the coding performed by the three coders, Krippendorff's Alpha was calculated using the 'irr' package in R with the 'kripp.alpha' function for all possible combinations of coder pairs as well as for all three coders together. The key strength of Krippendorff's Alpha lies in its ability to compare the observed agreement among coders to the agreement that would be expected purely by chance. A higher alpha value signifies greater agreement beyond what would be expected by chance, indicating more reliable coding overall. Interpretation guidelines for Krippendorff's Alpha vary, but generally, an alpha above 0.800 is considered to suggest excellent agreement, a value between 0.610 and 0.800 represents substantial agreement, and a score below 0.610 indicates questionable agreement among the coders (Krippendorff, 2011; Hallgren, 2012).

Given the binary nature of the data and the high agreement observed in preliminary analyses, bootstrapping was employed to estimate the 95% confidence intervals for Krippendorff's Alpha. After the first 1,000 iterations, the results showed robust confidence intervals. These steps were taken to ensure that the confidence intervals were not an artifact of too few bootstrap samples and demonstrated the stability of the estimates. Cohen's Kappa was calculated for each pair of coders to further confirm the inter-rater reliability. Cohen's Kappa is a more widely used method to measure the inter-rater reliability for categorical items, providing a robust measure of agreement adjusted for chance. This step was critical to compare with Krippendorff's Alpha and to ensure the reliability of pairwise coder agreements.

Finally, Fleiss' Kappa was measured to understand the agreement among all three coders simultaneously. This statistical method is an extension of Cohen's Kappa for more than two raters. While Krippendorff's Alpha offers versatility and effectively handles missing data, Fleiss' Kappa is specifically designed for multiple raters assigning items to mutually exclusive categories, making it more robust for this study's binary coding data. Multiple reliability measures are used to confirm the initial results, strengthening the coding reliability conclusion.

2.2.5 - Signal Detection Analysis Data Preparation

The present study utilized an equal-variance SDT model, which categorizes participants' responses to ambiguous stimuli into one of four categories: hits (H; correctly identifying a stimulus that was present), misses (M; not identifying a stimulus that was present), false alarms (FA; identifying a stimulus that was not present) or correct rejections (CR; not identifying a stimulus that was not present). The signal detection metrics of hit rate (*HR*), false alarm rate (*FAR*), discriminability (*d'*), and criterion placement (*c*) were calculated for each participant using the following equations:

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

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

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

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

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

where *z* represents the density of the standard normal distribution (Macmillan & Creelman, 2005).

The hit rates and false alarm rates were determined by the proportion of agreed upon ratings by the three qualitative coders. These rates were then converted to z-scores to compute the sensitivity index, d-prime, which represents the participant's ability to report category items that are considered signal (correct items) and noise (false alarm items). The decision threshold, criterion, was also calculated to understand the participant's tendency to freely report "signal" versus "noise" information.

2.3 Results

2.3.1 - Inter-rater Reliability

The analysis revealed substantial to almost perfect agreement among the coders, as demonstrated by the high values of Krippendorff's Alpha, Cohen's Kappa, and Fleiss' Kappa. The result of Krippendorff's Alpha for all three coders was 0.758, indicating substantial agreement. This level of agreement suggests that the coders were consistent in their coding and that the coding scheme was reliably applied across all subjects. To further investigate the reliability of the coding, pairwise comparisons of coders were also conducted using Krippendorff's Alpha. The pairwise comparisons reveal that the highest agreement was between Coder 1 and Coder 2, with an alpha of 0.905, suggesting an almost perfect agreement between these two coders. The agreement between Coder 2 and Coder 3 was substantial, with an alpha of 0.728, while the agreement between Coder 1 and Coder 3 was also substantial, with an alpha of 0.724. These results indicate that while there is strong agreement among all coders, some pairs of coders were more consistent in their coding than others. See Table 4.

Table 4
Krippendorff's Alpha for Coders

Coders	Subjects	Ratings	Raters	α	95% CI (Low)	95% CI (High)
All Coders	56	5936	3	0.786	0.773	0.799
Coder 1 & Coder 2	56	5936	2	0.905	0.893	0.915
Coder 1 & Coder 3	56	5936	2	0.728	0.709	0.745
Coder 2 & Coder 3	56	5936	2	0.724	0.705	0.742

Krippendorff's Alpha interpretation: > 0.800 (almost perfect agreement), 0.610 - 0.800 (substantial agreement), < 0.610 (moderate agreement)

Cohen's Kappa statistical tests were calculated to further understand the pairwise reliability estimates and determine whether the observed agreement is significantly higher than what would be expected by chance. Cohen's Kappa values indicated substantial agreement between all coder pairs. The Kappa values were 0.905 for Coders 1 and 2, 0.728 for Coders 2 and 3, and 0.724 for Coders 1 and 3. The Cohen's Kappa high values reinforce the reliability of the coders' ratings, adjusting for chance agreement, suggesting that the pairwise comparisons are robust. Regarding the agreement between each coder and the consensus ratings, Cohen's Kappa results indicate almost perfect agreement with Coder 2 exhibiting the highest agreement with the consensus at 0.945, followed by Coder 1 also with almost perfect agreement at 0.938, and last is Coder 3 at 0.753 which is substantial agreement. The z-scores and p-values for all comparisons indicate that the kappa values are highly significant, demonstrating strong reliability in the coding process. See Table 5.

Table 5
Cohen's Kappa for Pairwise Comparisons and Individual Coders Against Consensus

Coders	κ	*SE*	*z*	*p*
Coder 1 & Coder 2	0.905	0.003	319.788	< 0.0001
Coder 1 & Coder 3	0.728	0.005	153.911	< 0.0001
Coder 2 & Coder 3	0.724	0.005	152.101	< 0.0001
Coder 1 & Consensus	0.938	0.002	409.607	< 0.0001
Coder 2 & Consensus	0.945	0.002	439.535	< 0.0001
Coder 3 & Consensus	0.753	0.005	166.962	< 0.0001

Cohen's Kappa interpretation: > 0.80 (almost perfect agreement), 0.60 - 0.80 (substantial agreement), < 0.60 (moderate agreement)

To confirm the measure of consistency observed in the pairwise comparisons and extend it to the overall agreement among all coders, a Fleiss' Kappa was calculated. The high Fleiss' Kappa value at 0.786 supports the conclusion that the coders' ratings are reliable and consistent. See Table 6.

Table 6
Fleiss' Kappa for All Coders

Coders	Subjects	Ratings	Raters	κ	95% CI (Low)	95% CI (High)
All Coders	56	5936	3	0.786	0.778	0.793

Fleiss' Kappa interpretation: > 0.800 (almost perfect agreement), 0.610 - 0.800 (substantial agreement), < 0.610 (moderate agreement)

The inter-rater reliability analysis demonstrated substantial to almost perfect agreement among the coders, indicating that the coding scheme was reliably applied. The individual comparisons of each coder against the agreed rating underscore the robustness of the coding methodology. Despite the varying levels of agreement, all coders exhibited substantial consistency with the agreed ratings, affirming the validity of the coding outcomes. The highest reliability was observed between Coder 1 and Coder 2, suggesting that these two coders were particularly consistent in their coding. The overall high levels of agreement across all coders and pairs of coders provide strong evidence for the reliability of the coding process.

2.3.2 - Frequency of Participant Reported Information

Linear mixed-effect models were fitted to compare the frequency of reported information in the Baseline versus the Criterion Condition (Conservative vs. Liberal). This model was chosen because of its handling of repeated measures data structures and ability to account for within-subject correlations by including a random intercept for individual subjects. This will allow the model to account for the multiple data points each participant provides, thus a more accurate estimate of the fixed effects can be calculated.

The first linear mixed-effects model was used to examine the relationship between the three sessions (Baseline, Conservative, and Liberal) and the frequency of reported information for all information (regardless of correct or false alarm information coding categories). Individual differences were also controlled by including random intercepts for each participant. Additionally, the model was fit using restricted maximum likelihood estimation and this converged successfully at 8152.6. The random intercept variance was not statistically significant ($\sigma^2 = 0.008$, $SD = 0.087$), suggesting that there was minimal variation in the reporting of information in the Baseline condition across participants. Further model diagnostics such as the scaled residuals did not show any substantial deviations from the assumptions of linearity. The fixed effects of session on the frequency of reported information are presented in Table 7. The average frequency score for the Baseline condition was 0.409 (SE = 0.015, $t(52)$ = 28.052, p < .001). Participants in the Conservative condition had significantly lower frequency scores (β = -0.100, SE = 0.017, $t(52)$ = -5.989, p < .001) compared to the Baseline condition. However, there was no significant difference in the frequency scores between the Liberal condition and the Baseline condition (β = -0.002, SE = 0.017, $t(52)$ = -0.106, p = .916). See Figure 1 for a scatter plot showing the mean frequency score per subject for both their Baseline condition and criterion manipulation.

Table 7

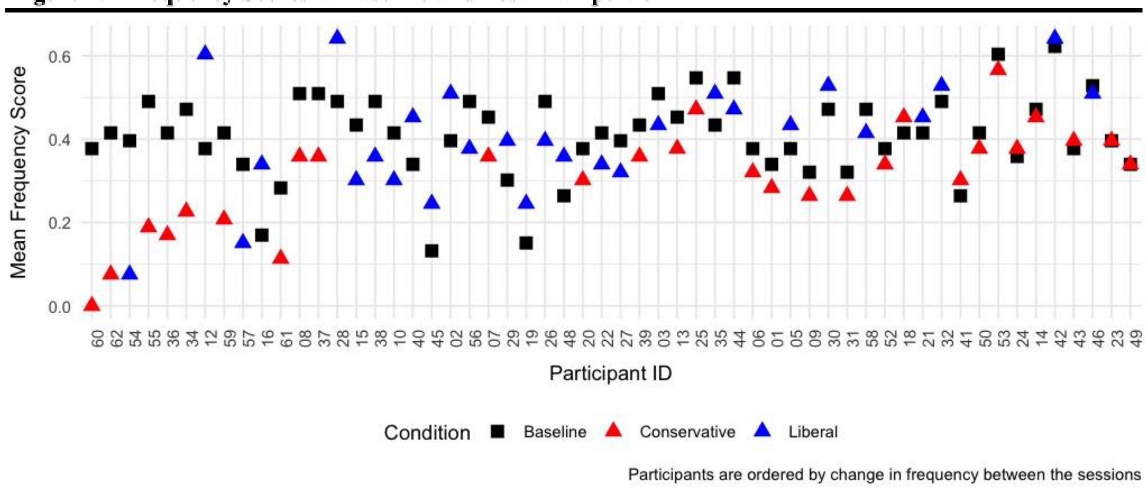
Model Summary: Frequency of All Reported Information - Baseline Reference

	Frequency of Reported Information				
Terms	β	*SE*	*95% CI*	*t*	*p*
(Intercept)	0.409	0.015	0.380 - 0.437	28.052	**< 0.001**
Condition: Baseline vs. Conservative	-0.100	0.017	-0.133 - -0.067	-5.989	**< 0.001**
Condition: Baseline vs. Liberal	-0.002	0.017	-0.034 - 0.031	-0.106	0.916

Random Effects

σ^2	0.23
$\tau_{00\ Subject\_ID}$	0.01
ICC	0.03
$N_{Subject\_ID}$	56
Observations	5936
Marginal R² / Conditional R²	0.008 / 0.040

Figure 1. Frequency Scores at Baseline and Post-Manipultion



To examine the impact of the criterion manipulations on the frequency of correctly reported information, two linear mixed-effects models were calculated with different reference conditions. One is to compare the Baseline condition against the criterion manipulation conditions (Conservative or Liberal) and the second is to understand the direct difference between those criterion manipulations. The first linear mixed-effects model that used the Baseline as a reference, showed that the frequency of correctly reported information has a significant main effect of Condition ($F(2, 5934) = 20.875$, $p < .001$). As seen in Table 8, participants in the Conservative condition ($\beta = -0.123$, $SE = 0.019$, $t(52) = -6.463$, $p < .001$) reported significantly fewer correct items compared to the Baseline condition ($\beta = 0.502$, $SE = 0.018$). However, there was no significant difference in the frequency of correct information reported between the Liberal condition ($\beta = -0.004$, $SE = 0.019$, $t(52) = -0.130$, $p = .847$) and the Baseline condition.

Random intercepts for participants were included in the model to account for individual differences in Baseline reporting tendencies. The variance of the random intercepts ($\sigma^2 = 0.012$, $SD = 0.110$) suggests that there was some variability in baseline reporting frequencies across participants, but this variability was relatively small.

Table 8
Model Summary: Frequency of Correct Information Reported - Baseline Reference

	Frequency of Reported Information				
Terms	β	*SE*	*95% CI*	*t*	*p*
(Intercept)	0.502	0.018	0.467 - 0.537	28.420	**< 0.001**
Condition: Baseline vs. Conservative	-0.123	0.019	-0.160 - -0.085	-6.463	**< 0.001**
Condition: Baseline vs. Liberal	-0.004	0.019	-0.041 - 0.034	-0.130	0.847

Random Effects

σ^2	0.23
$\tau_{00\ Subject\_ID}$	0.01
ICC	0.05
$N_{Subject\_ID}$	56
Observations	4816
Marginal R^2 / Conditional R^2	0.011 / 0.059

In the second model for correct information reported, the Liberal condition was used as the reference to directly evaluate the criterion manipulation relationship. As seen in Table 9, the average frequency of reported correct information in the Conservative condition (β = -0.119, *SE* = 0.026, *t*(52) = -4.635, *p* < .001) was significantly lower than in the Liberal condition (β = 0.498, *SE* = 0.022). For the Liberal and Baseline conditions (β = 0.004, *SE* = 0.019, *t*(52) = 0.193, *p* = .847), no significant difference was calculated (similar to the Baseline as reference model).

Table 9
Model Summary: Frequency of Correct Information - Liberal Reference

	Frequency of Reported Information				
Terms	β	*SE*	*95% CI*	*t*	*p*
(Intercept)	0.498	0.022	0.456 - 0.541	22.828	**< 0.001**
Condition: Baseline vs. Liberal	0.004	0.019	-0.034 - 0.041	0.193	0.847
Condition: Conservative vs. Liberal	-0.119	0.026	-0.169 - -0.069	-4.635	**< 0.001**

Random Effects	
σ^2	0.23
$\tau_{00\ Subject\_ID}$	0.01
ICC	0.05
$N_{Subject\_ID}$	56
Observations	4816
Marginal R^2 / Conditional R^2	0.011 / 0.059

Similar to the reported correct information analysis, two linear mixed-effects models were calculated to evaluate the frequency of reported false alarms, one with the Baseline as a reference and the other with the Liberal condition. The fixed effects of the frequency of reported false alarms with the Baseline as a reference are presented in Table 10. The average frequency score for the Baseline condition was 0.007 (*SE* = 0.004, $t(52) = 1.858$, $p = .065$). Participants in the Conservative condition had slightly lower frequency scores (β = -0.004, *SE* = 0.007, $t(52) = -0.544$, $p = .587$) compared to the Baseline condition, but this difference was not statistically significant. Additionally, participants in the Liberal condition had higher frequency scores (β = 0.007, *SE* = 0.007, $t(52) = 1.088$, $p = .277$) compared to the Baseline condition, though this difference was also not statistically significant. These results suggest that neither the Conservative nor the Liberal conditions had a significant impact on the frequency of false alarms reported compared to the Baseline condition. The random intercept variance was minimal and not statistically significant (σ^2 = 3.347e-05, *SD* = 0.006), suggesting very little variation in the reporting of false alarms in the Baseline condition across participants. The residual variance ($\sigma^2 = 0.008$, SD = 0.089) indicated that most variability in false alarm reporting occurred within subjects rather than between subjects.

With the Liberal condition acting as the reference value, no significant differences were found in the frequency of reporting false alarms between the Baseline, Conservative, and Liberal conditions. As shown in Table 11, participants in the Liberal condition reported a statistically significant number of false alarms with an average frequency score of 0.014 (SE = 0.005, $t(52) = 2.628$, $p = .009$). Comparatively, participants in the Baseline condition reported slightly fewer false alarms than those in the Liberal condition ($\beta = -0.007$, SE = 0.007, $t(52) = -1.088$, $p = .277$), but this difference was not statistically significant. Similarly, participants in the Conservative condition reported fewer false alarms than those in the Liberal condition ($\beta = -0.011$, SE = 0.008, $t(52) = -1.394$, $p = .164$), yet this difference also did not reach statistical significance.

Table 10
Model Summary: Frequency of False Alarms Reported - Baseline Reference

	Frequency of Reported Information				
Terms	β	*SE*	*95% CI*	*t*	*p*
(Intercept)	0.007	0.004	0.000 - 0.015	1.858	0.063
Condition: Baseline vs. Conservative	-0.004	0.007	-0.016 - 0.0109	-0.544	0.587
Condition: Baseline vs. Liberal	0.007	0.007	-0.006 - 0.020	1.088	0.277

Random Effects	
σ^2	0.01
$\tau_{00\ Subject\_ID}$	0.00
ICC	0.00
$N_{Subject\_ID}$	56
Observations	1120
Marginal R^2 / Conditional R^2	0.002 / 0.006

Table 11
Model Summary: Frequency of False Alarms Reported - Liberal Reference

	Frequency of Reported Information				
Terms	β	*SE*	*95% CI*	*t*	*p*
(Intercept)	0.014	0.005	0.004 - 0.025	2.628	**0.009**
Condition: Baseline vs. Liberal	-0.007	0.007	-0.020 - 0.006	-1.088	0.277
Condition: Conservative vs. Liberal	-0.011	0.008	-0.026 - 0.004	-1.394	0.163

Random Effects	
σ^2	0.01
$\tau_{00\ Subject\_ID}$	0.00
ICC	0.00
$N_{Subject\_ID}$	56
Observations	1120
Marginal R^2 / Conditional R^2	0.002 / 0.006

For further information about individual item frequencies, please see the codebook Tables 1, 2, and 3.

2.3.3 - Signal Detection Analysis

Given the low false alarm rates observed in this study, a detailed Signal Detection Analysis (SDA) was not feasible. The false alarm rates were very low, with the average frequency score for the Baseline condition being 0.007 (*SE* = 0.004, *t*(52) = 1.858, *p* = .065). Participants in the Conservative condition had a false alarm frequency score of -0.004 (*SE* = 0.007, *t*(52) = -0.544, *p* = .587), and those in the Liberal condition had a false alarm frequency score of 0.007 (*SE* = 0.007, *t*(52) = 1.088, *p* = .277). These low frequencies rendered individual-level SDA unreliable, as the values would be predominantly driven by hit rates. Additionally, the variability in false alarm reporting was minimal, indicating very little variation in reporting false alarms in the Baseline condition across participants.

Due to these low false alarm rates, the frequency of hits was used as a proxy measure for shifts in criterion. This approach aligns with previous research that utilized the amount of detail recalled to indicate criterion shifts (Glanzer & Adams, 1985; Glanzer et al., 2009). The analysis revealed that participants in the Baseline condition reported an average frequency of 0.409 (*SE* = 0.015), while those in the Conservative condition reported significantly fewer details, with an average frequency of 0.309 (*SE* = 0.017). In contrast, the Liberal condition yielded an average frequency of 0.407 (*SE* = 0.017), closely mirroring the Baseline condition. As shown in *Table 8*, the significant reduction in the number of details reported in the Conservative condition (β = -0.123, *SE* = 0.019, *t*(52) = -6.463, *p* < .001) compared to the Baseline condition (β = 0.502, *SE* = 0.018) suggests a shift towards a stricter decision criterion. Participants in the Liberal condition did not show a significant difference in the frequency of correct information reported (β = -0.004, *SE* = 0.019, *t*(52) = -0.130, *p* = .847) compared to the Baseline condition, indicating that the liberal instructions did not significantly alter the reporting criterion from the baseline. Moreover, participants in the Liberal condition reported a statistically significant number of false alarms with an average frequency score of 0.014 (*SE* = 0.005, *t*(52) = 2.628, *p* = .009; *See Table 11*). Comparatively, participants in the Baseline condition reported slightly fewer false alarms than those in the Liberal condition (β = -0.007, *SE* = 0.007, $t(52) = -1.088$, $p = .277$), but this difference was not statistically significant. Similarly, participants in the Conservative condition reported fewer false alarms than those in the Liberal condition (β = -0.011, *SE* = 0.008, $t(52) = -1.394$, $p = .164$), yet this difference also did not reach statistical significance.

The use of hit rates as a proxy for criterion shifting is supported by the significant main effect of condition on the frequency of correct information reported. This shift towards a more conservative reporting criterion under the Conservative condition can be interpreted as participants being more cautious and selective in their recall, likely due to the instructions emphasizing accuracy over quantity. Conversely, the Liberal condition did not produce a significant change in recall behavior, suggesting that the instruction to report more freely did not substantially affect the participants' decision criterion. This indicates that participants may have been reporting details fairly liberally in the Baseline condition due to the use of the Cognitive Interview method, which encourages comprehensive recall. Therefore, the liberal instructions did not significantly alter their existing reporting behavior, as the baseline instructions had already set a relatively low threshold for recalling and reporting details.

2.4 Discussion

This study investigated the impact of criterion manipulations on the free recall of stressful autobiographical events using the Trier Social Stress Test (TSST) and the Cognitive Interview method. The results demonstrated that participants instructed to adopt a conservative reporting criterion reported significantly fewer details than those in the baseline and liberal conditions. These findings suggest that instructing participants to adopt a conservative reporting criterion can lead to a more selective recall process, reducing the number of reported details without necessarily improving accuracy.

Caution is necessary when interpreting these results because the decrease in reported details in the Conservative condition could also reflect changes in memory quality. However, this is considered unlikely based on previous findings in similar research contexts, which have consistently shown that instructions to be conservative typically affect the decision criterion rather than the memory trace itself (Wixted & Mickes, 2010; Benjamin et al., 2009). The fact that the Liberal condition did not produce a significant change in the frequency of reported information compared to the Baseline condition also provides further evidence for this interpretation. Since the Cognitive Interview method used at Baseline encourages comprehensive recall and operates under a liberal criterion, the similarity in the amount of information reported between the Baseline and Liberal conditions suggests that participants were already utilizing a liberal reporting strategy. This consistency supports the notion that the observed decrease in the Conservative condition is primarily due to a shift in reporting criterion rather than a deterioration in memory quality.

Furthermore, this interpretation aligns with the broader body of research on memory and decision-making, which suggests that manipulations in reporting criteria primarily influence the threshold for reporting details rather than the underlying memory trace (Mickes et al., 2011). Previous studies have shown that participants can adjust their decision criteria based on the instructions provided, leading to variations in the quantity of information reported without significant changes in memory accuracy (Hilford et. al., 2019). Therefore, the significant reduction in reported details under the Conservative condition, coupled with the stable reporting behavior in the Liberal condition, strongly indicates that the observed effects are due to shifts in criterion rather than changes in memory quality.

This analysis highlights the importance of considering both the quantity and quality of recalled information when evaluating memory performance under different criterion conditions. The findings suggest that while conservative instructions may reduce the quantity of recalled details, they do not necessarily improve the accuracy of the recalled information, as indicated by the unchanged false alarm rates. These low false alarm rates observed across the conditions suggest a floor effect, where participants are unlikely to report details that did not occur. Several factors could contribute to this floor effect, such as the immediacy of the test, the limited novelty of items, or the uniqueness of the situation, making it difficult for participants to confuse the event with other life experiences. This phenomenon may not be an issue with the paradigm but rather a natural tendency of participants to avoid reporting incorrect information.

Importantly, future iterations of this study should continue to address real-world problems. As seen in the research on John Dean's case, rich episodic memory is susceptible to errors such as false recall and distortion. While the research conducted since John Dean's Senate testimony has increased the understanding of how criterion thresholds interact with memories, more research is still needed to measure these thresholds and determine recall accuracy for rich episodic events. This research is critical because much of our societal functioning relies on human memory (Lacy & Stark, 2013). A striking example is the legal system's heavy use of eyewitness testimony in civil and criminal cases. Without compelling scientific research on memory, cases are often decided based on how well attorneys can play to jurors' misconceptions about memory. The Innocence Project has found that eyewitness misidentification played a role in about 71% of over 360 wrongful convictions overturned by post-conviction DNA testing. This is likely only a fraction of the total number of wrongfully convicted people, as DNA evidence is only available in a limited number of cases ("DNA Exonerations in the United States," 2020). Future episodic memory research could profoundly impact how the legal system treats and values eyewitness testimony and influence other fields where reliance on memory is underappreciated or misunderstood.

The in-person interviews used to deliver criterion manipulations in this study underscore the potential role of social components in influencing participants' reporting behavior. This observation inspired the next chapter's investigation into how different social contexts might affect criterion shifting during recognition memory tasks. By taking a more controlled laboratory approach, the subsequent research aims to isolate and examine various aspects of social influence, such as monetary rewards, social competition, and group impacts. This shift from real-world social interactions to controlled social manipulations seeks to provide nuanced insights into the interplay between social influences and cognitive processes, furthering our understanding of memory-based decision-making and the factors that drive criterion shifts.

Chapter 3: Experiments 2 and 3 - Social Pressure and Criterion Shifting

3.1 Introduction

Research on decision-making has generally attempted to understand our ability to process multiple alternatives and choose an optimal course of action. This has been true for the study of criterion shifting, which is the ability to adapt to the most favorable decision strategy when the relative rewards or consequences of different response types change. However, what is "optimal" for the individual is dependent on the context of the decision, and one important context to consider is the human need to be accepted by social groups. Previous research has found that the need to belong is a powerful, fundamental, and extremely pervasive motivation (Baumeister & Leary, 1995). Numerous studies have explored how humans view social acceptance as a means to secure a broad range of social, occupational, financial, recreational, and relational opportunities (DeWall & Bushman, 2011; Juvonen et al., 2005), and how social rejection can result in diverse negative outcomes (Leary, 2010), such as subpar performance on challenging intellectual tasks (Baumeister et al., 2002), increased aggression (DeWall et al., 2009), and a decrease in impulse control (Baumeister et al., 2005). The importance of social acceptance as a driver of human decision-making has even been documented neurologically, as researchers have found that neural activity across the ventromedial prefrontal cortex associated with reward increases when a person feels they have conformed to a group (Campbell-Meiklejohn et al., 2010; Nook & Zaki, 2015; Chen et al., 2023). This need to belong is tied to social decision-making, which is defined as making a decision that affects others where an individual considers the preferences of others in addition to their own (Fehr & Camerer, 2007). Humans make many important decisions within the context of social interactions, with these choices further contingent upon the decisions and reactions of others (Sanfey, 2007; Rilling & Sanfey, 2011).

The intent behind Experiments 2 and 3 was to examine the intersection between social pressure and criterion shifting, specifically regarding memory-based decision-making. Much of the previous research and theories on criterion shifting have suggested that this behavior reflects a distinctive and consistent cognitive trait in individuals. As discussed earlier, while previous research demonstrates that criterion shifting is quite stable across various situations (Aminoff et al., 2012; Kantner et al., 2015; Frithsen et al., 2018, Layher et al., 2020; Layher, Santander et al., 2023), it has not yet examined criterion shifting under the application of social pressure. The novel goal of these experiments was to explore whether social pressure impacts memory-based decision-making and see if individuals who did not typically shift their criteria would shift when they were exposed to social pressure. The first social pressure experiment (Experiment 2) examined the effect of a competitive social context (publicly known ranked performances) on participants' criterion shifting on a memory-based task. Participants were placed in four different conditions: receiving points or monetary compensation for their performance, which was then either unranked or publicly ranked against other participants. The second social pressure experiment (Experiment 3) analyzed the effect of known collective outcomes (group punishment and reward) and criterion manipulations on a memory-based task. Subjects completed the same task design twice. In one condition, their performance for a monetary bonus only affected them, and in the second condition, their individual performance could result in a loss of money or additional money for their fellow participants. These experiments investigated how either group-based pressure (induced by publicly shared performance rankings) or consequences (induced by the threat of collective punishment or the prospect of collective reward) would affect individual criterion shifting for memory-based decisions. It was hypothesized that individuals would shift their decision-making criteria and potentially improve performance on a recognition memory computer task when exposed to social pressure compared to a condition without such pressure.

3.1.1 - Social Pressure and Individual Outcomes

One way to analyze the social context of decision-making is to look at how individuals react to their perception of status within a group. Extensive research by Herbert Marsh and colleagues (1984; 2007) over the years has shown that relative position to a group matters for individual well-being. For example, Marsh and others have found that the "big-fish-little-pond" effect, where individuals tend to prefer to be a "big fish in a little pond" (e.g., a good student in an average school) rather than a "little fish in a big pond" (e.g., an average student in a high-achieving school), is a universal phenomenon across different countries and cultures (Marsh & Seaton, 2015; Marsh et al., 2017).

Subsequent studies have documented that the knowledge of relative positions also affects individual performance. This has been shown through numerous studies in the education space. In a natural experiment, Azmat and Iriberri (2010) found that students in the Basque region of Spain received better grades in subsequent classes after receiving report cards that provided them with their relative rank compared to the class average. In a controlled experiment, Tran and Zeckhauser (2012) showed that undergraduates in a one-semester English course at a university in Hanoi, Vietnam who received bi-monthly public updates on their rank in the class performed better on tests than the control group and a second test group that received their ranking information privately. In a professional work context, researchers have shown that the presence of highly productive workers can pressure their peers to be more productive as well (Mas & Moretti, 2009) and that just the announcement of a relative ranking scheme, regardless of whether it was conveyed publicly or privately, increased worker's performance on a standardized math task (Kuhnen & Tymula, 2012). Researchers have also found that participants will perform better on tasks when they know their performance will be measured against their peers even when better performance brings no monetary rewards (Falk & Ichino, 2006). However, the research on social pressure from status comparisons is not uniformly positive. Multiple studies have found that feedback during a task can weaken performance by causing them to make more mistakes (Eriksson et al., 2009; Hannan et al., 2008). Barankay (2011) found that workers recruited to analyze images on Amazon's Mechanical Turk website were less likely to do a task if ranks were given, those given ranks were less likely to sign up for subsequent tasks and ranked workers were less productive than their unranked peers.

Given the depth of the research suggesting that public rankings can induce changes in behavior, incorporating public rankings into the design of Experiment 2 was anticipated to provide a reliable way to exert social pressure on participants. For Experiment 2, participants experienced four different conditions for completing a memory-based task, two of which were unranked and two of which were ranked, and the results were publicly shared. It was predicted that the ranked conditions would motivate better performance and therefore result in significant criterion shifting on the recognition memory task.

3.1.2 - Social Decision-Making and Collective Outcomes

The effects of social pressure can also be seen in how individual decisions are tied to group outcomes in contemporary society. For example, in the education system, it is common for educators to threaten their entire class with punishment, such as taking away free time or assigning additional work, to motivate their students to behave more respectfully in the classroom (Heckathorn, 1990). Conversely, many teachers use a class-wide reward system to encourage good behavior or productivity (Little et al., 2015). Much of the literature on how group members are held accountable for one another's conduct breaks down into these two broad categories: collective punishment, where all group members are punished for one person's transgressions, regardless of individual culpability, and collective reward, where all group members benefit from the efforts of others regardless of their individual contributions (i.e., the "free-riding" problem) (Heckathorn, 1990).

Collective punishment is used in numerous situations where the goal is to encourage specific behavior or deter future offenses by punishing an entire group for one person's decision or mistake (Bolle, 2021). Unsurprisingly, numerous studies argue that collective punishment is a harmful practice, with some studies showing that collective punishment can harm cognition, including spatial reasoning (Heckathorn, 1990), perceptual speed (Reysen, 2007), and semantic memory (Cacioppo & Hawkley, 2009). Others have asserted that collective punishment can be morally fraught. Fabricant (2011), for example, argues that zero-tolerance policing, where crime by a few individuals can lead to indiscriminate search-and-seizure operations in the communities where those crimes occur, can be seen as a form of unwarranted collective punishment.

Interestingly, in laboratory settings, there have been mixed results regarding the effectiveness of collective punishment as a social motivator. Chapkovski (2021) found that collective punishment was generally less effective than individualized punishment in promoting cooperation in a standard public goods game (where participants had a choice to either cooperate or “free-ride”). Bolle (2021) designed a study to examine the effects of collective punishment on groups of people who transgress against an authority (the experimenter) for a collective good (e.g., a strike for higher wages) versus people who transgress for an individual gain (e.g., corporations polluting despite regulations or athletes taking performance-enhancing drugs). The study showed that collective punishment is entirely ineffective against actors working towards a collective good because, in those scenarios, agents can only avoid collective punishment by reaching a critical mass of resistance against an authority. In other words, because there is no individual benefit, collective punishment simply motivates more resistance. By contrast, when an individual is transgressing for personal benefit, if that benefit is not deemed to be worth the punishment, then collective punishment can serve as an effective deterrence. However, when compared to collective rewards, other studies utilizing similar laboratory-setting public goods games have found that collective punishment can be more effective for promoting cooperation. Gao and colleagues (2015) found that collective punishment was decidedly more effective for promoting cooperation, especially in smaller groups and when the need for cooperation was immediate, while collective reward was ineffective at promoting cooperation on its own with large groups. In a similar study, Milinski and Rockenbach (2012) found that collective punishment alone is more effective than a combined collective punishment and reward system in addressing the issue of “free-riding” in a public goods game.

Although less studied, the research on collective rewards is also mixed as to its efficacy as a motivating force. Collective rewards, which can be financial (Kreitner & Kinick, 2007) and non-financial (e.g., formal acknowledgments of achievement, free lunches or dinners, the first choice of vacation time; Luthen, 1998) are frequently employed in business settings to improve employee attitudes toward their work (Olubusayo et al., 2014) and increase job performance (Madu & Anyalebechi, 2016). Some manufacturing companies across Europe have transitioned away from purely individual performance-based pay schemes and towards offering collective pay schemes based on factory- or company-wide performance (Brown, 2020). Research shows that companies that offer some form of collective reward saw better work performance across a variety of metrics, such as product quality, production speed, and flexibility with manufacturing new products, compared to their previous individual-only pay structures (Kankaraš & van Houten, 2015; Codero et al., 2005). However, as with collective punishment, collective reward appears highly context dependent. Stubbs and Bentley (2024) found that the impact of collective rewards in work settings can vary greatly between supervisors and standard employees. In laboratory settings, researchers have found that the size of the reward can significantly influence people's behaviors in a public goods game, with higher rewards leading to higher gains but also inducing a larger "free-riding" problem (Tambunlertchai & Pongkijvorasin, 2020). Additionally, in situations where individual contributions are difficult to measure or where tasks require interdependent efforts, collective rewards can negatively impact the perceived fairness of the reward system (Nguyen et al., 2018).

This research highlights the degree to which the use of collective outcomes depends on understanding the context and goals of the specific social situation they are implemented (Heckathorn, 1990; Bolle, 2021). In a comprehensive meta-analysis of 187 studies on collective punishments and rewards, Balliet and colleagues (2011) found that the usage of punishments and rewards did not statistically differ in their impact on cooperation and that other variables, such as the cost or source of incentives, was more predictive of effectiveness.

The variability in the research on these collective outcomes may stem from the fact that most empirical observations come from interactions in tightly controlled laboratory settings (Balliet, 2011), which lack the real-world factors that can further influence decision-making (Molho et al., 2020). Collective outcomes also have to be measured against related consequences. For example, militaries routinely use collective punishment during basic training to reinforce their soldiers' identity as a part of a uniform, cohesive group, despite evidence that such tactics may lead to an increased chance of adverse mental health effects later in life, such as PTSD (Bonner & Ellender, 2022; Bowker & Levine, 2016; Langan, 2018). It is up to the individuals and group at large whether using an arguably effective collective punishment for one purpose (e.g., better combat effectiveness) is worth the societal and moral cost in other areas.

The design of Experiment 3 was meant to incorporate these collective outcomes into a recognition memory task, similar to how Experiment 2 integrated competition pressure into its design. Experiment 3 included two conditions, an individual condition where participants completed the classic recognition memory task, and the second entails a collective outcome where participants would complete a group bonding activity prior to the computer task. Even though Experiment 3 did not include direct group collaboration, participants shared the same goal to (1) avoid the negative social consequence of triggering the collective punishment (losing the monetary reward) for the group and (2) achieve the positive social consequence of earning a higher collective reward for the group through their performance. By building both motivations into the task, Experiment 3 aimed to maximize the chance of social pressure exerting influence on participants' memory-based decision-making and potentially discern which has the more significant impact through analyzing their performance against their responses on the Social Anxiety Questionnaire for Adults (Caballo et al., 2012), the Self-Report Altruism Scale (Philippe et al., 1981), and the Watts Connectedness Scale (Watts et al., 2022).

3.2 General Method

3.2.1 - Participant Recruitment

Participants across Experiments 2 and 3 enrolled through the UCSB SONA website. The experiments were approved by the UCSB Human Subjects Committee Institutional Review Board (IRB). All participants gave informed consent.

3.2.2 - Signal Detection Theory

Experiments 2 and 3 utilized an equal-variance SDT model, which categorizes participants' responses to ambiguous stimuli into one of four categories: hits (H; correctly identifying a stimulus that was present), misses (M; not identifying a stimulus that was present), false alarms (FA; identifying a stimulus that was not present) or correct rejections (CR; not identifying a stimulus that was indeed not there). From this, we can formulate scores into rates for each category. For example, the hit rate (*HR*) is obtained by dividing participants' total hits by the total possible correct categories, whereas the false alarm rate (*FAR*) is the total false alarms divided by the total possible incorrect categories. SDT measures of discriminability (*d'*) and criterion placement (*c*) were obtained using the following equations:

$$HR = (\text{number of H}) / (\text{number of H} + \text{number of M})$$

$$FAR = (\text{number of FA}) / (\text{number of total possible FA})$$

$$d' = z(HR) - z(FAR)$$

$$c = [z(HR) + z(FAR)] / 2$$

$$\boldsymbol{\Delta} c = c(\text{conservative}) - c(\text{liberal}),$$

where z represents the density of the standard normal distribution (Macmillan & Creelman, 2005).

3.2.3 - Materials

Both experiments were completed on the UCSB campus, in the Miller Memory Lab, with computers provided by the researchers. Experiment 2 used face stimuli from the 10k U.S. Adult Faces database. These faces were cropped out of their original image in an oval fashion and pasted onto a white square background to create an overall image size of about 256 x 256 pixels (Bainbridge, Isola, & Oliva, 2013). Experiment 3 utilized two different versions of 1,024 scene images, one containing a single person and another edited not to include a person (Layher, Santander et al., 2023). The resolution of the scene images was 500 by 500 pixels. Participants conducted all tasks on a computer using MATLAB with version R2017B (Experiment 2) or R2021A (Experiment 3), incorporating open-source code from Psychophysics Toolbox, v3 (Brainard, 1997). Both experiments configured the computer task to display instructions in white text on a black background.

3.3 Experiment 2

The objective of Experiment 2 was to investigate whether different types of motivations, specifically monetary rewards and social competition, influence criterion shifting behavior. Criterion shifting behavior is recognized as a distinctive and consistent cognitive trait, with significant variability observed across individuals (Aminoff et al., 2012; Kantner et al., 2015; Frithsen et al., 2018, Layher et al., 2020; Layher, Santander et al., 2023). Some individuals readily shift their criterion, while others show little to no change. Understanding the factors that drive these differences is crucial, particularly the potential impact of external motivations. Given the strong influence of social factors on decision-making documented in social psychology (see Baumeister & Leary, 1995; Sanfey, 2007; Rilling & Sanfey, 2011), it is plausible that social competition might also affect criterion shifting. Thus, Experiment 2 aimed to assess whether monetary rewards or social competition through public ranking could prompt individuals to alter their criterion shifting strategies during a recognition memory task.

3.3.1 - Participants

A total of 62 subjects were recruited to participate in this study. However, only 61 participants' data were ultimately used (*n* = 61 [42 female], *M* = 19.9 years, range = 18-29 years, *SD* = 2.10), as one participant encountered technical difficulties that prevented them from completing the task. Study participants self-identified their race on a demographic questionnaire: 2 participants (3%) identified as Black, 20 participants (33%) identified as Asian, 17 participants (28%) identified as Latine, 3 participants (5%) identified as Middle Eastern, 15 participants (24%) identified as White, 2 participants (3%) identified as Indian, and 2 participants (3%) identified as multiracial, which included 1 (2%) identifying as Asian and White and 1 (2%) as unidentified multiracial. Participants received a base payment of $5 and could earn up to an additional $21 based on their task performance.

3.3.2 - Procedure

Participants first completed the consent form, read through the task instructions, and conducted a practice task. During the task, participants completed four cycles of a study block followed by two test blocks, all conducted in a testing room alongside 6 or 7 other participants. Each study block featured 108 face images presented sequentially for 300 ms each, followed by a 100 ms crosshair display. The rapid stimulus presentation during the study phase aimed to reduce discriminability performance to near-chance levels. Each test block comprised 108 trials, with 54 items being previously studied (old) and 54 being novel (new). On each test trial, an image appeared in the center of the screen, and participants used the “f” and “j” keys to respond “old” or “new.” The mapping of each key to a specific response type was counterbalanced across subjects. After each response, the text indicating the response type (e.g., “f = old”) turned orange and was displayed for 300 ms before transitioning to the next trial. Participants needed to respond within 1.8 seconds, otherwise the next trial began automatically.

To manipulate the decision criterion, participants earned 6 points for a correct response and lost 18 points for a critical error, but did not lose points for noncritical errors. Additionally, participants lost 2 points for failing to respond on a test trial. In the conservative criterion condition, participants lost points for a false alarm, while in the liberal criterion condition, a miss served as the critical error. Before each test block, participants received instructions specifying the criterion condition and viewed a reminder of the point structure for correct and incorrect responses on every trial. After each study block, participants completed one conservative and one liberal test block, presented in a random order. Following each test block, participants viewed a feedback screen informing them of the total points earned during that test block.

There were four incentive conditions (points only, payment, social ranking, and social ranking with payment) that aimed to influence the extent of criterion shifting. Each study/test cycle involved one of these incentive conditions, presented in a random order. In the points-only condition, participants did not receive money based on their point total, nor was their point total shared with others in the testing room. In the payment condition, participants earned one cent for each point accumulated, but their point total was not shared with others. In the social ranking condition, participants' point totals were displayed to all participants in ranking order from highest to lowest. In the social ranking with payment condition, point totals were ranked and displayed to all participants, with the top 2 scores receiving a $15 bonus each.

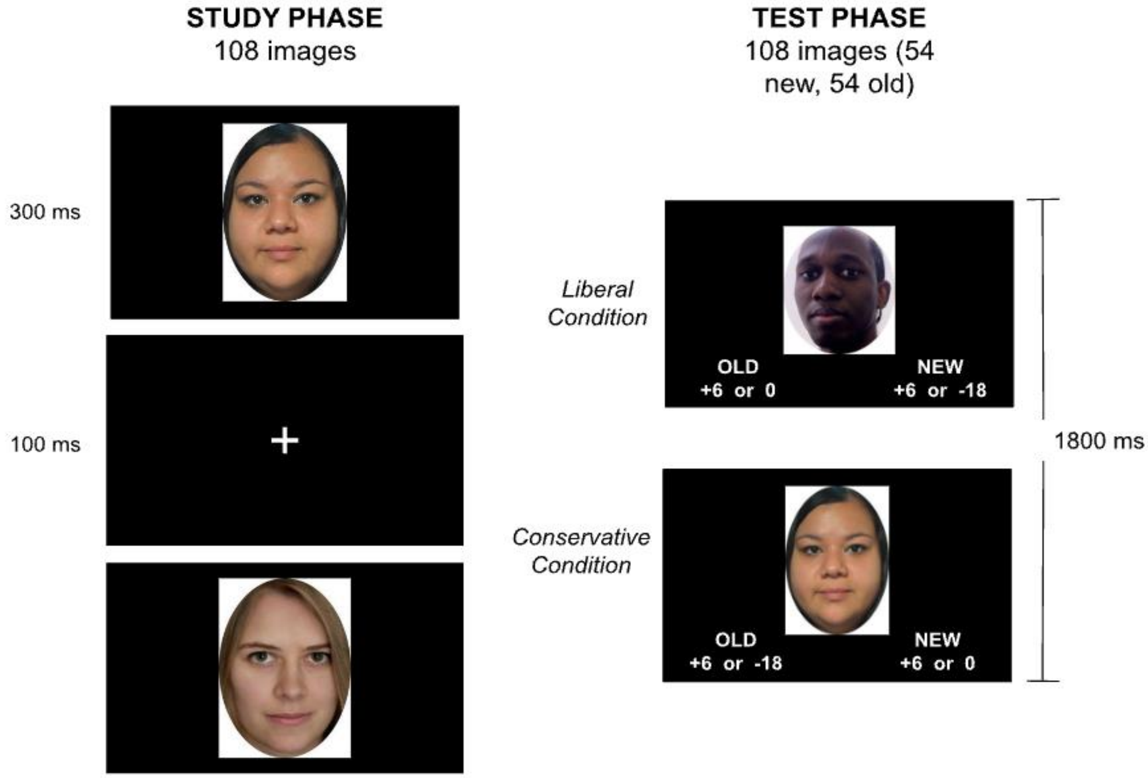


Figure 2: Experiment 2 social competition recognition task.

3.3.3 - Results

3.3.3.1 - Discriminability (d') Across Incentive Conditions

Mean *d'* remained quite low across the points only (*M* = 0.25, *SD* = 0.29, 95% CI [0.17, 0.32]), payment (*M* = 0.27, *SD* = 0.26, 95% CI [0.20, 0.33]), social ranking (*M* = 0.24, *SD* = 0.21, 95% CI [0.19, 0.30]), and social ranking with payment (*M* = 0.29, *SD* = 0.22, 95% CI [0.24, 0.35]) incentive conditions. There were no significant differences in *d'* across the incentive conditions ($F(3, 240) = 0.35$, $p = .79$). The recognition memory task was made intentionally difficult to encourage large criterion shifts. Table 12 presents the mean *d'* values for all incentive and criterion conditions.

3.3.3.2 - Criterion Shifting (Δc) Across Incentive Conditions

On average, criterion shifting (**Δ***c*) was quite robust across the points only (*M* = 2.54, *SD* = 0.81, 95% CI [2.34, 2.75]), payment (*M* = 2.71, *SD* = 0.66, 95% CI [2.54, 2.88]), social ranking (*M* = 2.38, *SD* = 0.61, 95% CI [2.22, 2.53]), and social ranking with payment (*M* = 2.46, *SD* = 0.68, 95% CI [2.29, 2.64]) incentive conditions, though there were vast individual differences (Figure 3). However, there were no significant differences in **Δ***c* across incentive conditions ($F(3, 240) = 0.55$, $p = .65$), indicating that the social pressure and payment manipulations did not substantially affect criterion shifting strategies. In fact, mean **Δ***c* was primarily influenced by test block order, as participants tended to shift criteria more significantly between test blocks 1 (*M* = 2.09, *SD* = 0.67, 95% CI [1.92, 2.26]), 2 (*M* = 2.51, *SD* = 0.47, 95% CI [2.39, 2.75]), 3 (*M* = 2.67, *SD* = 0.63, 95% CI [2.51, 2.84]), and 4 (*M* = 2.81, *SD* = 0.74, 95% CI [2.62, 3.00]), regardless of the incentive condition ($F(3, 240) = 7.46$, $p < .01$). This indicates that *some* participants learned to achieve more optimal decision outcomes by making larger criterion shifts, especially between test blocks 1 and 2. The point feedback received at the end of each test block may have helped participants better understand the consequences of not making substantial shifts.

Pearson correlations revealed strong consistency in **Δ***c* across all six test condition comparisons ($r(59)$ range: 0.73 – 0.85), indicating that despite substantial individual differences in the extent of criterion shifting, participants remained remarkably consistent within themselves.

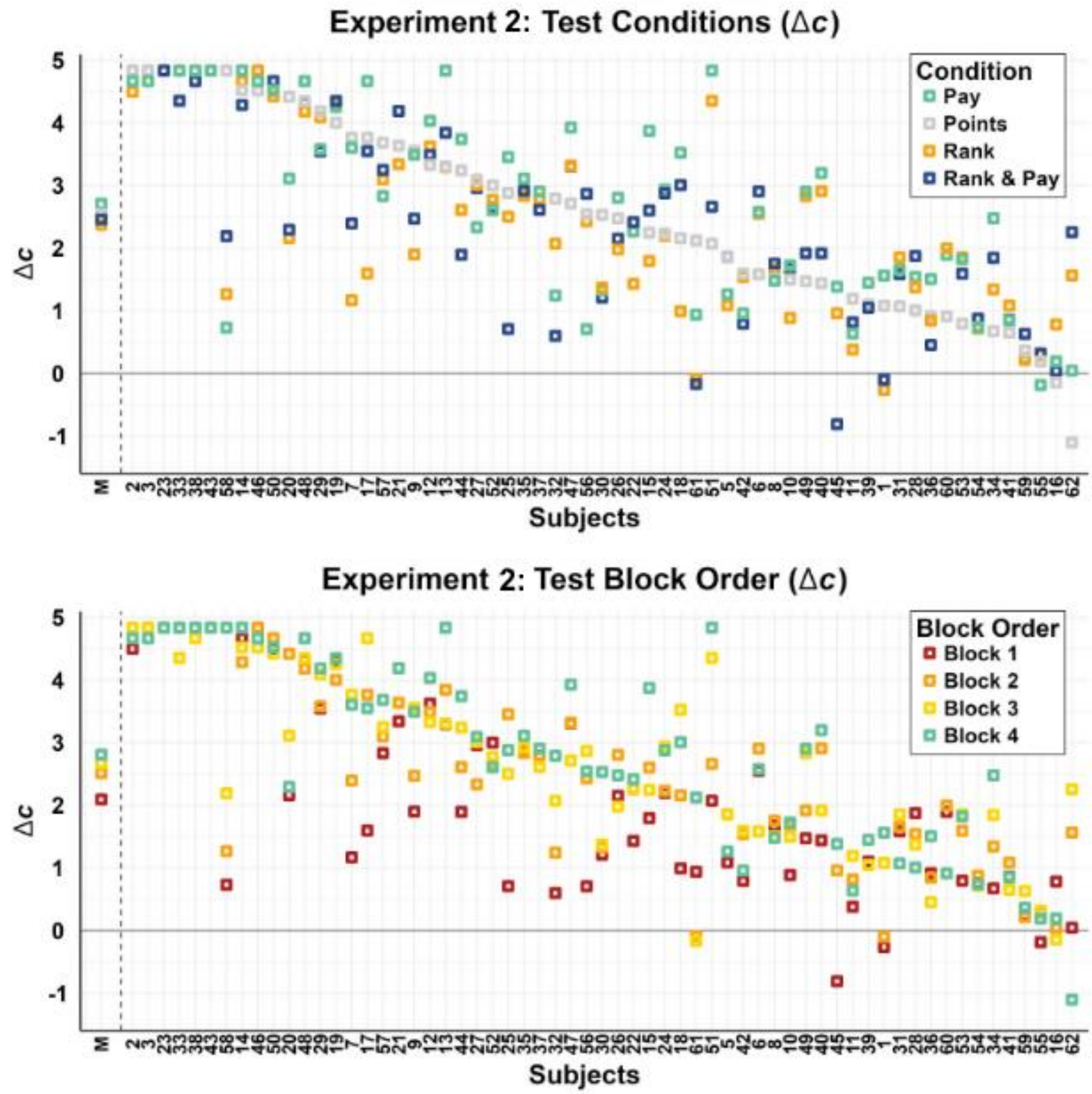


Figure 3: Experiment 2 individual differences in criterion shifting (**Δ***c*) across test conditions (top) and test block order (bottom). The mean (M) and individual subjects are displayed along the x-axis, ordered from left to right according to the extent of criterion shifting in the points-only incentive condition.

Table 12
Mean and standard deviation values (in parentheses) of criterion placement (c) and discriminability (d') across incentive conditions (paid vs. unpaid, ranked vs. unranked) and criterion conditions (conservative vs. liberal)

Incentive Condition	Criterion Condition	*c*	*d'*
Pay	Conservative	1.44 (0.87)	0.37 (0.37)
Pay	Liberal	-1.26 (0.82)	0.17 (0.32)
Points	Conservative	1.42 (0.78)	0.28 (0.38)
Points	Liberal	-1.13 (0.93)	0.21 (0.36)
Social Rank	Conservative	1.32 (0.81)	0.29 (0.36)
Social Rank	Liberal	-1.06 (0.82)	0.20 (0.30)
Social Rank & Pay	Conservative	1.38 (0.79)	0.39 (0.41)
Social Rank & Pay	Liberal	-1.09 (0.88)	0.20 (0.26)

3.3.4 - Discussion

The goal of Experiment 2 aimed to assess whether incentive manipulations, such as payment and social ranking, could influence the extent to which individuals adjusted their decision criteria during recognition memory tasks. While the social pressure conditions did cause a criterion shifting effect, no significant differences in criterion shifting emerged between any of the incentive conditions. These results, combined with the consistent criterion shifting strategies observed across test conditions, further support the findings of Layher and colleagues (2020), who identified criterion shifting as an individual and stable cognitive trait. Although the incentive manipulations did not significantly affect criterion shifting or discriminability, test block order had a notable impact. Participants exhibited more extreme criterion shifts on average throughout the experiment, particularly between test blocks 1 and 2, regardless of the incentive condition. This pattern suggests that feedback from the initial test blocks might have helped some participants better understand the consequences of not adjusting their criteria. Future research is needed to explore why *some* individuals shift their criteria more significantly over the course of an experiment.

3.4 Experiment 3

Because the findings from Experiment 2 did not reveal significant effects of monetary rewards or social competition on criterion shifting behavior, Experiment 3 was designed to further investigate how social pressures may influence this cognitive process. Specifically, Experiment 3 introduced a social pressure condition where participants were informed that their performance would impact others, aiming to enhance the motivational aspect of the social pressure. Additionally, the task difficulty was adjusted to better elucidate the interaction between memory performance and criterion shifting. In Experiment 2, the low discriminability (*d'*) was intended to induce large criterion shifts by creating a difficult task with quick presentation times and numerous images per study block. However, this design made it challenging to assess changes in discriminability that may result from social pressure. To address this limitation, Experiment 3 incorporated a much easier recognition memory test, allowing for a clearer separation of social influences on memory performance versus criterion shifting. Additionally, Experiment 3 incorporated a "memory" task in which the images were *not* shown during the test phase (hidden condition). The purpose of this task was to assess criterion shifting performance in situations where memory evidence is completely absent. To achieve the most optimal outcome, participants needed to maximize their responses by either always responding "old" or always responding "new," depending on which response achieves the best decisional outcomes. This condition highlighted the critical decision-making aspect of the task, emphasizing that any deviation from maximizing was a suboptimal choice. The hidden condition served to isolate the impact of social pressure on decision strategies without the confounding influence of memory-based performance. By assessing criterion shifting between a relatively easy recognition memory task (shown condition) and a task where memory is unusable (hidden condition), we can better assess the impacts of social pressure on criterion shifting and discriminability.

3.4.1 - Participants

Fifty-three undergraduate students at UCSB were recruited to participate in this study through the University's online recruitment database (SONA). Due to technical difficulties, seven participants' data were incomplete and are excluded from the final analysis, so only 46 participants' data were used (n = 46 [32 female, 13 male, one non-binary], M = 19.02 years, range = 18-22 years, SD = 1.22). Study participants self-identified their race on a demographic questionnaire: 6 participants (13%) identified as Asian, 14 participants (30%) identified as Latine, 14 participants (30%) identified as White, and 9 participants (20%) identified as multiracial, which included 4 (9%) identifying as Asian and White, 2 (4%) identifying as Asian and Latine, 2 (4%) identifying as Asian and White, and 1 (2%) identifying as Black and White. 3 participants (7%) preferred not to answer questions about their race. Participants received class credit and earned a monetary bonus (up to $25.60) based on performance in the tasks.

3.4.2 - Measures

All participants completed the study on lab-provided computers within the Miller Memory Lab at UCSB. The first portion of the study was completed via a Qualtrics online survey (Qualtrics, Provo, UT). This included the consent form and questionnaires such as the

Social Anxiety Questionnaire for Adults (SAQ), The Self-Report Altruism Scale, and the Watts Connectedness Scale.

The SAQ consists of 30 items rated on a 5-point scale, from 1 = "Not at all or very slight level of unease, stress or nervousness" to 5 = "Very high or extremely high level of unease, stress or nervousness." The scale identifies five factors that contribute to a total score: (1) Speaking in public or talking with people in authority, (2) Interactions with the opposite sex, (3) Assertive expression of annoyance, disgust, or displeasure, (4) Criticism and embarrassment, and (5) Interactions with strangers. Each dimension consists of six items distributed randomly throughout the questionnaire. There are scores for each dimension and a global score for the entire questionnaire (see Appendix A; Caballo et al., 2012; Caballo et al., 2016).

The Positive and Negative Affect Schedule short form is a 20-item questionnaire to evaluate positive and negative emotional states. The present study instructed participants to answer all questions to the extent of how they felt at that moment (Watson, Clark, & Tellegen, 1988). The complete questionnaire, along with scoring instructions, can be found in Appendix A.

The Self-Report Altruism Scale consists of 20 items. However, two questions were removed based on participant location and age (see Appendix A). Participants would rate how often they have engaged in altruistic behaviors, marking "Never," "Once," "More Than Once," "Often," and "Very Often." Higher scores indicate a higher frequency of self-reported altruistic behavior (Philippe et al., 1981).

Finally, the Watts Connectedness Scale was utilized to measure an individual's sense of connectedness to the self, others, and the world. Participants were asked to reflect on their experience from the past two weeks and then rated 19 items on a 0–100 visual analog scale (VAS), where 0 corresponded to “not at all” and 100 to “entirely.” See Appendix A for scoring instructions for the three connectedness subscale categories: (1) to the self, (2) to others, and (3) to the world (Watts et al., 2022).

3.4.3 - Procedure

Participants first reviewed the consent form via the Qualtrics survey and answered a set of questionnaires, including the SAQ (Caballo et al., 2012), the Self-Report Altruism Scale (Philippe et al., 1981), the Watts Connectedness Scale (Watts et al., 2022), and demographics. The second portion of the study was conducted on a computer running MATLAB software. Demographic information was collected, including age, race, ethnicity, gender, handedness, and which region of the world they felt the most at home. Next, the participants began the recognition memory task, which incorporated three types of manipulations in a fully crossed 2x2x2 design: a criterion condition (monetary manipulations: liberal vs conservative), a memory condition (“shown” versus “hidden” images), and a social condition (group vs. individual consequences), all of which are described in further detail below.

3.4.3.1 - Recognition Memory Task

The recognition memory task was split into two phases: a study phase and a retrieval phase. In the study phase, participants were shown images of indoor and outdoor scenery that either included a person in the image or not. Participants answered whether a person was “absent” or “present” in the scene for each image. Participants had an equal chance of seeing an image with a person “present” or “absent.” In the retrieval phase, participants were shown images and asked to identify them as “old” (meaning they had seen the image in the most recent study phase) or “new” (they had not seen the image before). Participants then reported their confidence rating as “low,” “medium,” or “high” after each response. Participants had an equal chance of seeing an “old” or “new” image.

Every participant completed two practice blocks per condition: two eight-image study phases and two eight-image retrieval phases. After these practice blocks, participants were instructed that there would be four study blocks that were similar to the practice block. Each study phase of the actual task had participants encode sixteen images, and each retrieval phase had participants respond to four test blocks with sixteen images per test block. Every photo shown to participants in the study phase was set to appear for 800 ms. Each study phase image was followed by a crosshair shown for 500 ms before the next image. During the recall phase, participants were given an unlimited amount of time to respond “old” or “new” and report their confidence rating as “low,” “medium,” or “high” for each image in the test block. Participants also received feedback after each test block to see how much they earned in the recently completed block and their total amount earned in the task.

There were 24 practice trials (8 study trials and 16 test trials) and 256 test trials per condition per participant. A total of 1,024 unique scenery images were utilized across conditions (512 for the individual and 512 for the group conditions). There was also an equal split within the condition of images with a human present and a human absent across conditions (256 images with a human present in the image and 256 without a human present for the individual condition; 256 images with a human present in the image and 256 without a human for the group condition).

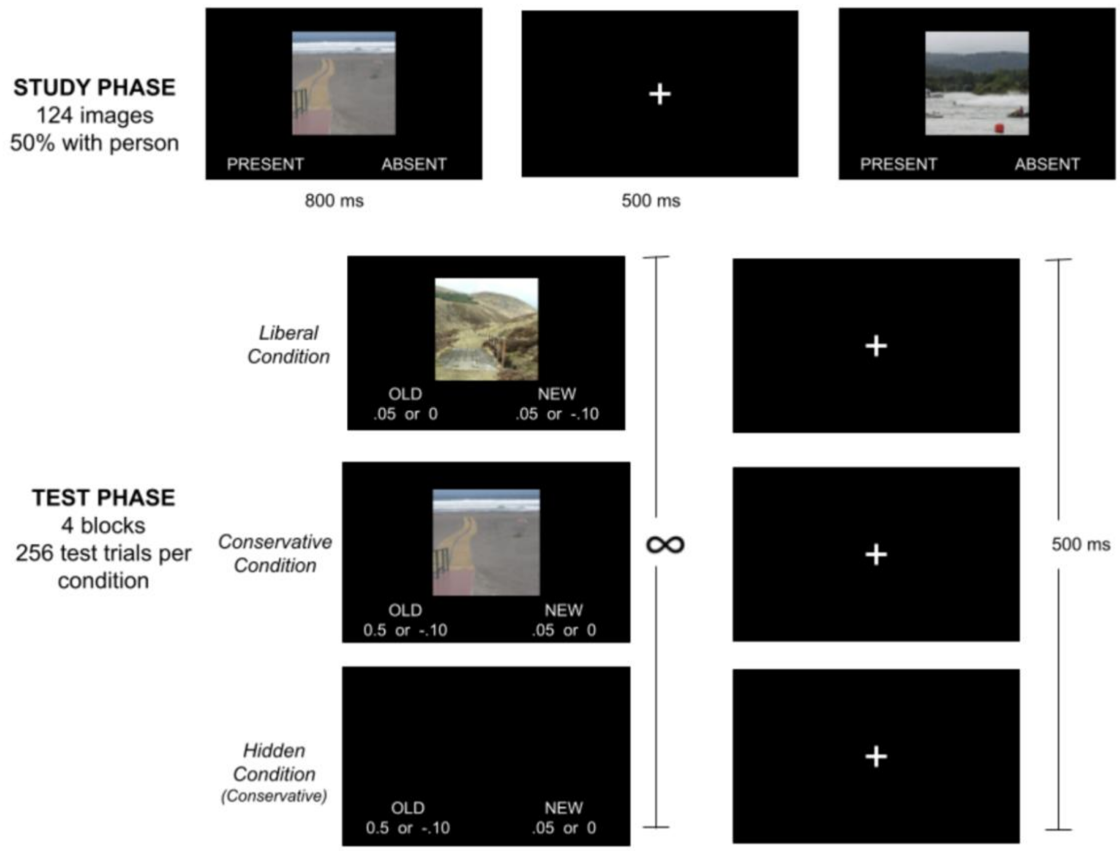


Figure 4: Experiment 3 social pressure recognition task.

3.4.3.2 - Criterion Condition

Monetary incentives were used in each test phase to influence participants to be more liberal or conservative with their decisions. In each test phase, participants received five cents for correctly responding “old” to a studied image (a hit) and “new” to an unstudied image (a correct rejection). In the liberal condition, participants were incentivized to choose the “old” response because an incorrect selection of “new” for a studied image (a miss) resulted in a $0.10 penalty, while an incorrect selection of “old” for an unstudied image (a false alarm) was unpenalized. Conversely, participants were incentivized to choose the "new" response in the conservative condition because they lost $0.10 for a false alarm, while misses were unpenalized.

3.4.3.3 - Memory Condition

Participants were randomly presented with four different trial conditions during the recognition memory task: liberal with shown images, liberal with hidden images, conservative with shown images, and conservative with hidden images. Participants were given the following instruction: "During half of the test blocks, the test images will be hidden, meaning you will NOT be able to view the image. However, there will still be a correct answer (either "old" or "new"). Participants were instructed to mark these blank screens as "old" or "new" and provide a confidence rating, just as with the shown images. Previous studies have examined how participants shift criteria in the absence of memory evidence, positing that participants may turn to other forms of non-probative, internally generated decision evidence outside the structure of the task (Kantner et al., 2015). The hidden image trial portion of this experiment was intended to encourage participants to maximize their monetary outcome based purely on the trial manipulations (liberal or conservative). This experiment differs from previous study designs because memory evidence was *completely* absent. All participants saw was a black screen, giving them *nothing* on which to make a memory-based decision.

3.4.3.4 - Social Condition

For this study, participants were recruited four at a time, and all had to confirm upon arrival that none of them knew each other before the experiment (i.e., everyone considered each other a "stranger"). Each participant completed two rounds of the recognition memory task, one under the "group" condition and one under the "individual" condition. In the individual condition, participants first answered two pre-task questions on a scale from “0” to “7,” with “0” signifying “Not at all true” and “7” signifying “Very true”: “In general, I am confident in my memory,” and “I am motivated to perform well during this task.” After completing the recognition memory task, participants completed a post-task questionnaire using the same 0-7 scale: “I put a lot of effort into this,” “I did NOT try very hard to do well at this task,” “I tried very hard on this task,” “It was important to me to do well at this task,” and “I did NOT put much energy into this task.” Participants earned a monetary reward for the individual condition task based on their performance.

In the group condition, participants first engaged in a group bonding activity facilitated by the researchers, which entailed having a seven-minute conversation to try to find a niche commonality between the four of them (e.g., did their parents all have the same birthday, did they unknowingly attend the same concert) to encourage connectedness (see Appendix B for examples). After the group bonding activity, participants rated their feelings of connectedness and sense of belonging in the group by answering the following questions on the same 0-7 scale: “I feel accepted by the group,” “I feel a sense of belonging in this group,” and “I feel good about being a part of this group.” Before doing the recognition memory task, researchers explained to the participants that their performance on the upcoming task would impact the payment outcome for each group member. Participants were informed that if they did not earn at least five dollars in this computer task, the whole group would lose their bonuses and receive nothing (i.e., a collective punishment). However, the participants were also told that if they all earned at least five dollars individually, they would each receive a bonus equal to the bonus of the highest earner in the group (i.e., a collective reward). At the end of the study, participants were informed that if they did not meet the $5 minimum in the group condition, they would not face collective punishment but would instead receive the individual bonus earned during the group task.

Upon arrival, participant groups were randomly assigned an order to complete the individual and group condition tasks, with six groups completing the individual condition task first and six groups completing the group condition task first. This was done to counterbalance the sample and avoid order effects.

3.4.4 - Results

3.4.4.1 - Discriminability (d') for Shown and Hidden Stimuli Across Conditions

Mean discriminability (*d'*) in the shown image task was significantly higher in the group condition (M = 2.55, SD = 0.56, 95% CI [2.38, 2.71]) compared to the individual condition (M = 2.02, SD = 0.59, 95% CI [1.84, 2.19]) on average (p = .002). Despite this difference, *d'* was strongly correlated across both conditions ($r(44) = 0.64$, $p < .001$; Figure 5, left). When considering test order effects of the individual vs. group condition in the shown images task, no significant differences in mean *d'* were observed between the first (M = 2.22, SD = 0.66, 95% CI [2.03, 2.42]) and second (M = 2.34, SD = 0.64, 95% CI [2.15, 2.53]) test blocks, regardless of the social condition (p = .52). This suggests that there were no substantial practice effects in improving discriminability from the first to the second test block.

The hidden image task showed participants' *d'* remaining at chance, on average, in both the individual (M = -0.04, SD = 0.42, 95% CI [-0.17, 0.08]) and group (M = 0.01, SD = 0.45, 95% CI [-0.13, 0.14]) conditions as anticipated. Additionally, as expected in the hidden task, no significant relationship in *d'* emerged between the individual and group conditions

($r(44) = .24$, $p = .10$), as discriminability performance relied on random chance (Figure 5, right). Table 13 presents mean *d'* values for all test conditions.

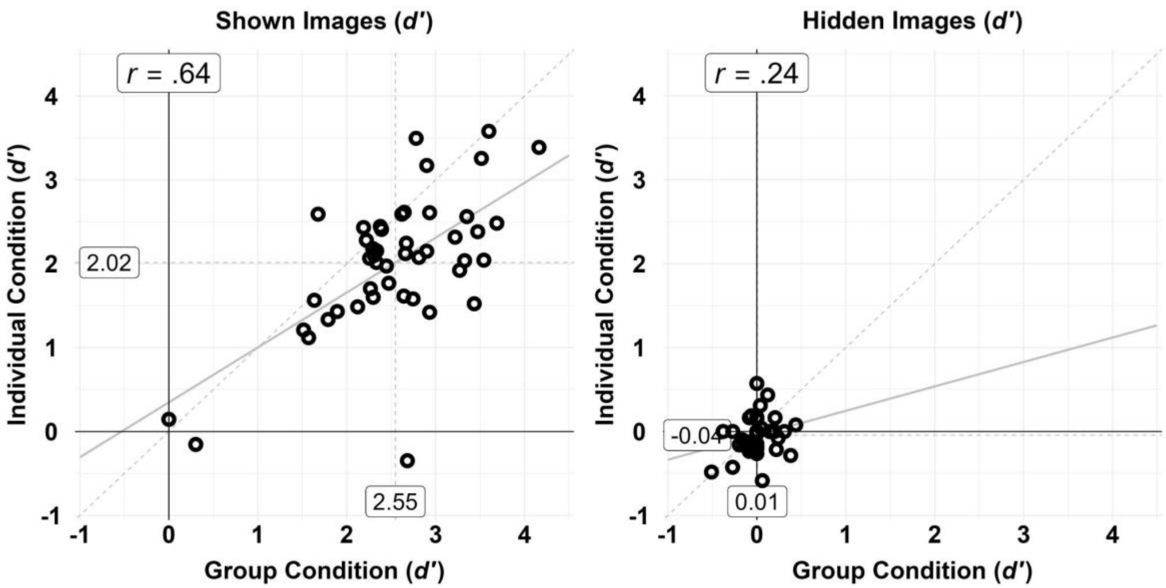


Figure 5: Pearson correlation plots depicting mean *d'* between the individual vs. group social conditions when images were shown (left) and hidden (right) in Experiment 3. Mean values are presented in boxes on each axis, with accompanying dashed gray lines. The Pearson *r* value is presented in the top left corner of each plot. Most participants demonstrated a higher *d'* in the group condition when images were shown, but could not control discriminability in the hidden condition since performance was based on chance alone.

3.4.4.2 - Criterion Shifting (Δc) for Shown and Hidden Stimuli Across Conditions

In the shown image condition, participants shifted their criteria (**Δ***c*) to modest extents on average in both the individual ($M = 0.43$, $SD = 0.93$, 95% CI [0.15, 0.70]) and group ($M = 0.41$, $SD = 0.98$, 95% CI [0.11, 0.70]) conditions with no significant differences between groups ($p = .91$). Strong correlations in **Δ***c* were observed in the shown images condition between the individual and group conditions ($r(44) = .61$, $p < .001$) suggesting consistent criterion shifting strategies (Figure 6, left). In the hidden image condition, participants shifted criteria to large extents on average in both the individual ($M = 2.55$, $SD = 1.27$, 95% CI

[2.18, 2.93]) and group (*M* = 2.70, *SD* = 1.30, 95% CI [2.32, 3.09]) condition, though there were no significant differences between groups ($p = .67$). The extent to which participants shifted their criteria between the individual and group condition was strongly correlated ($r(44) = 0.53$, $p < .001$) (Figure 6, right). Table 13 presents mean *c* values for all test conditions.

When considering test block order effects for the hidden image condition, there was a significant difference in the extent of criterion shifting between the first (*M* = 2.20, *SD* = 1.19, 95% CI [1.85, 2.56]) and second (*M* = 3.05, *SD* = 1.18, 95% CI [2.70, 3.41]) test block, regardless of the social condition ($p = .01$). This suggests that *some* participants learned to adjust their criterion more extremely during the second test block when they were unable to rely on memory to enhance their decisional outcomes. However, no significant differences occurred in the shown image condition between the first (*M* = 0.31, *SD* = 0.97, 95% CI [0.02, 0.60]) and second (*M* = 0.52, *SD* = 0.92, 95% CI [0.25, 0.79]) test block ($p = .22$).

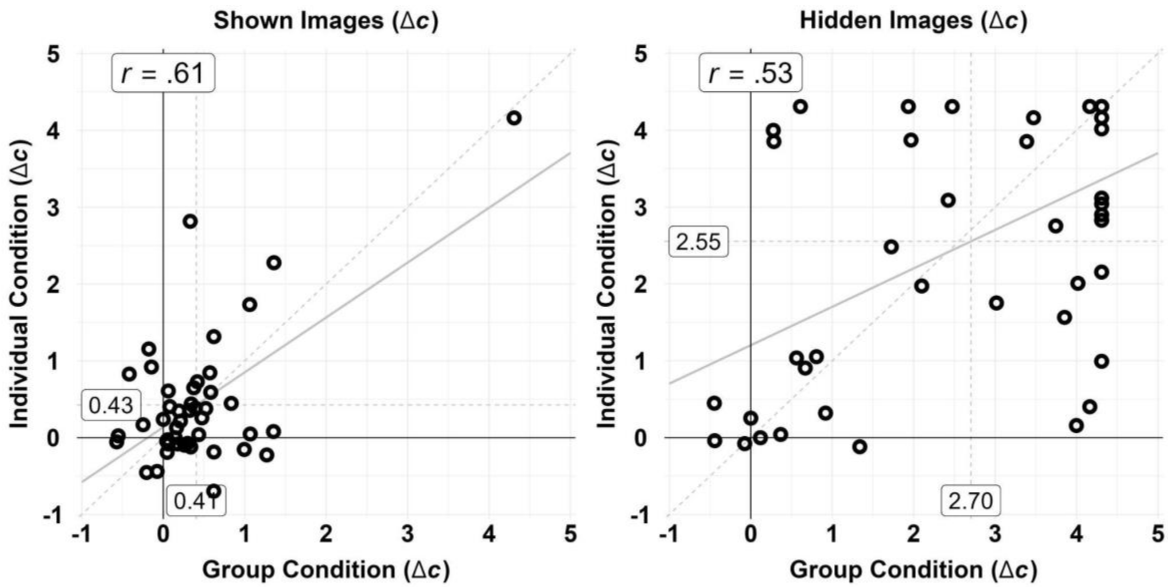


Figure 6: Pearson correlation plots depicting mean **Δ***c* between the individual vs. group social conditions when images were shown (left) and hidden (right) in Experiment 3. Mean values are presented in boxes on each axis, with accompanying dashed gray lines. The Pearson *r* value is presented in the top left corner of each plot. Participants shifted to much greater extents on average when images were hidden versus shown, but no significant differences emerged between the individual and group conditions.

Table 13
Mean and standard deviation values (in parentheses) of criterion placement (c) and discriminability (d') across social conditions (group vs. self), memory conditions (hidden vs. shown), and criterion conditions (conservative vs. liberal)

Social Condition	Memory Condition	Criterion Condition	*c*	*d'*
Group	Hidden	Conservative	1.36 (0.97)	0.02 (0.47)
Group	Hidden	Liberal	-1.34 (0.94)	-0.01 (0.46)
Group	Shown	Conservative	-0.07 (0.46)	2.60 (0.62)
Group	Shown	Liberal	-0.47 (0.45)	2.50 (0.59)
Self	Hidden	Conservative	1.30 (0.91)	-0.07 (0.42)
Self	Hidden	Liberal	-1.26 (0.94)	-0.02 (0.46)
Self	Shown	Conservative	0.11 (0.49)	2.10 (0.63)
Self	Shown	Liberal	-0.32 (0.54)	1.93 (0.67)

3.4.4.3 - Criterion Shifting (Δc) versus survey measures

Participants conducted a battery of surveys prior to completing the task. The main purpose of these surveys was to assess whether characteristics such as social anxiety, current emotional states, altruistic tendencies, and feelings of connection relate to criterion shift strategies across conditions. Table 14 presents the mean survey scores and the Pearson r correlations between each survey measure and criterion shifting (Δc) in both the individual and group conditions, with images either shown or hidden. The majority of the survey measures did not exhibit consistent relationships with criterion shift measures. However, most SAQ scores demonstrated a moderate negative relationship with criterion shifting, specifically in the hidden images condition. This suggests that individuals prone to stress and anxiety may be less likely to maximize their responses in the hidden image condition. This behavior could be influenced by demand characteristics, where participants feel compelled to vary their responses, perceiving that consistently maximizing responses may appear as if they are not putting in enough effort, despite it being the optimal strategy. However, future research is needed to replicate this finding and to investigate why individuals more susceptible to stress and anxiety are less likely to adopt a maximizing strategy, even when it is the optimal approach.

Table 14
Pearson Correlation Coefficients and 95% CIs (in Brackets) Between Δc in Shown vs. Hidden Condition and Individual vs. Group Condition Against the Questionnaire Measures

		Individual		Group	
Measure vs. **Δ***c*	Mean *(SD)*	Shown	Hidden	Shown	Hidden
SAQ Total	94.30 (16.58)	-.03 [-.32, .26] (5.32)	**-.37 [-.60, -.09] (0.23)**	.21 [-.08, .47] (2.03)	**-.35 [-.58, -.07] (0.34)**
SAQ (1)	20.00 (3.45)	.06 [-.24, .34] (5.07)	**-.40 [-.62, -.12] (0.14)**	.27 [-.02, .52] (1.13)	**-.44 [-.65, -.17] (0.06)**
SAQ (2)	19.22 (4.11)	.08 [-.22, .36] (4.82)	**-.33 [-.56, -.04] (0.51)**	.24 [-.05, .50] (1.52)	**-.30 [-.54, -.01] (0.75)**
SAQ (3)	20.52 (3.40)	.02 [-.28, .31] (5.41)	**-.35 [-.58, -.07] (0.35)**	.18 [-.11, .45] (2.67)	**-.35 [-.58, -.07] (0.35)**
SAQ (4)	17.61 (3.85)	-.05 [-.33, .25] (5.18)	-.22 [-.48, .08] (1.99)	.20 [-.09, .47] (2.22)	-.16 [-.43, .13] (3.10)
SAQ (5)	16.96 (4.22)	-.22 [-.48, .07] (1.89)	**-.34 [-.57, -.06] (0.40)**	.05 [-.24, .34] (5.14)	**-.30 [-.54, -.01] (0.75)**
PANAS Positive	26.04 (7.21)	-.16 [-.43, .13] (3.07)	-.10 [-.38, .20] (4.43)	-.01 [-.30, .28] (5.42)	.07 [-.23, .35] (4.92)
PANAS Negative	15.30 (4.67)	-.27 [-.52, .02] (1.14)	-.24 [-.49, .06] (1.63)	.01 [-.28, .30] (5.44)	-.17 [-.43, .13] (3.04)
Altruism Scale	28.72 (8.52)	-.10 [-.38, .19] (4.36)	**.31 [.03, .55] (0.62)**	-.12 [-.40, .18] (3.98)	.17 [-.13, .44] (2.98)
Watt General	53.91 (12.20)	-.16 [-.43, .14] (3.20)	-.07 [-.35, .23] (4.91)	-.10 [-.38, .20] (4.47)	-.06 [-.34, .23] (5.04)
Watt Self	63.50 (16.06)	.07 [-.22, .35] (4.90)	.01 [-.28, .30] (5.44)	.09 [-.20, .37] (4.52)	-.00 [-.29, .29] (5.44)
Watt Other	50.58 (18.45)	-.10 [-.38, .20] (4.49)	-.07 [-.35, .22] (4.89)	-.04 [-.32, .26] (5.29)	.05 [-.25, .33] (5.18)
Watt World	47.66 (19.58)	-.26 [-.51, .03] (1.20)	-.07 [-.35, .23] (4.94)	-.22 [-.48, .07] (1.88)	-.15 [-.43, .14] (3.28)

3.4.5 - Discussion

Experiment 3, like Experiment 2, showed a criterion shifting effect, but revealed no significant differences in the extent of criterion shifting as a result of social pressure. This finding indicates that social pressure is no larger of an influence on criterion shifting compared to any other manipulation that has been previously studied (such as monetary incentives). Criterion shifting tendencies were strongly correlated between the individual and group conditions, providing further support to the notion that such tendencies are a stable cognitive trait (Layher et al, 2020). In other words, people who have a tendency to employ criterion shifting strategies employ them just the same under social pressure as they do under other circumstances. An interesting relationship that did emerge is that criterion shifting tendencies in the hidden condition were negatively correlated with most of the SAQ measures. The hidden condition, specifically, requires participants to always respond “old” in the liberal condition and “new” in the conservative condition to maximize payouts. Sometimes individuals applied this maximizing strategy while others did not. Some participants may have been influenced by demand characteristics, possibly believing that consistently providing the same response during a test block meant they were not putting forth enough effort on the task. Since higher SAQ scores are associated with increased social anxiety, the negative relationships could indicate that those who are more socially anxious are less likely to implement the maximizing strategy in the hidden condition, regardless of the social condition. However, it is notable that SAQ measures did not relate to criterion shifting in the shown condition, indicating that this relationship might be specific to maximizing behavior as opposed to criterion shifting in general. Future research should better assess whether social anxiety is associated with the likelihood someone implements maximizing behavior, and whether such a relationship is due to demand characteristics.

While social pressure did not affect criterion shifting strategies, it did impact discriminability in the shown condition. The two ways in which participants can improve their decisional outcomes during memory tests in this paradigm, is to increase the number of correct responses and shift criteria to larger extents to avoid critical errors. In the group condition, *d’* substantially improved relative to the individual condition, indicating that participants put more effort into memorizing the images. It is possible that participants put greater effort to encode images in the study phase, or put more effort in their response at test, when their decisions impacted the group. Importantly, the group condition *did* improve decisional outcomes, but only by improving memory performance and not via larger criterion shifts. It is intriguing that participants did not alter criterion shifting strategies to improve decisional outcomes in the group condition, particularly in the hidden condition where it was impossible to utilize memory. This finding provides further evidence for the strong stability of criterion shifting strategies.

3.5 General Discussion

These two experiments sought to examine whether social influences can impact criterion shifting tendencies across individuals during recognition memory tasks. Experiment 2 aimed to investigate how different motivations, specifically monetary rewards and social competition, influence criterion shifting behavior. Experiment 3 was designed to clarify any potential effects of social influences on the interaction between memory performance and criterion shifting by introducing a different type of social pressure and adjusting task difficulty. Specifically, Experiment 3 aimed to determine whether participants' knowledge of their impact on others would influence criterion shifting behavior and memory performance.

The extent of criterion shifting varied substantially across individuals in both experiments, but the degree to which individuals shifted was not significantly affected by monetary incentives or social pressures. Criterion shifting tendencies remained quite consistent across all experimental manipulations, providing further support that such strategies are indicative of a stable cognitive trait (Layher et al., 2020). In fact, the test block order proved to be the biggest factor in altering the extent of criterion shifting. Participants, on average, shifted criteria to larger extents across subsequent test blocks, particularly between test blocks 1 and 2, a finding that Layher and colleagues (2020) also identified. This is due to *some* individuals shifting criteria more extremely across test blocks, possibly influenced by the end-of-test block feedback that heightened their awareness of the consequences of inadequate criterion shifts. Future research is necessary to identify the underlying factors that cause some individuals to shift more extensively across test blocks.

However, monetary incentives and social pressure manipulations did not significantly influence criterion shifting.

Although criterion shifting tendencies remained consistent across manipulations, the results of Experiment 3 indicate that social pressure with group consequences can enhance memory performance in recognition tasks, specifically by improving discriminability (*d'*). Participants in the group condition demonstrated significantly better recognition accuracy compared to those in the individual condition, likely due to the motivational impact of social pressure. It is possible that social pressures made participants more attentive during the study phase and/or they put greater effort in their memory responses during tests. Since participants were aware of the social condition before the study phase, improved discriminability may have resulted from improving either encoding or retrieval processes. Social pressures in Experiment 2, however, did not affect discriminability, which might be due to the extreme difficulty of the recognition task in that experiment. Since performance was at near chance levels, it is possible that extra effort to encode or retrieve items would prove fruitless. The findings of Experiment 3 underscore the powerful influence that social acceptance can have as a motivator to perform better. Whether driven by fear of consequences or potential for commendation, individuals perform better when their performance affects the entire group and is not constrained by extreme task difficulty. This indicates that while social pressure can enhance memory performance, it does not necessarily alter the underlying decision-making processes individuals employ.

The finding that social pressure enhances discriminability without affecting the extent of criterion shifting is particularly interesting when considering the results of the hidden image condition in Experiment 3. The hidden image condition creates a scenario where no memory evidence is available to inform participants' decisions, leading to performance inevitably at chance levels. The only way to achieve optimal decisional outcomes is to maximize responses by always responding "old" in the liberal criterion condition and "new" in the conservative criterion condition. While some individuals appropriately maximized, others did not, regardless of the social condition. Given that participants in Experiment 3 demonstrated a willingness to improve discriminability performance under social pressure in the shown image condition, it is curious why certain individuals were unwilling to adopt a maximizing strategy in the hidden image condition to optimize group outcomes. One possible explanation comes from significant relationships observed between maximizing behavior in the hidden image condition and SAQ scores.

The moderate negative relationship between SAQ scores and criterion shifting in the hidden image condition suggests that individuals with higher levels of social anxiety are less likely to maximize their responses. This reluctance to maximize, despite it being the optimal strategy, may be driven by demand characteristics. Individuals prone to stress and anxiety might feel compelled to vary their responses, possibly to avoid appearing disengaged or unmotivated to the experimenters. This behavior highlights how psychological factors can influence decision-making processes, particularly under conditions of absolute uncertainty. The observed reluctance among socially anxious individuals to adopt a maximizing strategy in the hidden image task warrants further exploration. Future research should aim to replicate these findings and investigate the underlying mechanisms that drive this behavior. Understanding why individuals with higher social anxiety are less likely to maximize their responses could provide valuable insights into how stress and anxiety influence cognitive strategies.

In summary, the results of Experiment 3 emphasize the significant impact of social pressure on enhancing memory performance in recognition tasks. Participants are motivated to improve discriminability to achieve greater decisional outcomes under social pressure, but are unwilling to adjust criterion shifting strategies to do so. The consistency in criterion shifting strategies across individual and group conditions suggests that decision-making processes remain stable, even under social pressure. Experiment 3's design, which allowed for better disentanglement of social influences on memory performance and criterion shifting, provided clearer insights compared to Experiment 2. This underscores the importance of task difficulty and social context in shaping cognitive strategies.

Chapter 4: Experiment 4 - Recognition and Cued Recall

4.1 Introduction

Recognition memory models are based primarily on the idea that people undertake two distinguishable processes when evaluating whether they have encountered a stimulus before: familiarity and recollection. Familiarity is viewed as an initial sense of a memory signal related to a target item previously seen and is likely based on the perceptual features of the target. Recollection, on the other hand, refers to the successful retrieval of source or contextual information of the target item that affirms a sense of familiarity (Yonelinas, 2002). A prime example of this is Mandler's (1980) "Butcher on the Bus" scenario, where a person gets on a bus and seemingly recognizes another passenger, a man whom they felt they "knew" prior. The person getting on the bus would likely conduct a mental search to figure out why the man seems familiar, examining various contexts that could explain the recognition. Eventually, the person concludes that the man is the butcher from the supermarket. Mandler (1980) states that familiarity and recollection can work in parallel, though he holds that familiarity is faster than recollection.

It is widely accepted in the memory literature that familiarity is a continuous process that is ranked based on the relationship between confidence and accuracy (Mickes, Wais, & Wixted, 2009). However, the process of recollection has been greatly debated, specifically, whether it is represented by a continuous or thresholded process. A thresholded process would assume that any item involving recollection would be reported as high confident 'old.' A continuous would not necessarily have this assumption, as recollected items can be reported with high or low confidence, which may be correct (hits) or incorrect (false alarms). This debate has led to the development of two prominent, competing models: the dual-process signal detection model (DPSD; which incorporates the equal variance signal detection model) and the unequal-variance signal detection (UVSD) model. The DPSD model characterizes recollection as a categorical procedure (or thresholded process) (Yonelinas, 1994; Yonelinas, 2001; Yonelinas & Parks, 2007). The UVSD model, by contrast, suggests that recollection is a continuous process that is graded like familiarity (Wixted, 2007a; Mickes, Wais, & Wixted, 2009).

Regarding the DPSD model, Yonelinas (2001) theorizes that familiarity and recollection lead to different thresholds for memory performance. He states that recollection involves retrieving qualitative information, while familiarity involves a signal-detection process where items that exceed a familiarity criterion are deemed previously viewed. Because familiarity is primarily based on introspection akin to the "gut feeling" about a memory (also known as a "familiarity value"), this is less complex than the recollection process, which involves searching for and retrieving specific information to determine if the stimulus was previously encountered. Based on Mandler's research and comments by Estes and Da Polito, recognition familiarity seems to rely on low levels of information storage. In contrast, recollection requires higher levels of information storage to determine if the target was previously seen.

Those in favor of the UVSD model have argued that it is a more accurate measure of discriminability and criterion placement because, for recognition memory tasks, target distributions will generally have a larger variance than lure distributions (see Figure 7; Egan, 1958; Mickes et al., 2007; Layher et al., 2020). To properly evaluate the UVSD model, experimental designs must include several criterion manipulations or confidence ratings to accurately assess the variance between the target and lure distributions (Macmillan &

Creelman, 2005; Layher et al., 2020). Neuroimaging studies have also supported the UVSD model's stance on recollection being a graded process because participants have retrieved details of encoded stimuli they claim not to recollect/remember. For example, in one study, remember responses increased activation in the retrosplenial cortex and posterior cingulate (Johnson, McDuff, Rugg, & Norman, 2009).

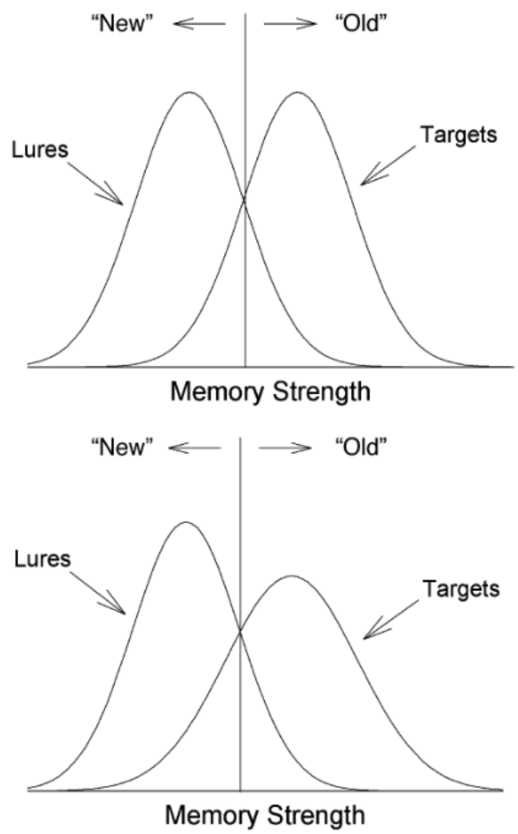


Figure 7: From Wixted (2007b): Equal variance (upper) and unequal variance (lower) signal detection models of recognition memory.

A more recent method to evaluate the recollection process within the DPSD model in recognition memory tasks is to use receiver operating characteristics (ROC) analysis. ROC analysis plots hit rates vs. false alarm rates across different levels of response bias (Wixted, 2020). Both models hold that familiarity behaves like a classic signal-detection process, where familiarity increases and produces a curved ROC as the response criterion relaxes (Yonelinas, 2001). Under the DPSD model, a confidence-based ROC plotting hits and false alarms for "remember" judgments would theoretically be linear if recollection were based on an extremely high threshold, representing pure recollection. However, in practice, such conditions are rarely observed. Because the DPSD model incorporates both familiarity and recollection judgments, the resulting ROC is curvilinear, making it virtually indistinguishable from data modeled by the UVSD framework. The key distinction lies in the underlying cause of this curvature: the UVSD model suggests that the curvilinear ROC arises due to a continuous signal detection process driven by familiarity, which results in unequal variances between old and new items (Wixted, 2007b; Wixted & Mickes, 2010).

Another way to see the difference between the two models is by looking at reaction times, which researchers have often used as a proxy for confidence (Starns & Ratcliff, 2014). In their 1998 study, Stretch and Wixted asked participants to make remember/know judgments and give confidence ratings for every recognition decision in the task. They found that participants reported "remember" more quickly and with higher confidence than "know" decisions, regardless of whether the judgment was correct or incorrect (see Table 15; Stretch & Wixted, 1998b; Wixted & Stretch, 2004).

Table 15:
Average Confidence Ratings and Average Reaction Times (in milliseconds) for Remember and Know Judgments to Targets and Lures From Experiment 1 of Stretch and Wixted (1998)

	Confidence		Reaction Time	
	Remember	Know	Remember	Know
Targets	4.64	3.03	718	797
Lures	3.82	2.80	734	812

The slower reaction time for "know" judgments may seem counterintuitive to the signal detection model (see Figure 8) because if an individual requires less information, then, in theory, they should be making these "know" decisions faster. A key premise of the DPSD model is that recollection is a slower process than familiarity (McElree, Dolan, & Jacoby, 1999; Yonelinas, 2002). For example, Gardiner (2001) defines recollection as "a relatively slow, effortful process, depending on conscious control, and that familiarity is a relatively fast, automatic process, not dependent upon conscious control." It may be that people are inclined to take more time to generate concrete (potentially autonoetic) memory evidence and see if recollection will "succeed" before they settle for making a response based on familiarity alone (Starns & Ratcliff, 2014). The confidence ratings from this study support this hypothesis because people tend to be less confident in their "know" judgments, thus potentially indicating that when making the decision, people take extra time to try to generate the evidence but are unable to do so, yet have enough of a feeling to commit to saying they had seen the stimuli before. It is imperative to note that the types of memory evidence used to make a decision may vary between feelings of familiarity and autonoetically generated episodic recollection. The ability to remain flexible and shift these criteria (a.k.a. to criterion shift), depending on the situation, is critical for successful decision-making (Wixted, 2020).

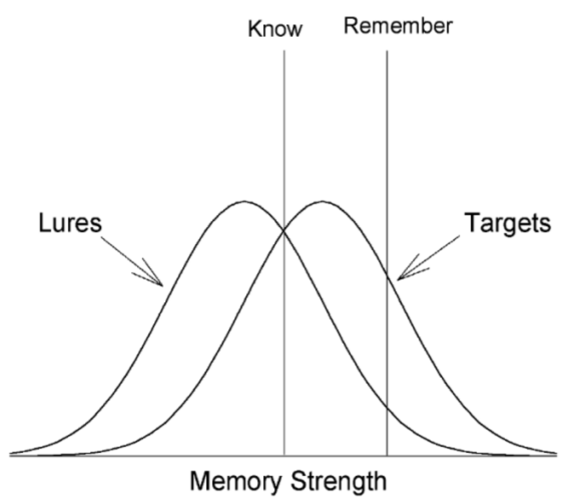


Figure 8: From Wixted (2007b): Signal detection interpretation of remember/know judgments.

Whether recollection systems rely more on a thresholded and binary process or one that is more continuous and graded is still an open question. Manipulating participants' criteria in a cued recall task design could support the DPSD or the UVSD model. The design of recognition memory tasks makes distinguishing a participant's reliance on either familiarity or recollection difficult, whereas a cued recall task design could prompt more reliance on recollection than recognition tasks. The impetus for Experiment 4 was to explore whether systematically increasing the reliance on recollection would test the presumptions of these two models. The DPSD model relies on recollection to a higher degree than the UVSD model. If the premise of the UVSD model is true, then criterion shift manipulations should be effective in the cued recall task despite the increased reliance on recollection. However, if the DPSD model is true, and a thresholded recollection process is impervious to a criterion shift, then it should be clear from the cued recall task that participants are relying on their recollection more than a decision strategy based on familiarity.

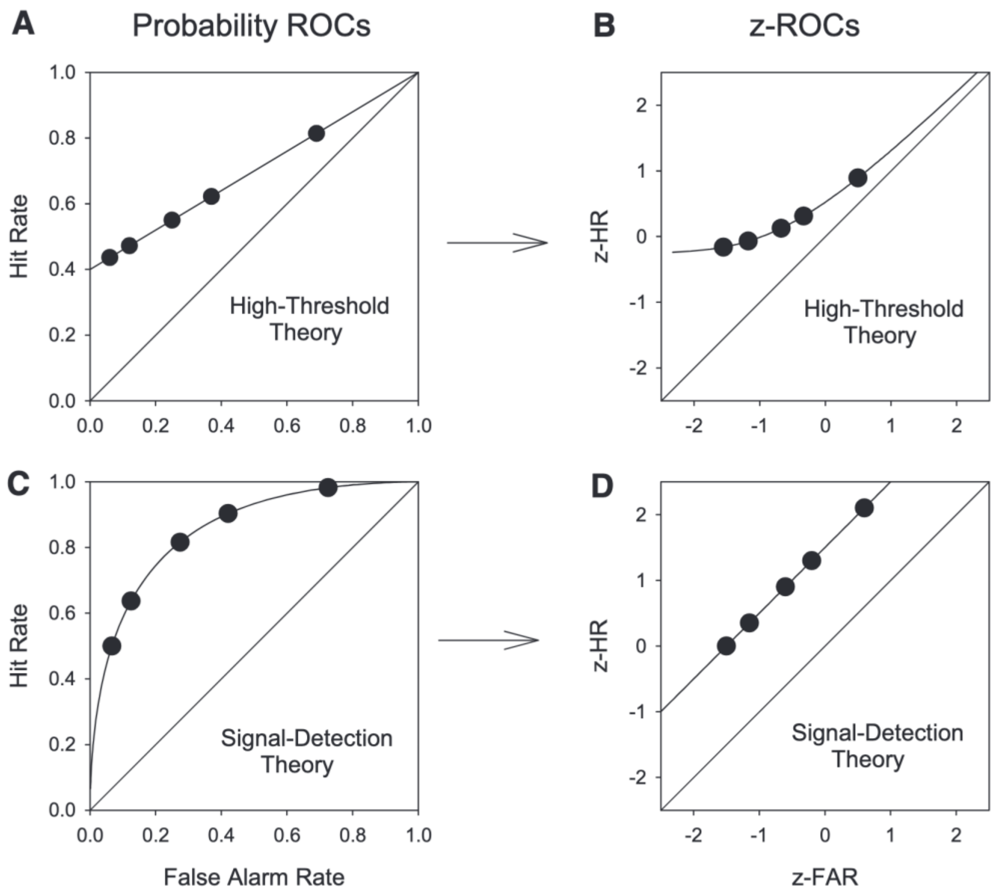


Figure 9: From Wixted (2007b): Idealized ROC data predicted by the high-threshold model (A) and by the signal-detection model (C). B and D show the corresponding ROCs in z-space.

To this end, Experiment 4 had a within-subject design, requiring participants to complete two tasks, one immediately after the other. First, participants completed a memory-based decision task, rating images as either "old" or "new," to establish a baseline for each participant's propensity to optimize their decisions (i.e., whether they are "good" or "bad" at shifting). This first task was modeled after the classic criterion shifting recognition memory paradigm that the Miller Memory lab is most well-known for, having established criterion shifting as a unique cognitive trait (Layher et al., 2018; Miller & Kantner, 2019; Layher et al., 2020). The recognition memory task was consistent with previous research using a blocked design, meaning that each experimental block contained the same criterion shift manipulation. If the criterion shift manipulations had been randomized within a recall block, then the burden of criterion shifting on a trial-by-trial basis may have been a factor that discouraged participants from using the most optimal decision strategy (Stretch & Wixted, 1998a). Then, each participant completed a more complex task that asked them to recall an image based on a question about a detail of that image, theoretically prompting participants to rely more on recollection than the earlier recognition memory task.

This research is an important step in determining if criterion shifting manipulations can be successfully applied to episodic free recall. If participants cannot criterion shift for the cued recall task, then this is likely an indicator that recollection plays a more prominent role in the decision-making process than familiarity, thus supporting the DPSD model. If criterion shifting is observed, a future experiment with a free recall component could investigate further if free recall aligns more closely with the thresholded DPSD model or the UVSD model (this idea will be discussed further in Chapter 5). Looking beyond the models themselves, knowing if free recall relies on a thresholded recollection system or one dependent on continuous familiarity and a strict recollection process could lead to improved interview techniques.

4.2 Method

4.2.1 - Participants

A total of 74 participants for Experiment 4 were recruited through the UCSB SONA website and the web-based recruitment platform, Prolific (Prolific, London, UK). Due to technical difficulties, 8 participants' data from the 42 SONA subjects were incomplete and were excluded from the final analysis, so only 66 participants' data were used (n = 66 [42 female, 23 male, one non-binary], *M* = 27.48 years, *SD* = 9.16, range = 18 - 45 years). Participants who enrolled through SONA received class credit, while participants from Prolific were compensated $10 for their time. All participants had the chance to earn a monetary bonus (up to $5) based on performance in the task.

Study participants self-identified their race on a demographic questionnaire: 10 participants (15%) identified as Black, 7 participants (10%) identified as Asian, 9 participants (14%) identified as Latine, 36 participants (55%) identified as White, and 4 participants (6%) identified as multiracial, which included 2 (3%) identifying as Latine and White and 2 (3%) as identifying as Asian and White. Experiment 4 was approved by the UCSB Human Subjects Committee Institutional Review Board (IRB). All participants gave informed consent via Qualtrics (Qualtrics, Provo, UT).

4.2.2 - Signal Detection Theory

Experiment 4 utilized an equal-variance SDT model, which categorizes participants' responses to ambiguous stimuli into one of four categories: hits (H; correctly identifying a stimulus that was present), misses (M; not identifying a stimulus that was present), false alarms (FA; identifying a stimulus that was not present) or correct rejections (CR; not identifying a stimulus that was indeed not there). From this, we can formulate scores into rates for each category. For example, the hit rate (*HR*) is obtained by dividing participants' total hits by the total possible correct categories, whereas the false alarm rate (*FAR*) is the total false alarms divided by the total possible incorrect categories. SDT measures of discriminability (*d'*) and criterion placement (*c*) were obtained using the following equations:

$$HR = (\text{number of H}) / (\text{number of H} + \text{number of M})$$

$$FAR = (\text{number of FA}) / (\text{number of total possible FA})$$

$$d' = z(HR) - z(FAR)$$

$$c = [z(HR) + z(FAR)] / 2$$

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

where *z* represents the density of the standard normal distribution (Macmillan & Creelman, 2005).

4.2.3 - Materials

Participants recruited through SONA completed Experiment 4 on the UCSB campus, in the Miller Memory Lab, with computers provided by the researchers, and participants enrolled through Prolific completed it online. Participants conducted all tasks using PsychoPy v2021.2.3 (Peirce et al., 2019). The image recognition task used stimuli from the THINGS Database, which contains a set of 1,854 diverse object concepts sampled systematically from concrete picturable and nameable nouns (e.g., cow, turtle, bicycle) in the American English language (Hebart et al., 2019). These images were displayed at a size of 500 x 500 pixels on a black background. The computer task for Experiment 4 was designed to have a black background with white lettering for instructions.

4.2.4 - Procedure

It was important to systematically evaluate individual differences within and across the image recognition and word cue tasks. Each task followed the same format, where participants (1) viewed an encoding block of images and (2) completed a memory test phase. In the encoding phase of the image recognition and word cue tasks, participants passively viewed 6 encoding blocks of 54 images each. After each encoding block of the image recognition and word cue tasks, participants were asked to complete 6 test blocks under three explicit condition instructions: conservative, liberal, and neutral. Participants used the “j” and “k” keys to respond “old” or “new” after each image appeared on the screen, depending on whether they believed they had seen the image in the most recent encoding phase. Each test trial ended after making an “old” or “new” decision, but if participants did not respond within 1.8 seconds, the next trial began automatically. Each test block contained 54 trials, 9 related to old stimuli and 9 related to new stimuli per condition (liberal, conservative, neutral) per block. In total, each task required the participant to make a determination about 324 stimuli, 162 of which were repeated from the encoding phase and 162 of which were new. After each trial, participants were asked to rate their confidence as “high,” “medium,” or “low.” Every image shown to participants was set to appear for 550 ms. Each image in the encoding and test phases was followed by a black screen for 200 ms before the next image. The brief exposure to stimuli during the encoding phase was intended to diminish discriminability performance to near-chance levels.

In the word cue task, instead of showing participants images, as in the image recognition task, the test phase only contained cue words in a question, and participants evaluated if the cue was “old” or “new.” For example, if a participant saw an image that included a turtle in the encoding phase, they would be asked in the testing phase, “Did you see a picture with a TURTLE?” and then they would select either “old” or “new.” The questions posed to participants were intentionally limited to the details of the image to avoid leading the participants to create false beliefs (i.e., implanting false memories or biasing the participant to believe that more details provided correlated to an increased likelihood that the image was studied). The description was written as a question as opposed to presented as a single word to increase ecological validity.

Criterion shift manipulations were the same for both tasks, utilizing the same monetary motivation: participants were rewarded 5 points (with each point being synonymous with one cent) for correctly responding "old" to a studied image (a hit) or "new" to an unstudied image (a correct rejection). In the conservative condition, participants were incentivized to choose the "new" response because an incorrect selection of "old" (a false alarm) would be penalized 20 points, while there would be no penalty for an incorrect selection of "new" (a miss). Conversely, in the liberal condition, participants were incentivized to choose the "old" response by penalizing misses, not false alarms. In the neutral condition, there were no penalties, only rewards for correct answers. At the end of each test phase block, the task screen displayed the bonus earnings for that block, along with an updated total of the participant's performance-based bonus earnings for the entire task.

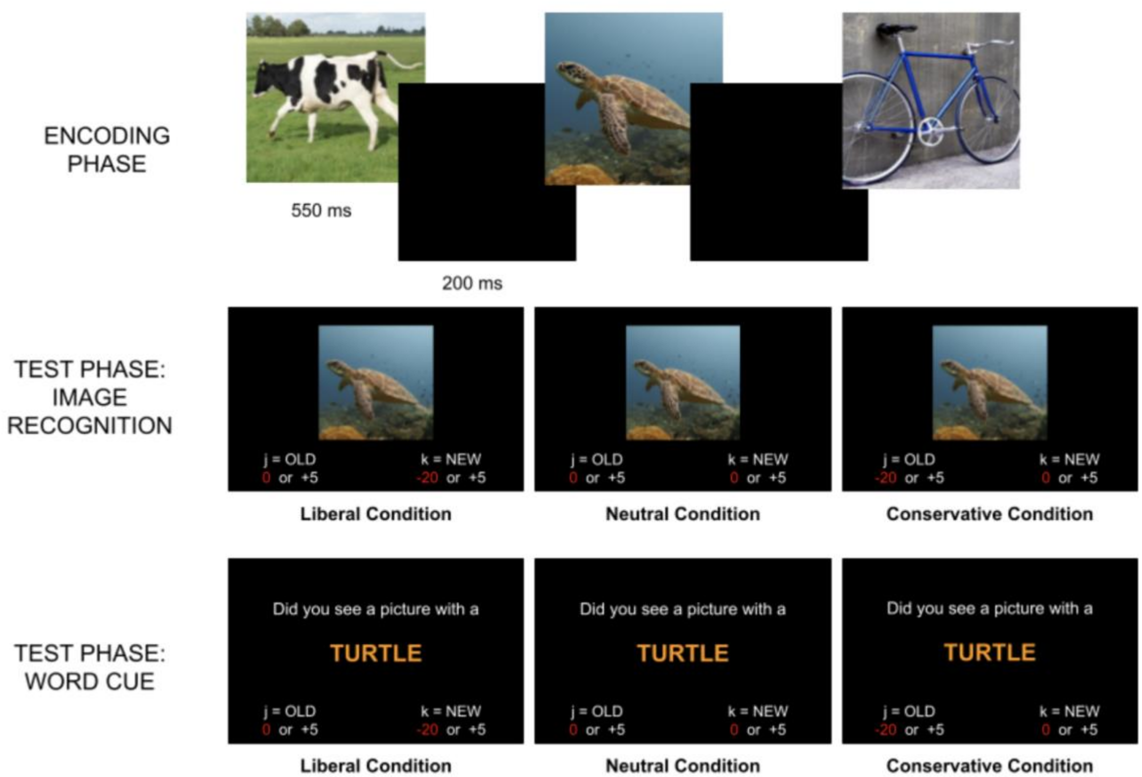


Figure 10: Experiment 4 image recognition and word cue memory tasks.

4.3 Results

4.3.1 - Discriminability (*d'*) for Image Recognition Task and Word Cue Task

Mean discriminability (*d'*) was significantly higher in the image recognition task (*M* = 2.44, *SD* = 1.07, 95% CI [2.29, 2.59]) compared to the word cue task (*M* = 1.07, *SD* = 0.78, 95% CI [0.96, 1.18]) on average ($p < .001$). Specifically, for the image recognition task, *d'* was highest in the neutral condition (*M* = 2.52, *SD* = 1.08), followed by the conservative condition (*M* = 2.46, *SD* = 1.06), and lowest in the liberal condition (*M* = 2.34, *SD* = 1.09). For the word cue task, *d'* was highest in the neutral condition (*M* = 1.14, *SD* = 0.79), followed by the conservative condition (*M* = 1.11, *SD* = 0.85), and lowest in the liberal condition (*M* = 0.95, *SD* = 0.71; see Table 16).

Despite the significant differences in discriminability, *d'* scores between the image recognition and word cue tasks were strongly correlated ($r(196) = 0.69$, $p < .001$), indicating that relative performance was consistent in both tasks. This relationship held across conditions, with correlations ranging from $r = 0.61$ in the liberal condition to $r = 0.74$ in the neutral condition, all highly significant ($p < .001$). As illustrated in Figure 11, the scatter plot reveals a strong positive correlation between *d'* scores for the image recognition and word cue tasks, with a consistent trend across all conditions. This substantial difference in discriminability, underscored by a large effect size (*Cohen's d* = 1.46), suggests that the image recognition task may be inherently easier or more reliant on familiarity processes, whereas the word cue task likely required more reliance on recollection. These findings indicate that, although the tasks differ in their overall difficulty, there is a consistent individual difference factor that influences performance across both tasks.

The disparity in *d'* between the image recognition and word cue tasks may reflect the differential reliance on cognitive processes such as familiarity and recollection. The higher discriminability in the image recognition task implies it might be more dependent on familiarity processes, which are generally faster and more automatic. In contrast, the word cue task likely demands a slower and more effortful process, leading to lower *d'* scores. The strong correlation across tasks, however, points to underlying cognitive abilities that are common to both tasks, such as general memory ability or attention to task-specific details.

Table 16

Discriminability (d'): mean, standard deviation values, and confidence intervals across criterion conditions (conservative vs. neutral vs. liberal)

Condition	Task	*M*	*SD*	95% CI (Low)	95% CI (High)
Conservative	Image Recognition	2.46	1.06	2.20	2.72
Conservative	Word Cue	1.11	0.85	0.90	1.32
Liberal	Image Recognition	2.34	1.09	2.07	2.60
Liberal	Word Cue	0.95	0.71	0.78	1.13
Neutral	Image Recognition	2.52	1.08	2.26	2.79
Neutral	Word Cue	1.14	0.79	0.95	1.34

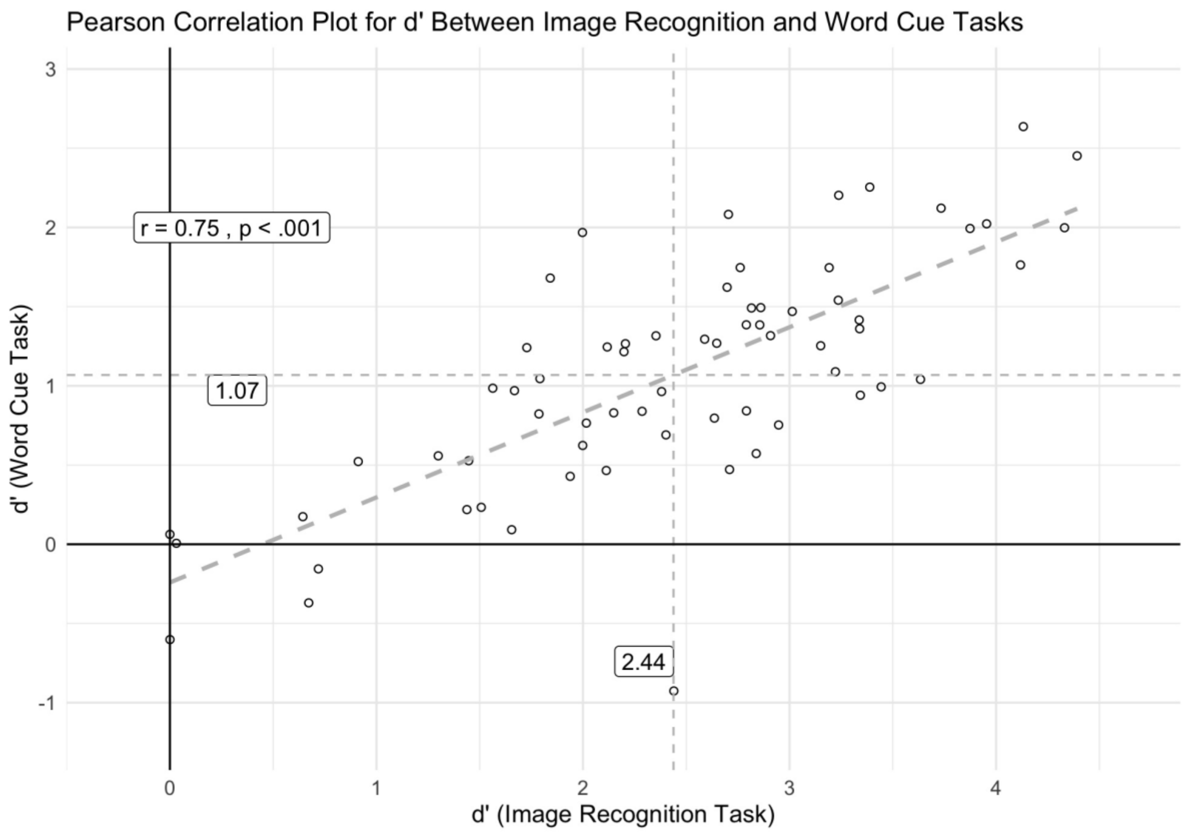


Figure 11: Pearson correlation plot depicting mean *d'* between the image recognition task vs. word cue task in Experiment 4. Mean values are presented in boxes on each axis, with accompanying dashed gray lines. The Pearson *r* value is presented in the top left corner.

4.3.2 - Criterion Shifting (**Δ***c*) for Image Recognition and Word Cue Tasks

Across Conditions

Participants exhibited modest to moderate shifts in their decision criteria (**Δ***c*) across different conditions in both the image recognition and word cue tasks, reflecting the varying cognitive demands associated with each task. In the image recognition task, where participants were more likely to rely on familiarity, criterion shifts were relatively modest. In the conservative condition, participants demonstrated a conservative bias (M = 0.26, SD = 0.48, 95% CI [0.14, 0.37]), suggesting a slightly more cautious approach when identifying images. The liberal condition (*M* = -0.10, *SD* = 0.51, 95% CI [-0.22, 0.03]) showed a small liberal bias (See Table 17).

Table 17
Criterion placement (c): mean, standard deviation values, and confidence intervals across criterion conditions (conservative vs. neutral vs. liberal)

Condition	Task	*M*	*SD*	95% CI (Low)	95% CI (High)
Conservative	Image Recognition	0.26	0.48	0.14	0.37
Conservative	Word Cue	0.39	0.89	0.31	0.48
Liberal	Image Recognition	-0.10	0.51	-0.22	0.03
Liberal	Word Cue	-0.30	0.97	-0.40	-0.21
Neutral	Image Recognition	0.17	0.27	0.11	0.24
Neutral	Word Cue	0.22	0.62	0.16	0.29

The word cue task resulted in more pronounced criterion shifts. In the conservative condition, participants exhibited a moderate conservative bias (*M* = 0.39, *SD* = 0.88, 95% CI [0.31, 0.48]), indicating a stronger bias toward a conservative response strategy when cues demanded recollection. The liberal condition also showed a moderate bias (*M* = -0.30, *SD* = 0.97, 95% CI [-0.40, 0.20]), suggesting that when the task encouraged a more liberal response criterion, participants were more willing to accept items as "old" even when they might not have been.

Across both tasks, the bias in the neutral condition was slightly conservative. In the word cue task (*M* = 0.22, *SD* = 0.62, 95% CI [0.16, 0.29]) participants demonstrated a slightly more conservative bias compared to the image recognition task (*M* = 0.17, *SD* = 0.27, 95% CI [0.10, 0.24]). These results imply that participants on average maintained a small bias even when explicit biasing cues were not provided.

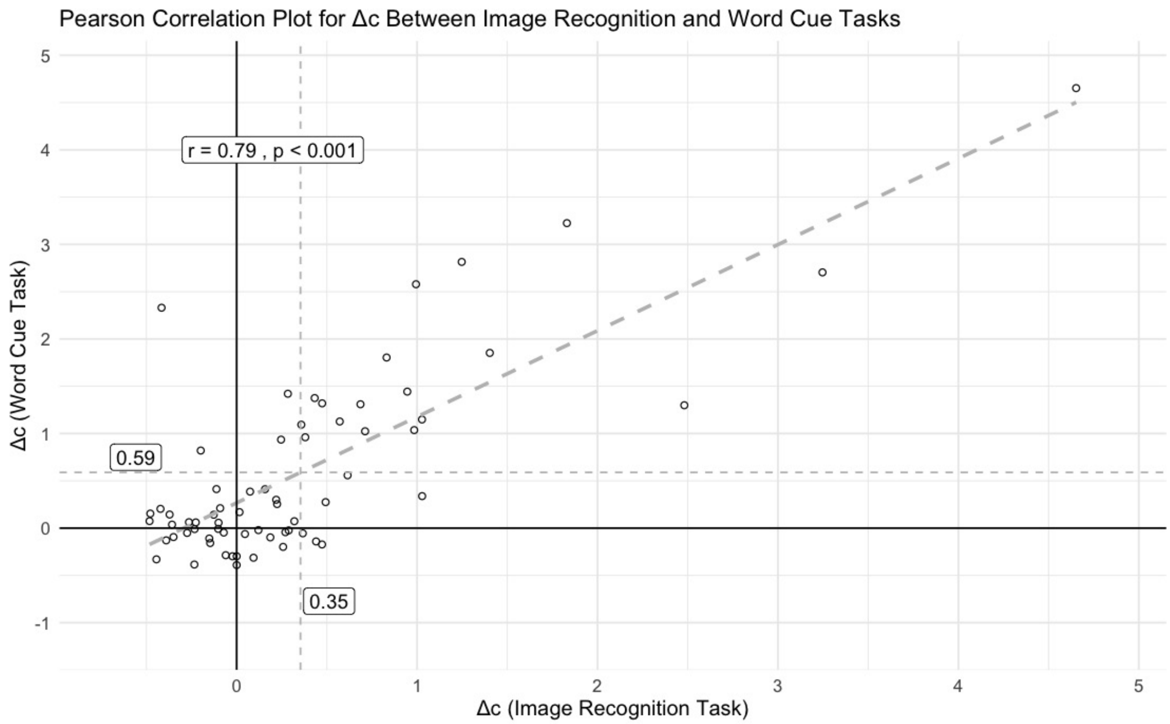


Figure 12: Pearson correlation plot depicting mean **Δ***c* between the image recognition task vs. word cue task. Mean values are presented in boxes on each axis, with accompanying dashed gray lines. The Pearson *r* value is presented in the top left corner of the plot.

A paired t-test comparing the first (blocks 1-3) and second (blocks 4-6) halves of the test blocks showed a significant shift towards a more conservative criterion in the second half, $t(65) = -2.99$, $p = .004$, with a mean difference of -0.10, 95% *CI* [-0.16, -0.03]. This shift likely reflects learning effects, fatigue, or strategic adjustments, highlighting a more flexible and adaptive response strategy in the word cue task compared to the image recognition task. These results suggest that while the image recognition task fostered stable decision criteria, the word cue task allowed for more adaptive, condition-dependent strategies, inferring different cognitive mechanisms are at play in these memory-based tasks.

4.3.3 - Confidence Ratings for Image Recognition and Word Cue Tasks Across Conditions

Further analysis focused on participants' confidence ratings and decision strategy, examining how these were influenced by the experimental conditions (liberal, neutral, conservative) within the image recognition and word cue tasks. Confidence ratings were converted into a numeric scale, where higher values indicated greater confidence.

A generalized linear mixed-effects model was used to examine the impact of confidence ratings and task type on response accuracy, accounting for random intercepts across participants. The analysis revealed significant main effects of both confidence ratings and task type and notable interactions between these variables. Specifically, higher confidence ratings were strongly associated with increased accuracy. Participants with medium confidence were significantly more accurate than those with low confidence, $\beta = 0.79$, $SE = 0.06$, $z = 13.72$, $p < .001$, and those with high confidence showed even greater accuracy, $\beta = 1.75$, $SE = 0.06$, $z = 31.79$, $p < .001$. Task type also played a critical role, with accuracy being lower in the word cue task compared to the recognition task, $\beta = -0.52$, $SE = 0.05$, $z = -10.14$, $p < .001$.

Significant interactions between confidence ratings and task type indicate that the type of task moderated the relationship between confidence and accuracy. The benefit of medium and high confidence on accuracy was reduced in the word cue task compared to the image recognition task, $\beta = -0.27$, $SE = 0.07$, $z = -3.98$, $p < .001$, and $\beta = -0.49$, $SE = 0.06$, $z = -7.61$, $p < .001$, respectively. These findings suggest that while higher confidence generally predicts greater accuracy, this effect may be more reflective of the overall ease of the recognition task rather than differences in the type of memory process. Specifically, the higher accuracy in the image recognition task could be attributed to the task's relative simplicity, where confidence more closely tracks performance levels, consistent with the notion that confidence-accuracy relationships typically strengthen as *d'* increases.

The ROC curves (see Figure 13) demonstrate a noticeable curvature across all conditions, providing robust evidence supporting a SDT framework, particularly the UVSD model. This model accounts for the observed asymmetry in the ROC curves by proposing that the variances of the old (target) and new (lure) item distributions differ, with the old item distribution typically showing greater variance. According to the UVSD model, the variance of the old item distribution is greater than that of the new item distribution and is often attributed to encoding variability. Specifically, some items are likely encoded more strongly or distinctively than others due to variations in factors such as attention, context, or the inherent properties of the stimuli themselves.

In the present study, the greater curvature seen in the image recognition task relative to the word cue task likely stems from the increased reliance on familiarity-based recognition in the image task, whereas the word cue task demands more effortful recollection processes. As a result, the memory strength of old items in the image recognition task appears more variable, leading to a more spread-out distribution, while new items maintain a more uniform distribution since they have not been previously encountered and lack such variability. Notably, when conducting a z-transformation on the hit and false alarm rates, the resulting z-ROC curves become linear (see Figure 14), indicating that the underlying memory strength distributions are Gaussian rather than rectangular. This linearity contrasts with threshold-based models like the DPSD model, which would predict that raw hit and false alarm rates would produce linear ROCs if recollection—a discrete, threshold-based process—were predominantly driving recognition. Instead, the straight lines observed after the z-transform confirm that the distributions are Gaussian, reinforcing the notion that recognition memory is better captured by a continuous, signal-detection-based framework rather than a threshold-based model.

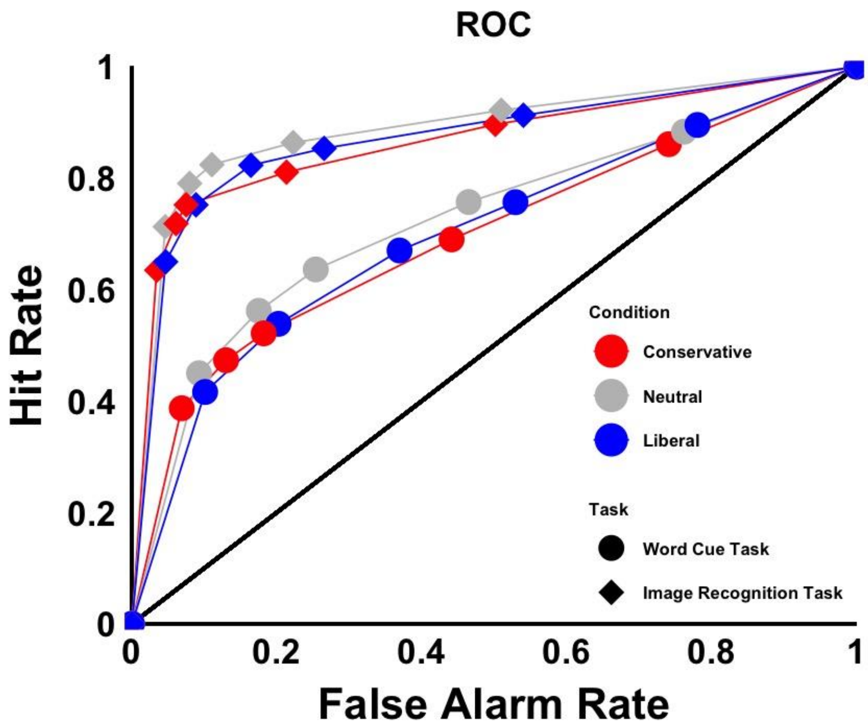


Figure 13: Receiver operator characteristic (ROC) curve for image recognition and word cue tasks across criterion conditions.

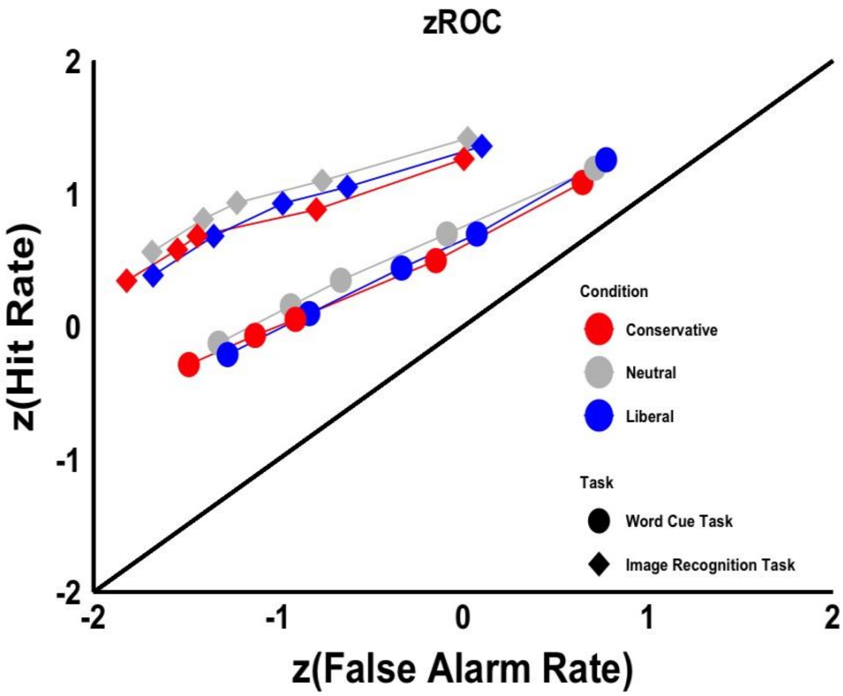


Figure 14: *z*-Receiver operator characteristic (ROC) curve for image recognition and word cue tasks across criterion conditions.

4.4 Discussion

Performance on memory tasks is generally believed to rely on familiarity and recollection. During recognition memory tasks, participants rely heavily on familiarity to inform their memory judgments, although recollection can play a substantial role. However, during cued recall tasks, participants are forced to use recollection to identify previously studied items. In most circumstances, recollection contributes to a greater extent during cued recall tasks compared to recognition memory tasks. Given that recollection differentially contributes to memory responses across these task types, cross-task assessments can help determine how recollection is represented.

According to the DPSD model, recollection is theorized to be represented by discrete states and, therefore, models recollection as a thresholded process. However, the UVSD model assumes that recollection is represented in a continuous manner, in which the underlying memory strength distributions are Gaussian in nature. Since recollection is presumed to play a substantial role in the cued recall task, the assumption under the DPSD model would be that the UVSD model is a poor fit to show recall performance. However, if memory distributions are continuous and best represented by Gaussian distributions, then the UVSD model will fit performance data on cued recall and recognition memory tasks. By incorporating criterion manipulations into each task, one can better assess the degree to which the UVSD model fits the data. The UVSD model would predict prominent criterion shifts across conditions regardless of whether familiarity or recollection contributed to the judgment, as the threshold for identifying images as 'old' should shift towards stronger memory evidence as incorrect 'old' responses become disadvantageous. However, the DPSD model would predict that people will always recognize recollected items with high confidence, a situation where criterion shifts would have little influence.

Criterion shifting (Δc) provides insight into participants' strategic adjustments in response to different cognitive demands across tasks. More complicated tasks are expected to result in more significant shifting, so it was not surprising to see the more substantial criterion shifts in the word cue task due to its increased difficulty. The consistency in participants' tendencies to criterion shift across both the word cue and image recognition tasks strongly suggests participants' use of common decision-making strategies that persist regardless of the task's cognitive demands.

The findings from the linear mixed-effects model demonstrate a robust positive correlation between confidence levels and accuracy, especially in the image recognition task, where high confidence was strongly linked to correct responses. This correlation implies that confidence is a reliable indicator of accuracy in tasks that depend on familiarity-based processes. Conversely, in the context of the word cue task, the diminished advantage of high confidence suggests that even when participants are confident, the complexity of recollection introduces greater uncertainty or potential for error, rendering confidence a less dependable predictor of accuracy. However, since the cued recall task was more difficult, the reduced accuracy to confidence ratings may be attributable to task difficulty rather than task type. These results highlight the importance of considering task type, difficulty level, and confidence when assessing performance. While confidence generally predicts accuracy well, its reliability varies depending on the cognitive demands of the task, emphasizing the need for careful experimental design.

The ROC curves demonstrate that the memory strength of both the recognition and cued recall task appear to be best represented by Gaussian distributions, as predicted by the UVSD model. According to the DPSD model, recollection is a thresholded process, which would presume that the memory strength distributions are rectangular instead of Gaussian. Since the cued recall task requires recollection for successful performance, the DPSD model would expect linear ROC curves. However, ROC curves were curvilinear for both tasks. The normalized ROC curves were also linear, indicating a solid fit for UVSD predictions relative to thresholded models.

Overall, these findings contribute to the ongoing debate between the DPSD and UVSD models by suggesting that while familiarity-driven processes in recognition tasks align well with the DPSD model, the more variable performance in tasks requiring recollection might be better explained by the UVSD model. The nuanced differences in how confidence relates to accuracy across tasks emphasize the complexity of memory processes and suggest that familiarity and recollection may not be as distinctly separate as the DPSD model proposes.

Chapter 5: Conclusion

The overarching goals of this dissertation are to explore the application of criterion shifting in ecologically valid situations, investigate the influence of social external factors on decision strategies, and identify the recollection processes that provide the best framework for future studies of memory and decision-making.

The results from **Experiment 1** demonstrate how criterion shifts, influenced by the instructions provided, played a significant role in how participants reported their recalled experiences. Specifically, instructions to be more conservative led participants to adopt a stricter decision criterion, resulting in fewer details being reported. This interpretation is bolstered by the lack of significant change in the frequency of correct information reported in the liberal condition, indicating that the liberal instructions did not further lower the already liberal reporting threshold established during the Baseline condition. Consequently, the data suggest that the conservative instructions effectively raised the decision threshold, leading to a more selective recall process without necessarily improving the accuracy of the recalled information. The findings from this research have the potential to impact the more complex scenarios in society that depend on the free recall of rich episodic memories, such as eyewitness testimony. Further research like this could also help to determine whether people utilize a threshold for memory strength during the recall of autobiographical memories, what the parameters of that threshold are, and to what extent that threshold can be manipulated.

Future research may also consider using stimuli with emotional valence in mind to improve ecological validity. Previous research has found that individuals rate negative emotional memories as more subjectively vivid and are more likely to remember details of negative events than neutral or positive experiences. Neuroimaging research has further supported this finding based on increased activation within the amygdala and orbitofrontal cortex during the encoding and retrieval of negative stimuli (Kensinger, 2007). There is an abundance of evidence that points to the stability of criterion shifting tendencies across stimuli, tasks, situations, and decision domains (Layher et al., 2020; Layher, Santander et al., 2023), but there has not been a study yet to confirm this stability in the presence of emotional stimuli.

Experiments 2 and 3 provide nuanced insights into how social contexts may influence individual decision-making strategies and discriminability during recognition memory tests. Neither experiment showed a significant impact of social pressure on individual criterion shifting tendencies, further supporting the notion that such tendencies are stable cognitive traits (Layher et al., 2020; Layher, Santander et al., 2023). While tasks designed to increase a sense of social competition (Experiment 2) did not significantly affect discriminability under conditions of nearly complete uncertainty, tasks with collective outcomes (Experiment 3) positively affected individual memory performance when the task was relatively easy. This demonstrates that participants are willing to improve discriminability under social pressure, but this improvement is only feasible when the task is easy enough to allow for enhanced performance. The new group consequences design in Experiment 3 addressed the potential limitation of Experiment 2 (that an indirect sense of social competition was not a strong enough pressure) and provided more precise insights into the interaction between social influences and cognitive processes. While these findings further support the general idea that social pressure can be leveraged to increase performance, that leverage might have a limit when it comes to specific cognitive tasks, such as shifting decision criteria. Ultimately, these findings contribute to the broader understanding of how social contexts influence cognitive processes and open avenues for future research to explore these effects in more detail.

The goal of **Experiment 4** was to begin a systematic exploration of how criterion shifting tendencies observed in recognition memory tasks extend to tasks that increasingly rely on recollection, offering a preliminary step toward understanding these dynamics in the context of the free recall of rich episodic events. This study focused on recognition tasks without a free recall component, using both a word cue and image recognition paradigm. While criterion shifts were expected during the recognition memory task, prior research on criterion shifting tendencies during cued recall is lacking. Since the DPSD model predicts recollection to be a thresholded process, it might be expected that participants may not shift their criteria at all during cued recall since recollected items are predicted to encompass very strong memories that should be virtually impervious to criterion setting. Interestingly, the consistency in participants' criterion shifts across both tasks, despite their differing cognitive demands, indicates a robust underlying strategy that participants employed regardless of whether the task relied more heavily on recollection or familiarity. This consistency suggests that while the cognitive demands of the tasks differ, participants may apply a generalized decision-making framework that adapts to the varying levels of difficulty but remains fundamentally similar across different memory tasks. This finding supports the idea that criterion shifting is a flexible yet consistent strategy employed by participants to optimize their performance across varying contexts, whether the task predominantly involves recollection or familiarity-based recognition.

Building on the observed consistency in criterion shifting across tasks, the z-transformed ROC curves provide further insights into the underlying memory processes.

Specifically, the linearity of the normalized curves supports the UVSD model's assumption that memory strength distributions are Gaussian, which contrasts with the rectangular distributions predicted by the DPSD model for tasks dominated by recollection. Although the results lean toward UVSD interpretations, the consistency in criterion shifts across both tasks suggests that participants apply similar decision-making strategies regardless of cognitive demand. These findings contribute to the broader discussion by highlighting that while familiarity-driven processes in recognition tasks align with the DPSD model, the variability introduced in recollection tasks may be better captured by a continuous signal detection framework.

Requiring participants to rely increasingly on recollection for the memory evidence they used to make decisions resulted in findings suggesting that the task's nature significantly influenced how participants adjusted their response criteria. The more pronounced shifts in the word cue task highlight the increased cognitive demand of recollection, where participants needed to exert greater control over their response criteria. In contrast, the modest shifts observed in the image recognition task are consistent with the idea that familiarity-based recognition is less demanding, allowing for more stable criterion placement across conditions. Additionally, confidence was a better predictor of accuracy in the image recognition task than in the word cue task, and high confidence was not reliably predictive of higher accuracy for either task. This finding challenges the DPSD model's assumption of a high-threshold process for recollection and suggests that familiarity and recollection may not be as distinct as the DPSD model assumes. However, it is important to note that this effect could be influenced by task difficulty rather than task type alone, as the *d'* values were unbalanced between tasks, potentially contributing to the observed differences in confidence-accuracy relationships. The greater difficulty of the word cue task might have led to lower accuracy and reduced the predictive power of confidence, confounding the interpretation of these results.

The next experiment to conduct in this line of inquiry should assess whether criterion shifting habits are present in a free recall task. As a precursor to the free recall task, a different set of participants would be recruited to complete a description task. Participants would view all the images that would be used in the free recall task and identify what they believe are the most important details to describe each image. Each participant will come up with four unique, one-to-four-word descriptors for the image, with explicit instructions not to use the word that would be used as the cue word in the free recall task. The top four details for each image would be used to assess hits in the subsequent free recall task. For example, participants may see a picture of a sea turtle underwater above a coral reef with the cue word "TURTLE" and come up with the following aggregate top-four detail list: green fins, hard shell, blue water, coral reef.

The free recall experiment could include the same type of description as seen in the word cue task, followed by a free recall task depending on the response. In both conditions, if the participant selects "new," they would complete the confidence rating before moving on to the next trial. If the participant selects "old" in either criterion shifting condition to the cued recall prompt, their response would be followed by a free recall prompt. The participant would be directed to a text box to describe the image and receive criterion shift manipulation instructions (to act more liberally or conservatively) in writing down what they recall. Participants will be informed about the precursor task and that the top four most reported details will count in the free recall evaluations for the criterion shift manipulations. For example, liberal instructions could be phrased as follows:

> *Please describe in detail what you remember. For every correct piece of information reported, you will receive 10¢, and for every piece of information you do not include, you will lose 20¢. There is no penalty for reporting incorrect information/guessing. Remember, you will be evaluated on the top four descriptors reported by the prior group of subjects when evaluating the most memorable aspects of the images.*

In this testing phase, one of the test blocks would have a criterion shift manipulation that is congruent with the initial manipulation condition (e.g., conservative during the word cue portion and in the free recall), and in the other block, the manipulation would be incongruent (e.g., conservative during the word cue portion and liberal in the free recall), with the order randomized. Participants would receive additional information to help establish a clear decision strategy. For example, if the free recall condition is congruent (liberal) with the liberal initial decision manipulation, an extra instruction could be included as follows:

> *If you choose "Old," you will have the opportunity to earn more money by completing a free recall task. In a previous study, other participants identified the most important details needed to describe each image, and your job is to report what you can remember about the image. You will need to recall the top four specific details about the image: for each correct detail (x; out of four total), you earn +20¢. However, for each detail you miss (y), -10¢ will be deducted. Incorrect details (z) do not affect your score and count as 0¢ each.*
>
> *For example, if you select "Old," at the initial word cue and your choice is correct, you will receive 10¢ for that choice, plus additional earnings based on how many details you recall. However, missing details will lower your total score. The Reward column in the table shows the maximum possible earnings if you recall all four correct details. The Risk column shows the potential losses if you miss all four details, which could result in a negative amount. If your initial selection of "Old" at the word cue is incorrect, then you will not earn or lose anything based on your responses.*

Initial Choice	Free Recall Rules	Equation	Reward	Risk
New	N/A	N/A	*10¢*	*-20¢*
Old	*+ 20¢ = Correct detail* *– 10¢ = Missed detail* *0¢ = Incorrect/false detail*	Correct "Old" Choice: 10 + 20x - 10y + 0z Incorrect "Old" Choice: 0	*90¢*	*-30¢*

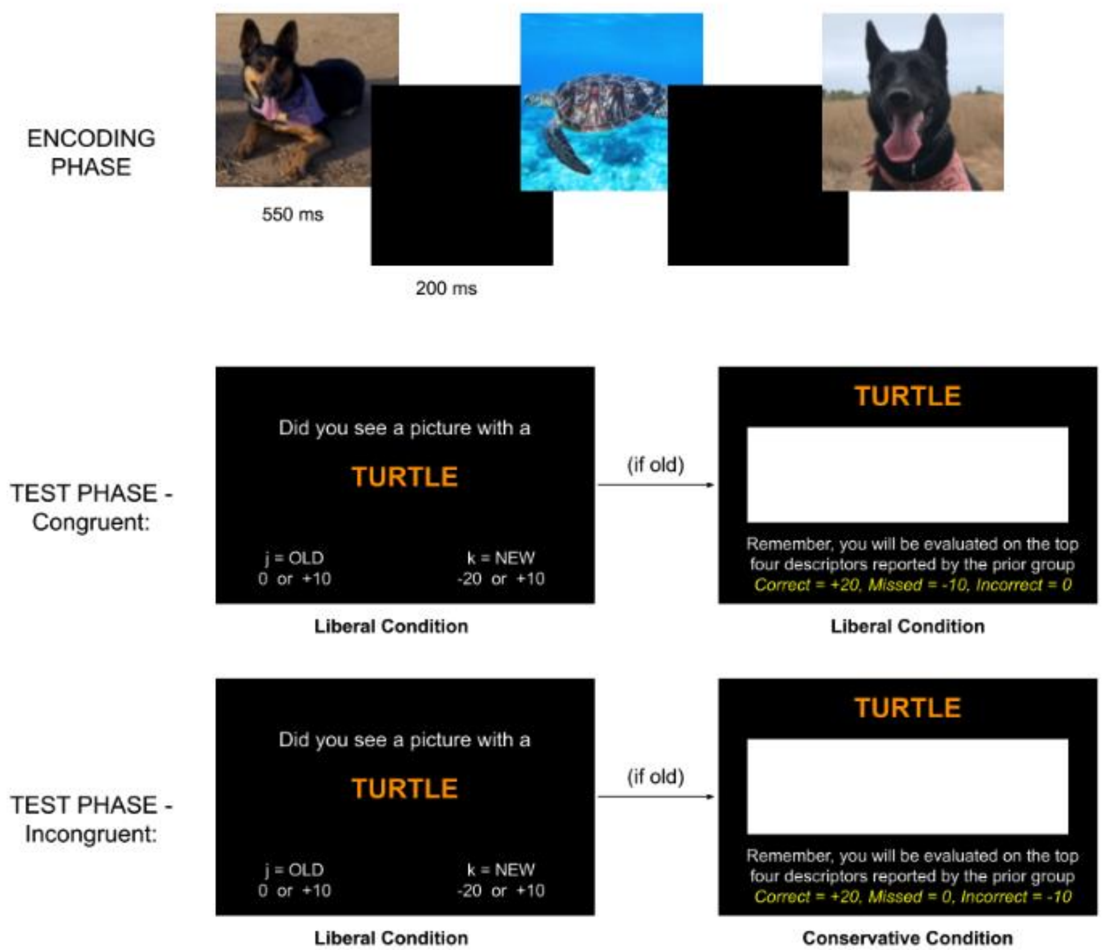


Figure 15: Future free recall task building on cued recall of Experiment 4.

The next step would be to look at the details of each free recall report and evaluate whether someone provided accurate details of the image based on the top four descriptors from the precursor study. Because the description information will be set for each stimulus prior to the free recall experiment, the reported information in the free recall can be automatically evaluated. Unfortunately, with spelling errors, reported details may not be given full credit, so two independent research assistant coders would later read each free recall report to confirm the task correctly evaluated all information. The research assistants will flag any discrepancies and assign a "1" if the participant mentions the detail or a "0" if they do not. This design allows participants to have hits, false alarms, and misses. A correct rejection is usually difficult to evaluate in a free recall context, but in this study design, participants will have the opportunity to complete correct rejections in the word cue phase of the task. Additionally, researchers could approach any false alarm detail in the free recall (which can be generated after the data collection from the final participant reports) as a correct rejection for those who did not report that detail.

Future studies could also evaluate the neural substrates of criterion shifting in cued recall tasks using fMRI. The DPSD and UVSD models suggest different theoretical perspectives on the neural mechanisms underlying criterion shifting in memory-based decisions. Proponents of the DPSD model believe that different brain networks subserve recollection and familiarity. More specifically, the lateral parietal region is related to recollection, while activity in the superior parietal region is related to familiarity (Yonelinas, Otten, Shaw, & Rugg, 2005). However, Wixted (2007b) suggests that recollection may not be an all-or-nothing phenomenon. Instead, varying levels of confidence for a memory could reflect different degrees of recollection, even in the absence of the full autonoetic experience (the feeling of re-living an event). This implies that recollection might occur implicitly, without the person fully realizing they are recalling specific details of a memory, thus blurring the lines between recollection and familiarity.

The remember-know memory literature suggests that remember responses are associated with a sense of autonoetic awareness, which may be a more threshold-like process than recollection. As Wixted (2007a) suggests, neural activity related to autonoetic awareness, rather than recollection itself, may be what DPSD model supporters have identified as correlated with activity in the posterior cingulate and other brain regions. The increases in brain activation associated with remember responses likely reflect the retrieval of strong memories rather than the exclusive presence of recollection. Therefore, the continuous memory strength perspective proposed by the UVSD model could explain the higher degrees of activation in the medial temporal lobe when completing remember-know judgments. Neuroimaging studies have also demonstrated that the prefrontal cortex and parietal cortex are involved in criterion shifting for memory-based decisions. For instance, the dorsolateral prefrontal cortex (DLPFC) was found to be positively correlated with the degree of criterion shifting in visual discrimination tasks, while increased activity in the superior parietal lobule, an area associated with working memory, was observed in recognition memory tasks (Kim & Cabeza, 2007). Carefully designed studies to explore the DPSD and UVSD models will ultimately shed light on the neural substrates involved in strategic memory-based decisions. This research is needed to fully elucidate the complex neural processes involved in familiarity and recollection for criterion shifting during memory-based decisions.

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. M., Freeman, S., Clewett, D., Tipper, C., Frithsen, A., Johnson, A., Grafton, S. G., & Miller, M. B. (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

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97–123. https://doi.org/10.1037/h0033773

Azmat, G., & Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. Journal of *Public Economics*, *94*(7-8), 435–452. https://doi.org/10.1016/j.jpubeco.2010.04.001

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, *142*(4), 1323–1334. https://doi.org/10.1037/a0033872

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, *137*(4), 594–615. https://doi.org/10.1037/a0023489

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

Barankay, I. (2011). Rankings and social tournaments: evidence from a crowd-sourcing experiment (Working Paper). Wharton School of Business, University of Pennsylvania. https://repository.upenn.edu/handle/20.500.14332/40407

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Twenge, J. M. (2005). Social exclusion impairs self-regulation. *Journal of Personality and Social Psychology, 88*(4), 589–604. https://doi.org/10.1037/0022-3514.88.4.589

Baumeister, R. F., Twenge, J. M., & Nuss, C. K. (2002). Effects of social exclusion on cognitive processes: anticipated aloneness reduces intelligent thought. *Journal of Personality and Social Psychology*, *83*(4), 817–827. https://doi.org/10.1037/0022-3514.83.4.817

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

Bolle, F. (2021). Deterrence by collective punishment may work against criminals but never against Freedom Fighters. *Games*, *12*(2), 41. https://doi.org/10.3390/g12020041

Bonner, M., & Ellender, G. (2022). Military training: Does it predispose service personnel to negative mental health issues?. *Journal of Mental Health Disorders*, *2*(1), 11–18. https://doi.org/10.33696/mentalhealth.2.012

Bowker, M. H., Levine, D. P. (2016). Beyond the battlefield: Moral injury and moral defense. *Organisational and Social Dynamics, 16*(1), 85–109.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brown, D. (2020). "We Are All in This Together:" Collective bonuses and incentives in the United Kingdom and Europe: The real performance-related pay?. *Compensation & Benefits Review*, *52*(4), 175–192. https://doi.org/10.1177/0886368720947328

Caballo, V. E., Arias, B., Salazar, I. C., Irurtia, M. J., Hofmann, S. G., & CISO-A Research Team (2015). Psychometric properties of an innovative self-report measure: The Social Anxiety Questionnaire for adults. *Psychological Assessment, 27*(3), 997–1012. https://doi.org/10.1037/a0038828

Caballo, V. E., Salazar, I. C., Irurtia, M. J., Arias, B., & Hofmann, S. G. (2012). The multidimensional nature and multicultural validity of a new measure of social anxiety: The Social Anxiety Questionnaire for adults. *Behavior Therapy*, *43*(2), 313–328. https://doi.org/10.1016/j.beth.2011.07.001

Cacioppo, J. T., & Hawkley, L. C. (2009). Perceived social isolation and cognition. *Trends in Cognitive Sciences*, *13*(10), 447–454. https://doi.org/10.1016/j.tics.2009.06.005

Campbell-Meiklejohn, K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology, 20*(13), 1165–1170. https://doi.org/10.1016/j.cub.2010.04.055

Chapkovski, P. (2021). Strike one hundred to educate one: Measuring the efficacy of collective sanctions experimentally. *PLoS ONE, 16*(4), e0248599. https://doi.org/10.1371/journal.pone.0248599

Chen, X., Liu, J., Luo, Y. J., & Feng, C. (2023). Brain systems underlying fundamental motivations of human social conformity. *Neuroscience Bulletin*, *39*(2), 328–342. https://doi.org/10.1007/s12264-022-00960-4

Cordero, R., Walsh, S. T., Kirchhoff, B. A. (2005). Motivating performance in innovative manufacturing plants. *The Journal of High Technology Management Research*, *16*(1), 89–99. https://doi.org/10.1016/j.hitech.2005.06.005

DeWall, C. N., & Bushman, B. J. (2011). Social acceptance and rejection. C*urrent Directions in Psychological Science*, *20*(4), 256–260. https://doi.org/10.1177/0963721411417545

DeWall, C. N., Twenge, J. M., Gitter, S. A., & Baumeister, R. F. (2009). It's the thought that counts: The role of hostile cognition in shaping aggressive responses to social exclusion. *Journal of Personality and Social Psychology, 96*(1), 45–59. https://doi.org/10.1037/a0013196

Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior, 19*(6), 695–704. https://doi.org/10.1016/S0022-5371(80)90379-5

Dywan, J. (1988). The imagery factor in hypnotic hypermnesia. *International Journal of Clinical and Experimental Hypnosis*, *36*(4), 312–326. https://doi.org/10.1080/00207148808410521

Dywan, J. (1995). The illusion of familiarity: An alternative to the report-criterion account of hypnotic recall. *International Journal of Clinical and Experimental Hypnosis, 43*(2), 194–211. https://doi.org/10.1080/00207149508409961

Dywan, J., & Bowers, K. (1983). The use of hypnosis to enhance recall. *Science, 222*(4620), 184–185. https://doi.org/10.1126/science.6623071

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. (H. A. Ruger & C. E. Bussenius, Trans.). Teachers College Press. https://doi.org/10.1037/10011-000

Egan, J. P. (1958). *Recognition memory and the operating characteristic*. USAF Operational Applications Laboratory Technical Note, 58–51, ii, 32.

Eriksson, T., Poulsen, A., & Villeval, M. C. (2009). Feedback and incentives: experimental evidence. *Labour Economics*, *16*(6), 679–688. https://doi.org/10.1016/j.labeco.2009.08.006

Fabricant, M. C. (2011). *War crimes and misdemeanors: Understanding "zero-tolerance" policing as a form of collective punishment and human rights violation*. Drexel Law Review, 3, 373–415.

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics, 24*(1), 39–58. https://doi.org/10.1086/497818

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences, 11*(10), 419–427. https://doi.org/10.1016/j.tics.2007.09.002

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Charles C. Thomas, Publisher.

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

Gao, L., Wang, Z., Pansini, R., Li, Y. T., & Wang, R. W. (2015). Collective punishment is more effective than collective reward for promoting cooperation. *Scientific Reports, 5*(1). https://doi.org/10.1038/srep17752

Gardiner J. M. (2001). Episodic memory and autonoetic consciousness: a first-person approach. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, *356*(1413), 1351–1361. https://doi.org/10.1098/rstb.2001.0955

Geiselman, E., Fisher, R., Cohen, G., Holland, H., & Surtes, L. (1986). Eyewitness responses to leading and misleading questions under the cognitive interview. *Journal of Police Science and Administration*, *14*(1), 31–39.

Geiselman, E., Fisher, R., MacKinnon, D., & Holland, H. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, *70*(2), 401–12. http://dx.doi.org/10.1037/0021-9010.70.2.401

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

Glanzer, M., Hilford, A. & Maloney, L. T. (2009). Likelihood ratio decisions in memory: Three implied regularities. *Psychonomic Bulletin & Review*, *16*(3), 431–455. https://doi.org/10.3758/PBR.16.3.431

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

Guynn, M. J., & McDaniel, M. A. (1999). Generate–Sometimes recognize, sometimes not. *Journal of Memory and Language*, *41*(3), 398–415. https://doi.org/10.1006/jmla.1999.2652

Hallgren K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

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

Hannan, R. L., Krishnan, R., & Newman, A. (2008). The effects of disseminating relative performance feedback in tournament and individual performance compensation systems. *Accounts Review*, *83*(4), 893–913. https://doi.org/10.2308/accr.2008.83.4.893

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

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLoS ONE, 14*(10), e0223792. https://doi.org/10.1371/journal.pone.0223792

Heckathorn, D. D. (1990). Collective sanctions and compliance norms: A formal theory of group-mediated social control. *American Sociological Review*, *55*(3), 366-384. https://doi.org/10.2307/2095762

Herten, N., Otto, T., & Wolf, O. T. (2017). The role of eye fixation in memory enhancement under stress–An eye tracking study. *Neurobiology of Learning and Memory, 140*, 134–144. https://doi.org/10.1016/j.nlm.2017.02.016

Hilford, A., Glanzer, M., Kim, K., & Maloney, L. T. (2019). One mirror effect: The regularities of recognition memory. *Memory & Cognition*, *47*, 266–278. https://doi.org/10.3758/s13421-018-0864-y

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

Hollingworth, H. L. (1913). Characteristic differences between recall and recognition. *The American Journal of Psychology*, *24*(4), 532–544. https://doi.org/10.2307/1413450

Innocence Project. (2024). *DNA Exonerations in the United States*. www.innocenceproject.org/dna-exonerations-in-the-united-states/

Jacoby, L. L., & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory and Language*, *29*(4), 433–454. https://doi.org/10.1016/0749-596X(90)90065-8

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big Five Inventory (BFI) [Database record]. *APA PsycTests*. https://doi.org/10.1037/t07550-000

Johnson, J. D., McDuff, S. G., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. *Neuron*, *63*(5), 697–708. https://doi.org/10.1016/j.neuron.2009.08.011

Juvonen, J., Gross, E. F. , Williams, K. D., Forgas, J. P., & von Hippel, W. (2005). The rejected and the bullied: Lessons about social misfits from developmental psychology

In Williams, K. D., Forgas, J. P., & von Hippel, W. (Eds.). *The social outcast: Ostracism, social exclusion, rejection, and bullying* (pp. 155–170). New York: Psychology Press.

Kankaraš, M., & van Houten, G. (2015). Third European company survey - overview report: Workplace practices - patterns, performance and well-being. *Publications Office*. http://dx.doi.org/10.2806/417263

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), 1–18. https://doi.org/10.3389/fpsyg.2015.01320

Kensinger, E. A. (2007). Negative emotion enhances memory accuracy: Behavioral and neuroimaging evidence. *Current Directions in Psychological Science*, *16*(4), 213–218. https://doi.org/10.1111/j.1467-8721.2007.00506.x

Kim, H., & Cabeza, R. (2009). Common and specific brain regions in high- versus low-confidence recognition memory. *Brain Research*, *1282*, 103–113. https://doi.org/10.1016/j.brainres.2009.05.080

Kintsch, W. (1970). *Learning, memory, and conceptual processes*. John Wiley & Sons.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*(1-2), 76–81. https://doi.org/10.1159/000119004

Klatzky, R. L., & Erdelyi, M. H. (1985). The response criterion problem in tests of hypnosis and memory. T*he International Journal of Clinical and Experimental Hypnosis, 33*(3), 246–257. https://doi.org/10.1080/00207148508406653

Klein, S. B. (2015). What memory is. *WIREs Cognitive Science, 6*(1), 1–38. https://doi.org/10.1002/wcs.1333

Kreitner, R., & Kinicki, A. (2007). *Organizational behavior* (7th edition). McGraw-Hill Inc.

Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication Methods and Measures, 5*(2), 93–112. https://doi.org/10.1080/19312458.2011.568376

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

Kudielka, B. M., Hellhammer, D. H., & Kirschbaum, C. (2007). Ten Years of Research with the Trier Social Stress Test--Revisited. In E. Harmon-Jones & P. Winkielman (Eds.), *Social neuroscience: Integrating biological and psychological explanations of social behavior* (pp. 56–83). The Guilford Press.

Kuhnen, C., & Tymula, A. (2012). Feedback,self-esteem and performance in organizations. *Management Science, 58*(1), 94–113. https://dx.doi.org/10.2139/ssrn.1333795

Lacy, J. W., & Stark, C. E. (2013). The neuroscience of memory: Implications for the courtroom. *Nature Reviews Neuroscience, 14*(9), 649–658. https://doi.org/10.1038%2Fnrn3563

Langan, J. M. (2018). PTSD in military service members. *DePaul Journal of Health Care Law, 19*(1).

Layher et al., 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), 1–14. https://doi.org/10.3389/fnins.2018.00705

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., Santander, T., Chakravarthula, P., Marinsek, N., Turner, B. O., Eckstein, M. P., & Miller, M. B. (2023). Widespread frontoparietal fMRI activity is greatly affected by changes in criterion placement, not discriminability, during recognition memory and visual detection tests. *NeuroImage, 279*, 120307. https://doi.org/10.1016/j.neuroimage.2023.120307

Leary, M. R. (2010). Affiliation, acceptance, and belonging: The pursuit of interpersonal connection. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 864–897). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002024

Little, S. G., Akin-Little, A., & O'Neill, K. (2015). Group Contingency Interventions With Children—1980-2010: A Meta-Analysis. *Behavior Modification, 39*(2), 322–341. https://doi.org/10.1177/0145445514554393

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366. https://doi.org/10.1101/lm.94705

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(1), 19–31. https://doi.org/10.1037/0278-7393.4.1.19

Luthans, F. (2007). *Organizational behavior: An evidence-based approach* (12th edition). McGraw-Hill Inc.

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., & Creelman, C. D. (2005). *Detection theory, a user's guide* (2nd ed.). Lawrence Erlbaum Associates, Inc.

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

Madu, C. C., & Anyalebechi, L. I. (2016). The effect of financial and non-financial rewards on performance of librarians in academic libraries in Imo State, Nigeria. *NG-Journal of Social Development*, *5*(5), 176–184.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*(3), 252–271. https://doi.org/10.1037/0033-295x.87.3.252

Marsh, H. W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education, 28*(2), 165–181. https://doi.org/10.1177/000494418402800207

Marsh, H. W. (2006). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. Vernon-Wall Lecture: British Psychological Society.

Marsh, H. W., & Seaton, M. (2015). Chapter Five - The Big-Fish-Little-Pond Effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation. *Advances in Motivation Science*, *2*, 127–184. https://doi.org/10.1016/bs.adms.2015.05.002

Marsh, H. W., Martin, A. J., Yeung, A., &. Craven, R. (2017) Competence self- perceptions. In Elliot, A. J., Dweck, C. S., & Yeager, D. S. (Eds.), *Handbook of competence and motivation: Theory and application* (2nd ed., pp. 85–115). The Guilford Press.

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, *99*(1), 112–145. https://doi.org/10.1257/aer.99.1.112

McCloskey, M., & Zaragoza, M. (1985). Postevent information and memory: Reply to Loftus, Schooler, and Wagenaar. *Journal of Experimental Psychology: General*, *114*(3), 381–387. https://doi.org/10.1037/0096-3445.114.3.381

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 563–582. https://doi.org/10.1037/0278-7393.25.3.563

Memon, A., Meissner, C. A., & Fraser, J. A. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, *16*(4), 340–372. https://doi.org/10.1037/a0020518

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

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science, 20*(4), 509–515. https://doi.org/10.1111/j.1467-9280.2009.02324.x

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

Milinski, M., & Rockenbach, B. (2012). On the interaction of the stick and the carrot in social dilemmas. *Journal of Theoretical Biology, 299*, 139–143. https://doi.org/10.1016/j.jtbi.2011.03.014

Miller, M. B., & Kantner, J. (2019). Not all people are cut out for strategic criterion shifting. *Current Direction in Psychological Science, 29*(1), 1–7. https://doi.org/10.1177/0963721419872747

Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review, 106*(2), 398–405. https://doi.org/10.1037/0033-295X.106.2.398

Miller, M. B., Guerin, S. A., & Wolford, G. L. (2011). The strategic nature of false recognition in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1228–1235. https://doi.org/10.1037/a0024539

Milne, B., & Bull, R. (2003). Interviewing by the police. In Carson, D., & Bull, R. (Eds.), *Handbook of psychology in legal contexts* (2nd ed., pp. 109–125). Wiley. https://doi.org/10.1002/0470013397.ch4

Molho, C., Tybur, J. M., Van Lange, P. A. M., Balliet, D. (2020). Direct and indirect punishment of norm violations in daily life. *Nature Communications*, *11*, 3432. https://doi.org/10.1038/s41467-020-17286-2

Neisser, U. (1981). John Dean's memory: A case study. *Cognition, 9*(1),1–22. https://doi.org/10.1016/0010-0277(81)90011-1

Neisser, U. (1982). *Memory observed: Remembering in natural contexts*. W. H. Freeman and Company.

Nguyen, D. T., Kumar, A., & Lau, H. C. (2018). Credit assignment for collective multiagent RL with global rewards. *Advances in Neural Information Processing Systems*.

Nook, E.C., & Zaki, J. (2015) Social norms shift behavioral and neural responses to foods. *Journal of Cognitive Neuroscience*, *27*(7), 1412–1426. https://doi.org/10.1162/jocn_a_00795

Olubusayo, H., Stephen, I. A., & Maxwell, O. (2014). Incentives Packages and Employees' Attitudes to Work: A Study Of Selected Government Parastatals In Ogun State, South-West, Nigeria. *International Journal of Research in Business and Social Science (2147- 4478), 3*(1), 63–74. https://doi.org/10.20525/ijrbs.v3i1.87

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

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., &bLindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Philippe Rushton, J., Chrisjohn, R. D., & Cynthia Fekken, G. (1981). The altruistic personality and the self-report altruism scale. *Personality and Individual Differences, 2*(4), 293–302. https://doi.org/10.1016/0191-8869(81)90084-2

Prolific. (2024). London, UK. https://www.prolific.com

Qualtrics. (2020). Provo, Utah. https://www.qualtrics.com

Reysen, M.B. (2007). The effects of social pressure on false memories. *Memory & Cognition, 35*, 59–65. https://doi.org/10.3758/BF03195942

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

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62*, 23–48. https://doi.org/10.1146/annurev.psych.121208.131647

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roediger, H. L., & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition*, *13*(1), 1–7. https://doi.org/10.3758/BF03198437

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–41). Lawrence Erlbaum Associates, Inc.

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

Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. *Science*, *318*(5850), 598–602. https://doi.org/10.1126/science.1142996

Sharman, S. J., & Powell, M. B. (2013). Do cognitive interview instructions contribute to false beliefs and memories? *Journal of Investigative Psychology and Offender Profiling*, *10*(1), 114–124. https://doi.org/10.1002/jip.1371

Smeets, T., Otgaar, H., Candel, I., & Wolf, O. T. (2008). True or false? Memory is differentially affected by stress-induced cortisol elevations and sympathetic activity at consolidation and retrieval. *Psychoneuroendocrinology*, *33*(10), 1378–1386. https://doi.org/10.1016/j.psyneuen.2008.07.009

Spielberger, C. D. (2010). State-Trait anxiety inventory. In Weiner, I. B., & Craighed, W. E. (eds.) *The Corsini encyclopedia of psychology*. https://doi.org/10.1002/9780470479216.corpsy0943

Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language, 70*, 36–52. https://doi.org/10.1016%2Fj.jml.2013.09.005

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

Stubbs, K. M., & Bentley, J. W. (2024), Show me the money! Supply and demand misalignment for tangible rewards in business. *Accounting Perspectives*. https://doi.org/10.1111/1911-3838.12364

Tambunlertchai, K., & Pongkijvorasin, S. (2020). The impacts of collective threshold requirements for rewards in a CPR experiment. *Environmental Economic and Policy Studies*, *22*, 537–554. https://doi.org/10.1007/s10018-020-00268-5

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

Tran, A., & Zeckhauser, R. (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics, 96*(9-10), 645–650. https://doi.org/10.1016/j.jpubeco.2012.05.004

Tulving, E. (1984). Précis of elements of episodic memory. *Behavioral and Brain Sciences, 7*(2), 223–238. https://doi.org/10.1017/S0140525X0004440X

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

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

Watts, R., Kettner, H., Geerts, D., Gandy, S., Kartner, L., Mertens, L., Timmermann, C., Nour, M. M., Kaelen, M., Nutt, D., Carhart-Harris, R., & Roseman, L. (2022). The Watts Connectedness Scale: A new scale for measuring a sense of connectedness to self, others, and world. *Psychopharmacology, 239*(11), 3461–3483. https://doi.org/10.1007/s00213-022-06187-5

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. J*ournal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Whitehouse, W. G., Dinges, D. F., Orne, E. C., & Orne, M. T. (1988). Hypnotic hypermnesia: Enhanced memory accessibility or report bias? *Journal of Abnormal Psychology, 97*(3), 289–295. http://dx.doi.org/10.1037/0021-843X.97.3.289

Wickelgren, W. A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology, 19*(2), 97–102. https://doi.org/10.1080/14640746708400077

Wiemers, U. S., Sauvage, M. M., & Wolf, O. T. (2014). Odors as effective retrieval cues for stressful episodes. *Neurobiology of Learning and Memory*, *112*, 230-236. https://doi.org/10.1016/j.nlm.2013.10.004

Wixted, J. T. (2007a). Spotlighting the probative findings: Reply to Parks and Yonelinas (2007). *Psychological Review*, *114*(1), 203–209. https://doi.org/10.1037/0033-295X.114.1.203

Wixted, J. T. (2007b). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. https://doi.org/10.1037/0033-295X.114.1.152

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(2), 201–233. https://doi.org/10.1037/xlm0000732

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

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*(4), 1025–1054. https://doi.org/10.1037/a0020874

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review*, *107*(2), 368–376. https://doi.org/10.1037/0033-295x.107.2.368

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*(4), 616–641. https://doi.org/10.3758/bf03196616

Wolf, O. T. (2019). Memories of and influenced by the Trier Social Stress Test. *Psychoneuroendocrinology*, *105*, 98–104. https://doi.org/10.1016/j.psyneuen.2018.10.031

Wright, D. B., Gabbert, F., Memon, A., & London, K. (2008). Changing the criterion for memory conformity in free recall and recognition. *Memory*, *16*(2), 137–148. https://doi.org/10.1080/09658210701836174

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1341–1354. https://doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1413), 1363–1374. https://doi.org/10.1098%2Frstb.2001.0939

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

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

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience, 25*, 3002–3008. https://doi.org/10.1523/jneurosci.5295-04.2005

Appendix

Appendix A: Questionnaires

Social Anxiety Questionnaire for Adults (SAQ-A30):

Below are a series of social situations that may or may not cause you UNEASE, STRESS, or NERVOUSNESS. Please place an "X" on the number next to each social situation that best reflects your reaction, where "1" represents no unease, stress, or nervousness and "5" represents very high or extreme unease, stress, or nervousness.

If you have never experienced the situation described, please imagine what your level of UNEASE, STRESS, or NERVOUSNESS might be if you were in that situation, and rate how you imagine you would feel by placing an "X" on the corresponding number.

Level of Unease, Stress, or Nervousness

Not at all or very slight	Slight	Moderate	High	Very high or extremely high
1	2	3	4	5

Please rate all the items and do so honestly; do not worry about your answer because there are no right or wrong ones.

1. Greeting someone and being ignored 1 2 3 4 5

2. Having to ask a neighbor to stop making noise 1 2 3 4 5

3. Speaking in public 1 2 3 4 5

4. Asking someone attractive of the opposite sex for a date 1 2 3 4 5

5. Complaining to the waiter about my food 1 2 3 4 5

6. Feeling watched by people of the opposite sex 1 2 3 4 5

7. Participating in a meeting with people in authority 1 2 3 4 5

8. Talking to someone who isn't paying attention to what I am saying 1 2 3 4 5

9. Refusing when asked to do something I don't like doing 1 2 3 4 5

10. Being mugged or robbed by an armed gang 1 2 3 4 5

11. Making new friends 1 2 3 4 5

12. Telling someone that they have hurt my feelings 1 2 3 4 5

13. Having to speak in class, at work, or in a meeting 1 2 3 4 5

14. Maintaining a conversation with someone I've just met 1 2 3 4 5

15. Expressing my annoyance to someone that is picking on me 1 2 3 4 5

16. Greeting each person at a social meeting when I don't know most of them 1 2 3 4 5

17. Being teased in public 1 2 3 4 5

18. Talking to people I don't know at a party or a meeting 1 2 3 4 5

19. Being asked a question in class by the teacher or by a superior in a meeting 1 2 3 4 5

20. Looking into the eyes of someone I have just met while we are talking 1 2 3 4 5

21. Being asked out by a person I am attracted to 1 2 3 4 5

22. Making a mistake in front of other people 1 2 3 4 5

23. Attending a social event where I know only one person 1 2 3 4 5

24. Starting a conversation with someone of the opposite sex that I like 1 2 3 4 5

25. Being reprimanded about something I have done wrong 1 2 3 4 5

26. While having dinner with colleagues, classmates or workmates, being asked to speak on behalf of the entire group 1 2 3 4 5

27. One of my parents getting seriously ill 1 2 3 4 5

28. Telling someone that their behavior bothers me and asking them to stop 1 2 3 4 5

29. Asking someone I find attractive to dance 1 2 3 4 5

30. Being criticized 1 2 3 4 5

31. Talking to a superior or a person in authority 1 2 3 4 5

32. Telling someone I am attracted to that I would like to get to know them better 1 2 3 4 5

Note. Items 10 and 27 are control items and do not count at all for the dimensions score or total score of the questionnaire. A score of 1 or 2 on both items is suspicious that the questionnaire could have been answered at random. From Caballo, Salazar, Arias, et al. (2010):

Scoring instructions for the SAQ and its dimensions:

Dimension 1: Interactions with strangers (sum of the items 10, 13, 15, 17, 19, & 22)

Dimension 2: Speaking in public/Talking with people in authority (sum of the items 3, 7, 12, 18, 25, & 29)

Dimension 3: Interactions with the opposite sex (sum of the items 4, 6, 20, 23, 27, & 30)

Dimension 4: Criticism and embarrassment (sum of the items 1, 8, 16, 21, 24, & 28)

Dimension 5: Assertive expression of annoyance, disgust or displeasure (sum of the items 2, 5, 9, 11, 14, & 26)

Total score: Sum of all items of the questionnaire

Cut scores for every dimension and the whole questionnaire are included in table 10.

The Self-Report Altruism Scale*:

Instructions: Check the category on the right that conforms to the frequency with which you have carried out the following acts.

	Never	Once	More than once	Often	Very often
I have given directions to a stranger.					
I have made change for a stranger.					
I have given money to a charity.					
I have given money to a stranger who needed it (or asked me for it).					
I have donated goods or clothes to a charity.					
I have done volunteer work for a charity.					
I have donated blood.					
I have helped carry a stranger's belongings (books, parcels, etc.).					
I have delayed an elevator and held the door open for a stranger.					
I have allowed someone to go ahead of me in line (in a store, to order food).					
I have pointed out a clerk's error (in a bank, at the supermarket) in undercharging me for an item.					
I have let a neighbour whom I didn't know too well borrow an item of some value to me (e.g., a dish, tools, etc.)					
I have bought "charity" Christmas cards deliberately because I knew it was a good cause.					
I have helped a classmate who I did not know that well with a homework assignment when my knowledge was greater than theirs.					
I have before being asked, voluntarily looked after a neighbour's pet or child without being paid for it.					

I have offered to help a handicapped or elderly stranger across a street.					
I have offered my seat on a bus or train to a stranger who was standing.					
I have helped an acquaintance to move households.					

*Note: The original questionnaire created by Ruston et al. (1981) included 20 total categories. Two of those categories, “I have helped push a stranger’s car out of the snow,” and “I have given a stranger a lift in my car,” were removed from the questionnaire for this study because they lacked relevance to the participant population. Language from other categories was also updated to be more inclusive of gender roles.

Watts Connectedness Scale:

Instructions

Reflecting on how you have felt over the past 2 weeks, please rate the following items on a scale from “Not at all” to “Entirely” according to how you have felt over this time period. Please answer every item, even if you are unsure or feel the item is unclear or poorly worded. Drag the indicator to a position on the scale that shows how much you agree or disagree with each of the following statements.

Response format

Each item is rated on a 0 – 100 visual analogue scale with the anchors 0 = Not at all, 100 = Entirely

Final items

1. I have felt trapped in my mind.
2. My mind has felt connected to my heart/emotion.
3. I have felt connected to my senses (touch, taste, sight smell, hearing).
4. I have felt connected to a range of emotions.
5. If I had chosen to, I could have “sat with” painful memories.
6. I have felt connected to my body.
7. I have been able to fully experience emotion, whether positive or negative.
8. I have felt alone.
9. I have felt connected to friends and/or family.
10. I have felt connected to a community.
11. I have felt connected to all humanity.
12. I have felt unwelcome amongst others.
13. I have felt separate from the world around me.
14. I have felt connected to a purpose in life.
15. I have felt connected to nature.
16. I have felt connected to a spiritual essence (in the secular or religious sense).
17. I have felt connected to a source of universal love.
18. I have seen things from a broad perspective, “the bigger picture.”
19. I have felt that everything is interconnected.

Scoring

Connectedness to Self (CTS): (WCS2 + WCS3 + WCS4 + WCS5 + WCS6 + WCS7) / 6

Connectedness to Others (CTO): ((100 – WCS1) + (100 – WCS8) + WCS9 + WCS10 + (100 – WCS12) + (100 – WCS13)) / 6

Connectedness to World (CTW): (WCS11 + WCS14 + WCS15 + WCS16 + WCS17 + WCS18 + WCS19) / 7

General Connectedness (WCS): (CTS + CTO + CTW) / 3

Positive and Negative Affect Schedule (PANAS):

Instructions: This scale consists of a number of words that describe different feelings and emotions. Please read each item and then click the circle from the scale below that best corresponds to how you feel. Indicate to what extent you feel this way right now, that is, at the present moment.

	Very Slightly or Not At All	A Little	Moderately	Quite a Bit	Extremely
1. Interested	1	2	3	4	5
2. Distressed	1	2	3	4	5
3. Excited	1	2	3	4	5
4. Upset	1	2	3	4	5
5. Strong	1	2	3	4	5
6. Guilty	1	2	3	4	5
7. Scared	1	2	3	4	5
8. Hostile	1	2	3	4	5
9. Enthusiastic	1	2	3	4	5
10. Proud	1	2	3	4	5
11. Irritable	1	2	3	4	5
12. Alert	1	2	3	4	5
13. Ashamed	1	2	3	4	5
14. Inspired	1	2	3	4	5
15. Nervous	1	2	3	4	5
16. Determined	1	2	3	4	5
17. Attentive	1	2	3	4	5
18. Jittery	1	2	3	4	5
19. Active	1	2	3	4	5
20. Afraid	1	2	3	4	5

Scoring

Positive Affect Score: Add the scores on items 1, 3, 5, 9, 10, 12, 14, 16, 17, and 19. Scores can range from 10 – 50, with higher scores representing higher levels of positive affect. Mean Scores: 33.3 (SD±7.2)

Negative Affect Score: Add the scores on items 2, 4, 6, 7, 8, 11, 13, 15, 18, and 20. Scores can range from 10 – 50, with lower scores representing lower levels of negative affect. Mean Score: 17.4 (SD ± 6.2)

Your scores on the PANAS: Positive: ____ Negative: ____

Appendix B: Bonding Script with Examples

Instructions:

As mentioned before, this experiment asks you to have a conversation with your fellow participants for **7 minutes** about things you could potentially have in common. I, the researcher, will facilitate this discussion and keep track of time.

Here is a list of potential similarities you all can discuss:

Low
Same:

- Major
- Year
- Eye color
- Hair color
- Height
- Most physical traits
- Color clothes (i.e. everyone has something black on)

Medium
Same:

- Music taste (genre)
- Taste in clothes
- Shoe size-**unless obscure size like 15 or 16, that would be a high level similarity**
- Dog breed
- Cat breed
- Favorite TV/Movie
- All first gen students
- Plans for future
    - Grad school
    - Work right after college
- Seen the same band in concert (not at the same time)
    - Saw the same band on the same tour
- Grew up playing same instrument
- Sports
    - Growing up or in high school

High
Same:

- Hometown-better if you went to same high school
- A brother that is the same age
- A sister that is the same age
- Parents have the same birthday-year included
- Plans to attend same grad school
  - Same program
  - Potentially same PI
- Went to the same concert (unknowingly of course)
  - Same band, tour, and showtime
- All homecoming/prom queen/king
  - Could have been on homecoming court/prom court
- Attending church/place of worship
  - Same pastor/priest/etc.
  - In same area
- Fluency in computer programing
- Preschool
- Parents met a similar way
  - In high school-High school sweethearts