JUST NOT THAT INTERESTED? DRIVERS OF THE GENDER GAP IN
SYSTEMIZING AND EMPATHIZING INTEREST

by

Audrey Aday

B.A., California State Polytechnic University, Pomona, 2016
M.A., University of British Columbia, 2018

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
in
THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2023

© Audrey Aday, 2023
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Just Not That Interested? Drivers of the Gender Gap in Systemizing and Empathizing Interest

submitted by Audrey Aday in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

in Psychology

Examining Committee:

Toni Schmader, Professor, Psychology, UBC
Supervisor

Steven Heine, Professor, Psychology, UBC
Supervisory Committee Member

Susan Birch, Professor, Psychology, UBC
Supervisory Committee Member

Connor Kerns, Associate Professor, Psychology, UBC
University Examiner

Michael Daniels, Assistant Professor, Sauder School of Business, UBC
University Examiner
Abstract

Gender gaps systemizing and empathizing interest are often cited to explain broader patterns of gender segregation. The present work examines situational affordances as a key mechanism for sustaining gender gaps in interest. Drawing on theories of psychological essentialism, I test how primarily biological (vs. sociocultural) explanations for gender differences in interest inform situational affordances in measurement (Chapter 2) and the workplace (Chapter 3) in ways that perpetuate the interest gap and contribute to gender segregation.

In Chapter 2 (Studies 1-3), I examine popular self-report measures of systemizing (SQ) and empathizing (EQ) that are often assumed to reflect biological differences. Study 1 ($N = 624$) first estimated gender differences on the EQ and SQ in large representative samples. Both lay coders (Study 2, $N = 199$) and psychology journal reviewers (Study 3, $N = 116$) rated SQ and EQ item activities as being more learned (vs. innate) and believed men are given more systemizing, and women given more empathizing (Study 3 only), affordances. Items showing the largest gender differences in Study 1 were those rated as having the most gendered affordances (more than gendered genetic advantages).

In Chapter 3 (Studies 4-6), I test how people’s explanations for gender gaps in interest drive situational affordances in the workplace. In Study 4 ($N = 285$), professionals of all genders who endorsed a primarily biological (vs. social) explanation for gender differences in interest were more likely to provide women with empathizing, and men with systemizing, affordances. I replicated Study 4’s results experimentally with men in Study 5 ($N = 379$). In Study 6 ($N = 300$), women who received gendered (vs. counter-gendered) situational affordances based on the selections of participants in Study 4 shifted their interests toward empathizing and chose more
empathizing work assignments. In contrast, women who received counter-gendered (vs.
gendered) situational affordances shifted their interests toward systemizing and chose
empathizing and systemizing work assignments equally. Together, this body of work highlights
situational affordances, driven by essentialist explanations for the interest gap, as a key
contributor to gender gaps in interest and occupational pursuits.
Lay Summary

Gender gaps in interest in systemizing and empathizing interest are often cited to explain men and women’s different career pursuits. Separate from the belief that gender differences on interest exist, I propose people’s explanations for the origin of gender differences play a central role in driving the interest gap through guiding affordances (that is, the learning opportunities people give to men and women). My dissertation finds that primarily biological (vs. social) explanations for the gender gap in interest tend to produce: (1) measures that are biased by gendered affordances, and (2) gender gaps in affordances provided to men and women in the workplace. In both cases, these different affordances magnify gender differences in interest. Together, this work highlights affordances, driven by people’s primarily biological or sociocultural explanations for the interest gap, as a key contributor to gender gaps in interest and occupational pursuits.
Preface

The research presented in this dissertation is the product of a collaboration between Dr Toni Schmader, Dr. Michelle Ryan, and myself. Under the supervision of Dr. Schmader, I was responsible for the formulation of the research questions, design, and implementation of surveys, statistical analyses of data, and composition of manuscripts. Dr. Ryan provided feedback on study conception and the manuscript for Studies 1-3 (Chapter 2). All projects and associated methods were approved by the University of British Columbia’s Behavioral Research Ethics Board (certificates H19-02067, H20-01712, H22-01880, H22-03011, and H22-03679).

Chapter 2 has been submitted for publication. For this paper, Dr. Schmader, Dr. Ryan, and I formulated the research question and designed the surveys. I collected the data and conducted all statistical analyses under the guidance of Dr. Schmader. The manuscript was written by all three authors.
Table of Contents

Abstract........................................................................................................................................... iii
Lay Summary ....................................................................................................................................... v
Preface ................................................................................................................................................... vi
Table of Contents ................................................................................................................................ vii
List of Tables ......................................................................................................................................... xv
List of Figures ......................................................................................................................................... xix
List of Abbreviations .......................................................................................................................... xxii
Acknowledgements ............................................................................................................................ xxiii
Chapter 1: Introduction ....................................................................................................................... 1
  1.1 Person-Level Explanations for Horizontal Gender Segregation .................................................. 3
      1.1.1 Can Gender Differences in Ability Account for Occupational Segregation? ............ 4
      1.1.2 Can Gender Differences in Interest Account for Occupational Segregation? ........ 6
  1.2 What Explains Gender Differences in interests? ........................................................................... 8
      1.2.1 Biological Perspectives ................................................................................................. 9
      1.2.2 Social Constructionist Perspectives ............................................................................. 10
  1.3 The Role of Psychological Essentialism ....................................................................................... 11
  1.4 Gender Essentialism: A Distinct Structure of Psychological Essentialism ............................... 12
      1.4.1 How Is Psychological Essentialism Different from Stereotyping? ............................ 13
  1.5 How Stereotypes Constrain Interests: Self-Fulfilling Prophecy ............................................... 14
  1.6 Gendered Situational Affordances ............................................................................................... 16
  1.7 Contribution and Overview of Current Research ......................................................................... 17
  1.8 Summary of Empirical Chapters ................................................................................................. 18
Chapter 2: Do Measures of Systemizing and Empathizing Reflect Perceptions of Gender Differences in Learning Affordances? ................................................................. 18

Chapter 3: Are Essentialized Explanations for Gender Differences in Interest Self-Fulfilling? ........................................................................................................ 19

Chapter 2: Do Measures of Systemizing and Empathizing Reflect Perceptions of Gender Differences in Learning Affordances? ................................................................. 21

2.1 Defining Systemizing and Empathizing: E-S Theory .................................................. 22

2.2 Evidence of Gender Differences on Systemizing and Empathizing ......................... 23

2.2.1 SQ and EQ Scales ................................................................................................. 23

2.3 Prior Efforts to Reduce Gender Bias on the SQ and EQ ........................................... 25

2.4 Overview of Current Research .................................................................................. 26

2.5 Study 1: Estimating SQ and EQ Gender Differences ................................................ 27

2.5.1 Method ................................................................................................................ 27

2.5.1.1 Participants ......................................................................................................... 27

2.5.1.1.1 Demographics ............................................................................................ 27

2.5.1.2 Procedure and Measures .................................................................................. 29

2.5.2 Results .................................................................................................................. 29

2.5.2.1 Gender Difference on Systemizing and Empathizing Composites ................. 29

2.5.2.2 Gender Difference on Systemizing and Empathizing Items ............................. 30

2.5.3 Discussion .............................................................................................................. 34

2.6 Study 2 ...................................................................................................................... 34

2.6.1 Method ................................................................................................................ 35

2.6.1.1 Participants ....................................................................................................... 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6.1.1.1 Demographics</td>
<td>35</td>
</tr>
<tr>
<td>2.6.1.2 Procedure and Measures</td>
<td>36</td>
</tr>
<tr>
<td>2.6.2 Results</td>
<td>37</td>
</tr>
<tr>
<td>2.6.2.1 Mean Participant Coder Ratings for SQ- and EQ-Short Activities</td>
<td>38</td>
</tr>
<tr>
<td>2.6.2.1.1 Estimated Gender Difference</td>
<td>38</td>
</tr>
<tr>
<td>2.6.2.1.2 Learned vs. Innate Attributions</td>
<td>38</td>
</tr>
<tr>
<td>2.6.2.1.3 Gendered Learning Affordances</td>
<td>39</td>
</tr>
<tr>
<td>2.6.2.1.4 Genetic Differences</td>
<td>39</td>
</tr>
<tr>
<td>2.6.2.2 Item-Level Analyses Predicting Observed Gender Differences</td>
<td>40</td>
</tr>
<tr>
<td>2.6.2.2.1 Analytic Strategy</td>
<td>41</td>
</tr>
<tr>
<td>2.6.2.2.2 Estimated Gender Difference</td>
<td>41</td>
</tr>
<tr>
<td>2.6.2.2.3 Learned vs. Innate Attributions</td>
<td>41</td>
</tr>
<tr>
<td>2.6.2.2.4 Gendered Learning Affordances</td>
<td>42</td>
</tr>
<tr>
<td>2.6.2.2.5 Genetic Differences</td>
<td>43</td>
</tr>
<tr>
<td>2.6.3 Discussion</td>
<td>43</td>
</tr>
<tr>
<td>2.7 Study 3</td>
<td>45</td>
</tr>
<tr>
<td>2.7.1 Method</td>
<td>45</td>
</tr>
<tr>
<td>2.7.1.1 Participants</td>
<td>45</td>
</tr>
<tr>
<td>2.7.1.1.1 Demographics</td>
<td>46</td>
</tr>
<tr>
<td>2.7.1.2 Procedure and Measures</td>
<td>48</td>
</tr>
<tr>
<td>2.7.1.2.1 Gendered Learning Affordances</td>
<td>48</td>
</tr>
<tr>
<td>2.7.1.2.2 Genetic Advantages</td>
<td>48</td>
</tr>
</tbody>
</table>
2.7.2 Results............................................................................................................................................. 49

2.7.2.1 Mean Expert Ratings for SQ- and EQ-Short Activities............................................................ 49

2.7.2.1.1 Estimated Gender Difference................................................................................................. 49

2.7.2.1.2 Learned vs. Innate Attributions............................................................................................. 50

2.7.2.1.3 Gendered Learning Affordances............................................................................................ 50

2.7.2.1.4 Genetic Advantages................................................................................................................. 50

2.7.2.2 Predicting Observed Gender Differences from Coder Ratings.............................................. 52

2.7.2.2.1 Estimated Gender Difference................................................................................................. 52

2.7.2.2.2 Learned vs. Innate Attributions............................................................................................. 52

2.7.2.2.3 Gendered Learning Affordances............................................................................................ 53

2.7.2.2.4 Genetic Advantages................................................................................................................. 54

2.7.2.3 Testing Competing Predictors: Learning Affordances vs. Genetic Difference
........................................................................................................................................................................ 55

2.7.3 Discussion........................................................................................................................................... 55

2.8 General Discussion ................................................................................................................................. 56

2.8.1 Implications......................................................................................................................................... 58

2.8.2 Limitations and Future Directions ..................................................................................................... 59

2.8.2.1 Difficulty of Determining Etiology............................................................................................. 59

2.8.2.2 Constraints on Generalizability.................................................................................................. 60

2.8.2.3 Alternative Sources of Gender Bias in the Activities .............................................................. 60

2.8.3 Conclusion ............................................................................................................................................ 61

Chapter 3: Are Essentialized Explanations for Gender Differences in Interest Self-Fulfilling?
............................................................................................................................................................................. 62
3.1 Overview of Studies

3.2 Study 4

3.2.1 Method

3.2.1.1 Participants

3.2.1.2 Procedure and Measures

3.2.2 Results

3.2.2.1 Men Endorsed More Essentialist Beliefs Than Women and Nonbinary Participants

3.2.2.1.1 People Generally Provide Gendered Affordances

3.2.2.1.2 People Who Primarily Endorse a Biological (Vs. Social) Explanation Provide More Gendered Affordances

3.2.2.2 The Relationship Between Explanation and Affordances is Mediated by Assignment Strategy

3.2.3 Discussion

3.3 Study 5

3.3.1 Method

3.3.1.1 Participants

3.3.1.2 Procedure and Measures

3.3.2 Results

3.3.2.1 Articles Changed Men’s Endorsement of Biological versus Social Determinism

3.3.2.1.1 In General, Men Provide Gendered Affordances
3.3.2.1.2 Men Primed with a Biological (vs. Social) Explanations Provide More Gendered Affordances ................................................................. 80

3.3.2.2 The Relationship Between Explanation and Systemizing Affordances is Mediated by Assignment Strategy ............................................. 81

3.3.3 Discussion .................................................................................. 82

3.4 Study 6 ......................................................................................... 82

3.4.1 Method ..................................................................................... 83

3.4.1.1 Participants ............................................................................ 83

3.4.1.2 Procedure and Measures ....................................................... 84

3.4.2 Results ..................................................................................... 87

3.4.2.1 Affordances Changed Women’s Interest in Empathizing vs. Systemizing ...... 88

3.4.2.2 Affordances Changed Women’s Empathizing vs. Systemizing Possible Selves .................................................................................. 88

3.4.2.3 Women Provided with Gendered (vs. Counter-Gendered) Affordances Selected More Empathizing vs. Systemizing Teams ................................................................. 89

3.4.2.4 Women Imagined Staying at the Company to the Degree it Matched Their Trait-Level Empathizing vs. Systemizing Interests ......................................................... 92

3.4.3 Discussion ................................................................................ 93

3.5 General Discussion ........................................................................ 94

Chapter 4: Conclusion ....................................................................... 98

4.1 Summary of Findings .................................................................... 99

4.1.1 Chapter 2: Do Measures of Systemizing and Empathizing Reflect Perceptions of Gender Differences in Learning Affordances? ......................................................... 99
4.1.2 Chapter 3: Are Essentialized Explanations for Gender Differences in interest Self-Fulfilling? ................................................................. 100

4.2 Implications .................................................................................. 101

4.2.1.1 Interest and Stereotype Accuracy ........................................ 101

4.2.1.2 Self-Fulfilling Prophecy ....................................................... 102

4.2.1.3 Potential Interventions ........................................................ 103

4.2.2 Equating Situational Affordances Through Policy and Practice .......... 104

4.2.2.1 Equating Situational Affordances Through Education ............. 106

4.3 Limitations and Future Directions .................................................. 108

4.3.1 Constraints on Generalizability ................................................ 108

4.3.2 Asymmetries Between Empathizing and Systemizing .................... 111

4.3.3 Interactions Between Biological and Social Explanations .............. 114

4.4 Conclusion .................................................................................. 116

Bibliography ..................................................................................... 117

Appendices ....................................................................................... 145

Appendix A ....................................................................................... 145

A.1 Study 1: Estimating SQ and EQ Gender Differences ........................ 145

A.2 Study 2 ...................................................................................... 149

A.3 Study 3 ...................................................................................... 160

A.4 Studies 2-3 ................................................................................ 178

Appendix B ....................................................................................... 180

B.1 Stimuli Creation and Pilot Testing ............................................... 180

B.2 Study 4 ...................................................................................... 205
| B.3  | Study 5 .................................................................................................................. 235 |
|------|----------------------------------------------------------------------------------------------------------------------------------|------|
| B.4  | Study 6 .................................................................................................................. 255 |
| B.5  | Studies 4-6 ............................................................................................................. 270 |
List of Tables

Table 2.1 Sample demographics by region (Study 1) ................................................................. 28
Table 2.2. Items, descriptive statistics by gender, and effect size for gender difference (Study 1)
...................................................................................................................................................... 31
Table 2.3. Sample demographics (Study 2) ................................................................................ 35
Table 2.4. Descriptive statistics, effect size for difference from midpoint, and N items
significantly below or above midpoint (Study 2) ........................................................................ 39
Table 2.5. Response rates by subdiscipline and gender (Study 3) ............................................. 46
Table 2.6. Sample demographics (Study 3) ................................................................................ 47
Table 2.7. Descriptive statistics, effect size for difference from midpoint, and N items
significantly below or above midpoint (Study 3) ........................................................................ 51
Table 3.1. Sample demographics (Study 4) ................................................................................ 66
Table 3.2. Descriptive statistics by gender (Study 4) ................................................................ 72
Table 3.3. Sample demographics (Study 5) ................................................................................ 76
Table 3.4. Descriptive statistics by condition (Study 5) ............................................................. 79
Table 3.5. Participant Demographics (Study 6) ....................................................................... 83
Table 3.6. Descriptive statistics by condition (Study 6) ............................................................. 90

Table A.1. Item-level information on coder ratings for each measure and the p-value for t-tests
comparing ratings to scale midpoint (Study 2) ............................................................................. 153
Table A.2. Mean coder ratings among men and women coders separately, as well as the effect
size for the gender difference between men and women coders (Study 2) ............................... 158
Table A.3. Effects predicting the magnitude of the gender difference on each item from mean
coder ratings among men and women coders separately (Study 2) ........................................... 159
Table A.4. Item-level information for coder ratings for each measure and the p-value for t-tests comparing ratings to scale midpoint (Study 3) ................................................................. 166
Table A.5. Mean coder ratings among men and women coders separately, as well as the effect size for the gender difference between men and women coders (Study 3) ...................... 171
Table A.6. Mean coder ratings by coder subdiscipline (Study 3) ........................................ 173
Table A.7. Effects predicting the magnitude of the gender difference on each item from mean coder ratings, among men and women coders separately (Study 3) ......................... 175
Table A.8. Standardized betas for effects for ratings of estimated gender difference by subdiscipline (Study 3) ........................................................................................................................................ 176
Table A.9. Standardized betas for effects for ratings of learned versus innate attributions by subdiscipline (Study 3) ........................................................................................................................................ 176
Table A.10. Standardized betas for effects for ratings of gendered learning affordances by subdiscipline (Study 3) ........................................................................................................................................ 177
Table A.11. Standardized betas for effects for ratings of genetic advantage by subdiscipline (Study 3) ........................................................................................................................................ 177
Table A.12. Single predictor analyses comparing GLM and MLM estimates (Studies 2 and 3) ........................................................................................................................................ 178
Table A.13. Competing predictor analysis comparing GLM and MLM estimates (Study 3) .... 179
Table B.1. Compilation of empathizing and systemizing definitions for team descriptions in Studies 4, 5, and 6 ........................................................................................................................................ 182
Table B.2. Compilation of empathizing and systemizing verbs and subjects for team descriptions in Studies 4, 5, and 6 ........................................................................................................................................ 184
Table B.3. Initial project teams developed for Studies 4, 5, and 6 ........................................ 186
Table B.4. Ratings of teams’ benefit from face-to-face meetings (Pilot Study 1)...................... 189
Table B.5. Ratings of team realism (Pilot Study 1) ................................................................. 190
Table B.6. Participant team assignments by instruction set (Pilot Study 1) ....................... 191
Table B.7. Ratings of teams’ benefit from face-to-face meetings (Pilot Studies 2A and 2B) .... 196
Table B.8. Ratings of team realism (Pilot Studies 2A and 2B) ............................................ 198
Table B.9. Cell sizes by condition (Pilot Study 3)..................................................................... 199
Table B.10. Descriptive Differences by Condition on Key Variables (Pilot Study 3) .......... 200
Table B.11. Overview of team assignment task trials (Study 4; order of trials randomized).... 217
Table B.13. Key effects modeled in GLM and MLM (Study 4) .................................. 218
Table B.14. Key effects for front-end/empathizing and back-end/systemizing teams separately (Study 4).................................................................................................................. 222
Table B.15. Selections by target for all participants (Study 4)............................................. 223
Table B.16. Selections by target for participants highest in biological vs. social determinism (Study 4).................................................................................................................. 224
Table B.17. Selections by target for participants highest in social vs. biological determinism (Study 4).................................................................................................................. 225
Table B.18. Key effects for biological and social determinism measures separately (Study 4) ...................................................................................................................... 226
Table B.19. Factor loadings for biological and social determinism items (Study 4)........ 228
Table B.20. Factor loadings for biological determinism items separately (Study 4) .......... 229
Table B.21. Factor loadings for social determinism items separately (Study 4) ............. 229
Table B.22. Effects for Immutability Items (Study 4)............................................................. 230
Table B.23. Effects for Biological and Social Determinism without Immutability Items (Study 4)
........................................................................................................................................231
Table B.25. Selections by target among all participants (Study 5)........................................249
Table B.26. Selections by target among participants in the biological condition (Study 5)......250
Table B.27. Selections by target among participants in the social condition (Study 5).........251
Table B.28. Process for translating perceiver selections (Study 4) to target work assignments
(Study 6)................................................................................................................................262
Table B.30. Results Comparing Difference Scores to Residuals (Study 6).............................264
Table B.33. Participant suspicion rates (Studies 4-6)................................................................271
List of Figures

Figure 2.1. Distribution of scores on the SQ-Short and EQ-Short (Study 1) .......................... 30
Figure 2.2. Lay coder ratings and comparisons against midpoint (Study 2) ......................... 40
Figure 2.3. Relationship between coder ratings of gendered learning affordances (Study 2) and observed gender difference (Study 1) .......................................................... 43
Figure 2.4. Expert (Study 3) and lay coder (Study 2) ratings and comparisons against midpoint .......................................................... 52
Figure 2.5. Relationship between expert ratings of gendered learning affordances (Study 3) and observed gender difference (Study 1) .......................................................... 54

Figure 3.1. Self-fulfilling process through which perceivers’ essentialized beliefs shape targets’ interests via situational affordances. Research relevant examples are provided in italics......... 64
Figure 3.2. Team sets (Study 4) ...................................................................................................... 69
Figure 3.3. Example of a critical trial in the team assignment task (Study 4) ......................... 70
Figure 3.4. Relationship between explanation for gender difference in interest and probability of choosing a woman intern, moderated by team type (empathizing vs. systemizing; Study 4) ..... 73
Figure 3.5. Indirect effect of task assignment strategy (Study 4) .............................................. 75
Figure 3.6. Probability of choosing a woman intern by condition and team type (empathizing vs. systemizing; Study 5) .................................................................................................. 80
Figure 3.7. Indirect effect of task assignment strategy (Study 5) .............................................. 82
Figure 3.8. Example of a trial selection (Study 6) ........................................................................ 86
Figure 3.9. Visual summary in gendered (left) and counter-gendered (right) conditions (Study 6) ......................................................................................................................... 86
Figure 3.10. Key outcomes by condition (Study 6). Error bars represent +/- 1 standard error .... 91
Figure 3.11. Relationship between condition and change in intentions to stay (post-pre),
moderated by trait empathizing vs. systemizing (Study 6) .................................................. 93

Figure B.1. Target photos shown to participants (Studies 4-5) ............................................. 181

Figure B.2. Both possible sets of project teams (Pilot Study 1) ........................................... 188

Figure B.3. Ratings of team relevance to empathizing and systemizing (Pilot Study 1) .......... 189

Figure B.4. Team assignment task choices (Pilot Study 2A) ................................................. 193

Figure B.5. Ratings of team relevance to empathizing and systemizing (Pilot Studies 2A and 2B)
................................................................................................................................................. 197

Figure B.6. Infographic for filler article (Study 4) ................................................................. 213

Figure B.7. Infographic for critical article (Study 4) ............................................................. 214

Figure B.8. Overview of team assignment (Study 4) ............................................................... 215

Figure B.9. Example of a critical trial in the team assignment task (Study 4) ....................... 207

Figure B.10. Scree plot of factors for all biological and social determinism items (Study 4).... 227

Figure B.11. Infographic accompanying biological article prime (Study 5) ......................... 242

Figure B.12. Photo accompanying biological article prime (Study 5). .................................. 243

Figure B.13. Infographic accompanying social article prime (Study 5) ............................... 245

Figure B.14. Photo accompanying social article prime (Study 5) ......................................... 246

Figure B.15. Project teams shown to participants (Study 5) ............................................... 248

Figure B.16. Mediation by personal endorsement of biological vs. social determinism (Study 5)
..................................................................................................................................................... 254

Figure B.17. Example of a trial selection (Study 6) ............................................................... 256

Figure B.18. Visual summary of personalized work assignments for gendered (left) and counter-
gendered (right) conditions (Study 6) ..................................................................................... 257
Figure B.19. Indirect effect of change in empathizing, systemizing team interest (Study 6)..... 260
Figure B.20. Indirect effect of empathizing, systemizing possible self (Study 6).................. 260
Figure B.21. Mediation by change in interest and possible self (Study 6).......................... 261
Figure B.22. Relationship between condition and team selections, moderated by trait empathizing vs. systemizing (Study 6)........................................................................... 269
Figure B.23. Participant suspicion rates (Studies 4-6)......................................................... 271
List of Abbreviations

E-S = Empathizing/Systemizing

HEED = Healthcare, Early Education, and Domestic roles

M = Mean

SD = Standard deviation

STEM = Science, Technology, Engineering, and Math
Acknowledgements

I would like to thank my advisor, Toni Schmader, for her generous time and wisdom as an advisor on this research, for supporting me throughout my graduate career, and for inspiring me to pursue research that makes a difference. I would also like to thank Michelle Ryan for her valuable input and collaboration on this line of work, especially the studies presented in Chapter 2. I thank my labmates and all affiliates of the Social Identity Lab, past and present, for their support at each stage of this work, as well as the members of my committee, Sue Birch and Steve Heine, for offering their valuable insight into this program of research. Finally, I owe a special thanks to my family and friends. Mom and Dad: thank you for inspiring and nurturing my curiosity of the cosmos and all things within it. Nathan, Megan, Milio, Stephanie, Auntie, Uncle Aaron, Michael, James, Aunt Chris, Uncle Tom, Megan, Andrew: thank you for believing in me and supporting me always. Gordon: thank you for being the most incredible partner I could ask for. This dissertation would not be possible without a cup of your coffee every morning (and, let’s be real, several more cups throughout the day). Finally, I would like to thank my cat, Walter, even though he contributed absolutely nothing to this dissertation. I love him anyway.
Chapter 1: Introduction

“And why, in over 100 years of the existence of the Fields Medal, maths' Nobel Prize, have none of the winners have ever been a woman?”

(Baron-Cohen, 2009)

At the time of writing his 2009 article for the BBC, Simon Baron-Cohen noted the Fields Medal, one of the highest honors a mathematician can receive, had never gone to a woman (although by 2023, two women have been awarded the prize: Maryam Mirzakhani in 2014 and Maryna Viazovska in 2022). Indeed, women remain underrepresented in computer and mathematical occupations, where they make up only 26.2% of the workforce (Bureau of Labor Statistics, 2022). In contrast, women are highly represented in occupations like preschool and kindergarten teaching, where they make up 96.8% of the workforce (Bureau of Labor Statistics, 2022). In fact, on the whole, women are less likely to pursue careers in Science, Technology, Engineering, and Math (STEM; Cheryan, Ziegler, Montoya, & Jiang, 2017), whereas men are less likely to pursue careers in Healthcare, Early Education, and Domestic roles (HEED; Croft, Schmader, & Block, 2015). These patterns reflect broader trends toward horizontal gender segregation, or the unequal representation of men and women in different professions that do not necessarily differ in their status (Block et al., 2023; Schmader & Nater, 2023). Indeed, although a large body of work has focused on addressing vertical gender segregation (i.e., the different positions of power and status held by men and women; see Schmader & Nater, 2023), relatively less work has sought to address causes of horizontal gender segregation. What might account for men’s and women’s different representation in careers like STEM and HEED?
Although prior work has identified features external to the environment that contribute to men and women’s differential selection into STEM and HEED careers, including cultural defaults (Cheryan & Markus, 2020), lack of role models (Croft et al., 2015; Stout et al., 2011), lack of belonging (Good, Rattan, & Dweck, 2012), gender bias (Moss-Racusin et al., 2018), and person-environment fit (Schmader, 2023; Schmader & Nater, 2023), we might also consider features internal to the person that guide people’s self-selection into different careers. As I will review, gender differences in interest are often cited as a key explanation for men and women’s unequal pursuit of STEM and HEED careers (e.g., Ceci, Williams & Barnett, 2009; Ceci & Williams, 2011). My dissertation aims to understand how interests are shaped by opportunities and constraints in the environment. In doing so, I challenge the notion that interest is purely a person-level phenomenon. Integrating theorizing on psychological essentialism and self-fulfilling prophecy, my dissertation examines how people’s essentialized theories of gender differences in interest as primarily caused by biological (vs. social) factors inform situational affordances that sustain (vs. change) gender gaps in interest.

Chapter 2 has the goal to explore this process in the context of measurement. As I review in this section, a key source of evidence for gender differences in interests are self-report measures such as the Systemizing Quotient (SQ; Baron-Cohen et al., 2003) and Empathizing Quotient (EQ; Baron-Cohen & Wheelwright, 2004). The large effect sizes found on the SQ and EQ are often interpreted by scholars as supporting biologically inherent sex differences on empathizing and systemizing (Archer, 2019; Baron-Cohen, 2010). Biologically essentialized interpretations of these measures, in turn, carry potential harms—they not only threaten women’s

---

1 Throughout my dissertation, I use the term *sex differences* to refer to (what are often assumed to be essential) differences based on sex assigned at birth. In contrast, I use the term *gender differences* to refer to differences based on people’s lived experience of gender identity.
sense of fit and ability to succeed in male-dominated fields (Bian et al., 2018; Cheryan & Markus, 2020; Dar-Nimrod & Heine, 2006), but are utilized by the public to discourage movements for gender parity (Damore, 2017). Given these concerns, Chapter 2 tests whether self-report scales designed to measure biological gender differences in interest ask about activities that men and women have different affordances to learn, and whether these different affordances magnify the observed gender gap in interest.

Chapter 3 has the goal to explore this process in the context of individual beliefs and behavior. Drawing on theorizing in self-fulfilling prophecy (Merton, 1948; Jussim, 1986), I test a process through which (1) perceivers believe gender differences in interest are biologically (vs. socially) determined, (2) perceivers provide gendered (vs. counter-gendered) situational affordances to targets, and (3) targets respond to these affordances in ways that confirm (vs. diverge from) gendered interest stereotypes. I examine this process in the context of a male-dominated work setting where gendered features of the environment are likely to be especially salient (Murphy, Steele, & Gross, 2007).

Having specified the empirical focus of my dissertation, I next review person-level explanations for horizontal gender segregation.

1.1 Person-Level Explanations for Horizontal Gender Segregation

When looking at person-level explanations for men and women’s different self-selection into occupations, we might consider whether there are: (1) gender differences in abilities relevant to these careers, and/or (2) gender differences in interest relevant to these careers. Although scholars historically proposed gender differences on abilities might explain men and women’s different rates of selection into STEM and HEED careers, more recent empirical evidence reveals ability alone cannot account for these effects. Instead, the academic discourse has shifted
toward gender differences in interest as a core explanation for men and women’s occupational segregation.

1.1.1 Can Gender Differences in Ability Account for Occupational Segregation?

Historically, scholars considered the possibility that gender differences on these abilities could at least partly explain men and women’s different pursuit of STEM and HEED careers. In 1974, psychologists Eleanor Maccoby and Carol Jacklin published a review citing reliable gender differences in mathematics, spatial, and verbal abilities, with men demonstrating better math and spatial abilities (skills relevant to STEM achievement; Stieff & Uttal, 2015), and women demonstrating better verbal abilities (skills that might be considered relevant to HEED achievement, given its broader focus on people and communal roles; Croft et al., 2015). However, in the four decades since this review, results across a multitude of studies yielded weak evidence for gender differences on these same abilities, with considerable evidence of cross-national variability (Hyde, 2005, 2014).

First considering STEM abilities described in Maccoby and Jacklin’s (1974) review, I consider evidence for gender differences on math and spatial ability. A meta-analysis investigating math performance among 7 million students in the US found no difference on boys’ and girls’ performance ($d$ ranged from -.02 to .06 across grades 2 through 11; Hyde et al., 2008). Although this effect is significant when examined cross-nationally ($d$ range from -.42 to .40; Else-Quest, Hyde, & Linn, 2010), cross-national differences are moderated by sociocultural factors relating to women’s status and welfare, such as girls’ access to formal education. On spatial ability, another skill relevant to STEM, meta-analyses yield moderate effect sizes that are larger than those found in math ($d = .56$ for mental rotation tasks, boys higher; Voyer, Voyer, & Bryden, 1995). Yet this effect size, too, depends on several environmental factors, including
previous training (Uttal, Miller, & Newcombe, 2013), experience playing video games (Feng, Spence, & Pratt, 2007), and time limits placed on the task (Voyer, 2011). In addition to sociocultural factors, the size of the gender difference on these abilities are changed by situational factors that bring to mind gender stereotypes (i.e., Eisenberg & Lennon, 1983). For example, when mental rotation tasks are framed as a test of spatial reasoning, men outperform women by a large margin ($d = 1.38$; Tarampi, Heydari & Hegarty, 2016). However, when the same task is framed as a test of empathic abilities, the size of the gender difference becomes smaller ($d = .42$, driven by women’s improved performance), suggesting that women’s performance is impaired when gender stereotypes are salient. Similarly, women’s performance on mental rotation tasks is equivalent to men’s performance when they believe women generally outperform men on these tasks (Moè, 2009). Literature on stereotype threat demonstrates that math abilities are similarly moderated by features of the situation that bring salient gender differences to mind (Schmader et al., 2008; Schmader & Nater, 2023).

Turning to HEED abilities described in Maccoby and Jacklin’s (1974) review, I consider evidence for gender differences on verbal ability. In addition, I consider evidence for gender differences on empathic accuracy (i.e., the degree to which people successfully infer others’ thoughts and feelings; Ickes, Gesn, & Graham, 2000), since, like verbal ability, empathic accuracy might be considered crucial for success in care-oriented roles focused on other people. Recent studies examining gender differences on verbal ability find a small gender difference favoring girls on reading ability in the US ($d = -.26$; Reilly, 2012). Like the gender difference on math, this effect is larger when examined cross-culturally ($d = -.44$) but is also correlated with gender equality at the country level, suggesting this gender difference is similarly sensitive to the sociocultural context (for a review and discussion of findings, see Hyde, 2014). And, as with the
STEM-related abilities reviewed above, differences on HEED-related abilities are sensitive to contextual factors that bring gender to mind. Turning to empathic accuracy, prior meta-analyses report an overall gender difference \((d = .26; \text{Ickes, Gesn,} \& \text{Graham, 2000})\). However, separating these studies by the instructions preceding the task, the same meta-analysis found that when participants were subtly reminded of gender stereotypes associating women with empathy prior to completing an empathic accuracy test, women outperformed men \((d = .56; \text{Ickes, Gesn,} \& \text{Graham, 2000})\). Yet when participants were not reminded of this difference, there was no gender difference on empathic accuracy \((d = .04)\).

In sum, the extant literature suggests gender differences on abilities that would be relevant to STEM and HEED careers are small and variable. In the few cases where gender differences on ability are documented, situational factors, whether embedded in the broader culture or present via contextual cues, can accentuate or mitigate this difference. We can contrast this with the large gaps in men and women’s relative pursuit of STEM and HEED careers (as highlighted at the start of this chapter, women comprise only 26.2% of the workforce in STEM occupations like computer and mathematical occupations but represent 96.8% of the workforce in HEED occupations like preschool and kindergarten teaching). Thus, gender gaps in ability by themselves do not offer a compelling explanation for broader patterns of horizontal gender segregation. Instead, scholars often argue that men and women’s different career pursuits reflect their different interests. I next review the evidence for this perspective.

### 1.1.2 Can Gender Differences in Interest Account for Occupational Segregation?

One of the most foundational observations in the psychological literature is that women are more interested in *people* and men are more interested in *things*. As early as the 1910s, Thorndike (1911) speculated the greatest gender difference is “the relative strength of the interest
in things and their mechanisms (stronger in men) and the interest in persons and their feelings (stronger in women)” (p. 31). Nearly a century later, in their meta-analysis of career interest inventories, Su and colleagues (2009) found a large gender difference on the people-things dimension, such that men preferred to work with things and women preferred to work with people (effect size for gender difference: \( d = .93, N = 503,188 \))^2, an effect that has been cited as one of the largest in the psychological literature to date (Zell, Krizan, & Teeter, 2015). These large gender differences map onto cultural stereotypes about men and women’s different interests. For instance, as early as age 8, children endorse the stereotype that boys are more interested in computer science and engineering than are girls (even more so than stereotypes about abilities; Master, 2021). How might these gender differences in interest account for patterns of horizontal gender segregation?

Interest plays a key role in informing and guiding people’s chosen careers. Perspectives like the expectancy-value model (Eccles, 1994) place interest as a key predictor of one’s academic and occupational choices. In a longitudinal study of 3,000 sixth graders in the US, Eccles and colleagues (1994, 2007) found that girls were more likely than boys to intend to pursue healthcare careers due to the greater value they placed on people-oriented jobs, regardless of their actual or self-perceived ability to succeed in STEM. These same interests predicted students’ college majors and professions over a decade later (Eccles & Vida, 2003). Similarly, in a longitudinal study of over 600 high school students from the US and Australia, boys’, but not girls’, interest in math predicted their intent to pursue a STEM career (Watt et al., 2017).

Regarding the lack of an effect for girls, the authors speculate that lowered perceptions of one’s ability to succeed (as is the case for girls in STEM; Else-Quest et al., 2010) might supersede the

---

2 The size of this effect, while reliably present, varies according to methodological decisions. I explore this topic further in Chapter 2.
effects of interest. Taken together, this work suggests that one’s interest, whether in people or things, is a powerful predictor of occupational choice. These large gender differences in interests, as reviewed above, may explain why women select into people-relevant HEED careers whereas men select into things-relevant STEM careers.

Some scholars prioritize gender differences in interest as better explanations of occupational segregation than contextual factors such as systemic bias or discrimination (Ceci, Williams & Barnett, 2009). But the focus on essentialized person-level explanations like interest can lead people to assume that environmental barriers do not also shape patterns of segregation, even if still through interest. In contrast to this perspective, I argue opportunities and constraints in the environment (i.e., situational affordances) directly contribute to occupational segregation by widening gender gaps in interest via disparate affordances. Moreover, I propose that situational affordances vary according to people’s lay theories of the root cause of gender differences in interest. To this point, I next review prominent perspectives that seek to explain the origin of gender differences in interest.

1.2 What Explains Gender Differences in Interests?

Given the central role interest plays in major career decisions, coupled with large gender differences on these interests, scholars have long been interested in understanding their origin. That is, how do men develop a greater interest in things, and how do women develop a greater interest in people? Most scholars and laypeople alike would likely agree the developmental trajectory of any psychological phenomenon is a complex interplay between biological and social factors (Eagly & Wood, 2012, 2013; Schmader & Nater, 2023). My dissertation specifically focuses on people’s primary explanation (whether biological or social) for the interest gap, with the aim of understanding how these different explanations guide situational
affordances. With this goal in mind, I next review academic perspectives that broadly prioritize: (1) biological, or (2) social factors as explaining gender differences in interest.

1.2.1 Biological Perspectives

One class of theories maintains that men and women are biologically predisposed toward roles and occupations that involve people versus things. Empathizing-Systemizing Theory (E-S Theory; Baron-Cohen, 2002, 2004) describes people’s orientation toward *people* versus *things* in terms of innate psychological drives toward empathizing, or “the ability to tune into how someone else is feeling, or what they might be thinking [...] understand the intentions of others, predict their behavior, and experience an emotion triggered by their emotion” (Baron-Cohen & Wheelwright, 2004, p. 163), and systemizing, or “the drive to analyze the variables in a system, to derive the underlying rules that govern the behaviour of a system [...] and the drive to construct systems” (Baron-Cohen et al., 2003, p. 361). Importantly, E-S Theory maintains that women have an innate advantage on empathizing, whereas men have an innate advantage on systemizing (Baron-Cohen, 2002, 2004).

How do gender differences in empathizing and systemizing develop? E-S Theory posits that gender differences on empathizing and systemizing are primarily rooted in biological processes (Baron-Cohen, 2002, 2008). The authors argue that exposure to greater testosterone, the hormone that biologically differentiates males from females in the womb, also causes sex differences in brain organization, which results in gender differences on empathizing and systemizing (Baron-Cohen, 2002, 2004, 2008, 2009; Baron-Cohen, Knickmeyer, & Belmonte, 2005). Taking these arguments a step further, some scholars argue that biological sex differences in empathizing and systemizing are further rooted in evolutionary pressures (Archer, 2019; though see Byrd-Craven et al., 2015 and Nettle, 2007). Together, these perspectives argue that
gender differences in interests (and abilities for that matter), particularly men and women’s orientation to *people* versus *things*, are primarily caused by biology.

### 1.2.2 Social Constructionist Perspectives

In contrast to these perspectives that emphasize the primacy of biology, social constructionist perspectives maintain that gender differences in interests are primarily produced via socialization into appropriate gender roles. In their 1990 book, *Gender Trouble*, Judith Butler argues gender differences are the product of the social environment, rather than an inherent truth that follows from biological processes. Butler maintains that gender is performed and constructed through shared social scripts in the environment. For example, a person might perform the gender *man* by adhering to a social script that requires them to seek out systemizing roles and avoid empathizing roles. Through repeated performance of these gender roles, individuals reinforce the notion of *men* and *women* as binary gender categories with unique psychological drives (Butler, 1990; Schmader & Nater, 2023). Similarly, Deaux and Major’s (1987) interactive model of gender-related behavior posits that gendered behavior arises from situational and contextual factors, including perceiver expectations, stereotypes in the environment, and targets’ own self-schemata. For instance, a woman might seek out empathizing roles to the degree that others expect her to pursue empathizing roles, situational cues activate stereotypes associating women with empathizing, and/or she holds gender stereotypical self-views.

How do gendered social scripts develop? By the age of 5, children have in-depth knowledge of gender, including their own gender identity and what traits are associated with which gender identities (Martin & Ruble, 2004; Olson et al., 2015). Martin and Ruble (2004) liken children to *gender detectives*, actively seeking out cues from their social environment that
provide them with information about gender. According to these social constructionist perspectives, children learn to perform their gender through a combination of external factors (e.g., being rewarded by parents for gender-congruent behavior; Mischel, 1966) and internal factors (e.g., learning to associate traits with gender and internalizing those most relevant to the self; Kohlberg, 1966). Additionally, children might learn gender by modeling the behavior of parents (Bandura, 1977, 1986) or peers (Birch, Vauthier, & Bloom, 2008), which entails a combination of internal and external factors. Together, these processes contribute to one’s sense of gender as a social identity that guides one’s expressed interests via stereotypes about masculinity and femininity (Biernat, 1991; Schmader & Block, 2015; Martin et al., 2002).

1.3 The Role of Psychological Essentialism

A key feature that distinguishes biological from social constructionist perspectives, and therefore a central conceptual framework for my dissertation, is the degree to which these perspectives tap into psychological essentialism: the belief that social categories have essences or underlying natures that make them what they are (Dar-Nimrod & Heine, 2011; Keller, 2005; Medin, 1989; Medin & Ortony, 1989). Although essentialism can apply to reasoning about any natural kind category, like tigers or water, recent work has been particularly interested in how these same beliefs are applied to social groups (Newman & Knobe, 2019). Past work finds endorsing essentialist beliefs about social groups is linked to greater endorsement of stereotypes and prejudice (Bastian & Haslam, 2006; Haslam & Levy, 2006), system justifying motives (Brescoll, Uhlmann, & Newman, 2013), acceptance of social inequities (Williams & Eberhardt, 2008), and less interest in interacting with outgroup members (Williams & Eberhardt, 2008). In my dissertation, I examine how psychological essentialism informs people’s explanations for
gender differences in interest. To this point, I next review the literature on gender essentialism as a distinct structure of psychological essentialism.

1.4 Gender Essentialism: A Distinct Structure of Psychological Essentialism

Although people can have essentialist beliefs about any social group, prior research has identified gender as the most essentialized social category (Diekman & Schmader, 2023; Gelman & Taylor, 2000; Prentice & Miller, 2006). Haslam and colleagues (2000) found that, of 40 social categories, gender was rated highest on dimensions of essentialism relating to natural kinds: discreteness (i.e., how clear-cut category membership is), naturalness (i.e., how natural the category is), immutability (i.e., how easy it is to change membership in the category), stability (i.e., how stable the category is over time), and necessity (i.e., whether the category has features that are necessary for membership). Moreover, essentialist beliefs about gender are present as early as age 4 (Taylor, Rhodes, & Gelman, 2009). Given this evidence, it is not surprising that influential perspectives of gender development have tapped into varying degrees of essentialism.

Building on past work that identified features of psychological essentialism common across social groups (see Dar-Nimrod & Heine, 2011), Lee, Reis, and Rogge (2020) proposed a gender-specific structure of essentialism that includes four factors: (1) *inductive potential*, or the belief that a person’s gender provides a rich source of information and inferences, (2) *immutability*, or the belief that gender is a discrete, sharply defined, and unchangeable category, (3) *biological determinism*, or the belief that gender differences are fundamentally biologically determined, and (4) *social determinism*, or the belief that gender differences are shaped by sociocultural factors. The first two facets of Lee et al.’s (2020) gender essentialism framework (*inductive potential* and *immutability*) map on most closely to gender identity (i.e., how a person defines their gender; York University, 2015; Diekman & Schmader, 2023). In contrast, the latter
two facets (biological and social determinism) map on most closely gender expression (i.e., the external attributes, behavior, appearance, dress, etc., by which a person expresses their gender; York University, 2015; Diekman & Schmader, 2023). Relevant to the current research, the expression of preferences associated with a given gender (i.e., men and women’s different interests) are included in the broader category of gender expression (Schmader & Nater, 2023). Given my focus on interest, my dissertation focuses on people’s contrasting beliefs in biological and social determinism.

1.4.1 How Is Psychological Essentialism Different from Stereotyping?

For conceptual clarity, it is important to distinguish between psychological essentialism and stereotyping. I argue that psychological essentialism is distinct from, but related to, stereotyping. Stereotypes in their broadest form encompass descriptive beliefs about groups (Ashmore & Del Boca, 1981), including what their different interests might be (Master, 2021). Some scholars posit that essentialism underlies stereotypes, in that it provides the “cognitive ingredients” necessary for stereotyping to occur (e.g., illusory correlation, accentuation, Bastian & Haslam, 2006; Dar-Nimrod & Heine, 2011). Other research examining the interaction between essentialism and stereotypes finds that stereotypes are especially harmful to the degree they are essentialized. For example, women who learned gender differences on math performance were due to genetic, as opposed to experiential, factors subsequently performed worse on a math test (Dar-Nimrod & Heine, 2006). Similarly, mothers (but not fathers) who learned gender differences on math were due to genetic factors reported less confidence in their daughters’ math abilities, which later predicted daughters’ own math interest and performance (Eccles & Jacobs, 1986). Other work finds that believing social differences are due to genetic, as opposed to
experiential, factors is linked with greater endorsement of gender stereotypes (Brescoll & LaFrance, 2004; Martin & Parker, 1995), and greater self-stereotyping (Coleman & Hong, 2008).

Those who endorse gender stereotypes often support them with biological accounts of hormonal, physical, and evolutionary sex differences (Ellemers, 2018). As reviewed above, this is reflected in biological perspectives that seek to explain gender differences in interest. Yet people can also take a social constructionist view of these same gender differences. Building on the prior work reviewed above, a key goal of my dissertation is to disentangle stereotypes about interest (e.g., “women are less interested in systemizing”) from beliefs about the root cause of the difference (e.g., “it is primarily because of biology,” or “it is primarily because of socialization”). Specifically, I aim to unpack how these essentialized beliefs lead perceivers to provide affordances that sustain or diverge from gender stereotypes about interest. To illustrate this process, I next review the social psychological research on self-fulfilling prophecy.

1.5 **How Stereotypes Constrain Interests: Self-Fulfilling Prophecy**

Self-fulfilling prophecy (Merton, 1948; for a review, see Jussim, 1986; Jussim & Harber, 2005) is the process through which perceivers’ expectations shape targets’ outcomes in three stages: (1) perceivers develop expectations, (2) perceivers treat targets differently depending on their expectations (i.e., they provide different situational affordances), and (3) targets react to this treatment in expectancy-confirming ways. American sociologist Robert Merton (1948) originally proposed self-fulfilling prophecy as a process to account for myriad social problems, ranging from the infamous Wall Street Crash of 1929 to racial tensions brought on by the exclusion of Black workers from labor unions during the 1940s. In the decades following Merton’s review, research on self-fulfilling prophecy spanned diverse contexts, including the classroom (Jussim & Harber, 2005), courtrooms (Rosenthal, 1994), and medical practice (Rosenthal, 2003). Relevant
to the current research, prior studies have documented self-fulfilling prophecy effects based on people’s stereotypes of gendered preferences (Hollingshead & Fraidin, 2003; Madon et al., 2018; Skrypnek & Snyder, 1982). As I will review below, gender stereotypes especially lend themselves to self-fulfilling prophecy research, as they reflect beliefs commonly held across perceivers.

Although self-fulfilling prophecies have been critiqued for their small effects, effects are larger when accumulated across multiple instances or perceivers (Jussim, Eccles, & Madon, 1996; Jussim & Harber, 2005; Madon et al., 2018) and especially when expectations are negative (Madon et al., 2004). Specifically, prior research suggests self-fulfilling prophecy effects are smallest in one-on-one dyadic settings where social agents have an idiosyncratic expectation of a target. Yet in settings where social agents’ expectations are based on stereotypes shared by others in the environment, these small effects can accumulate across multiple perceivers and magnify effects (Madon et al., 2018). Across four studies, Madon et al. (2018) found that perceivers provided different situational affordances to targets based on stereotypes about their weight (i.e., by providing more candy to targets they believed were heavy versus thin) and gender (i.e., by providing more gender stereotypical articles to prototypically masculine and feminine targets). Targets, in turn, responded to these disparate affordances by engaging in behavior that confirmed societal stereotypes (i.e., by eating more candy; by choosing more gender stereotypical articles). Given my dissertation’s focus on how gender gaps about interest are sustained via people’s shared beliefs about the root cause of these differences, self-fulfilling prophecy offers a compelling framework for understanding how people’s expectations might translate into self-fulfilling behavior.
1.6 Gendered Situational Affordances

A central mechanism at play in self-fulfilling prophecies is situational affordances, defined as the process by which perceivers create situations for targets that provide them with opportunities and/or constraints (Madon et al., 2018). Prior evidence suggests affordances are often guided by gender stereotypes. Indeed, although parents play an integral role in helping children develop empathizing- and systemizing-relevant drives (e.g., Haddock et al., 2017; Tenenbaum & Leaper, 2003), caregivers of young children have historically provided different learning opportunities to girls and boys. One meta-analysis of 172 studies conducted prior to 1990 found that of 19 areas of socialization, the only area with a significant effect was encouragement of gender-typical activities (Lytton & Romney, 1991). Another study conducted among parents of 865 elementary school children found parents were more likely to encourage daughters to learn cooking and homemaking skills (stereotypically feminine activities), whereas they were more likely to encourage sons to work on or play with a computer outside of school (stereotypically masculine activities; Eccles et al., 1993). In the context of STEM, prior research finds that boys are more likely to get early experience with coding, robotics, physics, and mechanical systems (Barron, 2004; Jones, Howe, & Rua, 2000; Schmader, 2023). And in the context of HEED, boys are more likely to be discouraged from expressing emotions (Eisenberg, Cumberland, & Spinrad, 1998) and receive less emotion-related language from caregivers (Dunn, Bretherton, & Munn, 1987).

Although these prior studies underscore historic gender gaps in affordances, there is a lack of more recent research examining whether disparities in affordances persist. On the one hand,

---

3 Interestingly, although this research finds that boys are more likely to receive early experience with prototypical systemizing domains like electric toys, fuses, microscopes, and pulleys; girls also received more early experience with bread-making, knitting, and sewing. I argue these, too, might be considered systemizing activities, though are typically not conceptualized as systemizing in the academic literature (see Chapter 2).
hand, we might expect gender gaps in affordances to lessen over time given cultural progress toward gender parity (Schmader & Nater, 2023) coupled with weakening gender stereotypes about women’s associations with STEM (Charlesworth & Banaji, 2022). On the other hand, recent work by Alison Master and colleagues (2021) demonstrates gender stereotypes associating boys with greater STEM interest remain prevalent. And given that gendered interest stereotypes guide people’s situational affordances (Hollingshead & Fraidin, 2003; Madon et al., 2018; Skrypnek & Snyder, 1982), we might continue to see gender gaps on affordances as a function of these interest stereotypes. Determining the current state of gender gaps in affordances is critical for understanding horizontal gender segregation, given that disparities in early experience are a key contributor to gender disparities in STEM and HEED participation (Cheryan et al., 2017; Croft et al., 2015). Thus, a central aim of my dissertation is to test whether gender gaps in affordances are still salient, and whether they might be especially present for those highest in essentialized beliefs about gender differences in interest.

1.7 Contribution and Overview of Current Research

Interests are often conceived of as reflecting personal choices freely made by individuals (Ceci et al., 2009, 2011; Ryan & Deci, 2006). Departing from this perspective, I contend that interests are directly shaped by opportunities and constraints in the environment (i.e., situational affordances). Integrating theorizing on psychological essentialism and self-fulfilling prophecy, I have proposed a process through which people’s essentialized explanations for gender differences in interest (as primarily biological vs. socially caused) leads them to provide disparate situational affordances that confirm those beliefs. In Chapter 2, I test whether scales designed to measure biological gender differences in interest ask about activities that men and women have different affordances to learn, and whether these gendered affordances partially
explain the size of the gender gaps observed using these measures. In Chapter 3, I test whether STEM professionals who believe gender differences in interest are biologically (vs. socially) determined provide men and women with disparate affordances, and whether these disparate affordances widen the gender gap in interest. I next provide a detailed summary of these empirical chapters.

1.8 Summary of Empirical Chapters

1.8.1 Chapter 2: Do Measures of Systemizing and Empathizing Reflect Perceptions of Gender Differences in Learning Affordances?

In Chapter 2, I consider whether scales designed to measure biological gender differences in interest ask about activities that men and women have different affordances to learn, and whether these gendered affordances partially explain the size of the gender gap. Across three studies, I leveraged a wisdom of crowds (Larrick et al., 2011) approach to examine these questions with E-S Theory’s self-report measures of systemizing (SQ) and empathizing (EQ). Study 1 ($N = 624$) first estimated gender differences on systemizing (SQ) and empathizing (EQ) scales in two large, nationally representative samples. Results of subsequent coding studies revealed both lay coders (Study 2, $N = 199$) and psychology journal reviewers (Study 3, $N = 116$) rated SQ and EQ item activities as being more learned (vs. innate), counter to how E-S theory tends to explain these differences. Coders also believed that men are given more systemizing, and women given more empathizing (Study 3 only), affordances. Relating item-level variation on these ratings to gender differences observed among participants in Study 1, items showing the largest gender differences in Study 1 were those rated as having the most gendered affordances (more than gendered genetic advantages) by coders in Studies 2 and 3. Together, this chapter concludes that claims about inherent sex differences in systemizing, and to
a lesser degree empathizing, might overlook the degree that self-report measures tap into
gendered learning affordances.

1.8.2 Chapter 3: Are Essentialized Explanations for Gender Differences in interest Self-Fulfilling?

In Chapter 3, I consider whether perceivers who believe gender differences in interest are
primarily biologically (vs. socially) determined provide men and women with gendered (vs.
counter-gendered) affordances, and whether these affordances drive the gender gap in interest. I
tested these questions across three studies that implemented a novel behavioral paradigm in the
context of an imagined tech company. In Studies 4 ($N = 285$) and 5 ($N = 379$), perceivers were
asked to imagine themselves in a project manager role. Those who primarily endorsed (Study 4)
or were experimentally primed with (Study 5) a biological (vs. social) explanation for interest
differences provided gendered affordances: they were more likely to assign women interns to
empathizing roles and men interns to systemizing roles. Translating these situational affordances
to target outcomes in a quasi-double randomization design (MacKinnon et al., 2007; Madon et
al., 2018; Word et al., 1974), in an imagined internship scenario (Study 6, $N = 300$), first- and
second-year undergraduate women who received gendered (vs. counter-gendered) affordances
based on perceivers’ selections in Study 4 shifted their interests to be more gendered (i.e.,
became more interested in empathizing than systemizing teams) and chose to work on more
empathizing than systemizing teams. In contrast, women who received counter-gendered (vs.
gendered) situational affordances shifted their interests to be counter-gendered (i.e., became
more interested in systemizing than empathizing teams) and chose to work on empathizing and
systemizing teams equally. I also found that women intended to stay at the company to the
degree affordances matched their trait-level interests. Together, this evidence highlights
situational affordances, driven by essentialist explanations for the interest gap, as a key contributor to gender gaps in interest and occupational pursuits.
Chapter 2: Do Measures of Systemizing and Empathizing Reflect Perceptions of Gender Differences in Learning Affordances?

In 2017, Google engineer James Damore published a memo titled *Google’s Ideological Echo Chamber*. In it he asserted, “On average, men and women biologically differ in many ways [...] that may explain why we don’t see equal representation of women in tech and leadership.” He cited women’s lower interest in *systemizing* and higher interest in *empathizing* as fundamental differences that make efforts toward equal representation in tech “unfair, divisive, and bad for business.” Damore was fired for his memo, but women and efforts toward gender equity were harmed by the event and the ensuing debate. One woman engineer voiced, “I’m exhausted by having this same damn argument over and over again [...] and the amount of time and energy I keep having to spend to counter it” (Cowansage, 2017). Indeed, Fine (2012) argues that biologically essentialized accounts of gender differences work not only to *explain*, but also to *sustain*, gender differences by threatening women’s sense of fit, belonging, and ability to be successful in male-dominant fields (Bian et al., 2018; Cheryan & Markus, 2020; Dar-Nimrod & Heine, 2006).

Damore’s memo illustrates the far-reaching consequences of empathizing-systemizing (E-S) theory (Baron-Cohen, 2002, 2004) and its claims of innate sex differences. Given the potential cost of these claims, a close look at the evidence supporting E-S theory is warranted. In the present work, we examine the degree to which measures developed to support E-S theory include activities that men and women are perceived to have different opportunities to learn, and whether perceived differences on learning opportunities correspond to the size of the gender
difference⁴. First, we review E-S theory and its central claims. Next, we consider the empirical evidence for gender/sex differences, with a particular focus on self-report measures. Finally, we review how our approach extends prior efforts to address sources of bias in these measures.

2.1 Defining Systemizing and Empathizing: E-S Theory

E-S theory, initially developed to describe the cognitive profile of individuals with autism spectrum conditions (ASC; Baron-Cohen, 2002), posits ASC can be defined by higher systemizing, or “the drive to analyze the variables in a system, to derive the underlying rules that govern the behaviour of a system [...] and the drive to construct systems” (Baron-Cohen et al., 2003, p. 361) and lower empathizing, or “the ability to tune into how someone else is feeling, or what they might be thinking [...] understand the intentions of others, predict their behavior, and experience an emotion triggered by their emotion” (Baron-Cohen & Wheelwright, 2004, p. 163).

E-S theory has been influential beyond its implications for understanding autism. In proposing “extreme male brain theory,” Baron-Cohen defined the male brain as having significantly more systemizing than empathizing ability (S > E) and the female brain as having significantly more empathizing than systemizing ability (E > S; Baron-Cohen, 2002, 2004, 2008, 2009). According to extreme male brain theory, the cognitive profile of an individual with ASC represents an exaggerated version of the typical male profile (S >> E; Baron-Cohen, 2002, 2004, 2008, 2009). Embedded in female/male brain terminology is an assumption that differences on systemizing and empathizing are rooted in biological differences by sex more than sociocultural differences by gender. Given the impact of these claims, what is the evidence for these fundamental differences?

⁴ We use the term gender differences throughout to refer to observed differences between women or men, acknowledging this distinction does not capture identities beyond these binary categories. We use sex differences to refer to what are sometimes assumed to be essentialized differences based on sex assigned at birth.
2.2 Evidence of Gender Differences on Systemizing and Empathizing

Evidence for gender differences on systemizing and empathizing includes behavioral and self-report measures (for reviews, see Fine, 2010; 2012; Gillis-Buck & Richardson, 2014; Grossi & Fine, 2012). Of these, self-report measures have yielded the clearest and largest effect sizes. Men score higher than women on the Systemizing Quotient (SQ) in both its original \( d = .59, N = 278; \) Baron-Cohen et al., 2003) and shortened form \( d = .95; N = 723, \) Wakabayashi et al., 2006). Likewise, women score higher than men on the Empathizing Quotient (EQ) in both its original \( d = -.50, N = 197; \) Baron-Cohen & Wheelwright, 2004) and shortened form \( d = -.63; N = 1038, \) Wakabayashi et al., 2006).

In contrast to these self-report measures, behavioral evidence is much weaker. Consider the Embedded Figures Task (EFT; Witkin et al., 1971) and the Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al., 2001) – two key measures identified by Baron-Cohen as behavioral indicators of systemizing and empathizing, respectively (Baron-Cohen, 2002, 2009; Chapman et al., 2006). Cross-cultural studies find no gender difference on EFT (systemizing) performance (Kühnen et al., 2001); the RMET (empathizing) also yields small effects in meta-analyses \( g = 0.18, \) Kirkland et al., 2013) and large-scale studies \( g = -0.10, \) Schroeter et al., 2022). Given that evidence of differences is primarily tied to self-report measures, we next review these scales.

2.2.1 SQ and EQ Scales

The systemizing quotient (SQ; Baron-Cohen et al., 2003) was developed to assess adults’ systemizing in the context of everyday activities (e.g., “When I look at a building, I am curious about the precise way it was constructed”). The empathy quotient (EQ; Baron-Cohen & Wheelwright, 2004), published the following year, was developed to measure adults’
empathizing (e.g., “I find it easy to put myself in somebody else’s shoes”). In each measure, participants rate their agreement with 40 statements and 20 filler items. Wakabayashi et al. (2006) developed short scales of each construct (the 25-item SQ-Short and 22-item EQ-Short) based on factor analyses of the SQ and EQ measures. Although other forms of the SQ and EQ have been developed by external research groups (e.g., revised EQ, Muncer & Ling 2006) and for different age ranges (e.g., the Children’s SQ and EQ, Auyeung et al., 2009), we focus on self-report measures developed by the Baron-Cohen research group for adult samples.

Gender differences found on the SQ and EQ are often interpreted as supporting E-S theory’s claims of inherent sex differences. For example, in Archer’s 2019 (p. 35) review entitled, “the reality and evolutionary significance of human sex differences,” the first piece of evidence provided under the heading of “Evidence for evolutionary origins” cites evidence from the SQ: “In a study of 53 nations, men consistently scored much higher than women on systemizing, (Manning et al., 2010).”

In contrast to this interpretation, we propose that such effects reflect perceived differences in opportunities afforded to men and women to learn the skills referenced in the items (e.g., SQ-Short Item 2: “If there was a problem with the electrical wiring in my home, I’d be able to fix it myself”), more than genetic advantages that men and women have for systemizing and empathizing activities. Indeed, prior research documents gender gaps in learning opportunities afforded to boys and girls during childhood (Lytton & Romney, 1991; Tenenbaum & Leaper, 2003). One study conducted among parents of 865 elementary school children found parents were more likely to encourage daughters to learn cooking and homemaking skills, whereas they were more likely to encourage sons to work on or play with a computer outside of
school (Eccles et al., 1993). Based on this prior research, we do not dispute that gender differences on SQ and EQ exist but question the prevailing interpretation.

2.3 Prior Efforts to Reduce Gender Bias on the SQ and EQ

Concurrent to Wakabayashi et al.’s (2006) development of the SQ-Short, Wheelwright et al. (2006) used the same dataset to create the SQ-R by adding systemizing activities from less gendered or explicitly feminine domains (e.g., grammar, animals, family). This revised scale yielded a reduced, but still moderate, gender difference \( (d = .49, N = 1,761; \text{Wheelwright et al., 2006}) \). Additionally, Allison et al. (2011, 2015) found no evidence that EQ items functioned differently for men and women, but 31 of 75 (41%) SQ-R items functioned differently across gender. After eliminating these items, the resulting measure still showed a moderate gender difference with men scoring higher \( (d = .53, N = 4,058) \).

Although prior efforts to debias systemizing measures present a promising step, we present three key reasons for continued examination of these scales. First, although published over a decade ago, Allison et al.’s debiased version of the SQ scale is not widely used; it has been cited 0 times in PsycINFO at the end of 2022 (compared to 134 and 191 citations for the SQ-Short and SQ-R, respectively). Second, the addition of gender-neutral or feminine activities on the SQ-R does not directly address concerns that items might be partly assessing variance in men’s and women’s different learning affordances. Third, despite these prior efforts, researchers continue to cite differences on the SQ and EQ to support conclusions about innate sex differences (Archer, 2019), despite the fact that self-reported measures can never reveal the etiology of observed differences on the constructs they assess. Given these concerns, we tested the degree to which SQ and EQ might measure gender differences that people perceive as reflective of learning affordances, rather than innate differences.
2.4 Overview of Current Research

We hypothesized that observed gender differences on systemizing and empathizing would be better predicted by people’s perceptions of men and women’s learning affordances for activities referenced in the SQ and EQ, more than their perceptions of innate gender differences on the same activities. We tested this with the SQ- and EQ-Short, the two scales that were psychometrically validated using factor analysis and widely used given their shorter format. We addressed two questions: (1) Do the SQ- and EQ-Short ask about activities perceived to reflect men and women’s innate differences and/or different learning affordances? (2) Do perceived innate differences or learning affordances better predict the size of the gender difference observed on the SQ- and EQ-Short items?

To measure perceptions of learning affordances and genetic differences on SQ and EQ items, we adopted a “wisdom of crowds” approach to assess how diverse and expert samples perceive the activities assessed in these scales (Larrick et al., 2011). We first conducted an empirical target study to obtain an estimate of the gender differences on each SQ- and EQ-Short item among representative samples from the United States (US) and United Kingdom (UK; Study 1). We next conducted two coding studies to examine how lay coders (Study 2) and experts in human behavior (i.e., psychology journal reviewers, Study 3) estimate the learning affordances and genetic advantages for each of the activities referenced in the items. We then tested the hypothesis that those items with the largest perceived gender affordances (in Studies 2 and 3), more than perceived genetic advantages, would have the largest gender differences in the target sample (Study 1).
2.5 Study 1: Estimating SQ and EQ Gender Differences

Study 1 estimated the size of the gender difference on each item of the SQ- and EQ-Short scales (Wakabayashi et al., 2006) among two nationally representative samples of participants from the UK (where the scales were originally developed) and the US (where coders for Studies 2 and 3 were largely based). Since there were no significant country differences on item-level analyses after Bonferroni correction, we assigned each item a score based on the effect size (Cohen’s $d$ for the gender difference) observed in the combined sample. This item-level score was then used as the outcome measure in Studies 2 and 3. Given the descriptive nature of Study 1, we did not preregister hypotheses.

2.5.1 Method

All data, materials, and analysis code for Studies 1-3 are available at osf.io/pyfk4/?view_only=941f5eb11e5642e0b1bd8b6f3ec47cbe.

2.5.1.1 Participants

2.5.1.1.1 Demographics

Our final sample included 624 adults (302 men, 315 women) recruited through Prolific ($N_{UK} = 313$, $N_{US} = 306$). We utilized Prolific’s nationally representative sampling option, which employs a stratified sampling technique to match the demographic composition of the country on gender, age, and ethnicity. Descriptive information by country is provided in Table 2.1. Additional participants were excluded from the final sample for failing an instructional attention check ($n = 37$; Oppenheimer et al., 2009; see Appendix A). A sensitivity analysis indicated we were able to detect effects above $d = .23$ with 80% power (see Appendix A for details).
Table 2.1 Sample demographics by region (Study 1)

<table>
<thead>
<tr>
<th>Gender</th>
<th>US N = 313</th>
<th>UK N = 306</th>
<th>Combined Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)</td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>152 (48.56%)</td>
<td>150 (49.02%)</td>
<td>302 (48.40%)</td>
</tr>
<tr>
<td>Woman</td>
<td>159 (50.80%)</td>
<td>156 (50.98%)</td>
<td>315 (50.48%)</td>
</tr>
<tr>
<td>Non-binary</td>
<td>2 (0.64%)</td>
<td>0 (0%)</td>
<td>2 (0.32%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Racial/Ethnic Background</th>
<th>US N = 313</th>
<th>UK N = 306</th>
<th>Combined Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)</td>
<td></td>
</tr>
<tr>
<td>Black or African</td>
<td>43 (13.74%)</td>
<td>16 (5.23%)</td>
<td>59 (9.46%)</td>
</tr>
<tr>
<td>East Asian</td>
<td>12 (3.83%)</td>
<td>6 (1.96%)</td>
<td>18 (2.88%)</td>
</tr>
<tr>
<td>Hispanic or Latino/a</td>
<td>12 (3.83%)</td>
<td>1 (0.33%)</td>
<td>13 (2.08%)</td>
</tr>
<tr>
<td>Indigenous</td>
<td>3 (0.96%)</td>
<td>0 (0%)</td>
<td>3 (0.48%)</td>
</tr>
<tr>
<td>Middle Eastern or Arabic</td>
<td>1 (0.32%)</td>
<td>2 (0.65%)</td>
<td>3 (0.48%)</td>
</tr>
<tr>
<td>South Asian</td>
<td>6 (1.92%)</td>
<td>22 (7.19%)</td>
<td>28 (4.49%)</td>
</tr>
<tr>
<td>Southeast Asian</td>
<td>7 (2.24%)</td>
<td>1 (0.33%)</td>
<td>8 (1.28%)</td>
</tr>
<tr>
<td>White (^5)</td>
<td>208 (66.45%)</td>
<td>237 (77.45%)</td>
<td>445 (71.31%)</td>
</tr>
<tr>
<td>Not Listed</td>
<td>3 (0.96%)</td>
<td>19 (6.21%)</td>
<td>22 (3.53%)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>18 (5.75%)</td>
<td>2 (0.65%)</td>
<td>20 (3.21%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>US N = 313</th>
<th>UK N = 306</th>
<th>Combined Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>45.53 (16.10)</td>
<td>45.23 (15.45)</td>
<td>45.38 (15.77)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Political Orientation</th>
<th>US N = 313</th>
<th>UK N = 306</th>
<th>Combined Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 = Extremely Liberal, 7 = Extremely Conservative)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^5\) In the original measures for Studies 1-3, the option “White” was listed as “White or Caucasian.”
Procedure and Measures

After passing the initial attention check and providing consent, participants completed the SQ-Short (25 items) and EQ-Short (22 items; see Table 2.2), presented in random order. They rated their agreement with each statement on a scale ranging from 1 (Strongly disagree) to 7 (Strongly agree). Lastly, participants provided basic demographic information. Participants received £2.50 (UK) and $2.73 (US) for completing the survey.

Results

Gender Difference on Systemizing and Empathizing Composites

First, to estimate gender differences in systemizing and empathizing, we calculated an average score for each person based on their ratings of all systemizing items and all empathizing items. One participant was missing data on systemizing items and did not receive a composite score. All other participants responded to all items. Both composites showed good reliability, $\alpha_{SQ} = .90, \alpha_{EQ} = .92$. Systemizing and empathizing were uncorrelated for the whole sample ($r = .02, p = .702$) but were positively correlated for men ($r = .16, p = .006$) and women ($r = .15, p = .007$) separately. As shown in Table 2.2, men scored significantly higher than women on the systemizing quotient, whereas women scored significantly higher than men on the empathizing quotient. Figure 2.1 shows the distribution of scores on the SQ-Short and EQ-Short by participant gender.

<table>
<thead>
<tr>
<th></th>
<th>$M$ (SD)</th>
<th>$M$ (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.24 (1.65)</td>
<td>3.53 (1.31)</td>
</tr>
</tbody>
</table>

The original SQ- and EQ-Short employed a different scale and scoring method, which included 4 forced-choice response options, treated all “disagree” options as equivalent, and summed rather than averaged scores (see Wakabayashi et al., 2006). However, including more scale points and a “Neutral” response option has been shown to reduce response bias and improve reliability (see Croasmun & Ostrom, 2011).
2.5.2.2 Gender Difference on Systemizing and Empathizing Items

Next, we obtained an effect size estimate for the gender difference on each item (\( N = 617 \); see Table 2.2).\(^7\) We used an arbitrary cutoff of \( d = .30 \) (reflecting a small but meaningful effect size) to bin items into three categories: women scored higher than men \( (d < -.30) \), men and women scored relatively equal \( (-.30 < d < .30) \), and men scored higher than women \( (d > .30) \).\(^8\) Using these cutoffs, men scored higher than women on the majority (70%) of systemizing items and women scored higher than men on just over half (55%) of empathizing items. Within each scale, the size of the gender differences varied considerably. For example, on the SQ-Short, the item with the smallest effect size was Sys 15: “I am not very meticulous when I carry out D.I.Y.” (reverse-scored), \( d = .02 \) and the item with the largest effect size was Sys 2: “If there was a problem with the electrical wiring in my home, I’d be able to fix it myself,” \( d = 1.04 \). On the EQ-Short, the item with the smallest effect size was Emp 21: “I am good at predicting what someone will do,” \( d = -.15 \) and the item with the largest effect size was Emp 22: “I tend to get emotionally involved with a friend’s problems,” \( d = -.62 \). Additionally, the average size of the

---

\(^7\) These analyses do not include \( N = 2 \) participants who identified as non-binary and \( N = 5 \) participants who did not report their gender.

\(^8\) There were no items that showed an effect size of exactly \( d = \pm .30 \); thus, all items were binned into one of the three predefined categories.
gender differences on individual items (SQ-Short: M_d = .47, SD = .23; EQ-Short: M_d = -.33, SD = .16) tended to be smaller than the size of the gender difference on the composites (SQ-Short: d = .92; EQ-Short: d = -.57; see Eagly & Revelle, 2022).

Table 2.2. Items, descriptive statistics by gender, and effect size for gender difference (Study 1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Item Text</th>
<th>Women M (SD)</th>
<th>Men M (SD)</th>
<th>Gender Difference (d, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemizing (Composite)</td>
<td></td>
<td>4.07 (0.87)</td>
<td>4.85 (0.82)</td>
<td>0.92 [.75, 1.09]</td>
</tr>
<tr>
<td>Sys 5</td>
<td>I am fascinated by how machines work.</td>
<td>4.00 (1.63)</td>
<td>5.35 (1.37)</td>
<td>0.89 [.72, 1.05]</td>
</tr>
<tr>
<td>Sys 14</td>
<td>I find it easy to grasp exactly how odds work in betting.</td>
<td>3.32 (1.65)</td>
<td>4.48 (1.69)</td>
<td>0.70 [.54, .86]</td>
</tr>
<tr>
<td>Sys 25</td>
<td>I am not interested in understanding how wireless communication works. (R)</td>
<td>3.87 (1.78)</td>
<td>4.89 (1.62)</td>
<td>0.60 [.44, .76]</td>
</tr>
<tr>
<td>Sys 13</td>
<td>If I were buying a stereo, I would want to know about its precise technical features.</td>
<td>4.20 (1.83)</td>
<td>5.20 (1.71)</td>
<td>0.56 [.40, .72]</td>
</tr>
<tr>
<td>Sys 4</td>
<td>I do not enjoy games that involve a high degree of strategy. (R)</td>
<td>4.38 (1.66)</td>
<td>5.24 (1.57)</td>
<td>0.53 [.37, .69]</td>
</tr>
<tr>
<td>Sys 1</td>
<td>If I were buying a car, I would want to obtain specific information about its engine capacity.</td>
<td>4.51 (1.74)</td>
<td>5.36 (1.55)</td>
<td>0.52 [.36, .68]</td>
</tr>
<tr>
<td>Sys 23</td>
<td>When I'm in a plane, I do not think about the aerodynamics. (R)</td>
<td>3.26 (1.68)</td>
<td>4.10 (1.82)</td>
<td>0.48 [.32, .64]</td>
</tr>
<tr>
<td>System</td>
<td>Item</td>
<td>Mean</td>
<td>SD</td>
<td>Correlation</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>----</td>
<td>-------------</td>
</tr>
<tr>
<td>Sys 7</td>
<td>I find it difficult to understand instruction manuals for putting appliances together. (R)</td>
<td>4.90 (1.63)</td>
<td>0.45</td>
<td>.29, .61</td>
</tr>
<tr>
<td>Sys 16</td>
<td>When I look at a building, I am curious about the precise way it was constructed.</td>
<td>3.51 (1.70)</td>
<td>0.40</td>
<td>.24, .56</td>
</tr>
<tr>
<td>Sys 11</td>
<td>I find it difficult to learn my way around a new environment.</td>
<td>4.78 (1.64)</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>Sys 17</td>
<td>I find it difficult to understand information the bank sends me on different investment and saving systems. (R)</td>
<td>4.73 (1.65)</td>
<td>0.35</td>
<td>.19, .51</td>
</tr>
<tr>
<td>Sys 21</td>
<td>When I look at a mountain, I think about how it was formed.</td>
<td>3.23 (1.65)</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Sys 12</td>
<td>I do not tend to watch science documentaries on television or read articles about science and nature. (R)</td>
<td>4.83 (1.84)</td>
<td>0.25</td>
<td>.09, .41</td>
</tr>
<tr>
<td>Sys 24</td>
<td>I am interested in knowing the path a river takes from its source to the sea.</td>
<td>3.05 (1.76)</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>Sys 20</td>
<td>When I hear about the weather forecast, I am not very interested in the meteorological patterns. (R)</td>
<td>4.03 (1.70)</td>
<td>.21</td>
<td>.05, .37</td>
</tr>
<tr>
<td>Sys 19</td>
<td>If I were buying a camera, I would not look carefully into the quality of the lens. (R)</td>
<td>5.14 (1.66)</td>
<td>.16</td>
<td>.00, .32</td>
</tr>
<tr>
<td>Sys 15</td>
<td>I am not very meticulous when I carry out D.I.Y. (R)</td>
<td>4.86 (1.63)</td>
<td>.02</td>
<td>-.14, .18</td>
</tr>
</tbody>
</table>

**Empathizing (Composite)**

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emp 17</td>
<td>Other people often say that I am insensitive, though I don’t always see why. (R)</td>
<td>5.77 (1.30)</td>
<td>-0.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System</th>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sys 7</td>
<td>I find it difficult to understand instruction manuals for putting appliances together. (R)</td>
<td>4.90 (1.63)</td>
<td>0.45</td>
<td>.29, .61</td>
</tr>
<tr>
<td>Sys 16</td>
<td>When I look at a building, I am curious about the precise way it was constructed.</td>
<td>3.51 (1.70)</td>
<td>0.40</td>
<td>.24, .56</td>
</tr>
<tr>
<td>Sys 11</td>
<td>I find it difficult to learn my way around a new environment.</td>
<td>4.78 (1.64)</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>Sys 17</td>
<td>I find it difficult to understand information the bank sends me on different investment and saving systems. (R)</td>
<td>4.73 (1.65)</td>
<td>0.35</td>
<td>.19, .51</td>
</tr>
<tr>
<td>Sys 21</td>
<td>When I look at a mountain, I think about how it was formed.</td>
<td>3.23 (1.65)</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Sys 12</td>
<td>I do not tend to watch science documentaries on television or read articles about science and nature. (R)</td>
<td>4.83 (1.84)</td>
<td>0.25</td>
<td>.09, .41</td>
</tr>
<tr>
<td>Sys 24</td>
<td>I am interested in knowing the path a river takes from its source to the sea.</td>
<td>3.05 (1.76)</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>Sys 20</td>
<td>When I hear about the weather forecast, I am not very interested in the meteorological patterns. (R)</td>
<td>4.03 (1.70)</td>
<td>.21</td>
<td>.05, .37</td>
</tr>
<tr>
<td>Sys 19</td>
<td>If I were buying a camera, I would not look carefully into the quality of the lens. (R)</td>
<td>5.14 (1.66)</td>
<td>.16</td>
<td>.00, .32</td>
</tr>
<tr>
<td>Sys 15</td>
<td>I am not very meticulous when I carry out D.I.Y. (R)</td>
<td>4.86 (1.63)</td>
<td>.02</td>
<td>-.14, .18</td>
</tr>
</tbody>
</table>

**Empathizing (Composite)**

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emp 17</td>
<td>Other people often say that I am insensitive, though I don’t always see why. (R)</td>
<td>5.77 (1.30)</td>
<td>-0.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System</th>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sys 7</td>
<td>I find it difficult to understand instruction manuals for putting appliances together. (R)</td>
<td>4.90 (1.63)</td>
<td>0.45</td>
<td>.29, .61</td>
</tr>
<tr>
<td>Sys 16</td>
<td>When I look at a building, I am curious about the precise way it was constructed.</td>
<td>3.51 (1.70)</td>
<td>0.40</td>
<td>.24, .56</td>
</tr>
<tr>
<td>Sys 11</td>
<td>I find it difficult to learn my way around a new environment.</td>
<td>4.78 (1.64)</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>Sys 17</td>
<td>I find it difficult to understand information the bank sends me on different investment and saving systems. (R)</td>
<td>4.73 (1.65)</td>
<td>0.35</td>
<td>.19, .51</td>
</tr>
<tr>
<td>Sys 21</td>
<td>When I look at a mountain, I think about how it was formed.</td>
<td>3.23 (1.65)</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Sys 12</td>
<td>I do not tend to watch science documentaries on television or read articles about science and nature. (R)</td>
<td>4.83 (1.84)</td>
<td>0.25</td>
<td>.09, .41</td>
</tr>
<tr>
<td>Sys 24</td>
<td>I am interested in knowing the path a river takes from its source to the sea.</td>
<td>3.05 (1.76)</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>Sys 20</td>
<td>When I hear about the weather forecast, I am not very interested in the meteorological patterns. (R)</td>
<td>4.03 (1.70)</td>
<td>.21</td>
<td>.05, .37</td>
</tr>
<tr>
<td>Sys 19</td>
<td>If I were buying a camera, I would not look carefully into the quality of the lens. (R)</td>
<td>5.14 (1.66)</td>
<td>.16</td>
<td>.00, .32</td>
</tr>
<tr>
<td>Sys 15</td>
<td>I am not very meticulous when I carry out D.I.Y. (R)</td>
<td>4.86 (1.63)</td>
<td>.02</td>
<td>-.14, .18</td>
</tr>
</tbody>
</table>

**Empathizing (Composite)**

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emp 17</td>
<td>Other people often say that I am insensitive, though I don’t always see why. (R)</td>
<td>5.77 (1.30)</td>
<td>-0.59</td>
</tr>
<tr>
<td>Employee</td>
<td>Statement</td>
<td>Score (Mean, Standard Deviation)</td>
<td>Correlation Coefficient</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Emp 15</td>
<td>Friends usually talk to me about their problems as they say that I am very understanding.</td>
<td>5.48 (1.24) 4.73 (1.50)</td>
<td>-0.54 [-.70, -.38]</td>
</tr>
<tr>
<td>Emp 18</td>
<td>I can tune into how someone else feels rapidly and intuitively.</td>
<td>5.24 (1.17) 4.61 (1.40)</td>
<td>-0.49 [-.65, -.33]</td>
</tr>
<tr>
<td>Emp 10</td>
<td>I am quick to spot when someone in a group is feeling awkward or uncomfortable.</td>
<td>5.57 (1.08) 5.09 (1.32)</td>
<td>-0.40 [-.56, -.24]</td>
</tr>
<tr>
<td>Emp 9</td>
<td>I am good at predicting how someone will feel.</td>
<td>4.99 (1.20) 4.59 (1.30)</td>
<td>-0.32 [-.48, -.16]</td>
</tr>
<tr>
<td>Emp 11</td>
<td>I can’t always see why someone should have felt offended by a remark. (R)</td>
<td>4.68 (1.66) 4.20 (1.46)</td>
<td>-0.31 [-.47, -.15]</td>
</tr>
<tr>
<td></td>
<td>In a conversation, I tend to focus on my own</td>
<td></td>
<td>0.22</td>
</tr>
<tr>
<td>Emp 1</td>
<td>I can easily tell if someone else wants to enter a conversation.</td>
<td>5.50 (1.04) 5.23 (1.24)</td>
<td>-0.24 [-.40, -.08]</td>
</tr>
<tr>
<td>Emp 3</td>
<td>I find it hard to know what to do in a social situation. (R)</td>
<td>4.75 (1.56) 4.45 (1.67)</td>
<td>-0.18 [-.34, -.02]</td>
</tr>
<tr>
<td>Emp 4</td>
<td>I often find it difficult to judge if something is rude or polite. (R)</td>
<td>5.58 (1.34) 5.35 (1.31)</td>
<td>-0.17 [-.33, -.01]</td>
</tr>
<tr>
<td>Emp 10</td>
<td>I can easily work out what another person</td>
<td>4.75 (1.22) 4.54 (1.33)</td>
<td>-0.16</td>
</tr>
<tr>
<td>Emp 21</td>
<td>I am good at predicting what someone will do.</td>
<td>4.72 (1.20) 4.54 (1.24)</td>
<td>-0.15</td>
</tr>
</tbody>
</table>
Note. Items marked with (R) are reverse scored. Negative $d$ scores indicate women scored higher; positive $d$ scores indicate men scored higher.

2.5.3 Discussion

Study 1 revealed the expected gender differences on widely used measures of systemizing and empathizing. Men scored significantly higher than women on systemizing and women scored significantly higher than men on empathizing. However, the size of this gender difference varied considerably across items within each scale. This observed variability introduces the possibility that the size of the gender difference might be affected by activities described in certain items. Specifically, we suspected that some items capture gender differences that people perceive as being due to learning affordances, more than genetic differences. If so, perceived affordances to learn these skills should predict the size of the gender difference observed, perhaps more than perception of genetic advantages. As reviewed above, we used a wisdom of crowds approach (Larrick et al., 2011) to test these hypotheses among a diverse sample of lay coders (Study 2) and experts in psychology (Study 3), as having diverse perspectives and expertise are the two conditions under which crowds are assumed to be wise.

2.6 Study 2

The goal of Study 2 was to assess the degree to which gender differences observed on SQ- and EQ-Short items in Study 1 are related to perceived patterns of gender-based learning affordances in the activities assessed by the SQ- and EQ-Short. We asked a diverse sample of lay coders to rate activities from the SQ- and EQ-Short on: (1) the estimated gender difference, (2) whether that activity is likely to be innate versus learned, (3) whether men or women have more affordances to learn that activity, and (4) how reflective of genetic sex differences each activity
is. Since this initial study was exploratory, we did not preregister our hypotheses nor analytic strategy.

2.6.1 Method

2.6.1.1 Participants

2.6.1.1.1 Demographics

Our final sample included $N = 199$ coders from the United States and Canada recruited through Amazon’s Mechanical Turk. Table 2.3 provides a summary of participant demographics. Participants were excluded from this final sample for failing the same attention check used in Study 1 ($n = 23$). Although a minimum of 2 coders is recommended for intercoder reliability (O’Connor & Joffe, 2020), given the subjective nature of our ratings, we aimed to collect a large and diverse sample of coders (about 50 coders per activity). In our final sample, each rating had between $N = 49$ and $77$ ($M = 62.85$, $SD = 9.94$) participant coders. Ratings showed good interrater reliability, ICC = .90, 95% CI [.88, .92] (one-way random effects computed with absolute agreement and multiple raters/measurements; Koo & Li, 2015; Shrout & Fleiss, 1979).

<table>
<thead>
<tr>
<th>Gender</th>
<th>$N$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man</td>
<td>99 (49.75%)</td>
</tr>
<tr>
<td>Woman</td>
<td>98 (49.25%)</td>
</tr>
<tr>
<td>Racial/Ethnic Background</td>
<td>N (%)</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Black or African</td>
<td>33 (16.58%)</td>
</tr>
<tr>
<td>East Asian</td>
<td>6 (3.02%)</td>
</tr>
<tr>
<td>Hispanic or Latino/a</td>
<td>11 (5.53%)</td>
</tr>
<tr>
<td>Indigenous</td>
<td>1 (0.50%)</td>
</tr>
<tr>
<td>South Asian</td>
<td>6 (3.02%)</td>
</tr>
<tr>
<td>Southeast Asian</td>
<td>2 (1.01%)</td>
</tr>
<tr>
<td>White</td>
<td>132 (66.33%)</td>
</tr>
<tr>
<td>Not Listed</td>
<td>1 (0.50%)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>5 (2.51%)</td>
</tr>
</tbody>
</table>

**SES (1 = Lowest SES, 10 = Highest SES)**  

<table>
<thead>
<tr>
<th></th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.04 (2.13)</td>
</tr>
</tbody>
</table>

**Age**  

<table>
<thead>
<tr>
<th></th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>33.89 (10.13)</td>
</tr>
</tbody>
</table>

**Political Orientation (1 = Extremely Liberal, 7 = Extremely Conservative)**  

<table>
<thead>
<tr>
<th></th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.10 (1.90)</td>
</tr>
</tbody>
</table>

2.6.1.2 Procedure and Measures

Participants received $1.50 to participate in a study titled “Activity, Interest, and Ability Ratings.” After consenting, we presented participants with a random subset of 15 (of 47 total) activities abstracted from the SQ and EQ-Short items. For example, for Item EQ 22, “I tend to get emotionally involved in a friend’s problems,” we presented coders with, “getting emotionally involved with a friend’s problems” (see Appendix A for materials). Of the 15 activities shown to each coder, 5 were activities that women tended to score higher on, 5 were activities that tended
to be gender neutral, and 5 were activities that men tended to score higher on (estimates using the cutoffs described in Study 1). Participants rated each activity on:

1) estimated gender differences: “To what degree do men and women differ on [activity]?” ranging from 1 (WOMEN are higher) to 4 (It is equal) to 7 (MEN are higher).

2) learned vs. innate attributions: “To what degree is [activity] reflective of a preference/skill one is born with vs. a preference/skill one learns through experience?” ranging from 1 (More of a skill one is BORN WITH) to 4 (It is equal) to 7 (More of a skill one LEARNS through experience).

3) gendered learning affordances: “Who has more opportunities to learn about [activity]?” ranging from 1 (WOMEN have more opportunities to learn this) to 4 (It is equal) to 7 (MEN have more opportunities to learn this).

4) assumed genetic sex differences: “To what degree is [activity] reflective of innate, genetic differences between men and women?” ranging from 1 (Not at all) to 4 (Somewhat) to 7 (Very much).

The order of the last two ratings was counterbalanced between coders. The full list of measures collected in Study 2 is available in Appendix A.9

2.6.2 Results

To perform analyses at the item level, we assigned each item a value based on the unweighted mean of coder ratings for that item10 (Lorenz et al., 2011). This yielded 47 datapoints

---

9 There was a technical error that affected 4 of the 100 ratings (see Appendix A). Reanalyzing data excluding these ratings only changed one result: the interaction between genetic differences and item type changes from marginal ($\beta = .27, p = .054$) to significant ($\beta = .31, p = .033$).

10 Using the unweighted mean is a common estimate of wisdom of crowds (Lorenz et al., 2011), though some studies (e.g., Sjöberg, 2009) use median coder ratings. In these data, several ratings showed a median of 4, resulting in low or zero variance. To maximize variance, we use the unweighted mean. Additionally, we did not analyze data in an MLM framework since our goal was to model aggregate item-level information. Results are unchanged by this decision unless noted otherwise. We report all effects by coder gender and modeled in MLM in Appendix A.
(25 SQ-Short items, 22 EQ-Short items) per rating dimension, each with a value that reflected the average rating for each activity on that dimension. Table 2.4 and Figure 2.2 provide a summary of coder ratings. Although men and women coders showed mean differences on several ratings, the relationships between coder ratings and observed gender difference was fairly consistent across coder gender (see Appendix A).

2.6.2.1 Mean Participant Coder Ratings for SQ- and EQ-Short Activities

Mean ratings on each activity and results for the corresponding test against scale midpoint are provided in the Appendix A.

2.6.2.1.1 Estimated Gender Difference

Coders rated men and women as differing significantly on systemizing and empathizing activities, $t(44.78) = -13.87, p < .001, d = 4.04$, with men rated as higher on systemizing activities and women rated as higher on empathizing activities (as revealed by one sample t-tests comparing means to the scale midpoint, see Table 2.4). At the item level, coders rated men significantly higher on all but one systemizing activity and rated women significantly higher on 27.27% of empathizing activities. Men were also rated significantly higher on two empathizing activities prior to reverse-scoring (“Being insensitive,” “Focusing on one’s own thoughts rather than what their listener might be thinking.”).

2.6.2.1.2 Learned vs. Innate Attributions

Coders rated systemizing activities as being more learned through experience than empathizing activities, $t(33.37) = -7.00, p < .001, d = 1.96$. Both systemizing and empathizing activities were rated as being relatively more learned through experience than a skill one is born with (Table 2.4, tests against scale midpoint). At the item level, coders rated all but one systemizing activity and most (68.18%) empathizing activities as significantly more learned
through experience than a skill one is born with. No activities were rated as being significantly more innate than learned through experience.

2.6.2.1.3 Gendered Learning Affordances

Coders rated men and women as differing significantly in their affordances to learn systemizing and empathizing activities, $t(44.08) = -10.01, p < .001, d = 2.93$. Coders rated men as having significantly more affordances than women to learn systemizing activities, whereas they assumed men and women have equal affordances to learn empathizing activities (Table 2.4, tests against scale midpoint). At the item level, coders rated all but one systemizing activity as reflecting more affordances to men. Unexpectedly, coders also rated 18.18% of empathizing activities as providing more affordances to men, and only two empathizing activities were rated as providing more affordances to women.

2.6.2.1.4 Genetic Differences

Finally, both systemizing and empathizing activities were rated as being somewhat reflective of innate genetic differences (Table 2.4), but less so for systemizing than empathizing, $t(44.48) = 2.13, p = .039, d = 0.62$. Since this item was rated from not at all to very much, comparisons to the midpoint were ambiguous, a limitation we rectified in Study 3.

Table 2.4. Descriptive statistics, effect size for difference from midpoint, and $N$ items significantly below or above midpoint (Study 2)

<table>
<thead>
<tr>
<th>Rating Dimension</th>
<th>SQ-Short Activities</th>
<th></th>
<th></th>
<th></th>
<th>EQ-Short Activities</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Difference from Midpoint</td>
<td>Items Below Midpoint</td>
<td>Items Above Midpoint</td>
<td></td>
<td>Difference from Midpoint</td>
<td>Items Below Midpoint</td>
<td>Items Above Midpoint</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$M$ (SD)</td>
<td>$(d)$ [95% CI]</td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
<td>$M$ (SD)</td>
<td>$(d)$ [95% CI]</td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
</tr>
</tbody>
</table>
### Estimated Gender Difference

<table>
<thead>
<tr>
<th></th>
<th>Mean (SE)</th>
<th>Lower Bound (95% CI)</th>
<th>Upper Bound (95% CI)</th>
<th>Percentage Below Midpoint</th>
<th>Percentage Above Midpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learned vs. Innate Attributions</td>
<td>4.80 (0.27)</td>
<td>[2.56, 3.37]</td>
<td>(0%)</td>
<td>24 (96%)</td>
<td>3.74 (0.25)</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>5.38 (0.59)</td>
<td>[1.97, 2.69]</td>
<td>(0%)</td>
<td>24 (96%)</td>
<td>4.47 (0.25)</td>
</tr>
<tr>
<td>Genetic Differences</td>
<td>3.85 (0.31)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4.04 (0.30)</td>
</tr>
</tbody>
</table>

*Note. Below Midpoint = Women Higher, More Innate; Above Midpoint = Men Higher, More Learned. ***p < .001.*

#### Figure 2.2
Lay coder ratings and comparisons against midpoint (Study 2)

### 2.6.2.2 Item-Level Analyses Predicting Observed Gender Differences from Coder Ratings

Having established that systemizing items, in particular, were perceived to capture learnable skills that men have more opportunities to learn, we next tested whether the item-level
perceptions measured by this independent sample of lay coders (Study 2) predicted the size of the gender differences across items measured in Study 1.

2.6.2.2.1 Analytic Strategy

As in the results above, each item on the SQ- and EQ-Short was assigned a value (for each of the four measures) based on aggregated coder ratings in this sample. We also assigned each item a value based on the effect size for the gender difference observed on that item from Study 1 (positive values = men higher, negative values = women higher). Each model tested the main effect of coder rating dimension (continuous) and subscale (categorical: systemizing vs. empathizing), as well as their interaction, as predictors of the observed gender difference in Study 1. For all models, continuous predictors were standardized. To derive main effects of coder ratings on observed gender differences across subscale, we contrast coded our subscale variable (systemizing = 0.5, empathizing = -0.5). To derive interaction terms and simple slopes, we dummy coded categorical predictors (for models with systemizing as the reference group: systemizing = 0, empathizing = 1; for models with empathizing as the reference group: empathizing = 0 systemizing = 1) to examine slope by subscale separately.

2.6.2.2.2 Estimated Gender Difference

To validate the wisdom of the crowds approach, we first tested whether participant coders (Study 2) accurately estimated the true gender difference on each activity (Study 1). As expected, coder ratings of gender differences positively predicted the observed gender differences, $\beta = .64$, $p < .001$, 95% CI [.40, .88], such that coders accurately tracked the size of the gender difference on each activity. There was no interaction by subscale, $\beta = .20$, $p = .411$, 95% CI [-.28, .68].

2.6.2.2.3 Learned vs. Innate Attributions
Next, we examined how coders’ ratings of learned versus innate attributions for each activity (Study 2) predicted the observed gender difference on the items (Study 1). There was a main effect of coder ratings on the observed gender difference, $\beta = 0.31, p = .022$, 95% CI [.05, .57], but no significant interaction by subscale (systemizing vs. empathizing), $\beta = -.42, p = .115$, 95% CI [-.95, .11]. Activities that coders judged to be less learned through experience were those on which women scored higher. Recall that all items were coded as being equally or more learned than innate.

2.6.2.2.4 Gendered Learning Affordances

We next examined the relationship between coders’ perceptions of men and women’s relative affordances to learn systemizing and empathizing activities (Study 2) and the observed gender difference on corresponding items (Study 1). There was a significant main effect of affordances, $\beta = .47, p < .001$, 95% CI [.27, .67]. Although there was no moderation by subscale (systemizing vs. empathizing), $\beta = .04, p = .830$, 95% CI [-.35, .44], Figure 2.3 reveals that for systemizing, activities that men have more (perceived) affordances to learn are those that men scored higher on (beta coefficients represent simple slopes for systemizing and empathizing separately. Axis bands represent a gender difference with $d > .30$; blue) or women scoring higher $d < -.30$; red) in Study 1). Although coders rated men and women as having equal affordances to learn empathizing activities on average, specific empathizing activities that women have more (perceived) affordances to learn are those on which women scored higher.

---

11 Exploratory simple slope analyses revealed this pattern was largely driven by empathizing activities, $\beta = 0.51, p = .038$, and not systemizing activities, $\beta = 0.10, p = .308$. 
2.6.2.2.5 Genetic Differences

We next examined the relationship between coders’ ratings of genetically based sex differences (Study 2) and observed gender differences on the items (Study 1). There was no significant main effect of coder ratings on the observed gender difference, $\beta = -.02$, $p = .804$, 95% CI [-.15, .12], and no significant interaction by subscale (systemizing vs. empathizing), $\beta = .27$, $p = .054$, 95% CI [.00, .55].

2.6.3 Discussion

Study 2 provides initial evidence that the gender differences observed on the SQ and EQ-Short scales could be partly assessing men’s and women’s different affordances to learn the activities referenced in the items. First, all activities included in the SQ and EQ-Short measures are assumed by lay coders to be equally or more learnable than innate. Coders also said that men had significantly greater affordances to learn systemizing activities and believed men and women

![Figure 2.3. Relationship between coder ratings of gendered learning affordances (Study 2) and observed gender difference (Study 1)]
have equal affordances to learn empathizing activities. Coders were able to predict the size of the
gender difference on each activity with relative accuracy, providing some validity for a wisdom
of crowds approach.

Across both subscales, there was an intuitive correspondence between the gendered
affordances of these activities (as rated by coders in this study) and the observed gender
differences assessed in Study 1. Coders rated all systemizing activities as being those that men
have greater affordances to learn, and those activities rated as having stronger affordances for
men were those where men scored higher than women (Study 1). Although empathizing items
were not rated as generally providing more affordances to women, those items where women are
seen as having greater affordances were the items where women scored higher than men (Study
1). In contrast, coder attributions to genetic differences did not significantly predict the
magnitude of gender difference observed on the items.

Although these findings provide some evidence that the SQ-Short and (to a lesser extent)EQ-Short are, in part, indexing patterns of learning affordances perceived to vary across gender,
there were three main limitations of this study. First, although it is informative to track lay
perceptions, those with expertise in human psychology might be in a better position to estimate
the causal factors underlying gender differences. We reasoned that a sample of reviewers at
influential psychology journals across a variety of disciplinary subfields would satisfy both
conditions (diversity and expertise) under which crowds can be wise (Larrick et al., 2011).

Second, although we found no evidence that coders’ ratings of genetic influence
predicted observed gender differences, our measure lacked parallelism to our item about learning
affordances. In Study 3, we reworded this item to ask about genetic advantages for men or
women to directly test these two ratings as competing predictors of the observed gender
difference in Study 1. Finally, based on Study 2 results, we preregistered our analytic plan and hypotheses prior to conducting analyses in Study 3.

### 2.7 Study 3

The goal of Study 3 was to conduct a preregistered replication of Study 2 among a sample of expert coders recruited from editorial boards of influential psychology journals. Experts completed the same rating activities from Study 2. Our preregistered hypotheses are as follows:

**Primary Hypothesis 1:** Experts are able to accurately estimate gender differences in systemizing and empathizing items.

**Primary Hypothesis 2:** The items that experts believe that men have greater opportunities to learn will be those that men score higher on relative to women (and vice versa).

**Primary Hypothesis 3:** The items that experts believe focus on more learnable activities/behaviors will show larger gender differences.

**Exploratory Hypothesis 4:** The items for which experts believe men have a greater genetic advantage will be those that men score higher on relative to women (and vice versa).

#### 2.7.1 Method

##### 2.7.1.1 Participants

We emailed study invitations to $N = 612$ experts who at the time were editorial board members of influential journals spanning clinical, developmental, evolutionary, gender, general, neuroscience/cognitive, and social/personality psychology. Details about our recruitment strategy and target sample size is provided in Appendix A. Table 2.5 provides information on contact and response rates by subdiscipline and gender.
Table 2.5. Response rates by subdiscipline and gender (Study 3)

<table>
<thead>
<tr>
<th>Area</th>
<th>Women</th>
<th></th>
<th></th>
<th>Men</th>
<th></th>
<th></th>
<th>Grand Total</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N Con</td>
<td>N Res</td>
<td>% Res</td>
<td>N Con</td>
<td>N Res</td>
<td>% Res</td>
<td>N Con</td>
<td>N Res</td>
<td>% Res</td>
</tr>
<tr>
<td>Clinical</td>
<td>50</td>
<td>6</td>
<td>12.00</td>
<td>53</td>
<td>6</td>
<td>11.32</td>
<td>103</td>
<td>12</td>
<td>11.65</td>
</tr>
<tr>
<td>Developmental</td>
<td>50</td>
<td>17</td>
<td>34.00</td>
<td>50</td>
<td>8</td>
<td>16.00</td>
<td>100</td>
<td>25</td>
<td>25.00</td>
</tr>
<tr>
<td>Evolutionary</td>
<td>27</td>
<td>7</td>
<td>25.93</td>
<td>26</td>
<td>8</td>
<td>30.77</td>
<td>53</td>
<td>15</td>
<td>28.30</td>
</tr>
<tr>
<td>Gender</td>
<td>26</td>
<td>3</td>
<td>11.54</td>
<td>25</td>
<td>4</td>
<td>16.00</td>
<td>51</td>
<td>7</td>
<td>13.73</td>
</tr>
<tr>
<td>General</td>
<td>51</td>
<td>4</td>
<td>7.84</td>
<td>50</td>
<td>8</td>
<td>16.00</td>
<td>101</td>
<td>12</td>
<td>11.88</td>
</tr>
<tr>
<td>Neuro/Cognitive</td>
<td>51</td>
<td>4</td>
<td>7.84</td>
<td>50</td>
<td>11</td>
<td>22.00</td>
<td>101</td>
<td>15</td>
<td>14.85</td>
</tr>
<tr>
<td>Social/Personality</td>
<td>51</td>
<td>11</td>
<td>21.57</td>
<td>52</td>
<td>19</td>
<td>36.54</td>
<td>103</td>
<td>30</td>
<td>29.13</td>
</tr>
<tr>
<td>Grand Total</td>
<td>306</td>
<td>52</td>
<td>16.99</td>
<td>306</td>
<td>64</td>
<td>20.92</td>
<td>612</td>
<td>116</td>
<td>18.95</td>
</tr>
</tbody>
</table>

*Note. Con = Contacted, Res = Responded*

2.7.1.1.1 Demographics
The final sample of 116 experts was fairly evenly split by gender (see Table 2.6 for a summary of participant demographics). Additional participants were excluded from the final sample for not finishing the survey or indicating on a single item at the end of the survey that we should not use their data ($n = 35$). Participants who expressed interest were sent a report of findings post-analysis.

Table 2.6. Sample demographics (Study 3)

<table>
<thead>
<tr>
<th>Gender</th>
<th>$N$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man</td>
<td>64 (55.17%)</td>
</tr>
<tr>
<td>Woman</td>
<td>52 (44.83%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Racial/Ethnic Background</th>
<th>$N$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black or African</td>
<td>4 (3.45%)</td>
</tr>
<tr>
<td>East Asian</td>
<td>1 (0.86%)</td>
</tr>
<tr>
<td>Hispanic or Latino/a</td>
<td>2 (1.72%)</td>
</tr>
<tr>
<td>South Asian</td>
<td>2 (1.72%)</td>
</tr>
<tr>
<td>White</td>
<td>98 (84.48%)</td>
</tr>
<tr>
<td>Not Listed</td>
<td>2 (1.72%)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>5 (4.31%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SES (1 = Lowest SES, 10 = Highest SES)</th>
<th>$M$ (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.84 (0.99)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>$M$ (SD)</th>
</tr>
</thead>
</table>

---

12 In $N = 2$ cases, participants’ self-identified gender was not the same gender recorded in our recruitment list and we defaulted to participants’ self-identified gender.
### Political Orientation

<table>
<thead>
<tr>
<th>Political Orientation (1 = Extremely Liberal, 7 = Extremely Conservative)</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>52.31 (12.70)</td>
<td></td>
</tr>
</tbody>
</table>

2.7.1.2 Procedure and Measures

After consenting to the study procedures, similar to Study 2, we showed experts a random subset of 24 (of 47 total) activities abstracted from the SQ- and EQ-Short items. We increased the number of activities presented to experts (from 15 in Study 2 to 24 in Study 3) to maximize data with this difficult to recruit sample. Of the 24 activities shown to coders, 8 were activities that women tended to score higher on, 8 were activities that tended to be neutral, and 8 were activities that men tended to score higher on (estimates based on data from Study 1 using the same cutoffs described previously). We asked experts to rate each activity on the same dimensions as in Study 2, with two adjustments made to gendered learning affordances and genetic differences, as described below (see Appendix A for all Study 3 measures). Similar to Study 2, experts in Study 3 showed excellent interrater reliability, ICC = .97, 95% CI [.96, .97] (one-way random effects with absolute agreement and multiple raters/measurements; Koo & Li, 2015; Shrout & Fleiss, 1979).

#### 2.7.1.2.1 Gendered Learning Affordances

In Study 3, we adapted this measure to say: “When it comes to [activity]… Who has more opportunities to learn?” ranging from 1 (WOMEN have more) to 4 (It is equal) to 7 (MEN have more).

#### 2.7.1.2.2 Genetic Advantages

In Study 3, experts rated the perceived gender-based genetic advantage of each activity (presented in the same block as gendered learning affordances, order counterbalanced between
coders): “When it comes to [activity]… Who has a genetic advantage?” ranging from 1 (WOMEN have more) to 4 (It is equal) to 7 (MEN have more).

2.7.2 Results

The same analytic approach in Study 2 was used in Study 3. As in Study 2, several mean ratings on several dimensions differed by coder gender and subdiscipline, but effects were similar to aggregate coder results when we examined effects for men and women coders separately, and also separate by subdiscipline (see Appendix A). All effects for the estimated gender difference and gendered learning affordances held across coder gender and subdiscipline.

2.7.2.1 Mean Expert Ratings for SQ- and EQ-Short Activities

We did not preregister hypotheses related to mean ratings across SQ and EQ-Short activities. Table 2.7 and Figure 2.4 provide a summary of coder ratings. Mean ratings on each activity and results for the corresponding test against scale midpoint is provided in Appendix A.

2.7.2.1.1 Estimated Gender Difference

As in Study 2, experts rated men and women as differing significantly on systemizing and empathizing activities, \( t(45) = -14.46, p < .001, d = 4.20 \), with men higher on systemizing activities and women higher on empathizing activities (Table 2.7, tests against scale midpoint). At the item level, experts rated men significantly higher on all but one systemizing activity and, unlike lay coders, experts rated women significantly higher on the majority (81.82%) of empathizing activities. As in Study 2, experts also rated men as significantly higher on two empathizing activities prior to reverse scoring (“Being insensitive,” “Finding social situations confusing.”).

---

13 We preregistered we would test hypotheses with MLM if there was substantial dependency within subdiscipline or coder gender. As specified in Study 2, we later reasoned it was more sensible to model the unweighted item-level mean. Effects are unchanged by this decision unless noted otherwise. We report all effects by coder gender, by subdiscipline, and modeled in MLM in Appendix A.
2.7.2.1.2  Learned vs. Innate Attributions

Like lay coders, experts rated systemizing activities as being significantly more learned through experience compared to empathizing activities, $t(37.31) = -5.51, p < .001, d = 1.55$. Also mirroring effects with lay coders, experts rated both systemizing and empathizing activities as being more learned through experience than an innate ability one is born with (Table 2.7, tests against midpoint). At the item level, experts rated most (80%) systemizing activities and half (50%) of empathizing activities as being significantly more learned through experience than a skill one is born with. Unlike in Study 2, experts rated one systemizing activity (“Being intrigued by the rules and patterns governing numbers in math”) and two empathizing activities (“Finding social situations confusing,” “Being insensitive;” both items reverse-scored) as being significantly more innate than learned through experience.

2.7.2.1.3  Gendered Learning Affordances

As in Study 2, experts rated men and women as differing significantly in their affordances to learn systemizing and empathizing activities, $t(43.14) = -12.46, p < .001, d = 3.66$. Experts rated men as having significantly more affordances to learn systemizing activities, and unlike lay coders in Study 2, experts also rated women as having significantly more affordances to learn empathizing activities (Table 2.7, tests against midpoint). At the item level, experts rated every systemizing activity as providing significantly more affordances to men and all but two empathizing activities as providing more affordances to women. One empathizing activity ("Being insensitive;” reverse-scored) was rated as providing more affordances to men prior to reverse-scoring.

2.7.2.1.4  Genetic Advantages
Experts rated men and women as differing significantly in their genetic advantage on systemizing and empathizing activities, \( t(33.70) = -11.50, p < .001, d = 3.47 \). Experts rated men as having a significantly greater genetic advantage on systemizing activities, and women as having a significantly greater genetic advantage on empathizing activities (Table 2.7, tests against midpoint). At the item level, experts rated the majority (68%) of systemizing activities as affording a genetic advantage to men and all but two empathizing activities as affording a genetic advantage to women.

Table 2.7. Descriptive statistics, effect size for difference from midpoint, and \( N \) items significantly below or above midpoint (Study 3)

<table>
<thead>
<tr>
<th>Rating Dimension</th>
<th>SQ-Short Activities</th>
<th></th>
<th>EQ-Short Activities</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Difference from Midpoint</td>
<td>Items Below Midpoint</td>
<td>Items Above Midpoint</td>
<td>Difference from Midpoint</td>
</tr>
<tr>
<td></td>
<td>( M ) (SD)</td>
<td>( (d) ) [95% CI]</td>
<td>( N ) (%)</td>
<td>( N ) (%)</td>
</tr>
<tr>
<td>Estimated Gender Difference</td>
<td>4.53 (0.29)</td>
<td>1.85*** [1.41, 2.28]</td>
<td>0 (0%)</td>
<td>24 (96%)</td>
</tr>
<tr>
<td>Learned vs. Innate Attributions</td>
<td>5.16 (0.72)</td>
<td>1.62*** [1.20, 2.04]</td>
<td>1 (4%)</td>
<td>20 (80%)</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>4.66 (0.31)</td>
<td>2.10*** [1.64, 2.55]</td>
<td>0 (0%)</td>
<td>25 (100%)</td>
</tr>
<tr>
<td>Genetic Advantages</td>
<td>4.15 (0.09)</td>
<td>1.71*** [1.28, 2.13]</td>
<td>0 (0%)</td>
<td>17 (68%)</td>
</tr>
</tbody>
</table>

*Note. Below Midpoint = Women Higher, More Innate; Above Midpoint = Men Higher, More Learned. \(*p < .05\). \(**p < .01. \(***p < .001.\)
2.7.2.2 Predicting Observed Gender Differences from Coder Ratings

Next, as preregistered and replicating Study 2, we examined the relationship between expert ratings and gender differences observed in Study 1.

2.7.2.2.1 Estimated Gender Difference

As in Study 2 and supporting H1, there was a main effect of expert ratings on observed gender differences, $\beta = .71, p < .001, 95\% \text{ CI } [.48, .95]$, such that experts accurately tracked the size of the gender difference on each activity. There was no interaction by subscale (systemizing vs. empathizing), $\beta = .19, p = .419, 95\% \text{ CI } [-0.28, .66]$.

2.7.2.2 Learned vs. Innate Attributions.

In Study 2, we found no evidence that gender differences would be larger for items that experts believe focus on learned (vs. innate) activities/behaviors. We reasoned, however, that experts might show this effect. However, similar to Study 2 and counter to H2, there was no significant effect of coder ratings on the observed gender difference, $\beta = .20, p = .057, 95\% \text{ CI } [-
and no significant interaction by subscale (systemizing vs. empathizing), $\beta = -.38, p = .077$, 95% CI [-.79, .04].

2.7.2.2.3 Gendered Learning Affordances

As in Study 2 and supporting H3, there was a significant main effect of gendered learning affordances on the observed gender difference in Study 1, $\beta = .58, p < .001$, 95% CI [.36, .80] (Figure 2.5; beta coefficients represent simple slopes for systemizing and empathizing separately. Axis bands represent a gender difference with $[d > .30; \text{blue}]$ or women scoring higher $[d < -.30; \text{red}]$ in Study 1). Systemizing activities experts rated as providing more affordances to men were also those that men scored higher on; empathizing activities experts rated as providing more affordances to women were also those that women scored higher on. Unlike in Study 2, there was a significant interaction by subscale (systemizing vs. empathizing) $\beta = .45, p = .042$, 95% CI [.02, .89], revealing that this effect was stronger for systemizing ($\beta = .80, p < .001$, 95% CI [.49, 1.11]) than empathizing ($\beta = .35, p = .026$, 95% CI [.04, .66]).

---

14 When modeled with MLM, the main effect of coder ratings on the observed gender difference is significant (see Appendix).
15 Preregistered simple slopes analyses showed the relationship between expert ratings and observed gender differences was not significant for systemizing ($\beta = .01, p = .872$) and significant in the opposite expected direction for empathizing ($\beta = .39, p = .042$). As in Study 2, empathizing activities that experts judge to be less learned through experience were those that women scored higher on.
2.7.2.2.4 Genetic Advantages

Using the revised measure of perceived genetic advantages in Study 3, we also found support for our exploratory H4. There was a main effect of expert ratings of genetic advantages on the observed gender difference, $\beta = .51, p < .001, 95\% \text{ CI } [.25, .77]$, and a significant interaction by subscale (systemizing vs. empathizing), $\beta = .60, p = .025, 95\% \text{ CI } [.08, 1.12]$. Systemizing items rated by experts as providing a genetic advantage to men were also those that men score higher on ($\beta = .81, p < .001, 95\% \text{ CI } [.37, 1.24]$), but this relationship was not significant for empathizing items ($\beta = .21, p = .146, 95\% \text{ CI } [-.07, .49]$).
2.7.2.3 Testing Competing Predictors: Learning Affordances vs. Genetic Difference

Finally, we preregistered that, if both gendered learning affordances and genetic advantages emerged as significant predictors of the observed gender difference in Study 1, we would test these two variables as competing predictors. Since we found support for both predictors among systemizing activities, we tested a model where gendered learning affordances and genetic advantages were entered as simultaneous predictors of the observed gender difference on systemizing items only in Study 1.\(^{16}\) The two predictors were moderately correlated, \(r(23) = .56, p = .003\). Perceptions of having gendered learning affordances emerged as a significant predictor (\(\beta = .57, p = .003, 95\% \text{ CI } [.21, .93]\)), whereas the relationship for perceived genetic advantage became nonsignificant (\(\beta = .25, p = .158, 95\% \text{ CI } [-.11, .61]\)).\(^{17}\)

2.7.3 Discussion

As a preregistered replication and extension of Study 2, the results of Study 3 provide a stronger test of the degree to which measures of systemizing and empathizing assess gender differences that are perceived to reflect learning affordances. Like lay coders, experts rated systemizing and empathizing activities as more likely to be learned through experience, rather than innate abilities. Notably, experts believed that men have significantly more affordances to learn every systemizing activity, and unlike in Study 2, they also indicated that women have significantly more affordances to learn empathizing activities. Supporting our preregistered hypothesis and replicating Study 2, activities that experts rate men and women as having different affordances on are also the activities that show the largest gender differences in Study

\(^{16}\) Although we preregistered modeling all items with subscale as a moderating factor, analyses revealed item-level average ratings of gendered and genetic advantage were highly correlated, \(r(45) = .93\). Since the systemizing items showed a more reasonable correlation magnitude and demonstrate the strongest effects, we report these analyses. Results are robust to either methodological decision.

\(^{17}\) When modeled in MLM, although the ratio of magnitude between predictors is similar (gendered learning affordances: \(b = .07, p < .001\); genetic advantage: \(b = .03, p = .013\)), genetic advantage also emerged as a significant predictor.
1. For systemizing, perceived difference in affordances predicted observed differences more than perceived genetic advantages. As hypothesized, experts were able to accurately estimate which systemizing and empathizing activities show the greatest gender differences.

One unexpected finding in our data is that empathizing activities coders rated as being relatively less learned through experience were also those that women tended to score higher on. Although suggestive of a somewhat more biologically based theory of empathizing, the interaction testing the difference in slopes was only marginally significant and expert ratings of women’s genetic advantages on empathizing activities did not predict gender differences. Given this, we hesitate to draw strong conclusions from this mixed pattern of findings.

Taken together, the results of Study 3 suggest that large gender differences observed on the SQ- and EQ-Short scales are related to experts’ perceptions of men’s and women’s different affordances to learn activities referenced in the items. Importantly, there is little evidence that even highly established experts in the field believe most items on these scales capture skills that are more innate than learned. Although most experts likely endorse more complex interactional influences on complex behavior, it is notable that each and every activity on the systemizing scale was seen as affording more learning opportunities to men. A belief that the SQ-Short measures an innate drive to construct and analyze systems typical of a “male brain” would seem to be out of step with a consensus view from psychological science.

2.8 General Discussion

Gender differences measured using the SQ and EQ are often cited as evidence for innate sex differences in systemizing and empathizing (Archer, 2019), consistent with theorizing about essentialized male and female brains (Baron-Cohen, 2002, 2004, 2008, 2009). Although prior revisions have attempted to debias these measures, no research to date has addressed whether the
activities assessed in these scales are seen as valid indicators of gender differences that are perceived to be innate or socially learned. Accordingly, this work set out to address two questions: (1) Do the SQ- and EQ-Short ask about activities perceived to reflect men and women’s innate differences and/or different learning affordances? (2) Do perceived innate differences or learning affordances better predict the size of the gender difference observed on the SQ- and EQ-Short items?

First, the items used to measure systemizing are especially geared toward activities perceived as learnable and for which men are seen as having a greater opportunity to learn. Across both coding studies, lay coders and experts rated the majority (96% in Study 2, 80% in Study 3) of systemizing activities as being more reflective of interests that are learned through experience than innate ability. They also consistently estimated that men have significantly more affordances to learn activities referenced in the SQ-Short. In Study 3, but not Study 2, experts perceived that women have significantly more affordances to learn activities referenced in the EQ-Short.

Turning to the relationship between perceived learning affordances and observed gender differences, results supported preregistered predictions. Activities that expert (and lay) coders assume men have more affordances to learn were the items men scored higher on in Study 1. These effects were sometimes stronger for systemizing and replicated when we analyzed data by subdisciplines (see Appendix A). Importantly, there was little indication that experts or lay coders believe sex-based genetic advantages better explain observed differences in systemizing and empathizing. Assuming the ratings across Studies 2 and 3 capture a wisdom of crowds regarding these skills and how they develop, the gender difference on self-report measures of systemizing (and to a lesser degree, empathizing) can be said to reflect a consensus perception of
men and women’s different opportunities to learn these activities, perhaps more than perceived genetic advantages.

2.8.1 Implications

E-S theory’s claims of male and female brains have far-reaching consequences outside of the ivory tower. As reviewed in the Introduction, claims about inherent gender differences on systemizing and empathizing have the potential to fuel gender discrimination, undermine women’s sense of fit and belonging, and guide gendered career selections. Given these repercussions, a responsible science must closely examine the evidence used to support these claims.

Our results suggest measurement of systemizing might often be biased by activities that men are seen as having more opportunities than women to learn. Similar, albeit weaker, effects were observed for measurement of empathizing. Our work extends prior efforts to debias the SQ (i.e., adding systemizing items in feminine domains in the SQ-R, Wheelwright et al., 2006; removing items that function differently by gender, Allison et al., 2015) by empirically pinpointing perceived affordances in SQ and EQ items as a possible explanation for the size of the observed gender difference. A self-report measure that is not biased by perceived affordances would need to assess systemizing entirely in the context of activities that men and women have relatively equal opportunities to learn. Similar efforts have been taken in development of career interests (Su, Rounds, & Armstrong, 2009). Until that time, measured effect sizes on the SQ in particular should not be interpreted as evidence of innate sex differences.

Beyond informing the SQ and EQ measures, our findings illuminate areas of practical focus for researchers. Subsequent efforts to develop self-report measures might incorporate a step in the validation process that aims to equate items on the degree to which they provide
learning affordances across gender. Possible strategies to this end could include: (1) reflecting on one’s own positionality during item generation, (2) crowdsourcing wisdom from diverse groups to yield a large pool of potential items, and/or (3) using more abstract items that focus on the process rather than specific activities tied to learning affordances. Most importantly, one must recognize that observed differences between men and women on any self-report measure cannot provide clear evidence of the origin of those differences.

2.8.2 Limitations and Future Directions

2.8.2.1 Difficulty of Determining Etiology

One key limitation is the difficulty of determining the true etiology of gender differences on any construct. Any true variation on systemizing and empathizing is likely due to complex interactions between biological and social factors. Because our goal was to assess the legitimacy of claims that the SQ and EQ could assess innate sex differences, we asked coders to consider only the relative strength of genetic and environmental affordances; we did not ask about more complex causal forces. Although the wisdom of crowds approach suggests that diverse and expert coders can accurately estimate true effects of gender differences, their causal explanations for these estimates might still be biased by their own point of view. For example, even experts in human behavior had varied perceptions of the causal forces at work, as suggested by the evidence that expert coders’ explanations for gender differences (whether learned or genetic) differed by subdiscipline in Study 3 (see Appendix B). Our wisdom of crowds approach was designed to balance out these individual biases and errors, but we acknowledge such estimates cannot be assumed to reflect what is likely a complex interplay of nature and nurture in shaping people’s interests and abilities. Future research could perhaps consider more complex understanding of interactional effects.
2.8.2.2 Constraints on Generalizability

To bolster the replicability of our work, we acknowledge the contextual and population-level factors that likely present boundary conditions for our work (Simons et al., 2017). First, given that gender stereotypes tend to change over time (Charlesworth & Banaji, 2022) and are situated within culture (Miller, Eagly, & Lin, 2015), we might not expect the content of ratings (i.e., estimated gender differences on specific activities) to replicate beyond this time period nor cultural context. Another boundary condition presented by our work is our inability to speak to these effects as they apply beyond gender as a single-axis identity. Past work on intersectionality reveals that gendered phenomena often vary across the intersection of race, class, sexual orientation, and other social identities (Crenshaw, 1989). Additionally, since our focus is on men and women, we are unable to speak to whether effects would replicate beyond these binary gender identities (Morgenroth & Ryan, 2018). Scholars might consider these boundary conditions as important directions for further work on this topic.

2.8.2.3 Alternative Sources of Gender Bias in the Activities

Our work concludes that gender differences on EQ and SQ activities are magnified to the degree that they afford learning opportunities to men and women. This provokes a further question of whether this effect is more strongly driven by prescriptive norms, things that society generally believes men and women ought to be (Rudman & Glick, 2001), or proscriptive norms, things that society generally believes men and women ought not to be (Vandello & Bosson, 2013). For example, women might receive more opportunities than men to learn empathizing activities because: (1) women are encouraged to learn these activities via prescriptive norms, and/or: (2) men are actively discouraged from learning these same activities via proscriptive norms. Future research might work to disentangle these distinct contributors.
2.8.3 Conclusion

The SQ and EQ self-report measures have been employed for nearly two decades to provide evidence for biologically based accounts of sex differences on systemizing and empathizing. Although prior revisions to the SQ and EQ have addressed sources of gender bias in the measures, no work to date has addressed whether gender differences on the SQ and EQ are magnified by men and women’s different affordances to learn skills referenced in the measures. By zeroing in on affordances as a central source of gender bias in self-report measures of systemizing and (to a lesser degree) empathizing, researchers may move toward a more complete understanding of these constructs and how best to measure them.
Chapter 3: Are Essentialized Explanations for Gender Differences in Interest Self-Fulfilling?

Despite recent progress toward gender equity, occupational gender segregation persists across and within occupations (Charles & Grusky, 2004). Across occupations, women are underrepresented in Science, Technology, Engineering, and Math (STEM; Cheryan et al. 2017; Schmader, 2023) and men are underrepresented in Healthcare, Early Education, and Domestic roles (HEED; Croft, Schmader, & Block, 2015). Within occupations, men and women also take on different roles in fields such as software engineering (Campero, 2021), medicine (Ku, 2011), law (Kay & Gorman, 2008), and domestic labor (Biernat & Wortman, 1991). Achieving more balanced occupational gender representation would not only promote gender equity, but could mitigate labor shortages (Croft, Schmader, & Block, 2015), boost innovation (Galinsky et al., 2015), increase role models (Dasgupta & Asgari, 2004), and narrow wage gaps (Goldin, 2021).

Gender differences in interests are often cited as a primary driver of men and women’s different occupational choices (Ceci, Williams, & Barnett, 2009, 2011). One large meta-analysis of 47 career interest inventories reported a large gender difference in people’s relative interest in careers involving people versus things \( (d = .93, N = 503,188; \text{Su, Rounds, & Armstrong, 2009}) \). Although many assume gender differences in interest reflect ‘free choice’ (Ceci et al., 2009) or innate preferences (Baron-Cohen, 2009), we consider how interests are systematically shaped by affordances in the environment (Schmader, 2023). To this end, we report three studies using a novel behavioral paradigm to demonstrate that people’s essentialized beliefs about gender differences in interest lead them to provide self-fulfilling affordances that confirm those beliefs.

Distinct from gender stereotypes about ability, gender stereotypes about interest are prevalent and are often essentialized. By 8 years of age, children stereotype boys as more
interested in computer science and engineering than girls (Master et al., 2021; see also Cvencek et al., 2011). Separate from the accuracy of these stereotypes is the tendency to attribute these differences more to biologically essentialized or sociocultural factors (i.e., gender essentialism, Lee, Reis, & Rogge, 2020). Gender essentialism beliefs vary among lay perceivers and academic psychologists alike. For instance, empathizing-systemizing theory (Baron-Cohen, 2002, 2004) proposes women are innately interested in empathizing (tuning into another’s feelings or thoughts; Baron-Cohen & Wheelwright, 2004) and men are innately interested in systemizing (analyzing underlying rules governing systems; Baron-Cohen et al., 2003, see also Archer, 2019). And yet, in Chapter 2, our findings suggest that self-report measures used to support these claims might partly reflect gender differences in learning affordances more than genetic advantages.

Guided by self-fulfilling prophecy as a theoretical framework (Merton, 1948; Jussim, 1986), we reasoned those who primarily endorse biologically essentialized (vs. sociocultural) explanations for gender differences in interest might, ironically, be the most likely to provide social learning affordances that deepen gender divides in occupational interests. Past research has linked psychological essentialism to greater stereotyping and prejudice (Bastian & Haslam, 2006; Haslam & Levy, 2006; Lee et al., 2020) and justification of the status quo (Lee et al., 2020; Brescoll et al., 2013; Williams & Eberhardt, 2008). However, clear experimental evidence linking gender essentialism to behavioral affordances is lacking. Traditionally, self-fulfilling prophecy research has focused on how perceivers’ ability stereotypes predict target performance, yielding only small and inconsistent effects (Jussim, Eccles, & Madon, 1996; Jussim & Harber, 2005). Yet more recent work demonstrates self-fulfilling affordances can constrain behavior when they accumulate across perceivers (Madon et al., 2018). Building on this work, we tested a
causal process through which: (1) perceivers believe empathizing and systemizing interest are primarily biologically (vs. socially) determined, (2) perceivers provide different learning affordances to women and men based on these beliefs, and (3) targets respond to perceivers’ affordances in belief-confirming ways (Figure 3.1).

Figure 3.1. Self-fulfilling process through which perceivers’ essentialized beliefs shape targets’ interests via situational affordances. Research relevant examples are provided in italics.

Additionally, we propose and test *affordance strategy* as a mechanism between interest beliefs and situational affordances (Figure 3.1, path *a*). That is, are perceivers who believe gendered interests are primarily biologically (vs. socially) determined more likely to say they provide targets with opportunities that match their inherent interests (vs. allow them to develop new interests)? Does affordance strategy, in turn, predict affordance behavior (i.e., providing gendered vs. counter-gendered opportunities to men and women)? This builds on prior work that conceives of immutability as a unique facet of gender essentialism, alongside biological and social determinism (Lee et al., 2020). In their measurement of gender essentialism, these authors find the factor corresponding to immutability is positively correlated with the factor
corresponding to biological determinism ($r = .63$, $N = 2,996$) and negatively correlated with the factor corresponding to social determinism ($r = -.28$, $N = 2,996$). As discussed in Chapter 1, Lee and colleagues’ (2020) formulation of immutability refers to immutability of gender identity. Here, our affordance strategy measure taps into perceivers’ beliefs about the immutability versus a growth mindset orientation toward gender expression (i.e., interests), by assessing a tendency to match targets to environments that afford their inherent interests versus helping them develop new interests. Layered onto the assumption that gendered interests are immutable versus malleable, our affordance strategy measure primarily taps into the motivation to inhibit or facilitate changing interests. Thus, this construct also builds on prior work looking at the relation between essentialism and system justification beliefs (Lee et al., 2020; Brescoll et al., 2013; Williams & Eberhardt, 2008). Finally, we hypothesized that affordance strategy, as with biological (vs. social) determinism, should operate independently of perceivers’ estimation of the size of the gender gap in abilities or interests. Thus, we control for these variables statistically and experimentally in studies measuring this construct.

3.1 Overview of Studies

Three studies employed a novel behavioral paradigm to test the process depicted in Figure 3.1 (preregistrations and materials for all studies available at osf.io/ae3fr). First, we tested whether measured (Study 4) and manipulated (Study 5) essentialized explanations for gender differences in interest predict gendered situational affordances (Figure 3.1, path $a$). Study 6 employed a quasi-double randomization design (MacKinnon et al., 2007; Madon et al., 2018; Word et al., 1974) to test whether targets who receive biased affordances (based on observed effect sizes in Study 4) endorse different interests than those who receive non-biased affordances (Figure 3.1, path $b$). We tested these questions in a managerial role-playing task where STEM
professionals assigned interns to client-facing and technical roles in a tech company, given tech is a male-dominated work setting where essentialized beliefs about empathizing and systemizing could shape these assignments (Cheryan et al., 2017; Damore, 2017; Jungert et al., 2019).

3.2 Study 4

The goal of Study 4 was to test whether people’s explanations for gender differences in interest (primarily biologically or socially determined, measured as a difference score), distinct from their knowledge of those differences, predict the affordances they provide. We developed a behavioral paradigm in which STEM professionals first reported their attributions for gender differences in interest before completing a mock management task requiring them to assign women and men interns (first-year university students participating in an intern exchange program) to teams responsible for empathizing (4 trials) or systemizing (4 trials) tasks.

3.2.1 Method

3.2.1.1 Participants

\( N = 285 \) STEM professionals received £2.15 to participate in an online study called “Project Manager Study.” As shown in Table 3.1, the majority of participants in Study 4 were men (70.10%) and White (59.30%).

<table>
<thead>
<tr>
<th>In a Manager Role</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>141 (49.8%)</td>
</tr>
<tr>
<td>No</td>
<td>142 (50.2%)</td>
</tr>
<tr>
<td>Gender</td>
<td>N (%)</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Man</td>
<td>199 (70.10%)</td>
</tr>
<tr>
<td>Woman</td>
<td>80 (28.20%)</td>
</tr>
<tr>
<td>Non-binary</td>
<td>5 (1.80%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Racial/Ethnic Background</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>169 (59.30%)</td>
</tr>
<tr>
<td>East Asian</td>
<td>32 (11.23%)</td>
</tr>
<tr>
<td>Hispanic / Latinx / Central or Spanish Origin</td>
<td>23 (8.07%)</td>
</tr>
<tr>
<td>African American, African, or Black</td>
<td>19 (6.67%)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>17 (5.96%)</td>
</tr>
<tr>
<td>South Asian</td>
<td>12 (4.21%)</td>
</tr>
<tr>
<td>Southeast Asian</td>
<td>5 (1.75%)</td>
</tr>
<tr>
<td>American Indian, Indigenous, or Alaska Native</td>
<td>2 (0.70%)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>1 (0.35%)</td>
</tr>
<tr>
<td>Middle Eastern or North African</td>
<td>1 (0.35%)</td>
</tr>
<tr>
<td>Not listed</td>
<td>3 (1.05%)</td>
</tr>
<tr>
<td>Missing response</td>
<td>2 (0.70%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SES (1 = Highest SES, 10 = Lowest SES)</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.00 (1.40)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35.40 (10.00)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Political Orientation (1 = Extremely Liberal, 7 = Extremely Conservative)</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.10 (1.70)</td>
</tr>
</tbody>
</table>
3.2.1.2 Procedure and Measures

Participants were asked to imagine they had been hired as a project manager at a fictitious tech company (“Pixlio”) and completed a manager onboarding task in which they were asked to read and answer questions about two articles, ostensibly selected from a larger pool of topics. After reading a filler article, participants read an article that presented gender differences in interest as an explanation for women’s underrepresentation in tech. Participants answered several questions reflecting on the article, including measures of biological determinism (5 items, $\alpha = .95$) and social determinism (5 items, $\alpha = .88$) adapted from Lee et al. (2020; e.g., biological determinism: “Many forms of gender-related interests are biologically determined;” social determinism: “The social background a person comes from is strongly reflected in the development of the person’s gender-related interests;” see Appendix B for materials and measures).

After this onboarding task, participants completed a behavioral task where they assigned a group of 6 interns (3 men, 3 women) to project teams (4 empathizing, 4 systemizing). In Study 4, participants were randomly assigned to see 1 of 2 possible sets of teams (Figure 3.2; results are robust to controlling for team set). For 4 of the 8 teams, depending on which set they received, participants saw either an empathizing (e.g., Communication Oversight) or systemizing (e.g., Information Oversight) version of each team (details on team development and pilot testing reported in Appendix B).
After learning about their project teams, participants read the following information:

A local university has partnered with Pixlio as part of a program for first-year students.

The program places students in a series of 4-week internships to get a variety of work experience across different fields. Students in this program come from a variety of backgrounds and are not necessarily interested in tech. 6 students from this program are starting an internship with Pixlio this week.

Below this statement, each intern was presented via a photo and quote about the internship program (quote order counterbalanced; see Appendix B). During each trial of the task, participants were presented with a project team (e.g., People Support) and asked to choose

---

18 E = Empathizing team, S = Systemizing Team. Participants were randomly assigned to team Set 1 or Set 2 (2 possible sets total). Across sets, the first four teams were either an empathizing or systemizing version of the same job (e.g., Communication Oversight vs. Information Oversight). The last four teams were identical across sets.
between 1 of 2 students to work on a task for this team. Participants repeated these selections for each project team across \( N = 8 \) same gender filler trials and \( N = 8 \) man-woman critical trials, for a total of \( N = 16 \) trials (example of a critical trial provided in Figure 3.3).\(^{19}\) After assigning interns to teams, participants responded to questions about their task strategy, completed comprehension checks, provided demographic information, and responded to a suspicion check. Finally, participants were thanked, debriefed, and compensated for their time.

![Client Strategy](Image)

Icon Credit (Client Strategy): Thinking © ArtWorkLeaf; retrieved from the Noun Project and licensed under CC-BY 3.0; icon was modified by the lead author from black and white to color icon. Intern images used with permission from Chicago Face Database (Ma et al., 2015).

**Figure 3.3. Example of a critical trial in the team assignment task (Study 4)**

### 3.2.2 Results

We preregistered testing hypotheses in a multilevel framework given that team assignments are nested within person, ICCs testing non-independence of observations were less than .05, indicating a GLM framework is preferred for these data (Thomas & Heck, 2001). Thus, we report standardized betas modeled in GLM. All estimates were identical to those obtained

\(^{19}\) We developed teams and the behavioral task used in Studies 4-5 based on results across 3 pilot studies (\( N = 605 \) total; see Appendix B for details).
when modeling the data using MLM (see Appendix B). In the preregistration for Study 5, we specified this ICC threshold as a decision rule for analyses.

3.2.2.1 Men Endorsed More Essentialist Beliefs Than Women and Nonbinary Participants

Compared to the combination of women (n = 80) and nonbinary participants (n = 5) 20, men (n = 199) attributed gender differences in interest more to biological (t(152.23) = 3.86, p < .001, d = .51) and less to social factors (t(178.23) = -3.19, p = .002, d = .40; Table 3.2). To create a measure of primary belief, we subtracted social determinism scores from biological determinism. There were no significant gender differences in participants’ explicit beliefs that they had assigned interns to tasks by matching interns’ natural interests: t(160.95) = 1.79, p = .075, d = .23; or helping interns’ develop new interests: t(153.83) = -0.56, p = .573, d = .07).

20 Given that our goal in this analysis was to compare responses between perceivers of the default gender identity in STEM (i.e., men; Cheryan & Markus, 2020; Schmader, 2023) and perceivers of non-default gender identities in STEM (i.e., women and non-binary people), we combined women and nonbinary participants into a single perceiver group (see De Souza & Schmader, 2022, for another example of this analytic approach).
Table 3.2. Descriptive statistics by gender (Study 4)

<table>
<thead>
<tr>
<th></th>
<th>Women &amp; Nonbinary (N = 85)</th>
<th>Men (N = 199)</th>
<th>Effect Size for Difference (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biological Determinism</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5 items, α = .95)</td>
<td>3.21 (1.65)</td>
<td>4.03 (1.57)</td>
<td>.51***</td>
</tr>
<tr>
<td><strong>Social Determinism</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5 items, α = .88)</td>
<td>5.26 (1.00)</td>
<td>4.83 (1.13)</td>
<td>.40**</td>
</tr>
<tr>
<td>Range: 1.4–7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biological vs. Social Determinism</strong></td>
<td>-2.05 (2.28)</td>
<td>-0.80 (2.25)</td>
<td>.55***</td>
</tr>
<tr>
<td>Range: -6–5.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Matched Natural Interests</strong></td>
<td>4.20 (1.75)</td>
<td>4.61 (1.78)</td>
<td>.23</td>
</tr>
<tr>
<td><strong>Helped Develop Interests</strong></td>
<td>4.56 (1.76)</td>
<td>4.44 (1.70)</td>
<td>.07</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05. **p** < .01. ***p*** < .001. For Biological vs. Social Determinism, we took the difference score between (Biological – Social); higher scores indicate favoring biological determinism; lower scores indicate favoring social determinism. All variable responses ranged from 1-7 unless otherwise specified.

### 3.2.2.1.1 People Generally Provide Gendered Affordances

First, as preregistered, we specified a binomial regression model predicting odds of selecting a woman (vs. man) intern from participants’ primary endorsement of biological or social determinism, moderated by team type (empathizing vs. systemizing). Supporting our primary hypothesis, at mean levels of biological vs. social determinism, participants generally provided gendered affordances by populating empathizing teams with more women (than men) interns, $\beta = -0.86$, $p < .001$ and systemizing teams with more men (than women) interns, $\beta = 0.86$, $p < .001$.  

3.2.2.1.2 People Who Primarily Endorse a Biological (Vs. Social) Explanation Provide More Gendered Affordances

In the same model specified above, also supporting our primary hypotheses, explanation significantly interacted with team type (empathizing vs. systemizing) to predict learning affordances, $\beta = -0.50, p < .001$ (Figure 3.4; bands around regression lines represent +/-1 standard error). Greater endorsement of biological (vs. social) determinism corresponded to more gendered learning affordances (more empathizing team assignments to women interns, $\beta = .20, p = .001$; more systemizing team assignments to men interns, $\beta = -.30, p < .001$). Effects were robust to controlling for participants’ general perceptions of gender differences in ability and interest, participant gender, and participant suspicion (see Appendix B). Thus, effects are driven not by the perceptions of gender differences in ability or interest, but by people’s beliefs about the essential nature of those differences.

Figure 3.4. Relationship between explanation for gender difference in interest and probability of choosing a woman intern, moderated by team type (empathizing vs. systemizing; Study 4)
3.2.2.2 The Relationship Between Explanation and Affordances is Mediated by Assignment Strategy

Finally, we tested a moderated mediation model predicting intern choice from explanation for interest difference (primarily biological vs. social), mediated by assignment strategy (matched interns to inherent interests vs. helped them develop new interests, measured as a difference score), with paths $b$ and $c$ moderated by team type (empathizing vs. systemizing; see Figure 3.5; coefficients are unstandardized). Supporting our secondary hypothesis, there was a significant indirect effect of task strategy for both empathizing and systemizing trials (for empathizing: $a*b = .005$, $p < .001$, for systemizing: $a*b = -.008$, $p < .001$). Participants who endorsed a biological (vs. social) explanation for interest differences tended to match interns to inherent interests (vs. help them develop new interests; $a$ path). Matching interns to inherent interests (vs. helping them develop new interests), in turn, predicted choosing more women for empathizing, and more men for systemizing, tasks ($b$ path). Like the results above, these indirect effects were robust to controlling for participants’ general perceptions of gender differences in ability and interest, participant gender, and participant suspicion.
3.2.3 Discussion

Study 4 provided initial evidence that STEM professionals’ gender essentializing beliefs relate to their tendency to provide gender biased learning affordances to women and men interns and is mediated by a desire to match interns to their presumed inherent interests. In contrast, those who endorsed more sociocultural explanations for interest differences made gender unbiased assignments. However, the quasi-experimental paradigm in Study 4 precludes causal inference. Thus, to test whether essentialism beliefs cause affordances, Study 5 manipulated gender essentialist beliefs before participants completed the mock managerial paradigm.

3.3 Study 5

In Study 5, we tested whether participants randomly assigned to read an article priming a biological (vs. social) explanation for gender differences in STEM interest provided more gendered affordances to men and women interns. In Study 5, we decided to focus on men for two reasons: (1) as the default cultural group in STEM, men play a particularly important role in
gatekeeping women’s opportunities (Cheryan & Markus, 2020; Schmader, 2023), and (2) in Study 4, men demonstrated a more equivalent endorsement of biological and social determinism compared to women and non-binary participants, and thus might be more sensitive to an experimental prime.

3.3.1 Method

3.3.1.1 Participants

\( N = 379 \) men with management experience in male-dominated fields (e.g., engineering, construction) received £2.15\(^{21}\) for their participation. As shown in Table 3.3, like in Study 4, most participants were White (64.38%). Additionally, most participants (61.4%) tended to currently occupy a manager role (though all had prior management experience).

Table 3.3. Sample demographics (Study 5)

<table>
<thead>
<tr>
<th>In a Manager Role</th>
<th>( N ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>231 (61.4%)</td>
</tr>
<tr>
<td>No</td>
<td>139 (37%)</td>
</tr>
<tr>
<td>Not sure</td>
<td>6 (1.6%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Racial/Ethnic Background</th>
<th>( N ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>244 (64.38%)</td>
</tr>
<tr>
<td>African American, African, or Black</td>
<td>30 (7.92%)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>27 (7.12%)</td>
</tr>
<tr>
<td>East Asian</td>
<td>24 (6.33%)</td>
</tr>
</tbody>
</table>

\(^{21}\) We initially planned to recruit men in STEM to match Study 4. Due to low availability of men in this participant pool, we expanded our criteria to male-dominated fields to capture a broader sample. To maximize realism of the study scenario, participants in Study 5 were required to have managerial experience. Finally, for a portion of participants, we later increased the payment rate to £2.78, as the study time exceeded our initial estimate.
### Racial/Ethnic Background (Cont.)

<table>
<thead>
<tr>
<th>Racial/Ethnic Background</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Asian</td>
<td>23 (6.07%)</td>
</tr>
<tr>
<td>Hispanic / Latinx / Central or Spanish Origin</td>
<td>19 (5.01%)</td>
</tr>
<tr>
<td>Southeast Asian</td>
<td>3 (0.79%)</td>
</tr>
<tr>
<td>Middle Eastern or North African</td>
<td>3 (0.79%)</td>
</tr>
<tr>
<td>American Indian, Indigenous, or Alaska Native</td>
<td>2 (0.53%)</td>
</tr>
<tr>
<td>Not listed</td>
<td>1 (0.26%)</td>
</tr>
<tr>
<td>Missing response</td>
<td>3 (0.79%)</td>
</tr>
</tbody>
</table>

### SES (1 = Highest SES, 10 = Lowest SES)

<table>
<thead>
<tr>
<th>SES</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.10 (1.60)</td>
</tr>
</tbody>
</table>

### Age

<table>
<thead>
<tr>
<th>Age</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>37.60 (10.90)</td>
</tr>
</tbody>
</table>

### Political Orientation (1 = Extremely Liberal, 7 = Extremely Conservative)

<table>
<thead>
<tr>
<th>Political Orientation</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.40 (1.60)</td>
</tr>
</tbody>
</table>

#### 3.3.1.2 Procedure and Measures

Study 5 employed a modified version of the same paradigm in Study 4. In Study 5, we adapted the manager onboarding task so that participants were randomly assigned to read and summarize an article ostensibly selected from a larger pool of topics (see Appendix B for article stimuli). In actuality, this article comprised the manipulation of biological or social determinism. Unlike in Study 4, there was no filler article in Study 5. After completing an article comprehension check, participants responded to an abbreviated measure of biological determinism (2 items, $r = .86$) and social determinism (2 items, $r = .82$), based on items that showed the highest factor loadings in Study 4 (see Appendix B).
Next, participants completed a modified version of the team assignment task used in Study 4. In Study 5, we reduced the number of project teams from 8 to 4 based on the teams that showed the strongest effects in Study 4 (2 empathizing: Client Strategy, People Support; 2 systemizing: Database Security, Server Admin) and only used one set of teams. Each participant completed eight trials (four critical man-woman selections; four filler same-gender selections) where each of the four project teams appeared twice, and their task was to select one of the two interns for the team. After the task, participants responded to questions about their task strategy, general perceptions of ability and interest differences, and article perceptions. They then provided demographic information and responded to a suspicion check. Finally, participants were thanked, debriefed, and compensated for their time.

3.3.2 Results

3.3.2.1 Articles Changed Men’s Endorsement of Biological versus Social Determinism

The article manipulation successfully influenced men’s beliefs; men in the biological (vs. social) article condition endorsed more biological determinism ($t(356.43) = 12.66, p < .001, d = 1.30$) and less social determinism ($t(335.46) = -9.53, p < .001, d = .99$; Table 3.4). Condition had no significant effect on task assignment strategy (matched to natural interests: $t(374.33) = 1.18, p = .239, d = .12$; helped develop new interests: $t(371.37) = -1.93, p = .055, d = .20$).
Table 3.4. Descriptive statistics by condition (Study 5)

<table>
<thead>
<tr>
<th></th>
<th>Biological Article</th>
<th>Social Article</th>
<th>Effect Size for Difference (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M (SD)</strong></td>
<td><strong>M (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biological Determinism</strong></td>
<td>(2 items, ( r = .86 ))</td>
<td>5.09 (1.20)</td>
<td>3.25 (1.61)</td>
</tr>
<tr>
<td><strong>Social Determinism</strong></td>
<td>(2 items, ( r = .82 ))</td>
<td>4.06 (1.52)</td>
<td>5.38 (1.12)</td>
</tr>
<tr>
<td><strong>Biological vs. Social Determinism</strong></td>
<td>Range: -6–6</td>
<td>1.02 (2.28)</td>
<td>-2.13 (2.34)</td>
</tr>
<tr>
<td><strong>Matched Natural Interests</strong></td>
<td>Range: -6–6</td>
<td>4.88 (1.66)</td>
<td>4.68 (1.66)</td>
</tr>
<tr>
<td><strong>Helped Develop Interests</strong></td>
<td>4.22 (1.68)</td>
<td>4.54 (1.60)</td>
<td>.20</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05. **p** < .01. ***p** < .001. For Biological vs. Social Determinism, like in Study 4, we took the difference score between (Biological – Social); higher scores indicate favoring biological determinism; lower scores indicate favoring social determinism. All variable responses ranged from 1-7 unless otherwise specified.

3.3.2.1.1 In General, Men Provide Gendered Affordances

As in Study 4, we preregistered using a binomial regression model predicting odds of selecting a woman (vs. man) intern from condition (biological vs. social), moderated by team type (empathizing vs. systemizing). Replicating Study 4 and supporting our primary hypothesis, men provided gendered affordances: across conditions, they populated empathizing teams with more women than men interns, \( \beta = -1.61, p < .001 \) and systemizing teams with more men than women interns, \( \beta = 1.61, p < .001 \).
3.3.2.1.2 Men Primed with a Biological (vs. Social) Explanations Provide More Gendered Affordances

As hypothesized, in the same model above, condition significantly interacted with team type to predict learning affordances, $\beta = .70$, $p = .002$, such that men primed with a biological (vs. social) explanation provided more gendered learning affordances (Figure 3.6; error bars represent +/- 1 standard error). Decomposing this interaction into its simple slopes, those primed with a biological (vs. social) explanation afforded more systemizing opportunities to men than women, $\beta = .44$, $p = .005$. Although in the predicted direction, condition did not significantly affect affordances for empathizing teams, $\beta = -.26$, $p = .103$. Effects were robust to controlling for participants’ general perceptions of gender differences in ability and interest, agreement with the article they read, and suspicion (see Appendix B).

![Figure 3.6. Probability of choosing a woman intern by condition and team type (empathizing vs. systemizing; Study 5)](image-url)
3.3.2.2 The Relationship Between Explanation and Systemizing Affordances is Mediated by Assignment Strategy

Next, we tested a moderated mediation model predicting intern choice from condition, as mediated by assignment strategy, with paths $b$ and $c$ moderated by condition (Figure 3.7; coefficients are unstandardized). We found support for this secondary hypothesis; the relationship between condition and intern choice was mediated by task strategy (empathizing, $a*b = -0.02, p < .001$; systemizing: $a*b = 0.03, p < .001$). Participants primed with a biological (vs. social) explanation for interest differences reported matching interns to inherent interests (vs. helping them develop new interests; $a$ path). Replicating our findings in Study 4, matching interns to inherent interests (vs. helping them develop new interests) predicted choosing more men for systemizing, and more women for empathizing, teams ($b$ path).\textsuperscript{22} As in prior analyses, these indirect effects were robust to controlling for participants’ general perceptions of gender differences in ability and interest, agreement with the article they read, and suspicion.

\textsuperscript{22} We found mixed support for our other secondary hypothesis testing a parallel moderated mediation model with personal endorsement of biological vs. social determinism as the mediator (supported for systemizing but not empathizing; see Appendix B).
3.3.3 Discussion

Study 5 established that essentialist explanations for gender differences in interest (distinct from the perception that such differences exist) causally shape the learning affordances men provide in a STEM context. Notably, effects were strongest and significant for systemizing teams. We revisit potential explanations for the asymmetry between empathizing and systemizing affordances in the General Discussion. The goal of Study 6 was to test the second link of our theoretical model. That is, when a person is provided with biased affordances, do these translate into their own pursued interests and change in the way they see themselves?

3.4 Study 6

In Study 6, we tested how receiving affordances based on essentialist beliefs shaped targets’ own interests and self-perceptions. Using a quasi-double randomization design (MacKinnon et al., 2007; Madon et al., 2018; Word et al., 1974), we randomly assigned first- and second-year undergraduate women to receive gendered or counter-gendered affordances, based on the selections of perceivers in Study 4. Participants engaged in a mock internship task.
in which they were asked to choose between 1 of 2 team options provided to them by a project manager. Before and after the task, participants were asked to rate their interest in each of teams and reflect on their possible selves in the context of the company.

3.4.1 Method

3.4.1.1 Participants

\(N = 300\) first- and second-year undergraduate women received partial course credit (for those recruited through our Departmental Subject Pool) or £2.15 (for those recruited through Prolific Academic) to complete an online study called “Project Intern Study.” As shown in Table 3.5, the majority of participants in Study 6 were White (32.67%) and East Asian (21.67%).

<table>
<thead>
<tr>
<th>Racial/Ethnic Background</th>
<th>(N) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>98 (32.67%)</td>
</tr>
<tr>
<td>East Asian</td>
<td>65 (21.67%)</td>
</tr>
<tr>
<td>South Asian</td>
<td>32 (10.67%)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>24 (8%)</td>
</tr>
<tr>
<td>African American, African, or Black</td>
<td>23 (7.67%)</td>
</tr>
<tr>
<td>Hispanic / Latinx / Central or Spanish Origin</td>
<td>22 (7.33%)</td>
</tr>
<tr>
<td>Southeast Asian</td>
<td>13 (4.33%)</td>
</tr>
<tr>
<td>Middle Eastern or North African</td>
<td>7 (2.33%)</td>
</tr>
<tr>
<td>American Indian, Indigenous, or Alaska Native</td>
<td>1 (0.33%)</td>
</tr>
<tr>
<td>Not listed</td>
<td>2 (0.67%)</td>
</tr>
<tr>
<td>Missing response</td>
<td>13 (4.33%)</td>
</tr>
</tbody>
</table>
### 3.4.1.2 Procedure and Measures

Participants were asked to imagine they were participating a 4-week internship at a fictitious tech company, “Pixlio.” Participants learned about the same project and teams as participants in Study 5. As part of their internship, participants learned a project manager would create a personalized work assignment for them in real time (this feature of the design was deceptive; participants were always randomly assigned to premade work assignments corresponding to the two experimental conditions). To ostensibly help the project manager create their work assignment, participants responded to 16 filler items about their work style (6 items from the Individual Work Performance Questionnaire, Koopmans et al., 2013 and 10 items from the Ten-Item Personality Inventory, Gosling et al., 2003; see Appendix B).

While the project manager was creating their assignment, participants completed baseline measures of their possible selves and interests. Prior to answering these items, participants read the following instructions: “Think a minute about what your experience in the internship at Pixlio will be like, and what you might be like at this time” (adapted from Brown & Diekman, 2010). There were four different possible selves items: empathizing (“I could see myself in a role working with emotions and people at Pixlio”), systemizing (“I could see myself in a role working with systems and things at Pixlio”), high-status (“I could see myself in a high-status role at Pixlio”), and staying at the company (“I could see myself staying at Pixlio after my internship is done,” items adapted from Brown & Diekman, 2010). Next, participants rated their baseline team interest (“I would be interested in working with [team] at Pixlio”) for all four teams (two
empathizing, averaged into a single empathizing team interest pre-measure; two systemizing, averaged into a single systemizing team interest pre-measure).

Next, participants were asked to complete a selection task that actually comprised the experimental manipulation of affordances. In this task, participants completed 8 trials in which they were presented with two possible team options, ostensibly selected for them by their manager, and were asked to choose between them (example trial provided in Figure 3.8). In reality, team options were populated based on affordances of perceivers in Study 4. In the gendered affordances condition, the 16 team options presented to participants were populated based on the selections of perceivers who most endorsed biological essentialism (i.e., 2 SD’s above the mean on biological vs. social essentialism). As a result, participants in the gendered affordances condition were given 12 empathizing and 4 systemizing teams to consider. In the counter-gender affordances condition, the 16 team options were populated based on the selections of perceivers who most endorsed social essentialism (i.e., 2 SD’s below the mean on biological vs. social essentialism). Thus, participants in the counter-gender affordances condition were given 6 empathizing and 10 systemizing teams to consider. After making all 8 selections, to emphasize that their manager had recommended either gendered or counter-gendered affordances, participants received a visual summary of their team options (Figure 3.9).

23 When possible, team options were presented so that participants chose between an empathizing and systemizing team. Since the experimental manipulation included unequal empathizing and systemizing options across condition, in the gendered affordances condition, there were four empathizing-only choices and in the counter-gendered affordances condition, there were two systemizing-only choices (see Table B.28 for team pairings across all 8 trials).
Icon Credits (Left to Right): Thinking © ArtWorkLeaf; Server © Rolas Design. Both icons retrieved from the Noun Project and licensed under CC-BY 3.0; both icons were modified by the lead author from black and white to color icons.

Figure 3.8. Example of a trial selection (Study 6)

Icon Credits: People Support Icon, Team © HASTA ICON; Client Strategy Icon, Thinking © ArtWorkLeaf; Database Security Icon, Security © Setyo Ari Wibowo; Server Admin Icon, Server © Rolas Design. All icons retrieved from the Noun Project and licensed under CC-BY 3.0; all icons were modified by the lead author from black and white to color icons.

Figure 3.9. Visual summary in gendered (left) and counter-gendered (right) conditions (Study 6)
After making their selections, participants completed a post measure of their interest in each of the 4 project teams (2 empathizing, 2 systemizing; same measures as described previously) and possible selves (empathizing, systemizing, high-status, staying at the company; same measures as described previously), rated how well their project manager created their work assignment, responded to a suspicion check, and answered questions about their personal background (trait empathizing and systemizing, gender expression, demographics; items for all measures provided in Appendix B). After these questions, participants were thanked, debriefed, and dismissed.

3.4.2 Results

For each variable, we computed a difference score to capture people’s relative endorsement for empathizing (E) vs. systemizing (S) as follows:

\[ ES = S - E \]

For this computation, positive scores indicate \( S > E \), zero indicates \( S = E \), and negative scores indicate \( E > S \). This is in line with the analytic approach and conceptual framework of E-S Theory (Baron-Cohen, 2002, 2004, 2008, 2009).

Next, for variables with a baseline (T0) and post-task (T1) measure, to capture change in relative E-S endorsement, we subtracted participants’ post-measure E-S scores from their baseline E-S scores\(^{24}\) as follows:

\[ \Delta ES = ES_{T0} - ES_{T1} \]

For this computation, positive scores indicate change in favor of systemizing, zero indicates no change, and negative scores indicate change in favor of empathizing. This analytic approach provided a high-powered, parsimonious technique for testing central predictions (analyses

\(^{24}\) Computing change across time using residuals in place of a difference score did not change results (see Appendix B).
examining empathizing and systemizing separately as specified in the original preregistration are provided in Appendix B). Effects were not moderated by participants’ own gender expression (masculinity, femininity) and were robust to controlling for suspicion (see Appendix B).

3.4.2.1 Affordances Changed Women’s Interest in Empathizing vs. Systemizing

Supporting the expectation that affordances would shape women’s interests, women who received gendered (vs. counter-gendered) affordances reported an increased interest in empathizing vs. systemizing teams, \( t(293.40) = -5.69, p < .001, d = .66 \) (Table 3.6, Row 1c and Figure 3.6). Within each condition, women who received gendered affordances significantly changed their interests toward empathizing (\( M = -0.40, SD = 1.05; \) test against zero, \( t(146) = -4.60, p < .001, d = .38 \)), whereas women who received counter-gendered affordances significantly changed their interests toward systemizing (\( M = +0.33, SD = 1.15; \) test against zero, \( t(149) = 3.50, p < .001, d = .29 \)).

3.4.2.2 Affordances Changed Women’s Empathizing vs. Systemizing Possible Selves

Next, supporting the prediction that affordance would change the way women see themselves, women who received gendered (vs. counter-gendered) affordances reported an increased empathizing vs. systemizing possible self, \( t(276.51) = -3.17, p = .002, d = .37 \) (Table 3.6, Row 2c and Figure 3.10). Within each condition, women who received gendered affordances changed their relative possible selves significantly toward empathizing (\( M = -0.47, SD = 1.64; \) test against zero, \( t(146) = -3.47, p < .001, d = .29 \)), whereas women who received counter-gendered affordances changed their relative possible selves in the direction of (but not

---

25 In our original preregistration, we unintentionally misspecified the following hypothesis in Section 1: “Compared to those who receive equitable situational affordances, participants who receive gendered situational affordances will report: PH1A: Increased interest in systemizing teams, PH1B: Decreased interest in empathizing teams.” The hypothesis should have specified decreased interest in systemizing teams and increased interest in empathizing teams. All other preregistered hypotheses were phrased in the correct direction; the same hypothesis is correctly specified in sections of the analytic strategy in Section 10 of the preregistration.
significantly toward) systemizing ($M = +0.24$, $SD = 2.18$; test against zero, $t(149) = 1.35$, $p = .180$, $d = .11$).

3.4.2.3 Women Provided with Gendered (vs. Counter-Gendered) Affordances Selected More Empathizing vs. Systemizing Teams

Recall that women in the gendered condition were afforded more empathizing than systemizing team selections (12 empathizing, 4 systemizing), whereas women in the counter-gendered condition were afforded more systemizing than empathizing team selections (6 empathizing, 10 systemizing). Due to the distribution of these teams as pairs across trials (see Appendix B), the possible range of ES difference scores in the gendered condition was -8 (all possible empathizing selections made) to 0 (all possible systemizing selections made), whereas the range in the counter-gendered condition was -4 (all possible empathizing selections made) to 8 (all possible systemizing selections made).

Reflective of constrained team choices across the two conditions, women who received gendered (vs. counter-gendered) affordances selected significantly more empathizing vs. systemizing teams, $t(262.94) = -10.98$, $p < .001$, $d = 1.27$ (Table 3.6, Figure 3.10). Women who received gendered affordances selected significantly more empathizing teams ($M = 5.19$ more empathizing than systemizing teams, $SD = 3.36$; test against zero, $t(147) = -18.77$, $p < .001$, $d = 1.54$), whereas women who received counter-gendered affordances selected empathizing and systemizing teams equally ($M = 0.15$ more systemizing than empathizing teams, $SD = 4.88$; test against zero, $t(148) = 0.37$, $p = .712$, $d = .03$).
Table 3.6. Descriptive statistics by condition (Study 6)

<table>
<thead>
<tr>
<th></th>
<th>Gendered Affordances</th>
<th>Counter-Gendered Affordances</th>
<th>Effect Size for Difference ($d$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($N = 149$)</td>
<td>($N = 151$)</td>
<td></td>
</tr>
<tr>
<td>$M$ (SD)</td>
<td>$M$ (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. E-S Interest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Baseline</td>
<td>-1.23 (2.58)</td>
<td>-1.29 (2.85)</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>Range: -5.5—6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Post Measure</td>
<td>-1.61 (2.63)</td>
<td>-0.96 (3.02)</td>
<td>.23*</td>
</tr>
<tr>
<td></td>
<td>Range: -4.5—2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Change (Pre—Post)</td>
<td>-0.40 (1.05)</td>
<td>+0.33 (1.15)</td>
<td>.66***</td>
</tr>
<tr>
<td></td>
<td>Range: -4.5—2.5</td>
<td>Range: -3—4.5</td>
<td></td>
</tr>
<tr>
<td>2. E-S Possible Selves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Baseline</td>
<td>-0.91 (2.27)</td>
<td>-0.43 (2.56)</td>
<td>.20</td>
</tr>
<tr>
<td>b. Post Measure</td>
<td>-1.39 (2.59)</td>
<td>-0.19 (3.14)</td>
<td>.42***</td>
</tr>
<tr>
<td></td>
<td>Range: -9—5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Change (Pre—Post)</td>
<td>-0.47 (1.64)</td>
<td>+0.24 (2.18)</td>
<td>.37**</td>
</tr>
<tr>
<td></td>
<td>Range: -9—5</td>
<td>Range: -5—9</td>
<td></td>
</tr>
<tr>
<td>3. Team Selections</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Single timepoint)</td>
<td>-5.19 (3.36)</td>
<td>+0.15 (4.88)</td>
<td>1.27***</td>
</tr>
<tr>
<td>Actual Range: 8—0</td>
<td>Actual Range: -4—8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possible Range: -8—0</td>
<td>Possible Range: -4—8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *$p < .05$. **$p < .01$. ***$p < .001$. E-S = Empathizing vs. systemizing. For baseline and post-measures, we computed a difference score between systemizing and empathizing; thus, higher scores indicate more systemizing than empathizing. The response range for all variables was -6 to 6, unless otherwise specified.
Figure 3.10. Key outcomes by condition (Study 6). Error bars represent +/- 1 standard error
3.4.2.4 Women Imagined Staying at the Company to the Degree it Matched Their Trait-Level Empathizing vs. Systemizing Interests

Finally, we examined the effect of gendered affordances on the degree to which women saw themselves staying at the company (asked as part of their possible selves). Looking across the whole sample, women who received counter-gendered (vs. gendered) affordances reported a somewhat lowered tendency to see themselves staying at Pixlio after their internship, \( t(293.3) = 1.99, p = .048, d = .23 \). However, follow-up analyses revealed this effect was moderated by participants’ trait empathizing vs. systemizing score (E-S traits; taken as the difference score between the two measures), \( \beta = .56, p = < .001 \) (Figure 3.11). In the counter-gendered affordances condition, the higher women were on S > E traits, the more they saw themselves staying at the company (\( \beta = .39, p = < .001 \) Figure 3.11; bands around regression lines represent +/-1 standard error). Women highest in S > E traits (i.e., at a “6,” plotted on the rightmost side of Figure 3.11; \( n = 12 \)) reported an increased tendency to see themselves staying at the company in the counter-gendered (vs. gendered) condition, \( d = .25 \). Yet in the gendered affordances condition, the higher women were on S > E traits, the less they saw themselves staying at the company (\( \beta = -.17, p = .044 \); Figure 3.11). Women highest in E > S traits (i.e., at a “-6,” plotted on the leftmost side of Figure 3.11; \( n = 25 \)) reported an increased tendency to see themselves staying at the company in the gendered (vs. counter-gendered) condition, \( d = .36 \). Because women in Study 6 tended to be relatively higher on trait empathizing than systemizing (\( M = -1.18, SD = 3.22 \)), this explains the overall pattern of effects we observed above. Importantly, E-S traits did not moderate the effect of condition on change in team interest or empathizing and systemizing possible selves (see Appendix B). Thus, even though participants varied in their affective response to different affordances (intending to exit situations low in self-concept fit),
the effect affordances had on shaping participants’ empathizing and systemizing interests and possible selves did not depend on these pre-existing preferences.

![Figure 3.11. Relationship between condition and change in intentions to stay (post-pre), moderated by trait empathizing vs. systemizing (Study 6)](image)

**3.4.3 Discussion**

The results of Study 6 provided causal support for the second link of our theoretical model: women who received affordances based on essentialized beliefs (i.e., gendered affordances), versus non-essentialized beliefs (i.e., counter-gendered affordances), shifted their interests and possible selves toward a more gendered self that confirms perceptions of essentialized interests. Additionally, reflective of their constrained choices, women who received gendered affordances were more likely to select more empathizing than systemizing teams. In contrast, women who received counter-gendered affordances selected empathizing and
systemizing teams equally. Finally, an exploratory analysis of intentions to stay at the company by trait-level empathizing and systemizing interest revealed that participants could see themselves staying at the company to the degree affordances matched their existing self-concept. Together, these results provide support for the theoretical model depicted in Figure 3.1, in which perceivers’ essentialist beliefs lead to disparate target outcomes via affordances.

3.5 General Discussion

Despite recent progress toward gender equity, gender segregation persists inter-occupationally as well as intra-occupationally, with men and women taking on different roles within fields like software engineering (Campero, 2021), medicine (Ku, 2011), law (Kay & Gorman, 2008), and domestic labor (Biernat & Wortman, 1991). Addressing gender segregation has potential benefits for labor supply, innovation, and wage equity. Departing from the notion that men’s and women’s different interests in roles involving people vs. things reflect innate preferences or free choice (Ceci, Williams, & Barnett, 2009, 2011; Baron-Cohen, 2009), we present evidence that gender gaps in interests are systematically shaped by environmental affordances. In particular, those who endorse biologically essentialized explanations for gender differences play a role in providing gendered affordances (Studies 4 and 5) that, when manipulated, shape women’s relative interest in, self-concept for, and choice of systemizing over empathizing roles (Study 6). Together, these studies highlight a self-fulfilling process through which essentializing perceivers create affordances that inadvertently confirm their beliefs.

Across these studies, the relationship between essentialism and affordances was consistently stronger for systemizing than empathizing. There are several possible explanations for this asymmetric pattern of effects. First, given our focus on training opportunities in tech, perceivers may have been more sensitive to monitoring women’s systemizing opportunities (as
opposed to men’s empathizing opportunities) in a field where women are underrepresented.

Alternatively, these patterns might reflect more general asymmetric barriers to men’s communal roles, as compared with women’s agentic roles (Block et al., 2018; Croft et al., 2015), given that men’s adoption of stereotypically feminine roles presents a threat to masculinity (Vandello et al., 2008; Vandello & Bosson, 2013). Indeed, women are no longer stereotyped as being less competent or intelligent than men, but men continue to be stereotyped as less communal than women (Eagly et al., 2020). Thus, perceivers might be more willing to afford women’s entry into male-stereotypic roles (i.e., a systemizing task), than to afford men’s entry into female-stereotypic roles (i.e., an empathizing task). Future research might extend our paradigm to HEED contexts to address these possibilities.

Although receiving counter-gendered affordances shifted women’s interests as predicted and increased anticipated retention for women high in systemizing, women anticipated wanting to leave the company if their team assignments were misaligned with their traits (e.g., women high in empathizing who received more systemizing affordances). While unanticipated, such effects are consistent with the notion that a lack of self-concept fit can lead some women to avoid STEM fields (Schmader, 2023). And even though women anticipated disliking working at a company with these mismatched team assignments, such effects could be indicative of affective forecasting errors skewed by women’s own gender stereotypes (Moons, Chen, & Mackie, 2017). Future research is needed to test whether the actual experience of working on teams that do not align with one’s traits would in fact lead to higher attrition. Perhaps only those who themselves have more essentialized beliefs about gender differences would show such effects. For instance, in other contexts, multiracial individuals who believe race has biological meaning are especially sensitive to race-related social cues in their environment (Sanchez &
Garcia, 2009). Similarly, women who themselves believe gender is biologically (vs. socially) determined might be especially sensitive to gender-related affordances.

A strength of the current studies is the use of a behavioral choice paradigm to approximate real-world decision-making within an immersive role-playing scenario, where participants were relatively unaware of the study’s purpose (across all 3 studies, only 7.99% of participants’ open-ended responses suggested they were aware of the study’s purpose; moreover, key effects are robust to controlling for suspicion or excluding these participants, see Appendix B). However, the mock managerial task is still hypothetical, and it is possible participants would make different choices in real-world job decisions where other factors intervene. For instance, past research on third-party prejudice (Vial et al., 2019) highlights that even participants motivated to make egalitarian hiring decisions enact prejudice when they believe the CEO of the company is sexist. Additionally, hiring decisions are often influenced by how well candidates fit with the existing organizational culture (Rivera, 2012; Rynes, & Gerhart, 1990; Tsai, Chi, Huang, & Hsu, 2011). In real-world organizations, people rarely enter hiring decisions free of these external influences or with such limited information about candidates. Future research as part of a naturalistic data collection effort would provide more insight into this question.

As a whole, these findings have implications for educators, employers, and caregivers responsible for providing young people with opportunities to learn about, and become interested in, empathizing and systemizing (Haddock et al., 2017; Tenenbaum & Leaper, 2003). The results of our studies demonstrate the integral role that environmental affordances play in shaping interest, as informed by perceivers’ beliefs. When these repeated experiences accumulate across perceivers or over time (Madon et al., 2018), they can eventually steer men and women toward or away from certain careers and reinforce perceivers’ beliefs. Not only do these findings lead us
to question claims that men’s and women's interests are freely chosen or innately determined, but they also suggest those who promote essentialized explanations for interest differences might contribute to perpetuating gender segregation.
Chapter 4: Conclusion

Women and men continue to select into careers and roles at different rates. At scale, patterns of horizontal gender segregation have downstream consequences for labor supply, creativity and innovation, perpetuation of gender stereotypes, and wage equity (Croft, Schmader, & Block, 2015; Galinsky et al., 2015; Goldin, 2021; Schmader, 2023). One of the most pervasive explanations for horizontal gender segregation is women’s greater interest in careers involving people and men’s greater interest in careers involving things (Su, Rounds, & Armstrong, 2009). In contrast to perspectives that conceive of men and women’s different interests as reflecting free choice (Ceci et al., 2009), I contend gendered patterns of interests are shaped by opportunities and constraints in the environment (i.e., situational affordances). Guided by theory on psychological essentialism and self-fulfilling prophecy, I propose a process through which people’s essentialized explanations for gender differences in interest (as primarily biological vs. socially caused) leads them to provide disparate situational affordances that confirm those beliefs by shaping the interests, self-views, and experiences that men and women have. In Chapter 2, I tested whether scales designed to measure biological gender differences in interest ask about activities that men and women have different affordances to learn, and whether gendered affordances partly explain the size of the gender gap on these measures. In Chapter 3, I test whether perceivers who believe gender differences in interest are biologically (vs. socially) determined provide men and women with disparate affordances, and whether disparate affordances widen the gender gap in interest. I next review a summary of these empirical chapters (also included in Chapter 1).
4.1 Summary of Findings

4.1.1 Chapter 2: Do Measures of Systemizing and Empathizing Reflect Perceptions of Gender Differences in Learning Affordances?

In Chapter 2, I consider whether scales designed to measure biological gender differences in interest ask about activities that men and women have different affordances to learn, and whether these gendered affordances partially explain the size of the gender gap. Across three studies, I leveraged a wisdom of crowds (Larrick et al., 2011) approach to examine these questions with E-S Theory’s self-report measures of systemizing (SQ) and empathizing (EQ).

Study 1 ($N = 624$) first estimated gender differences on systemizing (SQ) and empathizing (EQ) scales in two large, nationally representative samples. Results of subsequent coding studies revealed both lay coders (Study 2, $N = 199$) and psychology journal reviewers (Study 3, $N = 116$) rated SQ and EQ item activities as being more learned (vs. innate), counter to how E-S theory tends to explain these differences. Coders also believed that men are given more systemizing, and women given more empathizing (Study 3 only), affordances. Relating item-level variation on these ratings to gender differences observed among participants in Study 1, items showing the largest gender differences in Study 1 were those rated as having the most gendered affordances (more than gendered genetic advantages) by coders in Studies 2 and 3.

Together, this chapter concludes that claims about inherent sex differences in systemizing, and to a lesser degree empathizing, might overlook the degree that self-report measures tap into gendered learning affordances.
4.1.2 Chapter 3: Are Essentialized Explanations for Gender Differences in interest Self-Fulfilling?

In Chapter 3, I consider whether perceivers who believe gender differences in interest are primarily biologically (vs. socially) determined provide men and women with gendered (vs. counter-gendered) affordances, and whether these affordances drive the gender gap in interest. I tested these questions across three studies that implemented a novel behavioral paradigm in the context of an imagined tech company. In Studies 4 (N = 285) and 5 (N = 379), perceivers were asked to imagine themselves in a project manager role. Those who primarily endorsed (Study 4) or were experimentally primed with (Study 5) a biological (vs. social) explanation for interest differences provided gendered affordances: they were more likely to assign women interns to empathizing roles and men interns to systemizing roles. Translating these situational affordances to target outcomes in a quasi-double randomization design (MacKinnon et al., 2007; Madon et al., 2018; Word et al., 1974), in an imagined internship scenario (Study 6, N = 300), first- and second-year undergraduate women who received gendered (vs. counter-gendered) affordances based on perceivers’ selections in Study 4 shifted their interests to be more gendered (i.e., became more interested in empathizing than systemizing teams) and chose to work on more empathizing than systemizing teams. In contrast, women who received counter-gendered (vs. gendered) situational affordances shifted their interests to be counter-gendered (i.e., became more interested in systemizing than empathizing teams) and chose to work on empathizing and systemizing teams equally. I also found that women intended to stay at the company to the degree affordances matched their trait-level interests. Together, this evidence highlights situational affordances, driven by essentialist explanations for the interest gap, as a key contributor to gender gaps in interest and occupational pursuits.
4.2 Implications

4.2.1.1 Interest and Stereotype Accuracy

My program of research pinpoints how affordances, driven by essentialized beliefs, can magnify gender differences in interest and contribute to broader patterns of horizontal gender segregation within, and across, careers. These findings inform a deeper understanding of how interest—a phenomenon often assumed to be inherent to the person (Ceci, Williams, & Barnett, 2009; Ceci & Williams, 2011)—can be shaped by forces in the environment. The effect of affordances on interests might, for instance, partly explain the so-called “gender equality paradox,” (Stoet & Geary, 2018) which suggests that in countries with greater gender equality, women pursue STEM degrees at a lower rate than men. Although these findings are still under debate (Block et al., 2023), some scholars suggest measures of gender equality are often conflated with cultural values of self-expression and interests in Western, Educated, Industrial, Rich, Democratic (WEIRD, Henrich et al., 2010) societies (Soylu Yalcinkaya & Adams, 2020). Thus, compared to collectivistic cultures, in more gender equal societies that perceive self-expression as the primary goal of career choices, choosing careers that affirm one’s interests is tantamount. Connecting these bodies of work together, my findings highlight how interests are shaped by gendered affordances in the environment (see also Siy et al., 2023). Interests that are shaped by gendered affordances, when operating in cultures that view the expression of interests as a primary driver of career choice, may then become evident in documented gender disparities in occupational pursuits.

In addition to considering how his work also informs a larger body of research on stereotype accuracy, particularly the position that stereotypes often correspond to true differences at the group level (Jussim, 2012; Jussim et al., 2016; Ryan, 2003). Separate from people’s
accuracy in picking up on group differences at large (e.g., “women are less interested in systemizing than men”), my research pinpoints people’s explanations for these differences as a critical distinction between stereotypes that justify the status quo (e.g., “…this is due to biology, therefore society is not obligated to address women’s underrepresentation in fields like STEM”) and stereotypes that acknowledge the possibility of societal change (e.g., “…this is due to sociocultural factors, therefore society can do something about women’s underrepresentation in fields like STEM”). Scholars have recently argued for more nuanced understandings of stereotype accuracy that go beyond all-or-nothing accounts (Hall & Goh, 2017). Through this work, I hope to offer a layer of nuance to the debate on stereotype accuracy by considering not just when stereotypes are accurate, but when their underlying explanations perpetuate societal harms, and when they motivate people to address systemic inequities.

4.2.1.2 Self-Fulfilling Prophecy

In addition to informing the broader academic discourse around interest and stereotype accuracy, these findings advance a broader understanding of self-fulfilling prophecy. Prior work on self-fulfilling prophecy has examined how people’s descriptive stereotypes lead targets to fulfill expectations. Similar to stereotype accuracy, I seek to take these studies a step further by considering, not only the “what” (i.e., people’s belief in group differences), but the why (i.e., the causal reason for group differences). Integrating theorizing on self-fulfilling prophecy and psychological essentialism, I suggest these causal explanations for group differences play an integral role in the self-fulfilling prophecy process. To illustrate how this might apply to existing self-fulfilling prophecy research, past research on self-fulfilling prophecy has operationalized perceiver stereotypes as: (1) teachers’ expectations of student performance (Madon, Jussim, & Eccles, 1987), (2) judges’ perceptions of a trial defendant’s guilt (Rosenthal, 2003), and (3)
strangers’ gendered expectations of knowledge (Hollingshead & Fraidin, 2003). Tying each of these examples to essentialism, we might consider the self-fulfilling affordances produced by: (1) teachers’ beliefs that student performance is innate versus learned (Dweck, 2006; Dweck & Yeager, 2019), (2) judges’ beliefs that criminality is genetic versus environmental (Cheung & Heine, 2015; Heine, 2017), or (3) strangers’ beliefs that gendered knowledge is innate versus socialized (Lee et al., 2020). Relevant to the third example, my work finds that even controlling for people’s belief in the magnitude of gender differences in interest (the “what”), their explanations for gender differences as primarily caused by biology or socialization (the “why”) produce divergent affordances that either sustain or reduce gender gaps on interest. In fact, in Study 4, perceivers who perceived gender differences as being caused by socialization provided affordances that reversed typical gender stereotypes. By overlooking the role of essentialist explanations for group differences, self-fulfilling prophecy paradigms may unintentionally attenuate effects that are larger when these perceptions are taken into account. This might, at least partly, explain the small effect sizes that have historically been problematic for self-fulfilling prophecy research (Jussim & Harber, 2005).

4.2.1.3 Potential Interventions

Finally, these findings might inform interventions designed to reduce gender gaps in men and women’s interest and participation in key sectors of the workforce (e.g., HEED and STEM) and roles within those sectors (e.g., front-end versus back-end roles in tech). As emphasized previously, horizontal gender segregation has myriad consequences for labor supply, creativity and innovation, perpetuation of gender stereotypes, and wage equity (Croft, Schmader, & Block, 2015; Galinsky et al., 2015; Goldin, 2021; Schmader, 2023). Interventions, when designed effectively, can provide a high-impact, low-cost way for leaders and policymakers to address
these social problems (Walton & Wilson, 2018). There are at least two areas of intervention suggested by this research: (1) implementing policies and practices that equate situational affordances across gender, and (2) educating caregivers, teachers, and other social agents in the environment about the critical role they play in shaping interests. I consider each of these implications and directions for further work below.

4.2.2 Equating Situational Affordances Through Policy and Practice

In his original work on self-fulfilling prophecy, Merton (1948) writes, “The self-fulfilling prophecy […] operates only in the absence of deliberate institutional controls” (p. 210). Prior research has identified institutional accountability as among the most successful strategies for sustaining organizational change (e.g., maintaining a diverse managerial workforce, Kalev, Dobbin, & Kelly, 2006). This evidence speaks to the possibility of institutional policy, especially that which invokes accountability, as a tool to address gender disparities in situational affordances. For example, drawing on the tech internship scenario employed in Chapter 3 (Studies 4-6), a real company adopting a new internship program like this one might monitor the gender distribution of front-end and back-end tasks assigned to each intern. Assuming organizational leadership is motivated to maintain gender equity in work assignments and sets target goals accordingly (e.g., at least 40% women in back-end roles), introducing a monitoring mechanism that helps project managers keep track of their progress toward these established goals might help the company achieve greater gender parity in work assignments. Moreover, given that cues to inclusion and stigma tend to transfer across social groups (Chaney et al., 2016; Sanchez et al., 2017), policies aimed at shoring up inclusion for women might also signal inclusion to other stigmatized groups in the environment.
Although the example above focuses on the workplace, this type of intervention might be especially impactful in early learning settings, where children’s interests are more malleable and thus not as likely to be misaligned with affordances that do not match their existing or already gender-constrained interests. Indeed, past research finds young children (i.e., ages 8 to 10) tend to be more open to trying new tasks and activities than older learners (Renninger, Sansone, & Smith, 2004). Converging with these findings, in Study 5 we observed that women who received affordances misaligned with their trait-level interests (i.e., women high in trait-level empathizing interest who received more systemizing affordances, or women high in trait-level systemizing interest who received more empathizing affordances) reported lowered intentions to remain at the company, but they intended to stay at the company if affordances matched their trait-level interests. If younger children are more open to trying activities misaligned with their typical interests, a version of this paradigm adopted in early learning contexts might elicit less misfit. However, that is not to say interests are set in stone after childhood; scholars maintain interests can develop at any age (Renninger & Su, 2012). In fact, Hidi and Renninger’s (2006) four-phase model of interest development identifies situational triggers (i.e., affordances typically provided through the external environment) as the first phase. Affective reactions experienced during this phase can be either positive or negative (as we saw in Study 5), yet these situational triggers set the stage for new interests to emerge. Indeed, even though women in Study 5 who received affordances misaligned with their trait-level interests reported lowered intentions to stay at the company, across all levels of pre-existing interests, women’s interests and possible selves changed in response to the affordances they received.

In addition to informing best practices for policymakers, this work also highlights the vital role that researchers play in disentangling different explanations for gendered interests. As I
have argued, measuring people’s broad-level gender stereotypes (i.e., their perceptions of gender differences by themselves) might not be predictive of how those stereotypes affect behavior. Yet, as highlighted above, self-fulfilling prophecy research often takes this approach, yielding small or inconsistent effects. Considering people’s explanations for the origin of gender differences might have stronger relationships to their behavior that constrains others’ options and interests. Finally, in addition to accounting for bias introduced by participants’ explanations for gendered interest differences, researchers might also consider bias stemming from their own assumptions around the origin of gender differences in interest. As noted in Chapter 2, scientific theories built on essentialist thinking have the potential to not only explain, but to sustain, gender differences (Fine, 2012). Reflecting on one’s own values, collaborating with diverse teams of scholars, and soliciting regular feedback outside of one’s immediate academic sphere could be effective strategies for examining these assumptions.

4.2.2.1 Equating Situational Affordances Through Education

Another potential intervention to reduce gender gaps in situational affordances is educating social agents about the critical role they play in shaping interests. Prior work reveals that awareness of one’s own bias is a necessary (but not sufficient) step in overcoming prejudice (Devine & Monteith, 1993; Plant & Devine, 2009). And in hiring contexts, members of scientific evaluation committees who are most aware of the potential role gender bias plays in the evaluation process are less likely to make gender-biased decisions (measured as the decision to promote men over equally qualified women for elite research positions; Régner et al., 2019; see also Begany et al., 2020). Based on this prior evidence, an educational intervention to promote equitable situational affordances might be especially effective. For example, one might educate teachers, caregivers, and other social agents about the potential impact of their bias (in this case,
essentialist thinking) on situational affordances, with a focus on the role situational affordances play in shaping young people’s interests. Notably, our findings identify perceivers who are least aware of endorsing social determinants of interest (i.e., those highest in biological determinism) as being the most likely to widen gender gaps in interest via disparate situational affordances. Thus, this group provides an especially important focus for educational intervention work.

Despite the promise of an educational intervention, a key concern with educational interventions--especially one aimed at communicating information that contradicts people’s beliefs--is reactance (Jackson et al., 2014; Rudman et al., 2001). Successful strategies for reducing reactance in this domain could include making participation voluntary (Bezrukova et al., 2016), adopting components to reduce negative emotions (Rudman et al., 2001; Stell & Farsides, 2016), and implementing cooperative learning strategies (Paluck & Green, 2009, Paluck et al., 2021). For example, in a potential intervention administered before the start of the school year, teachers might engage in a group exercise to affirm their values (see Sherman, 2013) prior to learning about how essentialist thinking can lead to disparate affordances and outcomes. In addition to raising awareness, as alluded to earlier, concern and efficacy are also necessary for overcoming prejudice (Devine & Monteith, 1993; Plant & Devine, 2009).

Emphasizing the broader, cumulative impact of disparate affordances (e.g., highlighting societal disadvantages of horizontal gender segregation) and providing perceivers with concrete strategies for providing equitable situational affordances in their daily lives (similar to Chang et al., 2019) might be successful strategies for establishing motivation and efficacy in this context.
4.3 Limitations and Future Directions

4.3.1 Constraints on Generalizability

To bolster the replicability of this work, I acknowledge several contextual and population-level factors that likely present boundary conditions for my work (Simons et al., 2017). First, given that stereotypes are often andocentric, ethnocentric, and heterocentric (Coles & Pasek, 2020; Purdie-Vaughns & Eibach, 2008), perceptions of men and women’s relative affordances (Chapter 2) and people’s general explanations for gendered interests (Chapter 3) are perhaps most generalizable to straight, White men and women. Past work on intersectionality reveals that gendered phenomena often vary across the intersection of race, class, sexual orientation, and other social identities (e.g., Crenshaw, 1989; Ghavami & Peplau, 2013; Sesko & Biernat, 2010; Schmader & Nater, 2023). In these studies, I would expect affordances to vary not only by target gender, but also by other identities such as targets’ perceived race and class, based on the unique stereotypes people hold about these intersections of social identity. For example, based on the studies presented in Chapter 2, research might also consider how affordances to learn empathizing and systemizing activities are assumed to vary across gender and socio-economic status (SES). Knowing about a stereo’s precise technical features (the activity referenced in SQ Item 13) or learning one’s way around a new city (the activity referenced in SQ Item 11) are stereotypically masculine activities which are also contingent on having expendable income for stereos (or the modern equivalent of high-tech audio gear) or travel. In contrast, activities like caring for other people (EQ Item 2) or understanding how people are feeling and what they are thinking (EQ Item 13) are stereotypically feminine activities but are also more in line with individuals from low-SES backgrounds’ documented tendency to be more sensitive to the needs of others (Kraus et al., 2012; Piff & Robinson, 2017). Disentangling the dynamic
contributions of stereotypes about gender and SES might be a fruitful area for future research attempting to debias empathizing and systemizing scales.

Applying the same theoretically guided intersectional lens to Chapter 3, research might also consider how sexual orientation informs gendered affordances based on essentialist thinking. Past work finds that gay men, especially those in traditionally feminine roles (e.g., hairdresser, parent), are stereotyped as being feminine (Fingerhut & Peplau, 2006). Might these stereotypes about femininity correspond with a stereotype that gay men are more interested in empathizing than systemizing? And might this relationship differ among perceivers who endorse relatively more biological explanations (i.e., those who might assume interests track gender as an innate category) versus those who endorse more social explanations (i.e., those who might assume interests track gender expression)? I consider these boundary conditions as important directions for further work on this topic.

Relatedly, since my dissertation is focused on understanding stereotypes about, and affordances for, men and women, these studies do not generalize to individuals outside the gender binary (Morgenroth & Ryan, 2018, 2021; Worthen, 2021). Indeed, the embedded nature of binary gender stereotypes across multiple levels of social institutions often perpetuate notions of gender as a binary, essential construct, even if gender identity itself is not binary (Diekman & Schmader, 2023). The studies in Chapter 3 incorporated a more dimensional view of gender by addressing variation in gender expression that exists within gender categories; in Studies 4 and 5, I tested whether affordances varied by gender prototypicality of targets (they did not, with one exception; see Appendix B), and in Study 6, I tested whether reactions to work assignments were moderated by participants’ own gender expression (they were not). Although these findings provide some insight into how these processes vary within gender identities, they are still
restricted to binary gender categories of men and women. Relatively less work has sought to understand how gender stereotypes, for instance, about inclination toward STEM or HEED careers might apply to nonbinary or genderqueer individuals (Morgenroth & Ryan, 2021). Some questions that might be asked in a future line work addressing this topic include: What stereotypes about empathizing and systemizing interest exist for nonbinary or genderqueer individuals? How do affordances based on binary gender essentialist beliefs invokes particular harm for nonbinary or genderqueer individuals? How might nonbinary or genderqueer individuals’ gender expression, measured as a single time point (e.g., masculinity vs. femininity) as well as variation across time points (e.g., degree of gender fluidity) relate to empathizing and systemizing interest? These questions might require shifting beyond existing paradigms and adopt new practices of operationalizing and measuring gender (e.g., Gulgöz, Edwards, & Olson, 2022).

A third constraint on the generalizability of our findings is that our studies are situated in Western cultural contexts. With respect to research reviewed in Chapter 2, previous work examining the cross-cultural stability of the EQ and SQ finds that gender differences on the EQ, but not SQ, vary by country (Groen et al., 2015). On the EQ, women’s advantage tends to be more robust in Western cultures (e.g., $d = .88$, Netherlands, $N = 685$, Groen et al., 2015) than Eastern cultures (e.g., $d = .11$, Korea, $N = 478$, Kim & Lee, 2010). One explanation for this cultural difference may be that processes central to empathy (e.g., personal distress, empathic concern, communal values) often vary across cultures (Cassels et al., 2010; Block et al., 2023). In contrast, past work finds men’s higher score on the SQ is consistent across Western (e.g., $d = 1.01$, UK, $N = 167$, Ling et al., 2009) and Eastern (e.g., $d = 1.15$, Japan, $N = 137$, Wakabayashi et al., 2007) cultures, though a more recent study with a large sample in China found a smaller
difference on the measure \( d = .51, N = 1,367, \) Zhou et al., 2020). Although it is possible that these differences can be accounted for by measurement features like different interpretation of items (Groen et al., 2015) or cultural difference in reference groups (Heine et al., 2002), it may also be the case that activities referenced in the items are located within a Western cultural context and do not generalize to other cultures. Indeed, if a key mechanism driving the gender difference on these measures is the extent to which activities provide different affordances to men and women, given that socialization practices are variable across cultures (Heine, 2010), this could translate into different effect sizes on the measures.

Similarly, with respect to people’s endorsement of psychological essentialism as considered in Chapter 3, given that Eastern cultures tend to emphasize more of an incremental than entity mindset (Heine et al., 2001) and less of a tendency to prioritize dispositional over situational information for explaining people’s behavior (Choi, Nisbett, & Norenzayan, 1999), we might not observe the same relationship between essentialism and affordances in non-Western contexts. For example, participants situated in Eastern cultures might demonstrate a shared endorsement for social (vs. biological) determinist explanations for gender differences in interest. Restricted variability on this belief, in turn, would translate into an overall smaller effect for the relationship between essentialist thinking and situational affordances. Future research might test these cross-cultural extensions empirically.

### 4.3.2 Asymmetries Between Empathizing and Systemizing

So far, I have treated empathizing and systemizing as parallel constructs that follow the same process. Indeed, I propose one commonality across empathizing and systemizing is that men and women’s interest (or lack thereof) in them is driven by gendered patterns of affordances. Prior research has identified other common barriers to men and women’s
participation in empathizing and systemizing roles, such as lack of role models and gender stereotypes associating men with systemizing-relevant, and women with empathizing-relevant, roles (Cheryan et al., 2017; Croft et al., 2015). Yet prior research also delineates unique barriers to men and women’s entry into empathizing and systemizing domains. First considering systemizing, barriers to women’s entry in systemizing-relevant domains like STEM tend to be characterized by factors that signal a lack of fit to the environment, such as masculine cultural defaults (Cheryan & Markus, 2020), cues in the environment that signal a lack of fit and inclusion (Cheryan et al., 2009, 2011; Hall et al., 2015, 2018; Schmader & Sedikides, 2018), and explicit gender-based discrimination (especially for women of color, Berdahl & Moore 2006; see also McCord et al. 2018). At the same time, past research finds men are relatively less affected (or even advantaged, in the case of masculine defaults) by these same factors (Master, Cheryan, & Meltzoff, 2016). In contrast, barriers to men’s entry in empathizing-relevant domains like HEED tend to be characterized by financial and status penalties. Men are more likely to be discouraged from pursuing empathizing roles due to the lower pay of these careers (Block et al., 2018; Cohen & Huffman, 2003a, 2003b; England, 1992), threats to their precarious masculine identity (Vandello et al., 2008; Vandello & Bosson, 2013), and perceived loss of status (Moss-Racusin, Phelan, & Rudman, 2010). Yet women tend to be penalized less harshly for gender role transgressions (Levy, Taylor, & Gelman, 1995) and do not face the same expectations to achieve higher salary and wealth (Jones, Howe, & Rua, 2000; Williams et al., 2010). Together, these lines of research identify possible asymmetries in barriers to men and women’s interest in systemizing and empathizing. How might these asymmetries in barriers play out in the studies presented across Chapters 2 and 3?
The studies in Chapter 2 focused on situational affordances as a common driver of men and women’s lower interest in empathizing and systemizing. Yet as I review above, a sense of fit to the environment might more strongly predict women’s interest in systemizing than men’s interest in empathizing (which may still be driven by a sense of fit to the environment, but less strongly). To test this, a follow-up study might ask women to forecast their sense of fit in environments where others know exact details about a computer’s hard disk drive capacity and processor speed (the activity referenced in SQ Item 8). The same study might ask men to forecast their fit in environments where others know exactly how to act around their friends and coworkers (an example matched on level of specificity, based on the more general activity referenced in EQ Item 3: “knowing what to do in a social situation”). We might then test whether women’s forecasted fit in systemizing environments (operationalized via the SQ items)—more strongly than men’s forecasted fit in empathizing environments (operationalized via the EQ items)—predicts one’s own endorsement of EQ and SQ items.

Alternatively, considering asymmetric barriers to men’s interest in empathizing, we might extend the project management paradigm in Chapter 3 to manipulate the status of the roles available to interns. For example, in addition to a focus on empathizing (“People Support,” “Customer Strategy”) or systemizing (“Database Security,” “Sever Admin”), we might also manipulate the status of the role (high status: “Director of People Support,” “Database Security Lead;” low status: “Customer Strategy Apprentice,” Junior Server Admin”). We could then test the main effect of empathizing vs. systemizing (e.g., do perceivers choose more men for systemizing vs. empathizing roles?), status (e.g., do perceivers choose more men for high status vs. low status roles), and their interaction (e.g., do perceivers especially assign men to high status, systemizing roles?), as predicting affordances. As in our prior studies, we would also test
whether effects are strongest for those who primarily endorse a biological (vs. social) explanation for gendered interests.

### 4.3.3 Interactions Between Biological and Social Explanations

Finally, as reviewed in Chapter 1, my program of research is focused on people’s primary explanation for gender differences in interest (whether biological or social) and therefore cannot address more complex biosocial interactions that might be suggested by perspectives like Social Role Theory (Eagly & Wood 2012, Wood & Eagly, 2012). Social Role Theory proposes a biosocial process in which socialization and biological processes interact to produce gender differences in affect, cognition, and behavior (Wood & Eagly, 2002, 2012), including women’s greater pursuit of people-focused, and men’s greater pursuit of things-focused, roles (Wood & Eagly, 2015). Given their focus on the interactive nature of biological and social processes, Eagly and Wood argue that people themselves can prioritize one essentialized explanation (whether biological or social) without necessarily rejecting the other view (Wood & Eagly, 2012). Indeed, in studies measuring perceivers’ essentialist beliefs reported Chapter 3, I find that biological and social determinism are significantly, but not perfectly, negatively correlated (r’s range from -.41 in Study 4 to -.54 in Study 5), suggesting that many people endorse both biological and social determinism as independent explanations. Similarly, coders in Chapter 2 endorsed both genetic and socialized features as contributing to gender differences on EQ and SQ activities. These patterns of dual endorsement for biological and social determinist explanations converge with past research on gender essentialism (Martin & Parker, 1995; Lee, Reis & Rogge, 2020), yet more research is needed to understand how people perceive their interaction. I highlight one possibility for future research on this topic below.
Separate from people’s general endorsement of biological or social explanations, Social Role Theory suggests biological and social processes might be viewed as either proximal or distal explanations for gendered interests. For example, a perceiver who endorses both biological and social determinist explanations for a particular gender difference (e.g., “boys play more competitive sports than girls”) might see biological factors (e.g., males’ greater exposure to prenatal testosterone in the womb) as setting the stage for socialization to occur (e.g., testosterone facilitates male infants’ greater competitiveness, which leads parents to provide them more opportunities to play sports). Yet a different perceiver who also endorses both biological and social determinist explanations for this gender difference might see social factors (e.g., parents enroll more boys in sports than girls due to socialized gender norms) as a precursor to biological changes (e.g., playing competitive sports leads to an increase in testosterone, thus facilitating boys’ further engagement in competitive sports). These two explanations differ in the degree to which they imply immutability, or the belief that gender is a discrete, sharply defined, and unchangeable category (Lee, Reis, & Rogge, 2020; see Chapter 1). Although Lee et al.’s (2020) conceptualization of immutability is focused on perceptions of gender identity as immutable, in the example above, we might also consider people’s perceptions of gender expression (i.e., gendered interests) as immutable.26 Explanations that place biological factors at the outset imply a high degree of immutability, yet explanations that prioritize social factors suggest these same gender differences are malleable to social change. Much like the present work seeks to disentangle people’s causal explanations for agreed-upon gender differences (i.e.,

---

26 In Study 4, within each 5-item scale measuring either biological or social determinism, we also included 1 item that tapped into perceptions of interest immutability (biological item: “Men and women's different interests are mostly due to their innate biologies and are thus difficult to change,” social item: “Men and women's different interests are mostly due to how they are socialized and thus can be changed by new experiences”). Each item loaded highly onto its respective scale in factor analyses (biological: .91, social: .74) and we observed parallel results when modeling hypotheses isolating these single-item measures (see Appendix B).
people’s biological or social attributions for gender differences on interest), future work might go a step further by parsing proximal and distal explanations for agreed-upon causes (i.e., people’s proximal or distal perceptions of biological and social causes), focusing on immutability as a potential mechanism.

4.4 Conclusion

Gender differences in interest continue to serve as a leading explanation for men and women’s pursuit of different careers and roles. Ceci and Williams (2011) write: “To the extent that women’s choices are freely made and women are satisfied with the outcomes, we have no problem” (p. 3161). In my dissertation, I have highlighted how men and women’s interests, often thought to reflect freely made choices, are shaped by affordances in the environment. I underscore essentialism as a driver of these affordances in the context of measurement (Chapter 2) and the workplace (Chapter 3). By pinpointing the mechanisms that sustain or lessen gender gaps in interest, we can move toward a richer understanding of horizontal gender segregation.
Bibliography


Williams, M. J., Elizabeth, L. P., & Spencer-Rodgers, J. (2010). The masculinity of money: 
Automatic stereotypes predict gender differences in estimated salaries. *Psychology of 
Women Quarterly, 34*(1), 7-20.

Test. Consulting Psychologists Press, Inc.

Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and 

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in 
Academic Press.

461-473.

prophecies in interracial interaction. *Journal of Experimental Social Psychology, 10*(2), 
109-120.

Worthen, M. G. (2021). Why can’t you just pick one? The stigmatization of non-
binary/genderqueer people by cis and trans men and women: An empirical test of norm-
centered stigma theory. *Sex Roles, 85*(5-6), 343-356.

York University, Centre for Human Rights, Equity and Inclusion. (2015). *Gender Identity & 
Gender Expression: A Guide for Students, Faculty, and Staff.* 
https://rights.info.yorku.ca/gender-identity-gender-expression-a-guide-for-students-
faculty-and-staff


Appendices

Appendix A

This section contains supplementary materials for Chapter 2: Do Measures of Systemizing and Empathizing Reflect Perceptions of Gender Differences in Learning Affordances?

A.1 Study 1: Estimating SQ and EQ Gender Differences

Study 1 - List of Measures

Below is the full list of variables included in Study 1 in the order they were presented to participants. Variables analyzed and reported in the main text are denoted with an asterisk.

- Attention Check
- SQ-Short* / EQ-Short*
- Gender*
- Age*
- Ethnic/Racial Background*
- Region*
- Political Orientation*
- SES

Study 1 - Items for Measures

Attention Check

Research in decision making shows that people, when making decisions and answering questions, prefer not to pay attention and minimize their effort as much as possible. Some studies show that over 50% of people don’t carefully read questions. If you are reading this question and have read all the other questions, please select the box marked ‘other’ and type ‘Decision Making’ in the box below. Do not select “predictions of your own behavior.” Thank you for participating and taking the time to read through the questions carefully!

What is this study about?
- Predictions of your own behavior
- Predictions of your friend’s behavior
- Political preferences
- Other: ______________

Attention Check
SQ-Short

1 = Strongly disagree  ...  7 = Strongly agree

(R) = Reverse-scored

- 1. If I were buying a car, I would want to obtain specific information about its engine capacity.
- 2. If there was a problem with the electrical wiring in my home, I’d be able to fix it myself.
- 3. I rarely read articles or web pages about new technology. (R)
- 4. I do not enjoy games that involve a high degree of strategy. (R)
- 5. I am fascinated by how machines work.
- 6. In math, I am intrigued by the rules and patterns governing numbers.
- 7. I find it difficult to understand instruction manuals for putting appliances together. (R)
- 8. If I were buying a computer, I would want to know exact details about its hard disk drive capacity and processor speed.
- 9. I find it difficult to read and understand maps. (R)
- 10. When I look at a piece of furniture, I do not notice the details of how it was constructed.
- 11. I find it difficult to learn my way around a new city. (R)
- 12. I do not tend to watch science documentaries on television or read articles about science and nature. (R)
- 13. If I were buying a stereo, I would want to know about its precise technical features.
- 14. I find it easy to grasp exactly how odds work in betting.
- 15. I am not very meticulous when I carry out D.I.Y.
- 16. When I look at a building, I am curious about the precise way it was constructed.
- 17. I find it difficult to understand information the bank sends me on different investment and saving systems.
- 18. When traveling by train, I often wonder exactly how the rail networks are coordinated.
- 19. If I were buying a camera, I would not look carefully into the quality of the lens.
- 20. When I hear about the weather forecast, I am not very interested in the meteorological patterns.
- 21. When I look at a mountain, I think about how precisely it was formed.
- 22. I can easily visualize how the motorways in my region link up.
- 23. When I’m in a plane, I do not think about the aerodynamics. (R)
- 24. I am interested in knowing the path a river takes from its source to the sea.
- 25. I am not interested in understanding how wireless communication works. (R)
EQ-Short

1 = Strongly disagree . . . 7 = Strongly agree

(R) = Reverse-scored

● 1. I can easily tell if someone else wants to enter a conversation.
● 2. I really enjoy caring for other people.
● 3. I find it hard to know what to do in a social situation. (R)
● 4. I often find it difficult to judge if something is rude or polite. (R)
● 5. In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking. (R)
● 6. I can pick up quickly if someone says one thing but means another.
● 7. It is hard for me to see why some things upset people so much. (R)
● 8. I find it easy to put myself in somebody else's shoes.
● 9. I am good at predicting how someone will feel.
● 10. I am quick to spot when someone in a group is feeling awkward or uncomfortable.
● 11. I can't always see why someone should have felt offended by a remark. (R)
● 12. I don't tend to find social situations confusing.
● 13. Other people tell me I am good at understanding how they are feeling and what they are thinking.
● 14. I can easily tell if someone else is interested or bored with what I am saying.
● 15. Friends usually talk to me about their problems as they say that I am very understanding.
● 16. I can sense if I am intruding, even if the other person doesn't tell me.
● 17. Other people often say that I am insensitive, though I don't always see why. (R)
● 18. I can tune into how someone else feels rapidly and intuitively.
● 19. I can easily work out what another person might want to talk about.
● 20. I can tell if someone is masking their true emotion.
● 21. I am good at predicting what someone will do.
● 22. I tend to get emotionally involved with a friend's problems.
Study 1 – Sensitivity Analysis

A sensitivity analysis performed in G*Power to detect a difference between two independent means (two groups) revealed that our sample of $N = 624$ could detect a $d$ score of .23 or greater with 80% power, alpha = .05. Given that the original SQ- and EQ-Short validation paper (Wakabayashi et al., 2006) reported effect sizes of $d = .95$ and $d = -.63$ for gender differences on the SQ and EQ respectively, this sample is well-powered to detect effects.
A.2 Study 2

Study 2 - List of Measures

Below is the full list of variables included in Study 2 in the order they were presented to participants. Variables analyzed and reported in the main text are denoted with an asterisk.

- Attention Check
- Learned vs. Innate Attributions*
- Estimated Gender Difference*
- The order of the following variables was counterbalanced:
  - Gendered Learning Affordances*
  - Genetic Differences*
- Fixed vs. Growth Mindset
- Gender*
- Age*
- Ethnic/Racial Background*
- Political Orientation*
- SES*

Study 2 - Items for Measures

Attention Check

Research in decision making shows that people, when making decisions and answering questions, prefer not to pay attention and minimize their effort as much as possible. Some studies show that over 50% of people don’t carefully read questions. If you are reading this question and have read all the other questions, please select the box marked ‘other’ and type ‘Decision Making’ in the box below. Do not select “predictions of your own behavior.” Thank you for participating and taking the time to read through the questions carefully!

What is this study about?

- Predictions of your own behavior
- Predictions of your friend’s behavior
- Political preferences
- Other: ________________
SQ Activities

- 1. Obtaining specific information about a car’s engine capacity.
- 2. Fixing a problem with the electrical wiring in one’s home.
- 3. Reading articles or web pages about new technology.
- 4. Enjoying games that involve a high degree of strategy.
- 5. Being fascinated by how machines work.
- 6. Being intrigued by the rules and patterns governing numbers in math.
- 7. Understanding instruction manuals for putting appliances together.
- 8. Knowing exact details about a computer’s hard disk drive capacity and processor speed.
- 9. Reading and understanding maps.
- 10. Noticing the details of how a piece of furniture was constructed.
- 11. Learning one’s way around a new city.
- 12. Watching science documentaries on television or reading articles about science and nature.
- 13. Knowing about a stereo’s precise technical features.
- 16. Being curious about the precise way a building was constructed.
- 17. Understanding information the bank sends on different investment and saving systems.
- 18. Wondering exactly how rail networks of trains are coordinated.
- 21. Thinking about how precisely a mountain was formed.
- 22. Visualizing how the motorways in one’s region link up.
- 23. Thinking about the aerodynamics of a plane.
- 24. Knowing the path a river takes from its source to the sea.
- 25. Understanding how wireless communication works.
EQ Activities

(R) = Reverse-scored

Although not preregistered, we reverse-scored EQ Activities 5, 12, and 17 for directional ratings of the gender difference and gendered learning opportunities, since these activities are negatively worded (i.e., higher values corresponded with less empathizing). Not reverse-scoring these activities only changes 2 results: for empathizing items, (1) the relationship between estimate of gender difference and the size of the gender difference becomes non-significant, and (2) the relationship between gendered learning affordances and the size of the gender difference becomes marginal.

- 1. Telling if someone else wants to enter a conversation.
- 2. Caring for other people.
- 3. Knowing what to do in a social situation.
- 4. Judging if something is rude or polite.
- 5. Focusing on one’s own thoughts rather than on what their listener might be thinking. (R)
- 6. Picking up quickly if someone says one thing but means another.
- 7. Seeing why some things upset people so much.
- 8. Putting oneself in somebody else's shoes.
- 9. Predicting how someone will feel.
- 10. Spotting when someone in a group is feeling awkward or uncomfortable.
- 11. Seeing why someone should have felt offended by a remark.
- 12. Finding social situations confusing. (R)
- 13. Understanding how other people are feeling and what they are thinking.
- 14. Telling if someone else is interested or bored with what one is saying.
- 15. Being very understanding.
- 16. Sensing if one is intruding.
- 17. Being insensitive. (R)
- 18. Tuning into how someone else feels.
- 19. Working out what another person might want to talk about.
- 20. Telling if someone is masking their true emotion.
- 21. Predicting what someone will do.
Study 2 – Technical Error Information

In the survey for Study 2, there was a technical error where 3 EQ ratings (EMP 11, EMP 14, and EMP 19) contained the correct rating scale but the incorrect question stem. Additionally, there were minor display issues for a group of SQ ratings that did not affect the content of the ratings themselves and one case where the SQ rating labels were reversed for half of participants (SYS 7). As specified in the main text, reanalyzing our data excluding items affected by these technical errors only changes one result, such that the interaction between genetic differences and item type changes from marginal ($\beta = .27, p = .054$) to significant ($\beta = .31, p = .033$). All other results are unchanged.
## Study 2 – Item-Level Descriptive Statistics

Table A.1. Item-level information on coder ratings for each measure and the p-value for t-tests comparing ratings to scale midpoint (Study 2)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Item Text</th>
<th>Gender Differences</th>
<th>Innate vs. Learned Attributions</th>
<th>Gendered Learning Affordances</th>
<th>Genetic Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Above MP (p)</td>
<td>Below MP (p)</td>
</tr>
<tr>
<td>EMP1</td>
<td>Telling if someone else wants to enter a conversation.</td>
<td>3.89</td>
<td>1.60</td>
<td>0.696</td>
<td>0.304</td>
</tr>
<tr>
<td>EMP10</td>
<td>Spotting when someone in a group is feeling awkward or uncomfortable.</td>
<td>3.47</td>
<td>1.38</td>
<td>0.996</td>
<td>0.004</td>
</tr>
<tr>
<td>EMP11</td>
<td>Seeing why someone should have felt offended by a remark.</td>
<td>3.69</td>
<td>1.59</td>
<td>0.951</td>
<td>0.049</td>
</tr>
<tr>
<td>EMP12</td>
<td>Finding social situations confusing.</td>
<td>4.33</td>
<td>1.49</td>
<td>0.058</td>
<td>0.942</td>
</tr>
<tr>
<td>EMP13</td>
<td>Understanding how other people are feeling and what they are thinking.</td>
<td>3.47</td>
<td>1.55</td>
<td>0.998</td>
<td>0.002</td>
</tr>
<tr>
<td>EMP14</td>
<td>Telling if someone else is interested or bored with what one is saying.</td>
<td>4.08</td>
<td>1.26</td>
<td>0.330</td>
<td>0.670</td>
</tr>
<tr>
<td>EMP15</td>
<td>Being very understanding.</td>
<td>3.81</td>
<td>1.71</td>
<td>0.828</td>
<td>0.172</td>
</tr>
<tr>
<td>EMP16</td>
<td>Sensing if one is intruding.</td>
<td>3.88</td>
<td>1.52</td>
<td>0.753</td>
<td>0.247</td>
</tr>
<tr>
<td>EMP17</td>
<td>Being insensitive.</td>
<td>4.76</td>
<td>1.33</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>EMP</td>
<td>Description</td>
<td>Value1</td>
<td>Value2</td>
<td>Value3</td>
<td>Value4</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>EMP18</td>
<td>Tuning into how someone else feels.</td>
<td>3.66</td>
<td>1.78</td>
<td>0.951</td>
<td>0.049</td>
</tr>
<tr>
<td>EMP19</td>
<td>Working out what another person might want to talk about.</td>
<td>4.04</td>
<td>1.34</td>
<td>0.418</td>
<td>0.582</td>
</tr>
<tr>
<td>EMP2</td>
<td>Caring for other people.</td>
<td>3.50</td>
<td>1.63</td>
<td>0.995</td>
<td>0.005</td>
</tr>
<tr>
<td>EMP20</td>
<td>Telling if someone is masking their true emotion.</td>
<td>3.97</td>
<td>1.57</td>
<td>0.558</td>
<td>0.442</td>
</tr>
<tr>
<td>EMP21</td>
<td>Predicting what someone will do.</td>
<td>4.06</td>
<td>1.29</td>
<td>0.374</td>
<td>0.626</td>
</tr>
<tr>
<td>EMP22</td>
<td>Getting emotionally involved with a friend's problems.</td>
<td>3.42</td>
<td>1.76</td>
<td>0.997</td>
<td>0.003</td>
</tr>
<tr>
<td>EMP3</td>
<td>Knowing what to do in a social situation.</td>
<td>4.23</td>
<td>1.28</td>
<td>0.099</td>
<td>0.901</td>
</tr>
<tr>
<td>EMP4</td>
<td>Judging if something is rude or polite.</td>
<td>3.67</td>
<td>1.46</td>
<td>0.943</td>
<td>0.057</td>
</tr>
<tr>
<td>EMP5</td>
<td>Focusing on one's own thoughts rather than on what their listener might be thinking.</td>
<td>4.59</td>
<td>1.36</td>
<td>0.002</td>
<td>0.998</td>
</tr>
<tr>
<td>EMP6</td>
<td>Picking up quickly if Someone says one thing but means another.</td>
<td>3.80</td>
<td>1.56</td>
<td>0.864</td>
<td>0.136</td>
</tr>
<tr>
<td>EMP7</td>
<td>Seeing why some things upset people so much.</td>
<td>3.74</td>
<td>1.58</td>
<td>0.923</td>
<td>0.077</td>
</tr>
<tr>
<td>EMP8</td>
<td>Putting oneself in somebody else's shoes.</td>
<td>3.74</td>
<td>1.45</td>
<td>0.942</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>Predicting how someone will feel.</td>
<td>Obtaining specific information about a car’s engine capacity.</td>
<td>Noticing the details of how a piece of furniture was constructed.</td>
<td>Learning one’s way around a new city.</td>
<td>Watching science documentaries on television or reading articles about science and nature.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>EMP9</td>
<td>0.011</td>
<td>0.989</td>
<td>3.99</td>
<td>1.64</td>
<td>0.528</td>
</tr>
<tr>
<td>SYS1</td>
<td>5.21</td>
<td>1.14</td>
<td>0.000</td>
<td>1.000</td>
<td>6.06</td>
</tr>
<tr>
<td>SYS10</td>
<td>5.08</td>
<td>1.33</td>
<td>0.000</td>
<td>1.000</td>
<td>5.30</td>
</tr>
<tr>
<td>SYS11</td>
<td>4.62</td>
<td>1.14</td>
<td>0.000</td>
<td>1.000</td>
<td>5.53</td>
</tr>
<tr>
<td>SYS12</td>
<td>4.68</td>
<td>1.28</td>
<td>0.000</td>
<td>1.000</td>
<td>4.70</td>
</tr>
<tr>
<td>SYS13</td>
<td>5.06</td>
<td>1.09</td>
<td>0.000</td>
<td>1.000</td>
<td>6.09</td>
</tr>
<tr>
<td>SYS14</td>
<td>4.73</td>
<td>1.11</td>
<td>0.000</td>
<td>1.000</td>
<td>5.50</td>
</tr>
<tr>
<td>SYS15</td>
<td>4.22</td>
<td>1.22</td>
<td>0.107</td>
<td>0.893</td>
<td>4.65</td>
</tr>
<tr>
<td>SYS16</td>
<td>5.08</td>
<td>1.17</td>
<td>0.000</td>
<td>1.000</td>
<td>4.54</td>
</tr>
<tr>
<td>SYS17</td>
<td>4.57</td>
<td>1.21</td>
<td>0.000</td>
<td>1.000</td>
<td>6.01</td>
</tr>
<tr>
<td>SYS18</td>
<td>5.00</td>
<td>1.37</td>
<td>0.000</td>
<td>1.000</td>
<td>5.37</td>
</tr>
<tr>
<td>SYS19</td>
<td>4.35</td>
<td>1.06</td>
<td>0.011</td>
<td>0.989</td>
<td>5.88</td>
</tr>
<tr>
<td>SYS</td>
<td>Description</td>
<td>Mean</td>
<td>StDev</td>
<td>Median</td>
<td>Min</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>SYS2</td>
<td>Fixing a problem with the electrical wiring in one’s home.</td>
<td>5.19</td>
<td>1.22</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS20</td>
<td>Being interested in meteorological patterns.</td>
<td>4.52</td>
<td>1.06</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS21</td>
<td>Thinking about how precisely a mountain was formed.</td>
<td>4.57</td>
<td>0.96</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS22</td>
<td>Visualizing how the motorways in one’s region link up.</td>
<td>4.73</td>
<td>1.25</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS23</td>
<td>Thinking about the aerodynamics of a plane.</td>
<td>4.85</td>
<td>1.12</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS24</td>
<td>Knowing the path a river takes from its source to the sea.</td>
<td>4.77</td>
<td>1.29</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS25</td>
<td>Understanding how wireless communication works.</td>
<td>4.77</td>
<td>1.09</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS3</td>
<td>Reading articles or web pages about new technology.</td>
<td>4.92</td>
<td>1.16</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS4</td>
<td>Enjoying games that involve a high degree of strategy.</td>
<td>4.69</td>
<td>1.16</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS5</td>
<td>Being fascinated by how machines work.</td>
<td>5.39</td>
<td>1.24</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS6</td>
<td>Being intrigued by the rules and patterns governing numbers in math.</td>
<td>4.74</td>
<td>1.42</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS7</td>
<td>Understanding instruction manuals for putting appliances together.</td>
<td>4.80</td>
<td>1.34</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Knowing exact details about a computer’s hard disk drive capacity and processor speed.</td>
<td>4.77</td>
<td>1.58</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>----</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>SYS9</td>
<td>Reading and understanding maps.</td>
<td>4.75</td>
<td>1.23</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*Note.* MP = Scale midpoint. Tests against midpoint is not provided for Genetic Difference ratings since comparison to scale midpoint had no clear meaning in Study 2.
**Study 2 - Descriptives by Coder Gender**

Table A.2. Mean coder ratings among men and women coders separately, as well as the effect size for the gender difference between men and women coders (Study 2)

<table>
<thead>
<tr>
<th>Rating Dimension</th>
<th>SQ-Short Activities</th>
<th>EQ-Short Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men (N = 99) (M (SD))</td>
<td>Women (N = 98) (M (SD))</td>
</tr>
<tr>
<td>Estimated Gender Difference</td>
<td>4.92 (0.27)</td>
<td>4.67 (0.39)</td>
</tr>
<tr>
<td>Learned vs. Innate Attributions</td>
<td>5.50 (0.54)</td>
<td>5.26 (0.71)</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>4.95 (0.27)</td>
<td>4.64 (0.35)</td>
</tr>
<tr>
<td>Genetic Differences</td>
<td>3.90 (0.40)</td>
<td>3.80 (0.36)</td>
</tr>
</tbody>
</table>

*Note.* Below Midpoint = Women Higher, More Innate; Above Midpoint = Men Higher, More Learned. *\(p < .05\). **\(p < .01\). ***\(p < .001\). There were no non-binary coders in Study 2.
### Study 2 - Effects by Coder Gender and for Overall Sample

**Table A.3. Effects predicting the magnitude of the gender difference on each item from mean coder ratings among men and women coders separately (Study 2)**

<table>
<thead>
<tr>
<th>Rating</th>
<th>Effects for Men ((N = 99))</th>
<th>Effects for Women ((N = 98))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Main Effect ((\beta))</td>
<td>Rating (\times) Type Interaction?</td>
</tr>
<tr>
<td>Estimated Gender Difference</td>
<td>.37*** (p &lt; .001)</td>
<td>Yes</td>
</tr>
<tr>
<td>Learned vs. Innate Attributions</td>
<td>.10 (p = .341)</td>
<td>No</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>.24* (p = .017)</td>
<td>No</td>
</tr>
<tr>
<td>Genetic Difference</td>
<td>-.05 (p = .538)</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rating</th>
<th>Overall Effect for Reference ((N = 199))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Main Effect ((\beta))</td>
</tr>
<tr>
<td>Estimated Gender Difference</td>
<td>.64*** (p &lt; .001)</td>
</tr>
<tr>
<td>Learned vs. Innate Attributions</td>
<td>.31* (p = .022)</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>.47*** (p &lt; .001)</td>
</tr>
<tr>
<td>Genetic Difference</td>
<td>-.02 (p = .804)</td>
</tr>
</tbody>
</table>
A.3 Study 3

Study 3 - List of Measures

Below is the full list of variables included in Study 3 in the order they were presented to participants. Variables analyzed and reported in the main text are denoted with an asterisk.

- Learned vs. Innate Attributions*
- Estimated Gender Difference*
- The order of the following variables was counterbalanced:
  - Gendered Learning Affordances*
  - Genetic Differences*
- Fixed vs. Growth Mindset
- Gender*
- Age*
- Year Received PhD
- Current Position
- Area of Expertise
- Ethnic/Racial Background*
- Political Orientation*
- SES*
- Honesty Check*
Study 3 - Items for Measures

SQ Activities

- 1. Obtaining specific information about a car’s engine capacity.
- 2. Fixing a problem with the electrical wiring in one’s home.
- 3. Reading articles or web pages about new technology.
- 4. Enjoying games that involve a high degree of strategy.
- 5. Being fascinated by how machines work.
- 6. Being intrigued by the rules and patterns governing numbers in math.
- 7. Understanding instruction manuals for putting appliances together.
- 8. Knowing exact details about a computer’s hard disk drive capacity and processor speed.
- 9. Reading and understanding maps.
- 10. Noticing the details of how a piece of furniture was constructed.
- 11. Learning one’s way around a new city.
- 12. Watching science documentaries on television or reading articles about science and nature.
- 13. Knowing about a stereo’s precise technical features.
- 16. Being curious about the precise way a building was constructed.
- 17. Understanding information the bank sends on different investment and saving systems.
- 18. Wondering exactly how rail networks of trains are coordinated.
- 21. Thinking about how precisely a mountain was formed.
- 22. Visualizing how the motorways in one’s region link up.
- 23. Thinking about the aerodynamics of a plane.
- 24. Knowing the path a river takes from its source to the sea.
- 25. Understanding how wireless communication works.
EQ Activities

(R) = Reverse-scored

Although not preregistered, we reverse-scored EQ Activities 5, 12, and 17 for directional ratings of the gender difference, gendered learning opportunities, and genetic advantage, since these activities are negatively worded (i.e., higher values corresponded with less empathizing). Not reverse-scoring these activities only changes 2 results: for empathizing items, (1) the relationship between estimate of gender difference and the size of the gender difference becomes non-significant, and (2) the relationship between gendered learning affordances and the size of the gender difference becomes non-significant.

● 1. Telling if someone else wants to enter a conversation.
● 2. Caring for other people.
● 3. Knowing what to do in a social situation.
● 4. Judging if something is rude or polite.
● 5. Focusing on one’s own thoughts rather than on what their listener might be thinking.
  (R)
● 6. Picking up quickly if someone says one thing but means another.
● 7. Seeing why some things upset people so much.
● 8. Putting oneself in somebody else's shoes.
● 9. Predicting how someone will feel.
● 10. Spotting when someone in a group is feeling awkward or uncomfortable.
● 11. Seeing why someone should have felt offended by a remark.
● 12. Finding social situations confusing. (R)
● 13. Understanding how other people are feeling and what they are thinking.
● 14. Telling if someone else is interested or bored with what one is saying.
● 15. Being very understanding.
● 16. Sensing if one is intruding.
● 17. Being insensitive. (R)
● 18. Tuning into how someone else feels.
● 19. Working out what another person might want to talk about.
● 20. Telling if someone is masking their true emotion.
● 21. Predicting what someone will do.
● 22. Getting emotionally involved with a friend's problems.
**Honesty Check**

Given that we are only contacting a few experts in psychology to participate in our survey, individual data quality is **extremely** important to our study. We would like to give you the opportunity to withdraw your response if you feel you did not answer these questions to your full abilities.

Should we include your response in our dataset?

Please note: Your response to this question has NO bearing whatsoever on your participation nor eligibility to receive a summary report of the study’s findings.

- **YES**, I answered all questions to the best of my ability. **Please use my data.**
- **NO**, I did not answer all questions to the best of my ability. **Do not use my data.**
Study 3 – Recruitment Strategy

Our preregistered goal was to collect data from $N = 140$ experts from 7 subdisciplines of psychology: (1) Social/Personality, (2) Developmental, (3) Neuro/Cognitive, (4) Clinical, (5) General, (6) Evolutionary, and (7) Gender ($N = 20$ per discipline, 10 men and 10 women per discipline). Although 2 to 3 coders are recommended for typical interrater reliability estimates (Gisev et al., 2013; Lavrakas, 2008), given the subjective nature of our ratings and our wisdom of the crowds approach, we opted to collect $N = 10$ per cross-section of gender and discipline. Assuming a 20% response rate, we originally planned to contact $N = 700$ experts ($N = 100$ per discipline), but actually sent invitations to only 600, as explained below.

We randomly sampled experts from editorial boards of the most influential journals in each subdiscipline of psychology. To start, we generated a list of top 20 journals from SCImago Journal & Country Rank database. To identify influential journals, although an imperfect proxy, we sorted journals by H index for 2019 and subdiscipline. The initial list was reviewed by the research team and journals that did not directly correspond to the subdiscipline in question (e.g., clinical journals in developmental psychology) were removed. We then shared our initial list with experts from each subdiscipline and adjusted our list according to their feedback (i.e., if there were any journals that did not belong on the list, or if there were key journals missing from the list). Next, we cross-checked our revised journal list against journals affiliated with two major psychological organizations: the American Psychological Association and the Association for Psychological Science and added any missing journals in each subcategory. Finally, in order to triangulate on another common journal metric, we sorted our entire list by impact factor and selected the 10 journals with the highest impact factor from each subdiscipline. In cases where
journals appeared twice on different lists, we assigned the journal to the discipline it was most related to. Our final list included $N = 60$ journals.\footnote{Because Gender and Evolutionary Psychology contained fewer journals overall, we opted to sample $N = 5$ journals from each of these specialized but highly relevant categories. All other subdisciplines contained $N = 10$ journals per category.}

Next, as preregistered, we compiled a contact list based on publicly available contact information listed on the editorial board of each journal’s page. Our research team recorded each expert’s name, institution, editorial board position, email address, gender (based on pronouns or other information, when possible), and whether they were a psychologist (verified through their professional webpage or faculty website). Entries were filtered for eligibility (is a psychologist, email located, is an editorial board member) and cross-checked for duplicate email addresses. This resulted in a contact list of $N = 2374$ eligible participants (1081 women, 1293 men). From this list, we randomly sampled $N = 100$ experts (50 women, 50 men) from each subdiscipline. For evolutionary and gender psychology, given that we had a lower number of journals overall, and given there were not enough eligible women in evolutionary psychology, we lowered our sampling to $N = 50$ experts (25 women, 25 men) for these two subdisciplines only (this sampling decision was unforeseen, and our final recruitment target of 600 experts deviated from our preregistered target of 700).

A personalized invitation to participate in the survey was sent out to $N = 600$ experts. A reminder was sent out one week later. $N = 8$ email addresses were unable to receive email, and $N = 4$ participants were unable to complete the survey. To ensure we met our target sample size, we substituted these contacts with $N = 12$ experts randomly sampled from the same discipline and gender as the original participants. In total, we contacted $N = 612$ experts.
### Study 3 – Item-Level Descriptive Statistics

Table A.4. Item-level information for coder ratings for each measure and the p-value for t-tests comparing ratings to scale midpoint (Study 3)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Item Text</th>
<th>Gender Differences</th>
<th>Innate vs. Learned Attributions</th>
<th>Gendered Learning Affordances</th>
<th>Genetic Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Above MP (p)</td>
<td>Below MP (p)</td>
</tr>
<tr>
<td>EMP1</td>
<td>Telling if someone else wants to enter a conversation.</td>
<td>3.29</td>
<td>0.86</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>EMP10</td>
<td>Spotting when someone in a group is feeling awkward or uncomfortable.</td>
<td>3.24</td>
<td>0.74</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>EMP11</td>
<td>Seeing why someone should have felt offended by a remark.</td>
<td>3.20</td>
<td>0.74</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>EMP12</td>
<td>Finding social situations confusing.</td>
<td>4.15</td>
<td>0.60</td>
<td>0.030</td>
<td>0.970</td>
</tr>
<tr>
<td>EMP13</td>
<td>Understanding how other people are feeling and what they are thinking.</td>
<td>3.34</td>
<td>0.67</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>EMP14</td>
<td>Telling if someone else is interested or bored with what one is saying.</td>
<td>3.38</td>
<td>0.68</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>EMP15</td>
<td>Being very understanding.</td>
<td>3.22</td>
<td>0.87</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>EMP16</td>
<td>Sensing if one is intruding.</td>
<td>3.49</td>
<td>0.82</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>EMP17</td>
<td>Being insensitive.</td>
<td>4.73</td>
<td>0.70</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>EMP</td>
<td>Activity Description</td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>t-statistic</td>
<td>p-value</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------</td>
<td>--------------------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>EMP18</td>
<td>Tuning into how Someone else feels.</td>
<td>3.18</td>
<td>0.75</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP19</td>
<td>Working out what another person might want to talk about.</td>
<td>3.62</td>
<td>0.64</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP2</td>
<td>Caring for other people.</td>
<td>3.09</td>
<td>0.73</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP20</td>
<td>Telling if someone is masking their true emotion.</td>
<td>3.48</td>
<td>0.73</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP21</td>
<td>Predicting what Someone will do.</td>
<td>3.90</td>
<td>0.45</td>
<td>0.933</td>
<td>0.067</td>
</tr>
<tr>
<td>EMP22</td>
<td>Getting emotionally involved with a friend's problems.</td>
<td>2.97</td>
<td>0.77</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP3</td>
<td>Knowing what to do in a social situation.</td>
<td>3.50</td>
<td>0.57</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP4</td>
<td>Judging if something is rude or polite.</td>
<td>3.40</td>
<td>0.68</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP5</td>
<td>Focusing on one's own thoughts rather than on what their listener might be thinking.</td>
<td>4.15</td>
<td>0.96</td>
<td>0.131</td>
<td>0.869</td>
</tr>
<tr>
<td>EMP6</td>
<td>Picking up quickly if Someone says one thing but means another.</td>
<td>3.61</td>
<td>0.76</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP7</td>
<td>Seeing why some things upset people so much.</td>
<td>3.14</td>
<td>0.71</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP8</td>
<td>Putting oneself in Somebody else's shoes.</td>
<td>3.27</td>
<td>0.65</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EMP9</td>
<td>Predicting how someone will feel.</td>
<td>3.14</td>
<td>0.69</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>SYS1</td>
<td>Obtaining specific information about a car’s engine capacity.</td>
<td>4.85</td>
<td>0.97</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS10</td>
<td>Noticing the details of how a piece of furniture was constructed.</td>
<td>4.48</td>
<td>0.73</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS11</td>
<td>Learning one’s way around a new city.</td>
<td>4.46</td>
<td>0.65</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS12</td>
<td>Watching science documentaries on television or reading articles about science and nature.</td>
<td>4.27</td>
<td>0.52</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS13</td>
<td>Knowing about a stereo’s precise technical features.</td>
<td>5.11</td>
<td>0.81</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS14</td>
<td>Grasping exactly how odds work in betting.</td>
<td>4.35</td>
<td>0.68</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS15</td>
<td>Being meticulous when carrying out D.I.Y.</td>
<td>4.00</td>
<td>0.58</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>SYS16</td>
<td>Being curious about the precise way a building was constructed.</td>
<td>4.49</td>
<td>0.69</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS17</td>
<td>Understanding information the bank sends on different investment and saving systems.</td>
<td>4.16</td>
<td>0.55</td>
<td>0.022</td>
<td>0.978</td>
</tr>
<tr>
<td>SYS18</td>
<td>Wondering exactly how rail networks of trains are coordinated.</td>
<td>4.87</td>
<td>0.84</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS19</td>
<td>Looking carefully into the quality of a camera lens.</td>
<td>4.35</td>
<td>0.73</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS2</td>
<td>Fixing a problem with the electrical wiring in one’s home.</td>
<td>4.90</td>
<td>0.82</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS20</td>
<td>Being interested in meteorological patterns.</td>
<td>4.24</td>
<td>0.55</td>
<td>0.002</td>
<td>0.998</td>
</tr>
<tr>
<td>SYS21</td>
<td>Thinking about how precisely a mountain was formed.</td>
<td>4.21</td>
<td>0.56</td>
<td>0.003</td>
<td>0.997</td>
</tr>
<tr>
<td>SYS22</td>
<td>Visualizing how the motorways in one’s region link up.</td>
<td>4.54</td>
<td>0.73</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS23</td>
<td>Thinking about the aerodynamics of a plane.</td>
<td>4.73</td>
<td>0.65</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS24</td>
<td>Knowing the path a river takes from its source to the sea.</td>
<td>4.20</td>
<td>0.45</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>SYS25</td>
<td>Understanding how wireless communication works.</td>
<td>4.52</td>
<td>0.79</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS3</td>
<td>Reading articles or web pages about new technology.</td>
<td>4.69</td>
<td>0.84</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS4</td>
<td>Enjoying games that involve a high degree of strategy.</td>
<td>4.51</td>
<td>0.74</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS5</td>
<td>Being fascinated by how machines work.</td>
<td>5.06</td>
<td>0.85</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS6</td>
<td>Being intrigued by the rules and patterns governing numbers in math.</td>
<td>4.68</td>
<td>0.83</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SYS7</td>
<td>Understanding instruction manuals for putting appliances together.</td>
<td>4.42</td>
<td>0.83</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Knowing exact details about a computer’s hard disk drive capacity and processor speed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SYS8</td>
<td>4.78</td>
<td>0.86</td>
<td>0.000</td>
<td>1.000</td>
<td>6.12</td>
</tr>
<tr>
<td>SYS9</td>
<td>Reading and understanding maps.</td>
<td>4.44</td>
<td>0.74</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*Note.* MP = Scale midpoint.
**Study 3 - Descriptives by Coder Gender**

Table A.5. Mean coder ratings among men and women coders separately, as well as the effect size for the gender difference between men and women coders (Study 3)

<table>
<thead>
<tr>
<th>Rating Dimension</th>
<th>SQ-Short Activities</th>
<th>EQ-Short Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men ( N = 52 ) ( M ) (SD)</td>
<td>Women ( N = 64 ) ( M ) (SD)</td>
</tr>
<tr>
<td>Estimated Gender Difference</td>
<td>4.60 (0.33)</td>
<td>4.44 (0.26)</td>
</tr>
<tr>
<td>Learned vs. Innate Attributions</td>
<td>5.18 (0.69)</td>
<td>5.15 (0.80)</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>4.59 (0.34)</td>
<td>4.75 (0.30)</td>
</tr>
<tr>
<td>Genetic Advantage</td>
<td>4.18 (0.12)</td>
<td>4.13 (0.09)</td>
</tr>
</tbody>
</table>

*Note.* Below Midpoint = Women Higher, More Innate; Above Midpoint = Men Higher, More Learned. *\( p < .05 \). **\( p < .01 \). ***\( p < .001 \).
Study 3 - Descriptives by Coder Subdiscipline

Below are mean coder ratings among each subdiscipline separately. A one-way ANOVA comparing ratings by subdiscipline revealed significant differences on all variables across subdiscipline. Notably, gender and evolutionary scholars tended to be most extreme in their perceptions of the etiology of gender differences, with ratings that were more often significantly different than those from other subdisciplines. For example, evolutionary scholars were significantly different than all other subdisciplines in their belief that gender differences in systemizing and empathizing are due to sex-linked genetic advantages. Gender scholars were significantly higher than most subdisciplines in believing that empathizing and systemizing skills were more learnable than innate.
### Table A.6. Mean coder ratings by coder subdiscipline (Study 3)

<table>
<thead>
<tr>
<th>Rating Dimension</th>
<th>Clin (N = 12)</th>
<th>Dev (N = 16)</th>
<th>Evo (N = 15)</th>
<th>Gend (N = 7)</th>
<th>Genr (N = 12)</th>
<th>Neuro (N = 15)</th>
<th>S/P (N = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M \text{ (SD)})</td>
<td>(M \text{ (SD)})</td>
<td>(M \text{ (SD)})</td>
<td>(M \text{ (SD)})</td>
<td>(M \text{ (SD)})</td>
<td>(M \text{ (SD)})</td>
<td>(M \text{ (SD)})</td>
</tr>
<tr>
<td>Estimated Gender Difference</td>
<td>4.40(b)</td>
<td>4.46(b)</td>
<td>4.85(a)</td>
<td>4.43(b)</td>
<td>4.60(ab)</td>
<td>4.43(b)</td>
<td>4.56(ab)</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td>(0.27)</td>
<td>(0.47)</td>
<td>(0.36)</td>
<td>(0.46)</td>
<td>(0.32)</td>
<td>(0.35)</td>
</tr>
<tr>
<td>Learned vs. Innate Attributes</td>
<td>4.68(def)</td>
<td>5.42(bc)</td>
<td>4.27(f)</td>
<td>6.33(a)</td>
<td>5.55(ab)</td>
<td>5.30(bcd)</td>
<td>5.04(bce)</td>
</tr>
<tr>
<td></td>
<td>(1.18)</td>
<td>(0.67)</td>
<td>(0.92)</td>
<td>(0.57)</td>
<td>(0.73)</td>
<td>(0.94)</td>
<td>(0.84)</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>4.37(a)</td>
<td>4.93(b)</td>
<td>4.83(b)</td>
<td>4.99(b)</td>
<td>4.59(ab)</td>
<td>4.66(ab)</td>
<td>4.61(ab)</td>
</tr>
<tr>
<td></td>
<td>(0.57)</td>
<td>(0.61)</td>
<td>(0.36)</td>
<td>(0.52)</td>
<td>(0.45)</td>
<td>(0.46)</td>
<td>(0.31)</td>
</tr>
<tr>
<td>Genetic Advantage</td>
<td>3.88(a)</td>
<td>4.15(bc)</td>
<td>4.53(c)</td>
<td>4.01(ab)</td>
<td>4.20(cd)</td>
<td>4.13(bc)</td>
<td>4.13(bc)</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(0.16)</td>
<td>(0.37)</td>
<td>(0.05)</td>
<td>(0.25)</td>
<td>(0.11)</td>
<td>(0.13)</td>
</tr>
</tbody>
</table>

**Note.** *\(p < .05\). **\(p < .01\). ***\(p < .001\). Clin = Clinical, Dev = Developmental, Evo = Evolutionary, Gend = Gender, Genr = General, Neuro = Neuroscience, S/P = Social/Personality. Within each row, means not sharing the same subscript differ significantly using a Tukey HSD test for multiple comparisons.
**EQ-Short Activities**

<table>
<thead>
<tr>
<th>Rating Dimension</th>
<th>Clin N = 12 M (SD)</th>
<th>Dev N = 16 M (SD)</th>
<th>Evo N = 15 M (SD)</th>
<th>Gend N = 7 M (SD)</th>
<th>Genr N = 12 M (SD)</th>
<th>Neuro N = 15 M (SD)</th>
<th>S/P N = 30 M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated Gender Difference</td>
<td>3.53\textsubscript{a} (0.29)</td>
<td>3.41\textsubscript{ab} (0.35)</td>
<td>3.14\textsubscript{b} (0.43)</td>
<td>3.31\textsubscript{ab} (0.29)</td>
<td>3.23\textsubscript{ab} (0.38)</td>
<td>3.43\textsubscript{ab} (0.35)</td>
<td>3.48\textsubscript{a} (0.24)</td>
</tr>
<tr>
<td>Learned vs. Innate Attributions</td>
<td>3.58\textsubscript{ef} (0.37)</td>
<td>4.59\textsubscript{b} (0.50)</td>
<td>3.22\textsubscript{f} (0.48)</td>
<td>5.44\textsubscript{a} (0.97)</td>
<td>4.55\textsubscript{bc} (0.45)</td>
<td>4.00\textsubscript{de} (0.62)</td>
<td>4.51\textsubscript{bcd} (0.45)</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>3.51\textsubscript{ab} (0.46)</td>
<td>3.34\textsubscript{ab} (0.56)</td>
<td>3.40\textsubscript{ab} (0.47)</td>
<td>3.04\textsubscript{a} (0.89)</td>
<td>3.52\textsubscript{b} (0.45)</td>
<td>3.71\textsubscript{b} (0.41)</td>
<td>3.53\textsubscript{b} (0.23)</td>
</tr>
<tr>
<td>Genetic Advantage</td>
<td>3.78\textsubscript{b} (0.30)</td>
<td>3.82\textsubscript{b} (0.15)</td>
<td>3.24\textsubscript{a} (0.44)</td>
<td>3.97\textsubscript{b} (0.08)</td>
<td>3.71\textsubscript{b} (0.20)</td>
<td>3.77\textsubscript{b} (0.27)</td>
<td>3.84\textsubscript{b} (0.09)</td>
</tr>
</tbody>
</table>

**Note.** *p < .05. **p < .01. ***p < .001. Clin = Clinical, Dev = Developmental, Evo = Evolutionary, Gend = Gender, Genr = General, Neuro = Neuroscience, S/P = Social/Personality. Within each row, means not sharing the same subscript differ significantly using a Tukey HSD test for multiple comparisons.
Study 3 - Effects by Coder Gender and for Overall Sample

Table A.7. Effects predicting the magnitude of the gender difference on each item from mean coder ratings, among men and women coders separately (Study 3)

<table>
<thead>
<tr>
<th>Rating</th>
<th>Effects for Men</th>
<th>Effects for Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(N = 52)$</td>
<td>$(N = 64)$</td>
</tr>
<tr>
<td>Rating Main Effect $(\beta)$</td>
<td>Rating $\times$ Type Interaction?</td>
<td>Simple Slope $(\beta)$</td>
</tr>
<tr>
<td>Estimated Gender Difference</td>
<td>.68*** $p &lt; .001$</td>
<td>No</td>
</tr>
<tr>
<td>Learned vs. Innate Attributions</td>
<td>.18 $p = .100$</td>
<td>No</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>.53*** $p &lt; .001$</td>
<td>No</td>
</tr>
<tr>
<td>Genetic Advantage</td>
<td>.52*** $p &lt; .001$</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rating</th>
<th>Overall Effect for Reference</th>
<th>$(N = 116)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Main Effect $(\beta)$</td>
<td>Rating $\times$ Type Interaction?</td>
</tr>
<tr>
<td>Estimated Gender Difference</td>
<td>.71*** $p &lt; .001$</td>
<td>No</td>
</tr>
<tr>
<td>Learned vs. Innate Attributions</td>
<td>.20 $p = .057$</td>
<td>No</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>.58*** $p &lt; .001$</td>
<td>Yes</td>
</tr>
<tr>
<td>Genetic Advantage</td>
<td>.51 $p &lt; .001$</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Study 3 - Effects by Coder Subdiscipline

Below are effects predicting the magnitude of the gender difference on each item from mean coder ratings among each coder subdiscipline separately, compared to the overall effect for all expert ratings as reported in the main text. In general, social-personality experts tended to have the most accurate predictions, however, the ability to significantly predict the observed gender differences across items and the predictive effect of gender learning affordances on these differences replicated in each of the seven subdisciplines.

Table A.8. Standardized betas for effects for ratings of estimated gender difference by subdiscipline (Study 3)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.31**</td>
<td></td>
<td>.47***</td>
<td>.58***</td>
<td>.36**</td>
<td>.50***</td>
<td>.38***</td>
<td>.58***</td>
<td>.71***</td>
</tr>
<tr>
<td>p = .005</td>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p = .006</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
</tr>
</tbody>
</table>

Note. *p < .05. **p < .01. ***p < .001. Clin = Clinical, Dev = Developmental, Evo = Evolutionary, Gend = Gender, Genr = General, Neuro = Neuroscience, S/P = Social/Personality

Table A.9. Standardized betas for effects for ratings of learned versus innate attributions by subdiscipline (Study 3)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.18</td>
<td>.14</td>
<td>.16</td>
<td>.08</td>
<td>.00</td>
<td>.14</td>
<td>.21*</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>p = .218</td>
<td>p = .113</td>
<td>p = .136</td>
<td>p = .327</td>
<td>p = .994</td>
<td>p = .146</td>
<td>p = .019</td>
<td>p = .057</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05. **p < .01. ***p < .001. Clin = Clinical, Dev = Developmental, Evo = Evolutionary, Gend = Gender, Genr = General, Neuro = Neuroscience, S/P = Social/Personality
Table A.10. Standardized betas for effects for ratings of gendered learning affordances by subdiscipline (Study 3)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>.24**</td>
<td>.35**</td>
<td>.30*</td>
<td>.41***</td>
<td>.33***</td>
<td>.29**</td>
<td>.69***</td>
<td>.58***</td>
</tr>
<tr>
<td>p</td>
<td>p = .005</td>
<td>p = .001</td>
<td>p = .029</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p = .003</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
</tr>
</tbody>
</table>

Note. *p < .05. **p < .01. ***p < .001. Clin = Clinical, Dev = Developmental, Evo = Evolutionary, Gend = Gender, Genr = General, Neuro = Neuroscience, S/P = Social/Personality

Table A.11. Standardized betas for effects for ratings of genetic advantage by subdiscipline (Study 3)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>.09</td>
<td>.23*</td>
<td>.23</td>
<td>.12</td>
<td>.08</td>
<td>.17</td>
<td>.31**</td>
<td>.51***</td>
</tr>
</tbody>
</table>

Note. *p < .05. **p < .01. ***p < .001. Clin = Clinical, Dev = Developmental, Evo = Evolutionary, Gend = Gender, Genr = General, Neuro = Neuroscience, S/P = Social/Personality
A.4 Studies 2-3

Studies 2-3 – Effects Modeled in MLM

The following tables compare effect sizes modeled in an item-level GLM (as reported in the main text) and MLM framework. MLM analyses predict the size of the gender difference (Study 1) from coder ratings (Studies 2-3), with ratings grand mean centered and nested within coder and subdiscipline (Study 3 only). Since prior models specifying random slopes failed to converge, MLM models specify random intercepts and fixed slopes. MLM analyses are conducted using the R package lmer4 (version 1.1-21; Bates et al., 2015).

While the magnitude of effects generally tracked across GLM and MLM analyses (i.e., in the competing predictor analysis, the effect size for genetic advantage was approximately half the size of the effect of gendered learning affordances in both GLM and MLM), in two cases the MLM analysis was significant where the GLM analysis was not significant. We suspect this is due to the higher sample size for the MLM data (N= 2,975 in Study 2, N= 2,784 in Study 3).

Table A.12. Single predictor analyses comparing GLM and MLM estimates (Studies 2 and 3)

<table>
<thead>
<tr>
<th></th>
<th>Study 2</th>
<th>Study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GLM (β)</td>
<td>MLM (b)</td>
</tr>
<tr>
<td>Estimated Gender Difference</td>
<td>.64*** p &lt; .001</td>
<td>.02*** p &lt; .001</td>
</tr>
<tr>
<td></td>
<td>.71*** p &lt; .001</td>
<td>.06*** p &lt; .001</td>
</tr>
<tr>
<td>Learned vs. Innate</td>
<td>.31* p = .022</td>
<td>.008*** p &lt; .001</td>
</tr>
<tr>
<td>Attributions</td>
<td>.20 p = .057</td>
<td>.008*** p &lt; .001</td>
</tr>
<tr>
<td>Gendered Learning</td>
<td>.47*** p &lt; .001</td>
<td>.02*** p &lt; .001</td>
</tr>
<tr>
<td>Affordances</td>
<td>.58*** p &lt; .001</td>
<td>.05*** p &lt; .001</td>
</tr>
<tr>
<td>Genetic Differences/Advantage</td>
<td>.02 p = .804</td>
<td>-.001 p = .594</td>
</tr>
<tr>
<td></td>
<td>.51*** p &lt; .001</td>
<td>.03*** p &lt; .001</td>
</tr>
<tr>
<td>N = 47</td>
<td>N = 2,975</td>
<td>N = 47</td>
</tr>
</tbody>
</table>
Table A.13. Competing predictor analysis comparing GLM and MLM estimates (Study 3)

<table>
<thead>
<tr>
<th></th>
<th>Study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GLM</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
</tr>
<tr>
<td>Gendered Learning Affordances</td>
<td>.57**</td>
</tr>
<tr>
<td>Genetic Advantage</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>$N = 47$</td>
</tr>
</tbody>
</table>

Appendix B

This section contains supplementary materials for Chapter 3: Are Essentialized Explanations for Gender Differences in interest Self-fulfilling?

B.1 Stimuli Creation and Pilot Testing

Target Stimuli Selection for Studies 4-5

We selected photos from the Chicago Face Database (Ma et al., 2015, 2020) that had been pre-rated on masculinity and femininity. For the $N = 3$ faces within each race and gender category, we selected $N = 1$ highly masculine face (highest masculine - feminine difference score), $N = 1$ neutral face (masculine - feminine difference score closest to zero), and $N = 1$ highly feminine face (lowest masculine - feminine difference score). Additionally, to ensure targets were believable as interns, all targets were rated as appearing between 20 and 30 years old. Finally, to control for perceptions of attractiveness, all targets were rated between 3 and 5 on attractiveness using a 7-point scale.
Intern images used with permission from Chicago Face Database (Ma et al., 2015, 2020).

Figure B.1. Target photos shown to participants (Studies 4-5)
Project Team Development

To develop empathizing- and systemizing-specific project team descriptions, we first compiled the Baron-Cohen research group’s definitions of empathizing and systemizing. For each definition, we identified key verbs and subjects, as shown in the table below.

Table B.1. Compilation of empathizing and systemizing definitions for team descriptions in Studies 4, 5, and 6

<table>
<thead>
<tr>
<th>Source</th>
<th>Year</th>
<th>Empathizing Definition</th>
<th>Systemizing Definition</th>
<th>Empathizing</th>
<th>Systemizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baron-Cohen (2003)</td>
<td>2003</td>
<td>---</td>
<td>Systemizing is the drive to analyse the variables in a system, to derive the underlying rules that govern the behaviour of a system. Systemizing also refers to the drive to construct systems. Systemizing allows you to predict the behaviour of a system, and to control it [...] A system is defined as something that takes inputs, which can then be operated on in variable ways, to deliver different outputs in a rule-governed way.</td>
<td>---</td>
<td>analyze</td>
</tr>
<tr>
<td>Baron-Cohen &amp; Wheelwright (2004)</td>
<td>2004</td>
<td>Empathy is the drive or ability to attribute mental states to another person/animal, and entails an appropriate affective response in the observer to the other person's mental state.</td>
<td>---</td>
<td>attribute</td>
<td>derive</td>
</tr>
<tr>
<td>Wheelwright et al. (2006)</td>
<td>2006</td>
<td>the drive to identify another person's emotions and thoughts, and to respond to these with an appropriate emotion</td>
<td>the drive to analyze, understand, predict, control and construct rule-based systems</td>
<td>respond</td>
<td>system</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wakabayashi et al. (2006)</td>
<td>2006</td>
<td>The drive to identify emotions and thoughts in others and to respond to these with an appropriate emotion [...] used for making sense of an agent’s behaviour.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenberg et al. (2018)</td>
<td>2018</td>
<td>The ability to recognize another person’s mental state (“cognitive empathy”) and the drive to respond to it with an appropriate emotion.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baron-Cohen (2009, 2016)</td>
<td>2009, 2016</td>
<td>ToM is just the cognitive component of empathy [it simply involves identifying someone else’s (or your own) mental states ... sometimes also referred to as requiring an attribution (since these are ultimately a postulate—mental states are not visible per se) or requiring recognition (if the mental state leaves cues in facial or vocal or postural expressions of emotion, for example)]. The second component of empathy is the response element: having an appropriate emotional reaction to another person’s thoughts and feelings. This is referred to as affective empathy.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table:**

- Identify
- Respond
- Make sense
- Emotions
- Thoughts
- Behavior
- Construct
- Predict
- Control
- Systems: mechanical
- Abstract
- Organizable
- Behavior
- Events
- Objects
Next, we compiled the verbs and subjects identified above into a dictionary. To include more variation in our descriptions, we also included synonyms of the original verbs and referents (synonyms located using thesaurus.com). Finally, we eliminated verbs that were redundant across both empathizing and systemizing (e.g., *identify*). We used the resulting verbs and referents shown in the table below to construct our team descriptions.

Table B.2. Compilation of empathizing and systemizing verbs and subjects for team descriptions in Studies 4, 5, and 6

<table>
<thead>
<tr>
<th>Empathizing</th>
<th>Systemizing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Verbs</strong> (Original)</td>
<td><strong>Subjects</strong> (Original)</td>
</tr>
<tr>
<td>attribute</td>
<td>animal</td>
</tr>
<tr>
<td><strong>identify redundant</strong></td>
<td>behavior</td>
</tr>
<tr>
<td>make sense of</td>
<td>emotions</td>
</tr>
<tr>
<td>react</td>
<td>expression</td>
</tr>
<tr>
<td>recognize</td>
<td>feelings</td>
</tr>
<tr>
<td>respond</td>
<td>mental state</td>
</tr>
<tr>
<td></td>
<td>person</td>
</tr>
<tr>
<td></td>
<td>thoughts</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Verbs</strong> (Synonyms)</td>
<td><strong>Subjects</strong> (Synonyms)</td>
</tr>
<tr>
<td>spot</td>
<td>perspective</td>
</tr>
<tr>
<td>characterize</td>
<td></td>
</tr>
<tr>
<td>depict</td>
<td>inspect</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>express</td>
<td>study</td>
</tr>
<tr>
<td>interpret</td>
<td>test</td>
</tr>
<tr>
<td>communicate</td>
<td>formulate</td>
</tr>
<tr>
<td>acknowledge</td>
<td>organize</td>
</tr>
<tr>
<td>reciprocate</td>
<td>invent</td>
</tr>
<tr>
<td>observe</td>
<td>classify</td>
</tr>
</tbody>
</table>
**Project Teams**

**Table B.3. Initial project teams developed for Studies 4, 5, and 6**

<table>
<thead>
<tr>
<th>Team Name</th>
<th>Empathizing</th>
<th>Systemizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client Strategy</td>
<td>Make sense of clients’ needs and interpret which deliverables are most important to them</td>
<td>Evaluate project requirements and build lists of deliverables organized by priority status</td>
</tr>
<tr>
<td>Information Oversight</td>
<td>Recognize how people share information in the web application and communicate how their process can be improved</td>
<td>Analyze information pathways in the web application and construct more efficient task response systems</td>
</tr>
<tr>
<td>Market Forecasting</td>
<td>Observe behavior and interpret what people want in a product so the company can react to customer needs</td>
<td>Build an algorithm to predict systematic trends in the market so the company can control its marketing strategy</td>
</tr>
<tr>
<td>Web Content</td>
<td>Recognize site visitors’ needs and express content that increases their desire to visit the product site</td>
<td>Inspect site requirements and build content that increases web traffic to the product site</td>
</tr>
<tr>
<td>Product Testing</td>
<td>Improve the web application by making sense of users’ perspectives and communicating them to the product designers</td>
<td>Improve the web application by evaluating usability errors and classifying core product issues</td>
</tr>
<tr>
<td>Database Security</td>
<td>Acknowledge clients’ security concerns and respond to employees’ requests for access to company databases</td>
<td>Analyze project security vulnerabilities and control access permissions to company databases</td>
</tr>
</tbody>
</table>

Icon Credits (Empathizing, Top to Bottom): Thinking © ArtWorkLeaf; Social Networking © I Putu Kharismayadi; Person Love © Round Icons; User © jamhuden; Usability © Adrien Coquet; Family © Widya Bayu W. Icon Credits (Systemizing, Top to Bottom): List © Adrien Coquet; Share © Three Six Five; Web Traffic © Saculp Nahwan; Website © ibrandify; Prototype © Aris Sunjay; Security © Setyo Ari Wibowo. All icons retrieved from the Noun Project and licensed under CC-BY 3.0; all icons were modified by the lead author from black and white to color icons.
Pilot Study 1

We pilot tested key features of our intern assignment paradigm to be used in Studies 4-5 among \( N = 218 \) STEM professionals.

Project Manager Description

Participants first received the following instructions:

You have been assigned as the Project Manager for Project Point. Project Point aims to develop a new web application that helps clients manage and track incoming orders. The project tasks include:

- Meeting with prospective clients
- Establishing the product's interaction design
- Evaluating the usability of the product
- Ensuring the project meets security standards
- Creating a dedicated website to host the product
- Marketing the product

Participants rated the project manager role as slightly more suitable for men (\( M = 6.01, \ SD = 1.14 \)) than women (\( M = 5.78, \ SD = 1.31; \ d = .19 \)) but follow-up analyses revealed this was driven by men (\( N = 116, \ d = .25 \)) compared to women (\( d = .06 \)), suggesting an in-group bias. Thus, we were comfortable retaining this job description for Studies 4 and 5.

Task Instructions

Next, participants learned a group of 30 interns would be joining the company and were instructed to either assign them to teams that provide them the opportunity to learn new skills or strengthen their existing skillsets (1 of 2 instruction sets randomly assigned). They then read about 30 fictional interns (15 User Experience, with empathizing-relevant skills; 15 Software Engineering, with systemizing-relevant skills). Interns were presented via gender-neutral icons.

Project Teams

Next, participants learned about 6 different project teams in 1 of 2 frames. In Frame 1, 3 teams (Client Strategy, Information Oversight, Market Forecasting) were framed as empathizing and 3 teams (Product Testing, Database Security, Web Content) were framed as systemizing. In Frame 2, teams were flipped, such that empathizing teams were framed as systemizing and vice versa.
<table>
<thead>
<tr>
<th>Client Strategy</th>
<th>Make sense of clients' needs and interpret which deliverables are most important to them</th>
<th>Client Strategy</th>
<th>Evaluate project requirements and build lists of deliverables organized by priority status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Oversight</td>
<td>Recognize how people share information in the web application and communicate how their process can be improved</td>
<td>Information Oversight</td>
<td>Analyze information pathways in the web application and construct more efficient task response systems</td>
</tr>
<tr>
<td>Product Testing</td>
<td>Improve the web application by evaluating usability errors and classifying core product issues</td>
<td>Product Testing</td>
<td>Improve the web application by making sense of users' perspectives and communicating them to the product designers</td>
</tr>
<tr>
<td>Database Security</td>
<td>Analyze project security vulnerabilities and control access permissions to company databases</td>
<td>Database Security</td>
<td>Acknowledge clients' security concerns and respond to employees' requests for access to company databases</td>
</tr>
<tr>
<td>Market Forecasting</td>
<td>Observe behavior and interpret what people want in a product so the company can react to customer needs</td>
<td>Market Forecasting</td>
<td>Build an algorithm to predict systematic trends in the market so the company can control its marketing strategy</td>
</tr>
<tr>
<td>Web Content</td>
<td>Inspect site requirements and build content that increases web traffic to the product site</td>
<td>Web Content</td>
<td>Recognizes site visitors' needs and expresses content that increases their desire to visit the product site</td>
</tr>
</tbody>
</table>

Figure B.2. Both possible sets of project teams (Pilot Study 1)

To test whether our framing manipulation was successful, we asked participants how much each team might benefit from face-to-face meetings (expected to be higher for empathizing teams). Results indicated that Product Testing and Market Forecasting teams were rated as more likely to benefit from meetings when framed as empathizing (vs. systemizing). All other teams (Client Strategy, Information Oversight, Database Security, Web Content) were rated in the predicted direction but were not significantly different.
Item: To what extent would the following teams benefit from face-to-face meetings?
Scale: 1 = Not at all, 7 = Very much

Table B.4. Ratings of teams’ benefit from face-to-face meetings (Pilot Study 1)

| Team                | Frame 1 | | Frame 2 | | Difference (d) |
|---------------------|---------|---------|---------|----------------|
|                     | Type    | M (SD)  | Type    | M (SD)         |                |
| Client Strategy     | Emp     | 5.46 (1.36) | Sys     | 5.21 (1.57)   | .17            |
| Information Oversight | Emp   | 4.44 (1.60) | Sys     | 4.20 (1.77)   | .15            |
| Product Testing     | Sys     | 4.52 (1.80) | Emp     | 5.10 (1.80)   | .32*           |
| Database Security   | Sys     | 3.63 (1.80) | Emp     | 4.08 (1.94)   | .24            |
| Market Forecasting  | Emp     | 4.81 (1.74) | Sys     | 4.27 (1.82)   | .31*           |
| Web Content         | Sys     | 3.79 (1.82) | Emp     | 4.09 (1.83)   | .17            |

Ratings made at the end of the study suggested that Client Strategy and Database Security were consistently rated as relevant to empathizing or systemizing, regardless of team framing (see chart below). Examining open-ended responses suggested that participants categorized these teams as “front end” and “back end” teams and were thus more influenced by this heuristic than the job description we provided.

Figure B.3. Ratings of team relevance to empathizing and systemizing (Pilot Study 1)
Finally, examining quantitative and open-ended responses confirmed that all teams were rated as realistic (4.8 or higher on a 7-point scale). There was one significant difference across team frames: information oversight was rated as more realistic when framed as a systemizing than empathizing team. However, the empathizing-framed version of the team was still rated as being fairly realistic.

**Rating:** Realistic  
**Scale:** 1 = Not at all, 7 = Very much

### Table B.5. Ratings of team realism (Pilot Study 1)

<table>
<thead>
<tr>
<th>Team Assignment Task</th>
</tr>
</thead>
</table>
| We presented participants with one intern at a time and asked them to select 2 of 6 teams that would provide that intern the opportunity to learn new skills or strengthen their existing skillsets (depending on their assigned instruction set). To score each trial, for each empathizing team selected, we assigned a score of +1; for each systemizing team selected, we assigned a score of -1. For example, if a participant selected 2 empathizing teams for an intern, their score for that trial would be +2. On this measure, positive scores mean they assigned more empathizing teams; negative scores mean more systemizing teams. There were 30 trials total (1 per intern).

Collapsing across frame, results supported our predictions: participants instructed to help interns learn new skills assigned both types of interns to both types of teams relatively equally, and those instructed to help interns strengthen existing skills assigned user experience interns to more empathizing, and software engineering interns to more systemizing, teams.

<table>
<thead>
<tr>
<th>Team</th>
<th>Frame 1</th>
<th>Frame 2</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type</td>
<td>M (SD)</td>
<td>Type</td>
</tr>
<tr>
<td>Client Strategy</td>
<td>Emp</td>
<td>5.08 (1.15)</td>
<td>Sys</td>
</tr>
<tr>
<td>Information Oversight</td>
<td>Emp</td>
<td>5.06 (1.15)</td>
<td>Sys</td>
</tr>
<tr>
<td>Product Testing</td>
<td>Sys</td>
<td>5.23 (1.21)</td>
<td>Emp</td>
</tr>
<tr>
<td>Database Security</td>
<td>Sys</td>
<td>5.41 (1.31)</td>
<td>Emp</td>
</tr>
<tr>
<td>Market Forecasting</td>
<td>Emp</td>
<td>4.83 (1.29)</td>
<td>Sys</td>
</tr>
<tr>
<td>Web Content</td>
<td>Sys</td>
<td>5.11 (1.18)</td>
<td>Emp</td>
</tr>
</tbody>
</table>
Table B.6. Participant team assignments by instruction set (Pilot Study 1)

<table>
<thead>
<tr>
<th></th>
<th>Learn New Skills</th>
<th>Strengthen Existing</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Experience Interns</td>
<td>-.03 (.80)</td>
<td>+.23 (.87)</td>
<td>.31*</td>
</tr>
<tr>
<td>Software Engineering Interns</td>
<td>-.03 (.85)</td>
<td>-.34 (.94)</td>
<td>.34*</td>
</tr>
</tbody>
</table>

However, these results were moderated by frame (moderation for effect with User Experience Interns, $\beta = 1.28, p < .001$; for Software Engineering Interns, $\beta = -1.73, p < .001$). The predicted effect emerged in Frame 1 (in which Client Strategy—a “front-end” team—was framed as an empathizing team but Database Security—a “back-end” team—was framed as a systemizing team) but was supported in the opposite direction in Frame 2. Analyzing Client Strategy and Database Security collapsing across frame supported our predictions: regardless of frame, participants instructed to assign interns to learn new skills assigned interns to Client Strategy and Database Security equally, and those instructed to assign interns to strengthen existing skills assigned user experience interns to Client Strategy more often, and software engineering interns to Database Security more often. This effect was not moderated by frame. Based on this follow-up analysis, we concluded the task worked as expected but we needed to revise the teams.

Summary and Key Decisions

The results of our first pilot study confirmed our behavioral paradigm was able to distinguish between participants instructed to give interns opportunities to learn new skills (vs. strengthen existing skills). Broadly, we conceive of these psychological markers as mapping onto the behavioral correlates of those lowest (vs. highest) in essentialism. Additionally, all teams were rated as being realistic. Yet the analyses above suggested we needed to revise our teams to distinguish between front-end and back-end teams more clearly. Additionally, we also changed the language in follow-up pilot studies to refer to interests rather than skills.
Pilot Study 2A

Pilot Study 2A was a replication of Pilot Study 1 with \( N = 190 \) STEM professionals. For parsimony, we note only changes to the procedure from Pilot Study 1.

**Task Instructions**

We reduced the number of interns from 30 to 6 (3 User Experience, 3 Software Engineering). Additionally, rather than being instructed to either help interns develop new skills or strengthen their existing skills, we revised language to refer to interests. Participants were instructed to either help interns develop new interests or match their existing interests.

**Project Teams**

We designated 4 teams (Client Strategy, People Support, Database Security, Server Admin) as front-end or back-end teams that did not vary across frame. We designated 4 teams (Communication/Information Oversight, Site Visitor/Traffic Auditing, Market/Customer Trends, Product/Human Testing) as empathizing or systemizing teams that varied in their name and description across frame. These are the same teams used in Pilot Study 2B. In Study 4, teams are identical except we changed “Site Visitor/Traffic Auditing” to “Site Visitor/Auditing Needs”

**Team Assignment Task**

We simplified the structure of the team assignment task in our second pilot study so that, for each intern, participants chose between 4 pairs of teams across 4 trials (choices were always between a front-end/back-end or empathizing/systemizing team).
Empathizing / Systemizing Team Choices

Front-End / Back-End Team Choices

Icon Credits (First Row, Left to Right): Social Networking © I Putu Kharismayadi; Web Traffic © Saepul Nahwan; Share © Three Six Five; Person Love © Round Icons. Icon Credits (Second Row, Left to Right): User © jamhuden; Prototype © Aris Sunjaya; Website © ibrandify; Usability © Adrien Coquet. Icon Credits (Third Row, Left to Right): Team © HASTA ICON; Security © Setyo Ari Wibowo. Icon Credits (Fourth Row, Left to Right): Thinking © ArtWorkLeaf; Server © Rolas Design. All icons retrieved from the Noun Project and licensed under CC-BY 3.0; all icons were modified by the lead author from black and white to color icons.

Figure B.4. Team assignment task choices (Pilot Study 2A)

Results in Pilot Study 2A collapse across empathizing/front-end and systemizing/back-end teams; results did not differ when looking at empathizing/systemizing and front-end/back-end choices separately.

When told to match to their existing interests, people tend to assign User Experience interns to more empathizing teams, but when told to help them develop new interests, people tend to assign them to more systemizing teams (difference between conditions: $d = 1.58$, $p < .001$).

When told to match to their existing interests, people tend to assign Software Engineering interns to more systemizing teams, but when told to help them develop new interests, people tend to assign them to more empathizing teams (difference between conditions: $d = 1.82$, $p < .001$).

Summary and Key Decisions

These results verified our task was working as planned in the context of User Experience and Software Engineering interns. In Pilot Study 2B, we tested whether participants would show the same patterns when presented with women and men interns (as we planned to do in Study 4).
Pilot Study 2B

Pilot Study 2B was a replication of Pilot Study 1 with $N = 197$ STEM professionals. All features of Pilot Study 2B were identical to Pilot Study 2A, except that instead of User Experience and Software Engineering interns, we presented participants with $N = 3$ women (1 masculine, 1 neutral, 1 feminine) and $N = 3$ men (1 masculine, 1 neutral, 1 feminine) interns.

We predicted that, like results for User Experience interns, participants would assign women to more empathizing teams when instructed to match interns to their existing interests (vs. help them develop new interests). We predicted a parallel pattern for Software Engineering interns and men.

Within All Teams

When told to match to their existing interests, people tend to assign men interns to more systemizing teams, but when told to help them develop new interests, people tend to assign them to marginally more empathizing teams (difference between conditions: $d = .44, p = .004$). This supported our prediction.

When told to match to their existing interests, people tend to assign women to both types of teams equally, and when told to help them develop new interests, people tend to assign them to slightly more empathizing teams (difference between conditions: $d = .11$, n.s.). This did not support our prediction.

Interestingly, separating these effects out into empathizing/systemizing and front-end/back-end teams revealed different patterns for men and women.

Within Empathizing vs. Systemizing Teams

When told to match to their existing interests, people tend to assign men interns to more systemizing teams, but when told to help them develop new interests, people tend to assign them to marginally more empathizing teams (difference between conditions: $d = .50, p < .001$). This supported our prediction.

When told to match to their existing interests, people tend to assign women to marginally more systemizing teams, and when told to help them develop new interests, people tend to both types of teams equally (difference between conditions: $d = .21$, n.s.). This did not support our prediction.

Within Front-End vs. Back-End Teams

When told to match to their existing interests, people tend to assign men interns to front-end and back-end teams equally, and when told to help them develop new interests, people tend to assign them to both types of teams equally (difference between conditions: $d = .18$, n.s.).
When told to match to their existing interests, people tend to assign women to significantly more front-end teams, and when told to help them develop new interests, people also tend to assign them to significantly more front-end teams (difference between conditions: $d = .01$, n.s.).

**Summary and Key Decisions**

The results of Pilot Study 2B were inconclusive. One limitation of this study was that we did not require participants to choose between a man or woman intern for the same team, as might be the case in real-world hiring contexts. In this version of the task, participants could technically complete the task by assigning the same options to all interns, making any potential effect of gender bias more difficult to detect. In Study 4, we remedied this by restructuring the task as a person-level, rather than team-level decision (i.e., for each team, they were asked to choose between a man and woman intern).
Pilot Study 2A and 2B Combined Team Ratings

To maximize statistical power, we combined team ratings across Pilot Study 2A and Pilot Study 2B (N = 387 total).

Indicating an improvement based on team changes after Pilot Study 1, empathizing/front-end teams were consistently rated as benefitting more from face-to-face meetings compared to systemizing/back-end teams. There were no significant differences by study (2A vs. 2B).

**Item:** To what extent would the following teams benefit from face-to-face meetings?

**Scale:** 1 = Not at all, 7 = Very much

**Table B.7. Ratings of teams’ benefit from face-to-face meetings (Pilot Studies 2A and 2B)**

<table>
<thead>
<tr>
<th>Team</th>
<th>Frame 1</th>
<th></th>
<th>Frame 2</th>
<th></th>
<th>Difference (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type</td>
<td>M (SD)</td>
<td>Type</td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>[Communication/Information] Oversight</td>
<td>Emp</td>
<td>4.38 (1.79)</td>
<td>Sys</td>
<td>3.73 (1.76)</td>
<td>.37***</td>
</tr>
<tr>
<td>[Market/Customer] Trends</td>
<td>Emp</td>
<td>4.82 (1.61)</td>
<td>Sys</td>
<td>3.97 (1.89)</td>
<td>.48***</td>
</tr>
<tr>
<td>Site [Visitor/Traffic] Auditing</td>
<td>Emp</td>
<td>4.42 (1.80)</td>
<td>Sys</td>
<td>3.24 (1.88)</td>
<td>.64***</td>
</tr>
<tr>
<td>[Product/Human] Testing</td>
<td>Emp</td>
<td>5.34 (1.38)</td>
<td>Sys</td>
<td>4.30 (1.88)</td>
<td>.63***</td>
</tr>
<tr>
<td>Client Strategy</td>
<td>Emp</td>
<td>5.04 (1.48)</td>
<td>Sys</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>People Support</td>
<td>Emp</td>
<td>5.16 (1.58)</td>
<td>Sys</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Database Security</td>
<td>Emp</td>
<td>N/A</td>
<td>Sys</td>
<td>3.31 (1.92)</td>
<td>N/A</td>
</tr>
<tr>
<td>Server Admin</td>
<td>Emp</td>
<td>N/A</td>
<td>Sys</td>
<td>3.42 (1.96)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

With one exception for Site Visitor Auditing (seen as more relevant to systemizing than empathizing, d = .31, p = .003), all empathizing (or front-end) teams were rated as significantly more relevant to empathizing than systemizing (all p’s < .001) and all systemizing (or back-end) teams were rated as significantly more relevant to systemizing than empathizing (all p’s < .001).

Additionally, there were two differences by study: participants in Pilot Study 2B rated Information Oversight and Site Traffic Auditing as being less relevant to systemizing than
participants in Pilot Study 2A (Information Oversight: $d = .30, p = .044$, Site Traffic Auditing: $d = .30, p = .046$). The same effects hold when looking at Pilot Study 2B separately.

![Figure B.5. Ratings of team relevance to empathizing and systemizing (Pilot Studies 2A and 2B)](image)

Finally, examining quantitative and open-ended responses confirmed that all teams were rated as realistic (5.5 or higher on a 7-point scale). Customer trends was rated as more realistic than market trends, and product testing was also rated as more realistic than human testing. Despite these descriptive differences, all teams were rated as being moderately realistic. There were no significant differences by study on these ratings.
Rating: Realistic
Scale: 1 = Not at all, 7 = Very much

Table B.8. Ratings of team realism (Pilot Studies 2A and 2B)

<table>
<thead>
<tr>
<th>Team</th>
<th>Frame 1</th>
<th>Frame 2</th>
<th>Difference (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Communication/Information] Oversight</td>
<td>Emp</td>
<td>Sys</td>
<td>.06</td>
</tr>
<tr>
<td>[Market/Customer] Trends</td>
<td>Emp</td>
<td>Sys</td>
<td>.32**</td>
</tr>
<tr>
<td>Site [Visitor/Traffic] Auditing</td>
<td>Emp</td>
<td>Sys</td>
<td>.08</td>
</tr>
<tr>
<td>[Product/Human] Testing</td>
<td>Emp</td>
<td>Sys</td>
<td>.50***</td>
</tr>
<tr>
<td>Client Strategy</td>
<td>Emp</td>
<td>Sys</td>
<td>N/A</td>
</tr>
<tr>
<td>People Support</td>
<td>Emp</td>
<td>Sys</td>
<td>N/A</td>
</tr>
<tr>
<td>Database Security</td>
<td>Emp</td>
<td>Sys</td>
<td>N/A</td>
</tr>
<tr>
<td>Server Admin</td>
<td>Emp</td>
<td>Sys</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Summary and Key Decisions

The results of Pilot Study 2A and 2B confirmed our revised teams were rated as more relevant to empathizing than systemizing (when framed as empathizing) and more relevant to systemizing than empathizing (when framed as systemizing). There was one exception, which was that Site Visitor Auditing (an empathizing team) was framed as more relevant to systemizing than empathizing. To make empathizing more salient, we changed this team’s name to “Site [Visitor/Traffic] Needs.”
Pilot Study 3

The goal of Pilot Study 3 was to test the efficacy of two different manipulations (article prime, third-party prejudice prime) to prime biological vs. social determinism in Study 5. In the article condition, we also tested whether the option to choose the article from a set including neutral controls (vs. being automatically assigned to the article) affected the efficacy of the manipulation. Thus, this pilot study also provided us insight into how the articles operated relative to a control article on an unrelated topic. We tested this with $N = 207$ men working in STEM, recruited from Prolific Academic.

Table B.9. Cell sizes by condition (Pilot Study 3)

<table>
<thead>
<tr>
<th>Final Manipulation</th>
<th>Assigned Manipulation</th>
<th>BIO Article</th>
<th>SOC Article</th>
<th>BIO TPP</th>
<th>SOC TPP</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO Article (Assigned)</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BIO Article</td>
</tr>
<tr>
<td>BIO Article (Chosen)</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$N = 34$</td>
</tr>
<tr>
<td>SOC Article (Assigned)</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SOC Article</td>
</tr>
<tr>
<td>SOC Article (Chosen)</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$N = 36$</td>
</tr>
<tr>
<td>Control Article (Chosen)</td>
<td>15</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td>Control Article</td>
</tr>
<tr>
<td>BIO TPP</td>
<td></td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td>BIO TPP $N = 50$</td>
</tr>
<tr>
<td>SOC TPP</td>
<td></td>
<td></td>
<td></td>
<td>52</td>
<td></td>
<td>SOC TPP $N = 52$</td>
</tr>
</tbody>
</table>

Note. TPP = Third-Party Prejudice

57.69% of people assigned to the Biological, and 60.71% of people assigned to the Social, article chose the control. Thus, there did not appear to be different rates of choosing the control when presented with either option. Participants who selected the non-relevant article were retained in the dataset as a control group.
Table B.10. Descriptive Differences by Condition on Key Variables (Pilot Study 3)

<table>
<thead>
<tr>
<th></th>
<th>Third-Party Prejudice (N = 102)</th>
<th>Article Assigned (N = 48)</th>
<th>Article Choice Prime Selected (N = 22)</th>
<th>Article Choice Prime NOT Selected (N = 32)</th>
<th>Control Mean (SD)</th>
<th>Omnibus Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Spent Reading (Seconds)</td>
<td>BIO Mean (SD)</td>
<td>SOC Mean (SD)</td>
<td>BIO Mean (SD)</td>
<td>SOC Mean (SD)</td>
<td>BIO Mean (SD)</td>
<td>F=13.465***</td>
</tr>
<tr>
<td></td>
<td>104.4 (109.82)</td>
<td>122.58 (137.74)</td>
<td>318.67 (159.03)</td>
<td>286.26 (157.5)</td>
<td>453.72 (338.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>318.67 (318.67)</td>
<td>286.26 (286.26)</td>
<td>453.72 (453.72)</td>
<td>275.54 (275.54)</td>
<td>240.09 (184.58)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F=1.989*</td>
<td>F=3.497***</td>
<td>F=1.713</td>
<td>F=1.256</td>
<td>F=1.827*</td>
<td></td>
</tr>
<tr>
<td>Perceptions of Gender Differences in Ability</td>
<td>4.14 (1.64)</td>
<td>3.96 (1.58)</td>
<td>3.45 (1.71)</td>
<td>3.24 (1.88)</td>
<td>2.91 (1.58)</td>
<td>3.18 (0.98)</td>
</tr>
<tr>
<td>Perceptions of Gender Differences in Interest</td>
<td>4.46 (1.28)</td>
<td>4.96 (1.29)</td>
<td>5.18 (1.59)</td>
<td>5.52 (1.19)</td>
<td>5.45 (1.44)</td>
<td>5.64 (0.92)</td>
</tr>
<tr>
<td>Own BD</td>
<td>3.93 (1.69)</td>
<td>3.57 (1.54)</td>
<td>4.59 (1.96)</td>
<td>3.8 (1.49)</td>
<td>4.64 (1.29)</td>
<td>3.36 (1.32)</td>
</tr>
<tr>
<td></td>
<td>F=1.713</td>
<td>F=1.713</td>
<td>F=1.713</td>
<td>F=1.713</td>
<td>F=1.713</td>
<td></td>
</tr>
<tr>
<td>Own SD</td>
<td>4.87 (1.21)</td>
<td>5.21 (1.09)</td>
<td>5.14 (1.26)</td>
<td>5.08 (1.17)</td>
<td>4.64 (0.55)</td>
<td>5.5 (1.28)</td>
</tr>
<tr>
<td></td>
<td>F=1.256</td>
<td>F=1.256</td>
<td>F=1.256</td>
<td>F=1.256</td>
<td>F=1.256</td>
<td></td>
</tr>
<tr>
<td>Own BD - SD</td>
<td>-0.94 (2.39)</td>
<td>-1.64 (1.96)</td>
<td>-0.55 (2.82)</td>
<td>-1.28 (2.28)</td>
<td>0 (1.69)</td>
<td>-2.14 (2.48)</td>
</tr>
<tr>
<td></td>
<td>F=1.827*</td>
<td>F=1.827*</td>
<td>F=1.827*</td>
<td>F=1.827*</td>
<td>F=1.827*</td>
<td></td>
</tr>
<tr>
<td>Perceptions of CEO’s BD</td>
<td>5.63 (1.38)</td>
<td>2.9 (1.67)</td>
<td>4.75 (1.83)</td>
<td>3.78 (1.6)</td>
<td>4.05 (1.49)</td>
<td>4 (1.57)</td>
</tr>
<tr>
<td>Perceptions of CEO’s SD</td>
<td>3.59 (1.56)</td>
<td>5.79 (1.1)</td>
<td>4.05 (1.79)</td>
<td>4.46 (1.34)</td>
<td>5.14 (0.98)</td>
<td>4.59 (1.56)</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Perceptions of CEO’s BD - SD</td>
<td>2.04 (2.47)</td>
<td>-2.89 (2.21)</td>
<td>0.7 (2.61)</td>
<td>-0.68 (2.6)</td>
<td>-1.09 (2.17)</td>
<td>-0.59 (2.99)</td>
</tr>
<tr>
<td>Perceptions of Other Men at the Company’s BD</td>
<td>4.75 (1.37)</td>
<td>4.31 (1.44)</td>
<td>5.18 (1.45)</td>
<td>4.66 (1.48)</td>
<td>4.73 (1.25)</td>
<td>4.41 (1.39)</td>
</tr>
<tr>
<td>Perceptions of Other Men at the Company’s SD</td>
<td>4.18 (1.31)</td>
<td>4.8 (1.24)</td>
<td>4.43 (1.54)</td>
<td>4.08 (1.39)</td>
<td>4.91 (0.8)</td>
<td>4.45 (1.21)</td>
</tr>
<tr>
<td>Perceptions of Other Men at the Company’s BD - SD</td>
<td>0.57 (2.06)</td>
<td>-0.49 (2.04)</td>
<td>0.75 (2.5)</td>
<td>0.58 (2.13)</td>
<td>-0.18 (1.66)</td>
<td>-0.05 (2.27)</td>
</tr>
<tr>
<td>Perceptions of Other Women at the Company’s BD</td>
<td>3.51 (1.61)</td>
<td>3.31 (1.53)</td>
<td>3.34 (1.64)</td>
<td>3.28 (1.63)</td>
<td>3.09 (1.77)</td>
<td>2.32 (1.25)</td>
</tr>
<tr>
<td>Perceptions of Other Women at the Company’s SD</td>
<td>4.77 (1.3)</td>
<td>4.99 (1.28)</td>
<td>4.93 (1.44)</td>
<td>5.42 (1.19)</td>
<td>4.82 (1.49)</td>
<td>5.77 (1.4)</td>
</tr>
<tr>
<td>Perceptions of Other Women at the Company’s BD - SD</td>
<td>-1.26 (2.36)</td>
<td>-1.68 (1.93)</td>
<td>-1.59 (2.31)</td>
<td>-2.14 (2.4)</td>
<td>-1.73 (2.47)</td>
<td>-3.45 (2.1)</td>
</tr>
<tr>
<td>Hiring Intentions, Front-End Role (Higher #s = Woman More Likely)</td>
<td>4.16 (1.02)</td>
<td>4.38 (0.83)</td>
<td>4.27 (1.12)</td>
<td>4.36 (0.95)</td>
<td>4.64 (0.67)</td>
<td>4.73 (0.65)</td>
</tr>
<tr>
<td>Hiring Intentions, Back-End Role (Higher #s = Woman More Likely)</td>
<td>3.04 (1.09)</td>
<td>2.76 (1.08)</td>
<td>2.59 (0.85)</td>
<td>2.76 (1.05)</td>
<td>2.73 (1.27)</td>
<td>3.18 (0.98)</td>
</tr>
<tr>
<td>Estimated % Women at Company in a Front-End Role</td>
<td>70.8 (14.63)</td>
<td>68.1 (14.09)</td>
<td>67.14 (18.79)</td>
<td>62.04 (19.02)</td>
<td>61.27 (20.41)</td>
<td>63.91 (13.16)</td>
</tr>
<tr>
<td>Estimated % Women at Company in a Back-End Role</td>
<td>30.32 (22.1)</td>
<td>34.14 (21.33)</td>
<td>28.77 (16.36)</td>
<td>28.08 (18.57)</td>
<td>34 (16.58)</td>
<td>28.36 (14.9)</td>
</tr>
<tr>
<td>Manipulation Believable</td>
<td>5.04 (1.41)</td>
<td>5.22 (1.28)</td>
<td>5.55 (1.18)</td>
<td>5.44 (1.19)</td>
<td>5.91 (1.22)</td>
<td>5.55 (1.37)</td>
</tr>
<tr>
<td>Manipulation Persuasive</td>
<td>4.06 (1.54)</td>
<td>4.68 (1.45)</td>
<td>5.32 (1.43)</td>
<td>4.96 (1.54)</td>
<td>5.64 (1.63)</td>
<td>5.45 (1.04)</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Manipulation Clear</td>
<td>5.66 (1.35)</td>
<td>5.82 (1.21)</td>
<td>6.14 (0.83)</td>
<td>5.32 (1.57)</td>
<td>6.09 (1.04)</td>
<td>5.82 (1.33)</td>
</tr>
<tr>
<td>Manipulation Positive Tone</td>
<td>5.2 (1.18)</td>
<td>5.78 (1.04)</td>
<td>4.36 (1.59)</td>
<td>4.84 (1.11)</td>
<td>4.73 (1.68)</td>
<td>4.82 (1.25)</td>
</tr>
<tr>
<td>Manipulation Negative Tone</td>
<td>2.68 (1.45)</td>
<td>2.18 (1.35)</td>
<td>2.86 (1.49)</td>
<td>2.68 (1.22)</td>
<td>1.64 (0.81)</td>
<td>2.36 (1.43)</td>
</tr>
<tr>
<td>Perceptions of CEO as Likable</td>
<td>4.44 (1.46)</td>
<td>5.16 (1.22)</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Perceptions of CEO as Competent</td>
<td>4.88 (1.3)</td>
<td>5.56 (1.09)</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

*Note.* BD = Biological Determinism, SD = Social Determinism.
Prior to conducting Pilot Study 3, we generated several tests to gauge whether each manipulation was successful and guide selection of a manipulation to be used in Study 5. Each of these points is described in further detail below.

**Effects of manipulation on beliefs**

A successful manipulation will:

- In the article manipulation – lead to greater personal endorsement of biological > social explanation in biological vs. social condition

This was supported, $\beta = -0.48, p = 0.046$. The effect was not moderated by whether people chose the article or not and was also not moderated by political ideology. Finally, the effect was stronger but marginal ($\beta = -0.51, p = 0.072$) when repeating this analysis only with those who passed the comprehension check the first time.

- In the third-party manipulation – lead to greater perceived CEO endorsement of biological > social explanation in the biological vs. social condition

This was supported, $\beta = -1.45, p < 0.001$. The effect was stronger ($\beta = -1.56, p < 0.001$) when repeating this analysis only with those who passed the comprehension check the first time.

**Effects of manipulation on hiring intentions** (main effect of biological vs. social condition, explore interaction by manipulation type – article vs. interview)

A successful manipulation will:

- Lead to significantly greater intentions to hire a woman (vs. a man) for a front-end role in the biological vs. social condition

This was not supported, $\beta = 0.18, p = 0.249$. The effect was not moderated by media type (article vs. interview).

- Lead to significantly greater intentions to hire a man (vs. a woman) for a back-end role in the biological vs. social condition (main effect of condition)

This was not supported, $\beta = -0.06, p = 0.688$. The effect was also not moderated by media type (article vs. interview) and was not present for article or interview conditions separately.

**Comparing predictors of hiring intentions**

- Which is a stronger predictor of intentions to hire above: perception of CEO’s endorsement of biological/social explanation in the third-party condition or own personal endorsement of biological/social explanation in the article condition?

  - If personal endorsement is a stronger predictor and the article manipulation is successful, then we should opt for this manipulation
Personal endorsement of biological (more than social) determinism predicted being less intended to hire a woman for a back-end role, $\beta = -0.10$, $p = 0.001$, though did not predict front-end hiring intentions, $\beta = -0.04$, $p = 0.194$.

- If **perception of CEO’s endorsement** is a stronger predictor and the interview manipulation is successful, then we should opt for this manipulation

Perception of CEO endorsement did not predict hiring intentions for back-end roles, $\beta = -0.02$, $p = 0.529$, nor for front-end roles, $\beta = -0.03$, $p = 0.245$.

Given these patterns of effects, we reasoned the article manipulation was the best option. We next examined manipulation quality across the two article primes (biological vs. social) to fine-tune any potential imbalances. As shown below, there was only one case of a significant difference across the article primes (the biological article was rated as being clearer than the social article). We reasoned this might have been due to the concrete, brain-based evidence highlighted in the biological article, compared to the socialization article.

**Comparing manipulation quality**

- **Believability** – $d = 0.16$ (n.s.), biological higher
- **Persuasive** – $d = 0.22$ (n.s.), biological higher
- **Clarity** – $d = 0.52$, $p = 0.032$, biological higher
- **Positive Tone** – $d = 0.25$ (n.s.), biological higher
- **Negative Tone** – $d = 0.10$ (n.s.), social higher

**Summary and Key Decisions**

Based on the results of Pilot Study 3, we concluded the article manipulation would likely be the most effective prime of biological vs. social determinism for Study 5. Additionally, since effects did not hinge on whether participants chose or were assigned to the article, to maximize cell sizes (and thus our statistical power to detect differences by condition), we assigned participants to the biological or social article in Study 5. Prior to Study 5, we made minor edits to language in the article to emphasize expert credibility, with the intent of increasing persuasiveness ratings.
B.2 Study 4

Study 4 - List of Measures

Below is the full list of variables included in Study 4 in the order they were presented to participants. Variables analyzed and reported in the main text are denoted with an asterisk.

- Employment Screener
- Attention Check
- Filler Items for Tech Trends Article
- The following measures were presented together with items intermixed:
  - Biological Determinism*
  - Social Determinism*
- Perceptions of Fairness
- Motivation to Take Action
- Behavioral Assignment Task*
- Task Strategy*
- Task Quality Checks
- Open-Ended Feedback
- Suspicion Check
- Article Comprehension Checks
- Site [Visitor/Traffic] Needs Ratings
  - Relevance to Empathizing
  - Relevance to Systemizing
- Gender
- Age
- Occupation
- Management Role
- Ethnicity
- Political Orientation
- SES
Study 4 - Items for Measures

Employment Screener

Which of the following best describes your current employment status?

- Employed (Full-Time)
- Employed (Part-Time)
- Contract / Self-Employed
- Unemployed
- Retired
- None of the options above

Which of the following best describes your occupation?

- Architecture, Construction, Extraction
- Arts, Sports, Entertainment
- Business, Financial Operations
- Community and Social Service
- Computers, Tech, Mathematics
- Education, Teaching
- Engineering, Life Science, Physical Science
- Farming, Fishing, Forestry
- Food Preparation, Serving
- Installation, Maintenance, Repair
- Legal, Politics
- Office and Administrative Support
- Personal Care and Service
- Transportation, Materials Moving
- (Something Else Not Listed Above)
Attention Check

Research in decision making shows that people, when making decisions and answering questions, prefer not to pay attention and minimize their effort as much as possible. Some studies show that over 50% of people don’t carefully read questions. If you are reading this question and have read all the other questions, please select the box marked ‘Other’ and type ‘Decision making’ in the box below. Do not select 'Predictions of your own behavior.' Thank you for participating and taking the time to read through the questions carefully!

What is this study about?
- Predictions of your own behavior
- Predictions of your friend's behavior
- Political preferences
- Other: __________________________________________________

Filler Items for Tech Trends Article

1 = Strongly disagree . . . 7 = Strongly agree

- Many tech trends naturally develop from the technology that came before them.
- When trying to understand new tech trends, one should always consider tech trends of the past.
- Tech trends are largely established by big industry leaders.
- In most cases, tech trends can be traced back to what is most profitable.
- Tech trends are largely the product of thinking outside-the-box.
- The current events of a given era are evident in what technologies are trending.

1 = Strongly disagree . . . 7 = Strongly agree

- The current trends in tech reflect the actual needs of society today.
- Trending topics in tech are popular because they are more important.
- I would support initiatives to invest more resources into trending technologies.
**Biological Determinism**

1 = *Strongly disagree* ... 7 = *Strongly agree*

- Many forms of gender-related interests are biologically determined. 
- When trying to understand differences in interests by gender, one should always look first to biology. 
- Gender differences in interests are largely determined by one’s genetic background. 
- In most cases, gender differences can be traced back to biological causes. 
- Men and women's different interests are mostly due to their innate biologies and are thus difficult to change.

**Social Determinism**

1 = *Strongly disagree* ... 7 = *Strongly agree*

- The social background a person comes from is strongly reflected in the development of the person’s gender-related interests. 
- A person’s gender-related interests are largely the product of their social origin. 
- In most cases, gender differences in interest can be traced back to socio-cultural causes. 
- The type of social environment a person grows up in is evidence in the development of the person’s gender-related interests. 
- Men and women's different interests are mostly due to how they are socialized and thus can be changed by new experiences.

**Perceptions of Fairness**

1 = *Strongly disagree* ... 7 = *Strongly agree*

- We live in a just society where the gender composition in tech reflects inherent gender differences in interest. 
- The current gender composition in tech is fair.

**Motivation to Take Action**

1 = *Strongly disagree* ... 7 = *Strongly agree*

- I would support policies aimed at achieving equal gender representation in tech
Task Strategy

1 = Not at all . . . 7 = Very much

- I chose the person who I thought was more naturally interested in each team.
- I chose the person who I thought would be more likely to develop new interests from working on each team.

Task Quality Checks

1 = Not at all . . . 7 = Very much

- I made my selections at random.
- I put thought into each one of my selections.

Open-Ended Feedback

- Did you encounter any technical issues in today's study? If yes, please tell us about them here.
- Did you find any parts of this study confusing or difficult to answer? If yes, please tell us about that here.

Suspicion Check

- What did you think the purpose of today's study was?

Article Comprehension Checks

1 = Not at all . . . 7 = Very much

- To what extent is Future of Clean Energy an emerging tech trend?
- To what extent is Bluetooth an emerging tech trend?

1 = Not at all . . . 7 = Very much

- To what extent do men and women differ in their abilities?
- To what extent do men and women differ in their interests?
Site [Visitor/Traffic] Needs Ratings

The question below asks about your perceptions of empathizing and systemizing. This information is being collected for pilot testing purposes.

- By empathizing, we mean the drive to identify another person's emotions and thoughts and respond to these with an appropriate emotion.
- By systemizing, we mean the drive to analyze, understand, predict, control, and construct rule-based systems.

How would you rate the Site [Visitor/Traffic] Needs team on each of the following?

1 = Not at all . . . 7 = Very much

- Relevance to empathizing
- Relevance to systemizing

Gender

Which of the following best describes you?
- Woman
- Man
- Non-binary person
- Another option (please specify):
  ______________________________________________________

Age

Please indicate your age in years.

  18  26  34  43  51  59  67  75  84  92  100

  Age  ______________________________________________________

Occupation

What is your current occupation? ________________________
Management Role

Are you currently in a management role?
- Yes
- No
- Not sure

Ethnicity

Which of the following ethnicities do you identify with? (Select all that apply)
- African American, African, or Black (e.g., Appendixali, Haitian)
- Hispanic / Latinx / Central or Spanish Origin (e.g., Brazilian, Mexican)
- American Indian, Indigenous, or Alaska Native (e.g., First Nations, Inuit, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village or Barrow Inupiat Traditional Government)
- Native Hawaiian or Pacific Islander (e.g., Samoan, Chamorro, Tonga, Fijian, Marshallese)
- Middle Eastern or North African (e.g., Egyptian, Iranian, Lebanese, Algerian)
- East Asian (e.g., Chinese, Japanese)
- South Asian (e.g., Indian, Pakistani)
- Southeast Asian (e.g., Filipino, Thai)
- White (e.g., Australian, Irish)
- Not listed in the options above (please specify):
  __________________________________________________________

Political Orientation

In general, to what extent do you consider yourself to be liberal or conservative?
- 1 = Extremely Liberal
- 2
- 3
- 4 = Moderate
- 5
- 6
- 7 = Extremely Conservative
Think about your social class, or where you stand, relative to others in society. Select an answer corresponding to the rung of the ladder where you feel you stand, where the top rung means you feel like you're at the top, and the bottom rung means you feel like you're at the bottom.

*Alt Text: An image of a ladder with ten rungs. The top rung is labeled "10" and the bottom rung is labeled "1" with the rungs in between labeled in descending order as "9," "8," "7," "6," "5," "4," "3," and "2."*
The Tech of Tomorrow: Outlook on the Fastest-Growing Tech Trends

Investing in technology trends early can help businesses gain a leading edge. For example, a report by Statista found that 39% of businesses reported improved agility and responsiveness after adopting cloud technology in 2017. What new trends should businesses be aware of in 2022? According to a recent report released by McKinsey Technology Council, future of clean energy and future of mobility are two key trends to watch. Other trends of note include advanced connectivity, applied AI, cloud and edge computing, Web3, and future of sustainable consumption.

![Figure B.6. Infographic for filler article (Study 4)](image-url)
It's Not About Ability: Could Different Interests Help Explain Women's Underrepresentation in Tech?

Women remain underrepresented in many sectors of tech. For example, according to the U.S. Bureau of labor statistics, women make up only 26.7% of computer and information systems managers. What explains this disparity? According to experts, it cannot be explained by different abilities. Instead, one of the strongest explanations of occupational choice is gender differences in interests. Leading research across a number of studies has found that women are more interested in tasks and careers involving empathizing (i.e., drive to identify another person's emotions and thoughts and respond to these with an appropriate emotion), whereas men are more interested in tasks and careers involving systemizing (i.e., the drive to analyze, understand, predict, control, and construct rule-based systems). Research suggests that women's relatively lower interest in systemizing might help explain why women are not as attracted to tech careers.

Figure B.7. Infographic for critical article (Study 4)
**Study 4 – Project Teams**

All participants saw the same $N = 4$ Front-End and Back-End teams; for the remaining $N = 4$ teams, participants saw an empathizing or systemizing version of the same team (e.g., Site Visitor Needs for empathizing frame vs. Site Traffic Needs for systemizing frame).

![Figure B.8. Overview of team assignment (Study 4)](image)

Icon Credits (FE/BE Set, Top to Bottom): Thinking © ArtWorkLeaf; Team © HASTA ICON; Security © Setyo Ari Wibowo; Server © Rolas Design. Icon Credits (E/S Sets, Top to Bottom): Social Networking © I Putu Kharismayadi; Web Traffic © Saepul Nahwan; User © jamhuden; Prototype © Aris Sunjayay; Share © Three Six Five; Person Love © Round Icons; Website © Ibrandify; Usability © Adrien Coquet. All icons retrieved from the Noun Project and licensed under CC-BY 3.0; all icons were modified by the lead author from black and white to color icons.
Study 4 – Behavioral Assignment Task Information

Participants were presented with a group of 6 interns from a variety of backgrounds that would be joining the company as part of a required 4-week internship for first-year university students. We presented participants with 6 photos of interns and a brief quote attributed to them about the internship (e.g., “These internships will be helpful for getting a glimpse of the day-to-day of different jobs,” quotes were counterbalanced in 1 of 2 orders across participants; effects hold controlling for the effect of quote orders). Photos included $n = 6$ of 30 total images selected from the Chicago Face Database (Ma et al., 2015, 2020). The study stimuli ($n = 6$) and full image set ($n = 30$) contained 50% women and 50% men of varying gender prototypicality (1/3 highly masculine, 1/3 neutral, 1/3 highly feminine; effects did not vary based on gender prototypicality for men or women targets). Our final photo set controlled for perceptions of age, attractiveness, and varied along ethnicity (6 White, 6 Asian, 6 Black, 6, Latinx, 6 Multiracial). Participants were told their goal as project manager was to ensure students are matched to tasks in a way that enhances their experience and training. Participants began with $N = 2$ practice selections to warm up. Next, participants completed $N = 16$ trials (2 per team; $N = 8$ matched-gender filler selections; $N = 8$ man-woman critical selections). There was no time pressure on the task, but participants were encouraged to make each selection as quickly as possible.
Table B.11. Overview of team assignment task trials (Study 4; order of trials randomized)

<table>
<thead>
<tr>
<th>Trial</th>
<th>Team</th>
<th>Target 1</th>
<th>Target 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Critical Trials</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT 1</td>
<td>[Comm / Info] Oversight</td>
<td>Man</td>
<td>Woman</td>
</tr>
<tr>
<td>CT 2</td>
<td>[Market / Customer] Trends</td>
<td>Man</td>
<td>Woman</td>
</tr>
<tr>
<td>CT 3</td>
<td>[Site Visitor / Traffic] Needs</td>
<td>Man</td>
<td>Woman</td>
</tr>
<tr>
<td>CT 4</td>
<td>[Product/ Human] Testing</td>
<td>Man</td>
<td>Woman</td>
</tr>
<tr>
<td>CT 5</td>
<td>People Support</td>
<td>Man</td>
<td>Woman</td>
</tr>
<tr>
<td>CT 6</td>
<td>Database Security</td>
<td>Man</td>
<td>Woman</td>
</tr>
<tr>
<td>CT 7</td>
<td>Client Strategy</td>
<td>Man</td>
<td>Woman</td>
</tr>
<tr>
<td>CT 8</td>
<td>Server Admin</td>
<td>Man</td>
<td>Woman</td>
</tr>
<tr>
<td><strong>Filler Trials</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filler T 1</td>
<td>[Comm / Info] Oversight</td>
<td>Man</td>
<td>Man</td>
</tr>
<tr>
<td>Filler T 2</td>
<td>[Market / Customer] Trends</td>
<td>Man</td>
<td>Man</td>
</tr>
<tr>
<td>Filler T 3</td>
<td>[Site Visitor / Traffic] Needs</td>
<td>Woman</td>
<td>Woman</td>
</tr>
<tr>
<td>Filler T 4</td>
<td>[Product/ Human] Testing</td>
<td>Woman</td>
<td>Woman</td>
</tr>
<tr>
<td>Filler T 5</td>
<td>People Support</td>
<td>Man</td>
<td>Man</td>
</tr>
<tr>
<td>Filler T 6</td>
<td>Database Security</td>
<td>Man</td>
<td>Man</td>
</tr>
<tr>
<td>Filler T 7</td>
<td>Client Strategy</td>
<td>Woman</td>
<td>Woman</td>
</tr>
<tr>
<td>Filler T 8</td>
<td>Server Admin</td>
<td>Woman</td>
<td>Woman</td>
</tr>
</tbody>
</table>
Study 4 – GLM and MLM Results Comparison

Below is a table of results comparing estimates obtained with GLM (as reported in the manuscript) to estimates obtained modeling data with MLM, with observations nested within participant. All models predict affordances from biological vs. social determinism (“BioSocDet”), team type (empathizing vs. systemizing), and their interaction. As can be seen below, MLM estimates yielded results identical to GLM estimates.

Table B.12. Key effects modeled in GLM and MLM (Study 4)

<table>
<thead>
<tr>
<th></th>
<th>GLM (Standardized Estimates)</th>
<th>GLM (Unstandardized Estimates)</th>
<th>MLM (Unstandardized Estimates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Slope, Empathizing</td>
<td>β = .20</td>
<td>β = .09</td>
<td>β = .09</td>
</tr>
<tr>
<td></td>
<td>p = .001</td>
<td>p = .001</td>
<td>p = .001</td>
</tr>
<tr>
<td>Simple Slope, Systemizing</td>
<td>β = -.30</td>
<td>β = -.13</td>
<td>β = -.13</td>
</tr>
<tr>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Interaction (BioSocDet x Team Type)</td>
<td>β = -.50</td>
<td>β = -.22</td>
<td>β = -.22</td>
</tr>
<tr>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
</tr>
</tbody>
</table>
Study 4 – Results Controlling for General Perceptions of Gender Interest and Ability Differences

In Study 4, we asked participants to estimate their perceptions of gender interest and ability differences. This item was originally developed as a screening comprehension check to test knowledge of the article participants read at the beginning of the study. However, since the item did not reference the article explicitly (e.g., “reflecting on the article you read today…”), it is possible participants responded to this item with their general beliefs. We reran analyses controlling for these variables. As shown below, all primary hypotheses were robust to controlling for perceptions of ability and interest differences (entered as simultaneous controls in the model).

- **PH1a:** People who endorse a primarily biological over social determinist explanation for gender differences on interest will afford more **empathizing** learning opportunities to **women**
  
  Original estimate: $\beta = .20, p = .001$
  Estimate including controls: $\beta = .25, p < .001$

- **PH1b:** People who endorse a primarily biological over social determinist explanation for gender differences on interest will afford more **systemizing** learning opportunities to **men**
  
  Original estimate: $\beta = -.30, p < .001$
  Estimate including controls: $\beta = -.26, p < .001$

- **PH2a:** In general, people will afford more **empathizing** learning opportunities to **women**
  
  Original estimate: $\beta = -.86, p < .001$
  Estimate including controls: $\beta = -.86, p < .001$

- **PH2b:** In general, people will afford more **systemizing** learning opportunities to **men**
  
  Original estimate: $\beta = .86, p < .001$
  Estimate including controls: $\beta = .86, p < .001$
• **PH3**: The interaction between explanation for gender differences on interest and empathizing vs. systemizing team will be significant, such that the relationship between team type (empathizing or systemizing) and learning opportunities afforded to men or women will depend on endorsement of a biological versus social determinist explanation.

  Original estimate: $\beta = -.50, p < .001$
  Estimate including controls: $\beta = -.50, p < .001$
Study 4 – Results Controlling for Participant Gender

Although there were marginal differences by participant gender in likelihood of selecting a woman for empathizing teams (men slightly but not significantly more likely, $\beta = .17, p = .206$) and systemizing teams (men marginally less likely, $\beta = -.22, p = .090$), effects were not moderated by participant gender and controlling for participant gender in this model does not change estimates.

- **PH1a:** People who endorse a primarily biological over social determinist explanation for gender differences on interest will afford more empathizing learning opportunities to women
  
  Original estimate: $\beta = .20, p = .001$
  Estimate controlling for participant gender: $\beta = .21, p = .001$

- **PH1b:** People who endorse a primarily biological over social determinist explanation for gender differences on interest will afford more systemizing learning opportunities to men
  
  Original estimate: $\beta = -.30, p < .001$
  Estimate controlling for participant gender: $\beta = -.30, p < .001$

- **PH2a:** In general, people will afford more empathizing learning opportunities to women
  
  Original estimate: $\beta = -.86, p < .001$
  Estimate controlling for participant gender: $\beta = -.86, p < .001$

- **PH2b:** In general, people will afford more systemizing learning opportunities to men
  
  Original estimate: $\beta = .86, p < .001$
  Estimate controlling for participant gender: $\beta = .86, p < .001$

- **PH3:** The interaction between explanation for gender differences on interest and empathizing vs. systemizing team will be significant, such that the relationship between team type (empathizing or systemizing) and learning opportunities afforded to men or women will depend on endorsement of a biological versus social determinist explanation
  
  Original estimate: $\beta = -.50, p < .001$
  Estimate controlling for participant gender: $\beta = -.50, p < .001$
Study 4 – Results with Front-End/Back-End and Empathizing/Systemizing Teams Separately

To guide team selections for Study 5, we compared effects aggregating front-end/empathizing teams and back-end/systemizing teams (as reported in the manuscript) with effects separating front-end (FE) vs. back-end (BE) and empathizing (E) vs. systemizing (S) teams. As shown in the table below, effects were driven by front-end vs. back-end teams. Thus, these are the teams we used in the team assignment task for Study 5.

Table B.13. Key effects for front-end/empathizing and back-end/systemizing teams separately (Study 4)

<table>
<thead>
<tr>
<th></th>
<th>Aggregate Estimates</th>
<th>FE/BE Estimates</th>
<th>E/S Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Slope, Empathizing</td>
<td>( \beta = .20 )</td>
<td>( \beta = .29 )</td>
<td>( \beta = .12 )</td>
</tr>
<tr>
<td></td>
<td>( p = .001 )</td>
<td>( p = .001 )</td>
<td>( p = .177 )</td>
</tr>
<tr>
<td>Simple Slope, Systemizing</td>
<td>( \beta = -.30 )</td>
<td>( \beta = -.33 )</td>
<td>( \beta = -.28 )</td>
</tr>
<tr>
<td></td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
</tr>
<tr>
<td>Interaction (Biological-Social Determinism x Team Type)</td>
<td>( \beta = -.50 )</td>
<td>( \beta = -.62 )</td>
<td>( \beta = -.40 )</td>
</tr>
<tr>
<td></td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p = .001 )</td>
</tr>
</tbody>
</table>
Study 4 – Choice by Target Intern

We tested whether effects in Study 4 were stronger for more gender prototypical targets (i.e., highly masculine men, highly feminine women). As shown in the table below, there were some slight descriptive variations, but chi-square tests revealed no significant differences for all participants, participants highest (1 SD above the mean) on biological vs. social determinism, or participants highest (1 SD above the mean) on social vs. biological determinism.

All Participants

Table B.14. Selections by target for all participants (Study 4)

<table>
<thead>
<tr>
<th>Target</th>
<th>N Emp Trials</th>
<th>% Emp Trials</th>
<th>N Sys Trials</th>
<th>% Sys Trials</th>
<th>N Total Trials</th>
<th>% Total Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria (Masculine Woman)</td>
<td>233</td>
<td>20.44</td>
<td>164</td>
<td>14.39</td>
<td>397</td>
<td>17.41</td>
</tr>
<tr>
<td>Abby (Neutral Woman)</td>
<td>247</td>
<td>21.67</td>
<td>152</td>
<td>13.33</td>
<td>399</td>
<td>17.50</td>
</tr>
<tr>
<td>Gabrielle (Feminine Woman)</td>
<td>245</td>
<td>21.49</td>
<td>173</td>
<td>15.18</td>
<td>418</td>
<td>18.33</td>
</tr>
<tr>
<td>Mark (Masculine Man)</td>
<td>129</td>
<td>11.32</td>
<td>233</td>
<td>20.44</td>
<td>362</td>
<td>15.88</td>
</tr>
<tr>
<td>Adam (Neutral Man)</td>
<td>140</td>
<td>12.28</td>
<td>213</td>
<td>18.68</td>
<td>353</td>
<td>15.48</td>
</tr>
<tr>
<td>Christopher (Feminine Man)</td>
<td>146</td>
<td>12.81</td>
<td>205</td>
<td>17.98</td>
<td>351</td>
<td>15.39</td>
</tr>
<tr>
<td>Grand Total</td>
<td>1140</td>
<td>1140</td>
<td>2280</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- A chi-square test examining selections by trial type (emp vs. sys) among women targets only (Maria, Abby Gabrielle) was not significant, $p = .554$

- A chi-square test examining selections by trial type (emp vs. sys) among men targets only (Mark, Adam, Christopher) was not significant, $p = .249$
Participants +1 SD in Biological > Social

Table B.15. Selections by target for participants highest in biological vs. social determinism (Study 4)

<table>
<thead>
<tr>
<th>Target</th>
<th>N Emp Trials</th>
<th>% Emp Trials</th>
<th>N Sys Trials</th>
<th>% Sys Trials</th>
<th>N Total Trials</th>
<th>% Total Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria (Masculine Woman)</td>
<td>40</td>
<td>22.73</td>
<td>22</td>
<td>12.50</td>
<td>62</td>
<td>17.61</td>
</tr>
<tr>
<td>Abby (Neutral Woman)</td>
<td>39</td>
<td>22.16</td>
<td>16</td>
<td>9.09</td>
<td>55</td>
<td>15.63</td>
</tr>
<tr>
<td>Gabrielle (Feminine Woman)</td>
<td>41</td>
<td>23.30</td>
<td>23</td>
<td>13.07</td>
<td>64</td>
<td>18.18</td>
</tr>
<tr>
<td>Mark (Masculine Man)</td>
<td>18</td>
<td>10.23</td>
<td>46</td>
<td>26.14</td>
<td>64</td>
<td>18.18</td>
</tr>
<tr>
<td>Adam (Neutral Man)</td>
<td>17</td>
<td>9.66</td>
<td>37</td>
<td>21.02</td>
<td>54</td>
<td>15.34</td>
</tr>
<tr>
<td>Christopher (Feminine Man)</td>
<td>21</td>
<td>11.93</td>
<td>32</td>
<td>18.18</td>
<td>53</td>
<td>15.06</td>
</tr>
<tr>
<td>Grand Total</td>
<td>176</td>
<td>176</td>
<td>352</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- A chi-square test examining selections by trial type (emp vs. sys) among women targets only (Maria, Abby Gabrielle) was not significant, $p = .686$

- A chi-square test examining selections by trial type (emp vs. sys) among men targets only (Mark, Adam, Christopher) was not significant, $p = .407$
Participants +1 SD in Social > Biological

Table B.16. Selections by target for participants highest in social vs. biological determinism (Study 4)

<table>
<thead>
<tr>
<th>Target</th>
<th>N Emp Trials</th>
<th>% Emp Trials</th>
<th>N Sys Trials</th>
<th>% Sys Trials</th>
<th>N Total Trials</th>
<th>% Total Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria (Masculine Woman)</td>
<td>37</td>
<td>17.13</td>
<td>41</td>
<td>18.98</td>
<td>78</td>
<td>18.06</td>
</tr>
<tr>
<td>Abby (Neutral Woman)</td>
<td>38</td>
<td>17.59</td>
<td>43</td>
<td>19.91</td>
<td>81</td>
<td>18.75</td>
</tr>
<tr>
<td>Gabrielle (Feminine Woman)</td>
<td>39</td>
<td>18.06</td>
<td>38</td>
<td>17.59</td>
<td>77</td>
<td>17.82</td>
</tr>
<tr>
<td>Mark (Masculine Man)</td>
<td>39</td>
<td>18.06</td>
<td>35</td>
<td>16.20</td>
<td>74</td>
<td>17.13</td>
</tr>
<tr>
<td>Adam (Neutral Man)</td>
<td>32</td>
<td>14.81</td>
<td>31</td>
<td>14.35</td>
<td>63</td>
<td>14.58</td>
</tr>
<tr>
<td>Christopher (Feminine Man)</td>
<td>31</td>
<td>14.35</td>
<td>28</td>
<td>12.96</td>
<td>59</td>
<td>13.66</td>
</tr>
<tr>
<td>Grand Total</td>
<td>216</td>
<td>216</td>
<td>432</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- A chi-square test examining selections by trial type (emp vs. sys) among **women** targets only (Maria, Abby Gabrielle) was not significant, \( p = .880 \)

- A chi-square test examining selections by trial type (emp vs. sys) among **men** targets only (Mark, Adam, Christopher) was not significant, \( p = .971 \)
Study 4 – Effects for Biological and Social Determinism Beliefs Separately

Given that our predictor measure in Study 4 tested people’s relative endorsement of biological vs. social determinism as a difference score (biological – social determinism; higher scores mean stronger endorsement of biological determinism), we tested whether effects were driven more strongly by biological or social determinism as separate variables. Replicating our main analytic model from Study 4 (a binomial regression predicting odds of choosing a woman intern from biological vs. social determinism, team type, and their interaction), results were descriptively stronger for biological determinism, but they emerged independently for biological and social determinism as separate predictors.

Table B.17. Key effects for biological and social determinism measures separately (Study 4)

<table>
<thead>
<tr>
<th></th>
<th>Difference Score (BD – SD)</th>
<th>Biological Determinism Only</th>
<th>Social Determinism Only</th>
</tr>
</thead>
</table>
| Simple Slope, Empathizing | **β = .20**  
                          | **p = .001**  
                          | **β = .19**  
                          | **p = .002**  
                          | **β = -.14**  
                          | **p = .022**  |
| Simple Slope, Systemizing | **β = -.30**  
                          | **p < .001**  
                          | **β = -.31**  
                          | **p < .001**  
                          | **β = .17**  
                          | **p = .006**  |
| Interaction (Attitude x Team Type) | **β = -.50**  
                          | **p < .001**  
                          | **β = -.50**  
                          | **p < .001**  
                          | **β = .31**  
                          | **p < .001**  |

*Note.* BD = Biological Determinism, SD = Social Determinism.
**Study 4 – Exploratory Factor Analysis of Biological vs. Social Determinism Scales**

We first ran a principal components analysis to determine the appropriate number of factors to analyze. As shown in the figure below, we found a 2-factor solution that explained 78.97% of variance.

![Scree plot of factors for all biological and social determinism items (Study 4)](image)

*Figure B.9. Scree plot of factors for all biological and social determinism items (Study 4)*
We next subjected all 10 items to a principal axis factor analysis with oblique, promax rotation (Costello & Osborne, 2005). As seen below, the biological and social determinism items showed excellent factor structure, with each composite’s items loading onto a separate factor.

<table>
<thead>
<tr>
<th>Item Name</th>
<th>Item Text</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio Det 4</td>
<td>In most cases, gender differences can be traced back to biological causes.</td>
<td>.92</td>
<td>.02</td>
</tr>
<tr>
<td>Bio Det 3</td>
<td>Gender differences in interests are largely determined by one’s genetic background.</td>
<td>.91</td>
<td>.08</td>
</tr>
<tr>
<td>Bio Det 2</td>
<td>When trying to understand differences in interests by gender, one should always look first to biology.</td>
<td>.90</td>
<td>.01</td>
</tr>
<tr>
<td>Bio Det 1</td>
<td>Many forms of gender-related interests are biologically determined.</td>
<td>.90</td>
<td>- .01</td>
</tr>
<tr>
<td>Bio Det 5</td>
<td>Men and women’s different interests are mostly due to their innate biologies and are thus difficult to change.</td>
<td>.88</td>
<td>- .06</td>
</tr>
<tr>
<td>Soc Det 3</td>
<td>In most cases, gender differences in interest can be traced back to socio-cultural causes.</td>
<td>-.10</td>
<td>.81</td>
</tr>
<tr>
<td>Soc Det 4</td>
<td>The type of social environment a person grows up in is evidence in the development of the person’s gender-related interests.</td>
<td>.11</td>
<td>.80</td>
</tr>
<tr>
<td>Soc Det 1</td>
<td>The social background a person comes from is strongly reflected in the development of the person’s gender-related interests.</td>
<td>.13</td>
<td>.77</td>
</tr>
<tr>
<td>Soc Det 2</td>
<td>A person’s gender-related interests are largely the product of their social origin.</td>
<td>-.04</td>
<td>.77</td>
</tr>
<tr>
<td>Soc Det 5</td>
<td>Men and women's different interests are mostly due to how they are socialized and thus can be changed by new experiences.</td>
<td>-.23</td>
<td>.65</td>
</tr>
</tbody>
</table>
Next, we conducted exploratory factor analyses for each separate construct (biological, social determinism; 5 items each) using principal axis factor analysis with oblique, promax rotation (Costello & Osborne, 2005) to inform which highest-loading items to use for the abbreviated measure of biological and social determinism in Study 5.

Table B.19. Factor loadings for biological determinism items separately (Study 4)

<table>
<thead>
<tr>
<th>Item Name</th>
<th>Item Text</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio Det 4</td>
<td>In most cases, gender differences can be traced back to biological causes.</td>
<td>.91</td>
</tr>
<tr>
<td>Bio Det 5</td>
<td>Men and women's different interests are mostly due to their innate biologies and are thus difficult to change.</td>
<td>.91</td>
</tr>
<tr>
<td>Bio Det 1</td>
<td>Many forms of gender-related interests are biologically determined.</td>
<td>.90</td>
</tr>
<tr>
<td>Bio Det 2</td>
<td>When trying to understand differences in interests by gender, one should always look first to biology.</td>
<td>.90</td>
</tr>
<tr>
<td>Bio Det 3</td>
<td>Gender differences in interests are largely determined by one’s genetic background.</td>
<td>.88</td>
</tr>
</tbody>
</table>

Table B.20. Factor loadings for social determinism items separately (Study 4)

<table>
<thead>
<tr>
<th>Item Name</th>
<th>Item Text</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soc Det 3</td>
<td>In most cases, gender differences in interest can be traced back to socio-cultural causes.</td>
<td>.85</td>
</tr>
<tr>
<td>Soc Det 2</td>
<td>A person’s gender-related interests are largely the product of their social origin.</td>
<td>.79</td>
</tr>
<tr>
<td>Soc Det 4</td>
<td>The type of social environment a person grows up in is evidence in the development of the person’s gender-related interests.</td>
<td>.75</td>
</tr>
<tr>
<td>Soc Det 5</td>
<td>Men and women's different interests are mostly due to how they are socialized and thus can be changed by new experiences.</td>
<td>.74</td>
</tr>
<tr>
<td>Soc Det 1</td>
<td>The social background a person comes from is strongly reflected in the development of the person’s gender-related interests.</td>
<td>.71</td>
</tr>
</tbody>
</table>
Study 4 - Effects for Immutability Items Separately

To test whether effects in Study 4 were present just for items that tapped into perceptions of immutability, we re-ran key analyses for the single item measuring immutability (1 item: “Men and women's different interests are mostly due to their innate biologies and are thus difficult to change;” embedded in the biological determinism scale), mutability (1 item: “Men and women's different interests are mostly due to how they are socialized and thus can be changed by new experiences;” embedded in the social determinism scale), and their difference score (immutability – mutability; higher scores mean more perceptions of immutability than mutability). All key effects emerged with these items separately.

Table B.21. Effects for Immutability Items (Study 4)

<table>
<thead>
<tr>
<th></th>
<th>Difference Score (IM – M)</th>
<th>Immutability Only</th>
<th>Mutability Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Slope,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empathizing</td>
<td>β = .21 ( p &lt; .001 )</td>
<td>β = .17 ( p = .005 )</td>
<td>β = -.19 ( p = .003 )</td>
</tr>
<tr>
<td>Simple Slope,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systemizing</td>
<td>β = -.32 ( p &lt; .001 )</td>
<td>β = -.33 ( p &lt; .001 )</td>
<td>β = .21 ( p &lt; .001 )</td>
</tr>
<tr>
<td>Interaction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Attitude x Team Type)</td>
<td>β = -.52 ( p &lt; .001 )</td>
<td>β = -.50 ( p &lt; .001 )</td>
<td>β = .40 ( p &lt; .001 )</td>
</tr>
</tbody>
</table>

*Note.* IM = Immutability, M = Mutability.
Study 4 - Effects without Immutability Items

To test whether effects in Study 4 were present without items that tapped into perceptions of immutability, we re-ran key analyses with a revised composite of biological determinism, social determinism, and their difference score (biological – social determinism; higher scores mean more biological determinism), without including the single item of immutability in each subscale (see previous section). With the exception of one effect (the simple slope of social determinism predicting empathizing affordances became marginal without the immutability item), all key effects emerged with these items separately.

Table B.22. Effects for Biological and Social Determinism without Immutability Items (Study 4)

<table>
<thead>
<tr>
<th></th>
<th>Difference Score (BD – SD)</th>
<th>Biological Determinism Only</th>
<th>Social determinism Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Slope, Empathizing</td>
<td>( \beta = .20 ) ( p = .001 )</td>
<td>( \beta = .19 ) ( p = .002 )</td>
<td>( \beta = .12 ) ( p = .056 )</td>
</tr>
<tr>
<td>Simple Slope, Systemizing</td>
<td>( \beta = -.28 ) ( p &lt; .001 )</td>
<td>( \beta = -.30 ) ( p &lt; .001 )</td>
<td>( \beta = .14 ) ( p = .020 )</td>
</tr>
<tr>
<td>Interaction (Attitude x Team Type)</td>
<td>( \beta = -.48 ) ( p &lt; .001 )</td>
<td>( \beta = -.49 ) ( p &lt; .001 )</td>
<td>( \beta = .26 ) ( p = .003 )</td>
</tr>
</tbody>
</table>

Note. BD = Biological Determinism, SD = Social Determinism
Study 4 – Effects Controlling for Participant Suspicion

To test whether effects emerged when controlling for participant suspicion, we first coded participants’ open-ended response to the question: “What did you think the purpose of today's study was?” Codes ranged from 0 (participants’ response explicitly stated they did not know or was completely irrelevant to the true purpose of the study) to 3 (participants’ response correctly identified the independent and dependent variables, or otherwise suggested they understood the true purpose of the study). Full descriptive statistics and coding scheme for participant suspicion across Studies 4-6 is reported at the end of this supplement. As shown below, key effects were unchanged when controlling for participant suspicion or when removing suspicious participants (i.e., those coded as “3”).

Analyses controlling for suspicion codes

- **PH1a**: People who endorse a primarily biological over social determinist explanation for gender differences on interest will afford more **empathizing** learning opportunities to **women**

  Original estimate: $\beta = .20, p = .001$

  Estimate controlling for suspicion: $\beta = .24, p < .001$

- **PH1b**: People who endorse a primarily biological over social determinist explanation for gender differences on interest will afford more **systemizing** learning opportunities to **men** (in robustness check → less systemizing opportunities to women)

  Original estimate: $\beta = -.30, p < .001$

  Estimate controlling for suspicion: $\beta = -.28, p < .001$

- **PH2a**: In general, people will afford more **empathizing** learning opportunities to **women**

  Original estimate: $\beta = -.86, p < .001$

  Estimate controlling for suspicion: $\beta = -.87, p < .001$
• **PH2b:** In general, people will afford more systemizing learning opportunities to **men** (in robustness check → less systemizing opportunities to women, I think this is also how it was coded in the original analysis reported here)

Original estimate: $\beta = .86, p < .001$
Estimate controlling for suspicion: $\beta = .87, p < .001$

• **PH3:** The interaction between explanation for gender differences on interest and empathizing vs. systemizing team will be significant, such that the relationship between team type (**empathizing** or **systemizing**) and learning opportunities afforded to **men** or **women** will depend on endorsement of a biological versus social determinist explanation

Original estimate: $\beta = -.50, p < .001$
Estimate controlling for suspicion: $\beta = -.51, p < .001$

**Analyses without suspicious participants (i.e., those coded as “3”)**

• **PH1a:** People who endorse a primarily biological over social determinist explanation for gender differences on interest will afford more **empathizing** learning opportunities to **women**

Original estimate: $\beta = .20, p = .001$
Estimate without suspicious participants: $\beta = .24, p < .001$

• **PH1b:** People who endorse a primarily biological over social determinist explanation for gender differences on interest will afford more **systemizing** learning opportunities to **men** (in robustness check → less systemizing opportunities to women)

Original estimate: $\beta = -.30, p < .001$
Estimate without suspicious participants: $\beta = -.29, p < .001$

• **PH2a:** In general, people will afford more **empathizing** learning opportunities to **women**

Original estimate: $\beta = -.86, p < .001$
Estimate without suspicious participants: $\beta = -.87, p < .001$

• **PH2b:** In general, people will afford more **systemizing** learning opportunities to **men** (in robustness check → less systemizing opportunities to women, I think this is also how it was coded in the original analysis reported here)

Original estimate: $\beta = .86, p < .001$
Estimate without suspicious participants: $\beta = .87, p < .001$
• **Ph3**: The interaction between explanation for gender differences on interest and empathizing vs. systemizing team will be significant, such that the relationship between team type (*empathizing* or *systemizing*) and learning opportunities afforded to *men* or *women* will depend on endorsement of a biological versus social determinist explanation

Original estimate: $\beta = -.50, p < .001$
Estimate without suspicious participants: $\beta = -.52, p < .001$
B.3 Study 5

Study 5 - List of Measures

Below is the full list of variables included in Study 5 in the order they were presented to participants. Variables analyzed and reported in the main text are denoted with an asterisk.

- Attention Check
- Article Comprehension Checks
- Social Determinism*
- Biological Determinism*
- Behavioral Assignment Task*
- Task Strategy*
- Task Quality Checks
- Absolute Perceptions of Gender Differences on Ability, Interest*
- Article Ratings (Agreement, Persuasiveness, Clarity)*
- Suspicion Check
- Open-Ended Feedback
- Gender
- Age
- Occupation
- Management Role
- Ethnicity
- Political Orientation
- SES
Study 5 - Items for Measures

Attention Check

Research in decision making shows that people, when making decisions and answering questions, prefer not to pay attention and minimize their effort as much as possible. Some studies show that over 50% of people don’t carefully read questions. If you are reading this question and have read all the other questions, please select the box marked ‘Other’ and type ‘Decision making’ in the box below. Do not select 'Predictions of your own behavior.' Thank you for participating and taking the time to read through the questions carefully!

What is this study about?
- Predictions of your own behavior
- Predictions of your friend's behavior
- Political preferences
- Other: __________________________________________________

Article Comprehension Checks

According to the article, what is the main cause of women’s underrepresentation in tech?
- Men and women’s different salaries
- Men and women’s different communication styles
- Men and women’s different interests (Correct answer in BD, SD condition)
- Men and women’s different abilities

What is the main argument of this article?
- There is a biological basis for men and women's different interests (Correct answer in BD condition)
- There is no reason to believe men and women have different interests
- There is a biological basis for men and women's different abilities
- There is a socialized basis for men and women's different interests (Correct answer in SD condition)
Social Determinism

1 = *Strongly disagree* . . . 7 = *Strongly agree*

- A person’s gender-related interests are largely the product of their social origin.
- In most cases, gender differences in interest can be traced back to socio-cultural causes.

Biological Determinism

1 = *Strongly disagree* . . . 7 = *Strongly agree*

- Gender differences in interests are largely determined by one’s genetic background.
- In most cases, gender differences can be traced back to biological causes.
- Men and women's different interests are mostly due to their innate biologies.

Task Strategy

1 = *Not at all* . . . 7 = *Very much*

- I chose the person who I thought was more naturally interested in each team.
- I chose the person who I thought would be more likely to develop new interests from working on each team.

Task Quality Checks

1 = *Not at all* . . . 7 = *Very much*

- I made my selections at random.
- I put thought into each one of my selections.

Absolute Perceptions of Gender Differences on Ability, Interest

1 = *Not at all* . . . 7 = *Very much*

In your estimation…

- To what extent do men and women differ in their abilities?
- To what extent do men and women differ in their interests?
Article Ratings

1 = Not at all . . . 7 = Very much

- How much did you agree with the article's conclusions?
- How persuasive was the article?
- How clear was the article?

Open-Ended Feedback

- What did you think the purpose of today's study was?
- (Optional) Did you encounter any technical issues in today's study? If yes, please tell us about them here.
- (Optional) Did you find any parts of this study confusing or difficult to answer? If yes, please tell us about that here.

Gender

Which of the following best describes you?
- Woman
- Man
- Non-binary person
- Another option (please specify):

Age

What is your age (in years)? ______________________

Occupation

What is your current occupation? ______________________

Management Role

Are you currently in a management role?
- Yes
- No
- Not sure
Ethnicity

Which of the following ethnicities do you identify with? (Select all that apply)

☐ African American, African, or Black (e.g., Appendixali, Haitian)
☐ Hispanic / Latinx / Central or Spanish Origin (e.g., Brazilian, Mexican)
☐ American Indian, Indigenous, or Alaska Native (e.g., First Nations, Inuit, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village or Barrow Inupiat Traditional Government)
☐ Native Hawaiian or Pacific Islander (e.g., Samoan, Chamorro, Tonga, Fijian, Marshallese)
☐ Middle Eastern or North African (e.g., Egyptian, Iranian, Lebanese, Algerian)
☐ East Asian (e.g., Chinese, Japanese)
☐ South Asian (e.g., Indian, Pakistani)
☐ Southeast Asian (e.g., Filipino, Thai)
☐ White (e.g., Australian, Irish)
☐ Not listed in the options above (please specify):

__________________________________________________

Political Orientation

In general, to what extent do you consider yourself to be liberal or conservative?

- 1 = Extremely Liberal
- 2
- 3
- 4 = Moderate
- 5
- 6
- 7 = Extremely Conservative
Think about your social class, or where you stand, relative to others in society. Select an answer corresponding to the rung of the ladder where you feel you stand, where the top rung means you feel like you're at the top, and the bottom rung means you feel like you're at the bottom.

Alt Text: An image of a ladder with ten rungs. The top rung is labeled "10" and the bottom rung is labeled "1" with the rungs in between labeled in descending order as "9," "8," "7," "6," "5," "4," "3," and "2."
Study 5 – Article Stimuli

Participants in Study 5 were randomly assigned to read and summarize 1 of 2 articles, ostensibly selected from a larger pool of articles (this feature was deceptive; participants were always presented with the same two articles). Each article primed either a Biological Determinist (BD) or Social Determinist (SD) explanation for interest differences. The articles were adapted from past research priming essentialist beliefs (Dar-Nimrod & Heine, 2006) and featured real research findings (BD article: Baron-Cohen, Knickmeyer, & Belmonte, 2005 and Auyeung et al., 2006; SD article: Eccles & Jacobs, 1986 and Tenenbaum & Leaper, 2003) and expert quotes to bolster the credibility of the articles. Results of a pilot study with \( N = 207 \) men STEM professionals confirmed that both articles were equated on believability, persuasiveness, positive tone, and negative tone. In this pilot study, the BD article was rated as clearer than the SD article \((d = .52, p = .032; \text{we speculated that this could perhaps be because the BD article relied on concrete brain mechanisms, whereas the SD article relied on abstract socialization concepts})\). For each article, participants were asked to summarize three key takeaways. After reading the articles, participants were provided with two opportunities to pass a comprehension check (if they fail the first time, they are allowed to review the article and try again; if they fail the second time, they were allowed to proceed but are excluded from our data).
CAMBRIDGE — Women remain underrepresented in many sectors of tech. For example, according to the U.S. Bureau of labor statistics, women make up only 26.7% of computer and information systems managers. What explains this disparity?

According to experts, it cannot be explained by different abilities. Instead, one of the strongest explanations of occupational choice is gender differences in interests. Leading research across several studies has found that women are more interested in tasks and careers involving empathizing (the drive to identify another person's emotions or thoughts) whereas men are more interested in tasks and careers involving systemizing (the drive to analyze or construct systems). Research suggests that women's relatively lower interest in systemizing might help explain why women are not as attracted to tech careers.

![Gender Differences by Area](image_url)

**Figure B.10. Infographic accompanying biological article prime (Study 5)**

*Women are less interested in systemizing (right-hand bar) than men, even though they are just as good at it.*

Researchers have long been trying to understand what explains this interest gap. The biological camp on this issue has received the most convincing support to date in results released today from an international group of clinical endocrinology researchers. These leading researchers find evidence for hormonal differences as an explanation for the gender gap in systemizing interest. The study shows a clear biological basis for men and women’s different interests.

The new research is the largest published study to test the effect of prenatal hormones on systemizing. The findings replicate prior research showing that biological differences, brought on by males’ greater
exposure to prenatal testosterone, explain why boys are typically more interested in systemizing than girls.

One of the main findings demonstrates an influence of prenatal testosterone levels on children’s systemizing interest years later. In general, prenatal testosterone levels are higher in males than females. These hormonal sex differences manifest in physical changes to the brain, which tend to boost boys’ interest in systemizing and inhibit girls’ interest in systemizing. Underscoring the real-world relevance of these findings, the researchers note that in clinical practice, boys are more likely to receive an autism diagnosis (indicated by a high degree of systemizing interest) than girls.

Figure B.11. Photo accompanying biological article prime (Study 5)

Biological differences in the brain explain why boys are more interested in systemizing than girls.

The research was supported by the National Institute of Health (NIH), which provided this prestigious international team of researchers, led by Dr. Thomas James-Anderson from the Harvard Biochemistry Research Institute, with a grant of an unprecedented 35 million dollars to fund a 6-year study of prenatal hormones on neural development. The results that appeared today in the Journal of American Medical Association, a leading journal of the AMA, are only the start of many that will follow in the coming years from this prolific team.

Dr. Robert Hedge, speaking for the team, concluded that “synthesizing hormonal states that are equivalent between males and females in the womb may enable us in the future to elevate women’s systemizing interest drive to be in line with that of men.”

"This study is both statistically and clinically significant," said the leading author, Dr. Karen Dinear, director of child and adolescent psychiatry at the University of Wisconsin Medical Branch. "Its magnitude sheds new light on a long discourse concerning the role that biology and socialization play in the finding that, in general, men have greater interest than women in systems and things. That isn’t to say
socialization doesn’t matter, but right now, there is clearly stronger evidence for the biological component."

Other experts said the study was important in adding to the limited knowledge about the effects of sex hormones on kids’ development of interests.
Article for Social Condition:

Socialization Explains Gender Gap in Interests, Researchers Say

By TOM SMITH, Tech Journalist

CAMBRIDGE—Women remain underrepresented in many sectors of tech. For example, according to the U.S. Bureau of labor statistics, women make up only 26.7% of computer and information systems managers. What explains this disparity?

According to experts, it cannot be explained by different abilities. Instead, one of the strongest explanations of occupational choice is gender differences in interests. Leading research across several studies has found that women are more interested in tasks and careers involving empathizing (the tendency to identify another person's emotions or thoughts) whereas men are more interested in tasks and careers involving systemizing (the tendency to analyze or construct systems). Research suggests that women's relatively lower interest in systemizing might help explain why women are not as attracted to tech careers.

![Figure B.12. Infographic accompanying social article prime (Study 5)](image)

**Figure B.12. Infographic accompanying social article prime (Study 5)**

*Women are less interested in systemizing (right-hand bar) than men, even though they are just as good at it.*

Researchers have long been trying to understand what explains this interest gap. The socialization camp on this issue has received the most convincing support to date in results released today from an international group of developmental psychology researchers. These leading researchers find evidence for learning differences as an explanation for the gender gap in systemizing interest. The study shows a clear socialized basis for men and women’s different interests.
The new research is the largest published study to test the effect of childhood socialization on systemizing. The findings replicate prior research showing that socialization differences, brought on by parents’ different expectations for boys and girls, explain why boys are typically more interested in systemizing than girls.

One of the main findings demonstrates an influence of parental expectations on children’s systemizing interest years later. In general, parents expect boys to be more interested in systemizing than girls. These different gender expectations manifest in changes to the learning opportunities parents provide to kids, which tend to boost boys’ interest in systemizing and inhibit girls’ interest in systemizing. Underscoring the real-world relevance of these findings, the researchers note that in interactive science exhibits, parents are more likely to explain science concepts (facilitating more systemizing interest) to boys than girls.

The research was supported by the National Institute of Health (NIH), which provided this prestigious international team of researchers, led by Dr. Thomas James-Anderson from the Harvard Gender Research Institute, with a grant of an unprecedented 35 million dollars to fund a 6-year study of parental factors on social development. The results that appeared today in Child Development, a leading journal of the APA, are only the start of many that will follow in the coming years from this prolific team.

Dr. Robert Hedge, speaking for the team, concluded that “helping parents hold social expectations that are equivalent between boys and girls during childhood may enable us in the future to elevate women’s systemizing interest to be in line with that of men.”

"This study is both statistically and clinically significant," said the leading author, Dr. Karen Dinear, director of child and adolescent psychiatry at the University of Wisconsin Medical Branch. "Its magnitude
sheds new light on a long discourse concerning the role that biology and socialization play in the finding that, in general, men have greater interest than women in systems and things. That isn’t to say biology doesn’t matter, but right now, there is clearly stronger evidence for the socialization component.

Other experts said the study was important in adding to the limited knowledge about the effects of parental expectations on kids’ development of interests.
Study 5 – Project Teams

In Study 5, we reduced the number of project teams to the $N = 4$ front-end and back-end teams in Study 4, given that results of Study 4 revealed strongest effects of biological vs. social determinism predicting affordances with these teams. Therefore, in Study 5, all participants saw the same 4 teams as shown below. Accordingly, the intern assignment task was identical to Study 4, except it only contained trials for these 4 teams (4 critical trials, 4 filler trials).

<table>
<thead>
<tr>
<th>Client Strategy</th>
<th>Make sense of clients' needs and interpret which deliverables are most important to them</th>
</tr>
</thead>
<tbody>
<tr>
<td>People Support</td>
<td>Support project team functioning by serving as a point of contact for employee project concerns</td>
</tr>
<tr>
<td>Database Security</td>
<td>Analyze project security vulnerabilities and control access permissions to company databases</td>
</tr>
<tr>
<td>Server Admin</td>
<td>Maintain project servers by regularly maintaining, upgrading, and developing new server systems if needed</td>
</tr>
</tbody>
</table>

Figure B.14. Project teams shown to participants (Study 5)

Icon Credits (Top to Bottom): Thinking © ArtWorkLeaf; Team © HASTA ICON; Security © Setyo Ari Wibowo; Server © Rolas Design. All icons retrieved from the Noun Project and licensed under CC-BY 3.0; all icons were modified by the lead author from black and white to color icons.
Study 5 – Choice by Target Intern

We tested whether effects in Study 5 were stronger for more gender prototypical targets (i.e., highly masculine men, highly feminine women). For all participants, and for participants in the biological condition separately, chi-square tests revealed no significant differences in selections. However, a chi-square test revealed participants in the social condition were most likely to assign empathizing trials to the most prototypically feminine woman target (Gabrielle) and systemizing trials to neutral and masculine woman targets (Maria, Abby), \( \chi^2(2, N = 401) = 6.60, p = .037 \).

All Participants, Collapsing Across Condition

Table B.23. Selections by target among all participants (Study 5)

<table>
<thead>
<tr>
<th>Target</th>
<th>N Trials</th>
<th>% Trials</th>
<th>N Trials</th>
<th>% Trials</th>
<th>N Total</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria</td>
<td>171</td>
<td>22.68</td>
<td>82</td>
<td>10.88</td>
<td>253</td>
<td>16.78</td>
</tr>
<tr>
<td>Abby</td>
<td>155</td>
<td>20.56</td>
<td>85</td>
<td>11.27</td>
<td>240</td>
<td>15.92</td>
</tr>
<tr>
<td>Gabrielle</td>
<td>200</td>
<td>26.53</td>
<td>75</td>
<td>9.95</td>
<td>275</td>
<td>18.24</td>
</tr>
<tr>
<td>Mark</td>
<td>75</td>
<td>9.95</td>
<td>176</td>
<td>23.34</td>
<td>251</td>
<td>16.64</td>
</tr>
<tr>
<td>Adam</td>
<td>71</td>
<td>9.42</td>
<td>166</td>
<td>22.02</td>
<td>237</td>
<td>15.72</td>
</tr>
<tr>
<td>Christopher</td>
<td>82</td>
<td>10.88</td>
<td>170</td>
<td>22.55</td>
<td>252</td>
<td>16.71</td>
</tr>
<tr>
<td>Grand Total</td>
<td>754</td>
<td>100.00</td>
<td>754</td>
<td>100.00</td>
<td>1508</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Descriptively, among the entire sample, people do assign empathizing more to feminine targets and systemizing more to masculine targets, but no significant differences.

- A chi-square test examining selections by trial type (emp vs. sys) among women targets only (Maria, Abby Gabrielle) was not significant, \( p = .130 \)
- A chi-square test examining selections by trial type (emp vs. sys) among men targets only (Mark, Adam, Christopher) was not significant, \( p = .765 \)
Participants in Biological Condition

Table B.24. Selections by target among participants in the biological condition (Study 5)

<table>
<thead>
<tr>
<th>Target</th>
<th>N Emp Trials</th>
<th>% Emp Trials</th>
<th>N Sys Trials</th>
<th>% Sys Trials</th>
<th>N Total Trials</th>
<th>% Total Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria</td>
<td>95</td>
<td>25.82</td>
<td>36</td>
<td>9.78</td>
<td>131</td>
<td>17.80</td>
</tr>
<tr>
<td>Abby</td>
<td>79</td>
<td>21.47</td>
<td>30</td>
<td>8.15</td>
<td>109</td>
<td>14.81</td>
</tr>
<tr>
<td>Gabrielle</td>
<td>93</td>
<td>25.27</td>
<td>34</td>
<td>9.24</td>
<td>127</td>
<td>17.26</td>
</tr>
<tr>
<td>Mark</td>
<td>30</td>
<td>8.15</td>
<td>100</td>
<td>27.17</td>
<td>130</td>
<td>17.66</td>
</tr>
<tr>
<td>Adam</td>
<td>30</td>
<td>8.15</td>
<td>85</td>
<td>23.10</td>
<td>115</td>
<td>15.63</td>
</tr>
<tr>
<td>Christopher</td>
<td>41</td>
<td>11.14</td>
<td>83</td>
<td>22.55</td>
<td>124</td>
<td>16.85</td>
</tr>
<tr>
<td>Grand Total</td>
<td>368</td>
<td>100.00</td>
<td>368</td>
<td>100.00</td>
<td>736</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Descriptively, among participants in the biological condition, people generally assign systemizing more to masculine targets, but no significant differences.

- A chi-square test examining selections by trial type (emp vs. sys) among women targets only (Maria, Abby Gabrielle) was not significant, \( p = .989 \)

- A chi-square test examining selections by trial type (emp vs. sys) among men targets only (Mark, Adam, Christopher) was not significant, \( p = .190 \)
Participants in Social Condition

Table B.25. Selections by target among participants in the social condition (Study 5)

<table>
<thead>
<tr>
<th>Target</th>
<th>N Emp Trials</th>
<th>% Emp Trials</th>
<th>N Sys Trials</th>
<th>% Sys Trials</th>
<th>N Total Trials</th>
<th>% Total Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria</td>
<td>76</td>
<td>19.69</td>
<td>46</td>
<td>11.92</td>
<td>122</td>
<td>15.80</td>
</tr>
<tr>
<td>Abby</td>
<td>76</td>
<td>19.69</td>
<td>55</td>
<td>14.25</td>
<td>131</td>
<td>16.97</td>
</tr>
<tr>
<td>Gabrielle</td>
<td>107</td>
<td>27.72</td>
<td>41</td>
<td>10.62</td>
<td>148</td>
<td>19.17</td>
</tr>
<tr>
<td>Mark</td>
<td>45</td>
<td>11.66</td>
<td>76</td>
<td>19.69</td>
<td>121</td>
<td>15.67</td>
</tr>
<tr>
<td>Adam</td>
<td>41</td>
<td>10.62</td>
<td>81</td>
<td>20.98</td>
<td>122</td>
<td>15.80</td>
</tr>
<tr>
<td>Christopher</td>
<td>41</td>
<td>10.62</td>
<td>87</td>
<td>22.54</td>
<td>128</td>
<td>16.58</td>
</tr>
<tr>
<td>Grand Total</td>
<td>386</td>
<td>100.00</td>
<td>386</td>
<td>100.00</td>
<td>772</td>
<td>100.00</td>
</tr>
</tbody>
</table>

- A chi-square test examining selections by trial type (emp vs. sys) among women targets only (Maria, Abby, Gabrielle) was significant, $p = .037$. Participants in the social condition were most likely to assign empathizing trials to the most prototypically feminine woman target (Gabrielle) and slightly more likely to assign systemizing trials to neutral and masculine woman targets (Maria, Abby).

- A chi-square test examining selections by trial type (emp vs. sys) among men targets only (Mark, Adam, Christopher) was not significant, $p = .682$
Study 5 – Effects Controlling for Participant Suspicions

To test whether effects emerged when controlling for participant suspicion, we first coded participants’ open-ended response to the question: “What did you think the purpose of today's study was?” Codes ranged from 0 (participants’ response explicitly stated they did not know or was completely irrelevant to the true purpose of the study) to 3 (participants’ response correctly identified the independent and dependent variables, or otherwise suggested they understood the true purpose of the study). Full descriptive statistics for participant suspicion and coding scheme across Studies 4-6 is reported at the end of this supplement. As shown below, key effects were unchanged when controlling for participant suspicion or when removing suspicious participants (i.e., those coded as “3”).

Results controlling for suspicion codes (0-3)

- **PH1a.** People who read a biological (vs. social) determinist explanation for gender differences on interest will afford more empathizing learning opportunities to **women**

  Original estimate: $\beta = -.26, p = .103$
  Estimate controlling for suspicion: $\beta = -.28, p = .082$

- **PH1b.** People who read a biological (vs. social) determinist explanation for gender differences on interest will afford more **systemizing** learning opportunities to **men**

  Original estimate: $\beta = .44, p = .005$
  Estimate controlling for suspicion: $\beta = .43, p = .007$

- **PH2a.** In general, people will afford more empathizing learning opportunities to **women**

  Original estimate: $\beta = -1.61, p < .001$
  Estimate controlling for suspicion: Identical estimate

- **PH2b.** In general, people will afford more **systemizing** learning opportunities to **men**

  Original estimate: $\beta = 1.61, p < .001$
  Estimate controlling for suspicion: Identical estimate
• PH3. The interaction between condition and team type will be significant, such that people who read a biological (vs. social) determinist explanation will afford the most disparate learning opportunities by team type (most empathizing learning opportunities to women, most systemizing learning opportunities to men).

Original estimate: $\beta = .70, p = .002$

Estimate controlling for suspicion: Identical estimate

Results without suspicious participants (i.e., those coded as “3”)

• PH1a. People who read a biological (vs. social) determinist explanation for gender differences on interest will afford more empathizing learning opportunities to women

Original estimate: $\beta = -.26, p = .103$

Estimate without suspicious participants: $\beta = -.27, p = .123$

• PH1b. People who read a biological (vs. social) determinist explanation for gender differences on interest will afford more systemizing learning opportunities to men

Original estimate: $\beta = .44, p = .005$

Estimate without suspicious participants: $\beta = .56, p = .002$

• PH2a. In general, people will afford more empathizing learning opportunities to women

Original estimate: $\beta = -1.61, p < .001$

Estimate without suspicious participants: $\beta = -1.63, p < .001$

• PH2b. In general, people will afford more systemizing learning opportunities to men

Original estimate: $\beta = 1.61, p < .001$

Estimate without suspicious participants: $\beta = 1.63, p < .001$

• PH3. The interaction between condition and team type will be significant, such that people who read a biological (vs. social) determinist explanation will afford the most disparate learning opportunities by team type (most empathizing learning opportunities to women, most systemizing learning opportunities to men).

Original estimate: $\beta = .70, p = .002$

Estimate without suspicious participants: $\beta = .83, p < .001$
Study 5 – Mediation by Own Endorsement of Biological vs. Social Determinism

As specified in our secondary hypothesis, we tested whether the relationship between condition (biological vs. social article prime) and intern choice was mediated by personal endorsement of biological vs. social determinism, with \( b \) and \( c \) paths moderated by team type (empathizing vs. systemizing). The indirect effect was significant for systemizing teams, \( a^b = .07, p < .001 \), but not for empathizing teams, \( a^b = -.02, p = .138 \).

![Diagram](image_url)

**Figure B.15. Mediation by personal endorsement of biological vs. social determinism (Study 5)**
B.4 Study 6

Study 6 - List of Measures

Below is the full list of variables included in Study 6 in the order they were presented to participants. Variables analyzed and reported in the main text are denoted with an asterisk.

- Attention Check
- Filler Work Style Questionnaire
  - Individual Work Performance Questionnaire
  - Ten-Item Personality Inventory
- Possible Selves (Baseline)
  - Empathizing*
  - Systemizing*
  - High-Status
  - Staying*
- Team Interest (Baseline)*
- Team Choice Task*
- Possible Selves (Post-Measure)
  - Empathizing*
  - Systemizing*
  - High-Status
  - Staying*
- Team Interest (Post-Measure)*
- Perceptions of Work Assignment
- Project Manager Rating
- Trait Empathizing, Systemizing*
- Open-Ended Feedback
- Suspicion Check
- Major
- Gender
- Gender Expression
  - Masculinity
  - Femininity
- Age
- Ethnicity
Study 6 - Items for Measures

Attention Check

Research in decision making shows that people, when making decisions and answering questions, prefer not to pay attention and minimize their effort as much as possible. Some studies show that over 50% of people don’t carefully read questions. If you are reading this question and have read all the other questions, please select the box marked ‘other’ and type ‘Decision Making’ in the box below. Do not select “predictions of your own behavior.” Thank you for participating and taking the time to read through the questions carefully!

What is this study about?
- Predictions of your own behavior
- Predictions of your friend’s behavior
- Political preferences
- Other: _______________

Filler Work Style Questionnaire

- **Individual Work Performance Questionnaire**

1 = Strongly disagree . . . 7 = Strongly agree

- I manage to plan my work so it gets done on time.
- I’m able to perform my work well with minimal time and effort.
- I’m open to criticism of my work.
- My professors / supervisors are satisfied with my work.
- I recover fast after difficulties or setbacks in my work.
• Ten-Item Personality Inventory

1 = Strongly disagree . . . 7 = Strongly agree

I see myself as…

• Extraverted, enthusiastic
• Critical, quarrelsome
• Dependable, self-disciplined
• Anxious, easily upset
• Open to new experiences, complex
• Reserved, quiet
• Sympathetic, warm
• Disorganized, careless
• Calm, emotionally stable
• Conventional, uncreative

Possible Selves (Baseline)

Think a minute about what your experience in the internship at Pixlio will be like, and what you might be like at this time. Please respond using the scale below.

1 = Not at all . . . 7 = Definitely

• Empathizing

I could see myself in a role working with emotions and people at Pixlio.

• Systemizing

I could see myself in a role working with systems and things at Pixlio.

• High-Status

I could see myself in a high-status role at Pixlio.
• Staying

I could see myself staying at Pixlio after my internship is done.

Team Interest (Baseline)

[Team name, icon, description; repeated for each of the 4 project teams]

1 = Not at all . . . 7 = Definitely

• I would be interested in working with [team name] at Pixlio.

Possible Selves (Post-Measure)

Once again, think a minute about what your experience in the internship at Pixlio will be like, and what you might be like at this time. Please respond using the scale below.

1 = Not at all . . . 7 = Definitely

• Empathizing

I could see myself in a role working with emotions and people at Pixlio.

• Systemizing

I could see myself in a role working with systems and things at Pixlio.

• High-Status

I could see myself in a high-status role at Pixlio.

• Staying

I could see myself staying at Pixlio after my internship is done.
Team Interest (Post-Measure)

[Team name, icon, description; repeated for each of the 4 project teams]

I = Not at all . . . 7 = Definitely

- I would be interested in working with [team name] at Pixlio.

Perceptions of Work Assignment

My personalized work assignment…

I = Strongly disagree . . . 7 = Strongly agree

- ...gave me opportunities to strengthen my existing interests.
- ...gave me opportunities to develop new interests.
- ...felt true to who I am.

I think my project manager gave me more opportunities to work on teams related to…

I = Systems and things . . . 4 = I was given equal opportunities . . . 7 = Emotions and people

Project Manager Rating

How well did your project manager do today? Please select a score out of 100 based on how well you think they created your personalized work assignment.

[0 – 100 Rating Scale]

Trait Empathizing, Systemizing

Please respond to the following statements about yourself in general, outside the context of this study.

I = Not at all . . . 7 = Definitely

- I’m interested in work that involves emotions and people. (Empathizing)
- I’m interested in work that involves systems and things. (Systemizing)
Major

What is your major? ________________________

Gender

Which of the following best describes you?

- Woman
- Man
- Non-binary person
- Another option (please specify): _________________________

Gender Expression

People have different ways of expressing themselves and their gender identity. Below, we ask about two possible ways you might express yourself / your gender identity.

At this moment…

1 = Not at all . . . 4 = Somewhat . . . 7 = Very much

- To what extent do you express yourself as masculine?
- To what extent do you express yourself as feminine?

Age

Please indicate your age in years. ______
Ethnicity

Which of the following ethnicities do you identify with? (Select all that apply)

- African American, African, or Black (e.g., Appendixali, Haitian)
- Hispanic / Latinx / Central or Spanish Origin (e.g., Brazilian, Mexican)
- American Indian, Indigenous, or Alaska Native (e.g., First Nations, Inuit, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village or Barrow Inupiat Traditional Government)
- Native Hawaiian or Pacific Islander (e.g., Samoan, Chamorro, Tonga, Fijian, Marshallese)
- Middle Eastern or North African (e.g., Egyptian, Iranian, Lebanese, Algerian)
- East Asian (e.g., Chinese, Japanese)
- South Asian (e.g., Indian, Pakistani)
- Southeast Asian (e.g., Filipino, Thai)
- White (e.g., Australian, European, Irish)
- Prefer to self-describe: __________________________________________________
Study 6 – Process for Translation Perceiver Selections to Target Choices

After completing the pre-measures, participants in Study 6 were directed to another loading screen that read, “Your personalized work assignment is ready. Please wait while we upload your results…” After 3 seconds, the page automatically progressed to a confirmation screen. Participants were then directed to a task in which they were asked to choose between 2 teams for 8 different work tasks. Importantly, these teams were populated based on the real selections of perceivers in Study 4 (see Table B.28). After making all 8 selections, participants received a summary graphic of their work assignment.

Table B.26. Process for translating perceiver selections (Study 4) to target work assignments (Study 6)

| Step 1. For perceivers 2 SD’s above/below the mean on relative biological vs. social determinism (Study 4), what percentage of intern selections for each team were women? |
|----------------------------------|------------------------------------------------|
| 2 SD’s Higher on Biological > Social Determinism (% Women Selected) | 2 SD’s Higher on Social > Biological Determinism (% Women Selected) |
| Client Strategy                  | 50.00%                                         | 28.57%                                         |
| People Support                   | 33.33%                                         | 57.14%                                         |
| Database Security                | 83.33%                                         | 28.57%                                         |
| Server Admin                     | 0.00%                                          | 28.57%                                         |

Step 2. Divide percentages by 10 and round to the nearest integer to determine how many times each team appears as an option in Study 6

<table>
<thead>
<tr>
<th>2 SD’s Higher on Biological &gt; Social Determinism</th>
<th>2 SD’s Higher on Social &gt; Biological Determinism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client Strategy (CS)</td>
<td>5</td>
</tr>
<tr>
<td>People Support (PS)</td>
<td>3</td>
</tr>
<tr>
<td>Database Security (DS)</td>
<td>8</td>
</tr>
<tr>
<td>Server Admin (SA)</td>
<td>1*</td>
</tr>
</tbody>
</table>

Step 3. Distribute choices across N=8 trials

<table>
<thead>
<tr>
<th>2 SD’s Higher on Biological &gt; Social Determinism</th>
<th>2 SD’s Higher on Social &gt; Biological Determinism</th>
</tr>
</thead>
</table>

262
<table>
<thead>
<tr>
<th></th>
<th>(Gendered Condition)</th>
<th></th>
<th>(Counter-Gendered Condition)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice 1</td>
<td>Choice 2</td>
<td>Choice 1</td>
<td>Choice 2</td>
</tr>
<tr>
<td>Trial 1</td>
<td>CS</td>
<td>PS</td>
<td>CS</td>
<td>DS</td>
</tr>
<tr>
<td>Trial 2</td>
<td>CS</td>
<td>PS</td>
<td>CS</td>
<td>SA</td>
</tr>
<tr>
<td>Trial 3</td>
<td>CS</td>
<td>PS</td>
<td>CS</td>
<td>SA</td>
</tr>
<tr>
<td>Trial 4</td>
<td>CS</td>
<td>PS</td>
<td>PS</td>
<td>DS</td>
</tr>
<tr>
<td>Trial 5</td>
<td>SA</td>
<td>CS</td>
<td>PS</td>
<td>DS</td>
</tr>
<tr>
<td>Trial 6</td>
<td>DS</td>
<td>PS</td>
<td>PS</td>
<td>DS</td>
</tr>
<tr>
<td>Trial 7</td>
<td>DS</td>
<td>PS</td>
<td>SA</td>
<td>DS</td>
</tr>
<tr>
<td>Trial 8</td>
<td>DS</td>
<td>PS</td>
<td>SA</td>
<td>DS</td>
</tr>
</tbody>
</table>

*We rounded 0 up to 1 to provide at least one opportunity for participants to select each team.*
**Study 6 – Results Using Alternative Analytic Approaches**

In Study 6, we computed change across time using a difference score (T1 – T0) and modeled results using a relative empathizing-systemizing difference score (systemizing – empathizing). Although sometimes criticized, difference scores can provide easy to analyze measures of change (Zumbo, 1999). However, I also provide a summary of alternative analyses to demonstrate that conclusions were not affected by these analytic decisions.

**Columns** in the table below compare results using a difference score (reported in the main text) to results using residuals from a regression model predicting T1 from T0. As shown below, results are robust to either methodological decision.

**Rows** in the table below compare results using an empathizing-systemizing difference score (reported in the main text) results for empathizing and systemizing separately. All primary hypotheses were robust to either methodological decision. With one exception (the effect of condition on systemizing possible self), all secondary hypotheses were supported for empathizing and systemizing separately.

**Table B.27. Results Comparing Difference Scores to Residuals (Study 6)**

<table>
<thead>
<tr>
<th>PH1: Condition Predicting Change in Interest in Empathizing/Systemizing Teams</th>
<th>Original Estimate Using a Difference Score (T1 – T0)</th>
<th>Estimate Using Residuals Predicting T1 from T0</th>
</tr>
</thead>
</table>
| Systemizing Only | \( \beta = .58 \)  
\( p < .001 \) | \( \beta = .57 \)  
\( p < .001 \) |
| Empathizing Only | \( \beta = -.33 \)  
\( p = .004 \) | \( \beta = -.35 \)  
\( p = .003 \) |
| Relative Measure  
(Systemizing – Empathizing) | \( \beta = .63 \)  
\( p < .001 \) | \( \beta = .63 \)  
\( p < .001 \) |
<table>
<thead>
<tr>
<th></th>
<th>Condition Predicting Empathizing and Systemizing Team Selections (Single Timepoint)</th>
<th>Systemizing Only (Proportion Score)</th>
<th>Empathizing Only (Proportion Score)</th>
<th>Relative Measure (Systemizing – Empathizing)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\beta = 1.07$</td>
<td>$\beta = -1.07$</td>
<td>$\beta = 1.07$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p &lt; .001$</td>
<td>$p &lt; .001$</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td><strong>SH2:</strong></td>
<td>Condition Predicting Change in Empathizing/Systemizing Possible Selves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Original Estimate Using a Difference Score (T1 – T0)</td>
<td>Estimate Using Residuals Predicting T1 from T0</td>
<td></td>
</tr>
<tr>
<td>Systemizing Only</td>
<td></td>
<td>$\beta = .20$</td>
<td>$\beta = .23$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p = .079$</td>
<td>$p = .045$</td>
<td></td>
</tr>
<tr>
<td>Empathizing Only</td>
<td></td>
<td>$\beta = -.36$</td>
<td>$\beta = -.45$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p = .002$</td>
<td>$p &lt; .001$</td>
<td></td>
</tr>
<tr>
<td>Relative Measure</td>
<td></td>
<td>$\beta = .36$</td>
<td>$\beta = .39$</td>
<td></td>
</tr>
<tr>
<td>(Systemizing – Empathizing)</td>
<td></td>
<td>$p &lt; .001$</td>
<td>$p &lt; .001$</td>
<td></td>
</tr>
</tbody>
</table>
Study 6 – Effects Controlling for Participant Suspicion

To test whether effects emerged when controlling for participant suspicion, we first coded participants’ open-ended response to the question: “What did you think the purpose of today's study was?” Codes ranged from 0 (participants’ response explicitly stated they did not know or was completely irrelevant to the true purpose of the study) to 3 (participants’ response correctly identified the independent and dependent variables, or otherwise suggested they understood the true purpose of the study). Full descriptive statistics for participant suspicion and coding scheme across Studies 4-6 is reported at the end of this supplement. As shown below, key effects were unchanged when controlling for participant suspicion or when removing suspicious participants (i.e., those coded as “3”).

Controlling for Coded Suspicion

Compared to those who receive counter-gendered situational affordances, participants who receive gendered situational affordances will report:

- **PH1A**: Decreased interest in systemizing teams  
  Original estimate: $\beta = .58, p < .001$  
  Estimate with control: $\beta = .60, p < .001$

- **PH1B**: Increased interest in empathizing teams  
  Original estimate: $\beta = -.33, p = .004$  
  Estimate with control: $\beta = -.30, p = .013$

- **PH1** with relative interest measure  
  Original estimate: $\beta = .63, p < .001$  
  Estimate with control: $\beta = .62, p < .001$

- **SH1A**: Lower number of systemizing team selections  
  Original estimate: $\beta = 1.07, p < .001$  
  Estimate with control: $\beta = 1.10, p < .001$

- **SH1B**: Greater number of empathizing team selections  
  Original estimate: $\beta = -1.07, p < .001$  
  Estimate with control: $\beta = -1.10, p < .001$
• **SH1** with relative count measure  
  Original estimate: $\beta = 1.07, p < .001$  
  Estimate with control: $\beta = 1.10, p < .001$

• **SH2A**: Decreased systemizing possible self at the company (pre-post)  
  Original estimate: $\beta = .20, p = .079$  
  Estimate with control: $\beta = .23, p = .064$

• **SH2B**: Increased empathizing possible self at the company (pre-post)  
  Original estimate: $\beta = -.36, p = .002$  
  Estimate with control: $\beta = -.35, p = .004$

• **SH2** with relative possible selves measure  
  Original estimate: $\beta = .36, p = .002$  
  Estimate with control: $\beta = .37, p = .003$

### Analyses without suspicious participants (i.e., those coded as “3”)

Compared to those who receive **counter-gendered** situational affordances, participants who receive **gendered** situational affordances will report:

• **PH1A**: Decreased interest in systemizing teams  
  Original estimate: $\beta = .58, p < .001$  
  Estimate without suspicious participants: $\beta = .60, p < .001$

• **PH1B**: Increased interest in empathizing teams  
  Original estimate: $\beta = -.33, p = .004$  
  Estimate without suspicious participants: $\beta = -.28, p = .020$

• **PH1** with relative interest measure  
  Original estimate: $\beta = .63, p < .001$  
  Estimate without suspicious participants: $\beta = .61, p < .001$

• **SH1A**: Lower number of systemizing team selections  
  Original estimate: $\beta = 1.07, p < .001$  
  Estimate without suspicious participants: $\beta = 1.06, p < .001$

• **SH1B**: Greater number of empathizing team selections  
  Original estimate: $\beta = -1.07, p < .001$  
  Estimate without suspicious participants: $\beta = -1.06, p < .001$

• **SH1** with relative count measure  
  Original estimate: $\beta = 1.07, p < .001$  
  Estimate without suspicious participants: $\beta = 1.06, p < .001$
• **SH2A**: Decreased systemizing possible self at the company (pre-post)
  Original estimate: $\beta = .20, p = .079$
  Estimate without suspicious participants: $\beta = .21, p = .075$

• **SH2B**: Increased empathizing possible self at the company (pre-post)
  Original estimate: $\beta = -.36, p = .002$
  Estimate without suspicious participants: $\beta = -.36, p = .003$

• **SH2** with relative possible selves measure
  Original estimate: $\beta = .36, p = .002$
  Estimate without suspicious participants: Identical estimate
Study 6 – Moderation of Other Variables by Trait Empathizing vs. Systemizing

As reported in the main text, change in empathizing vs. systemizing possible selves and change in empathizing vs. systemizing interests were not moderated by participants’ trait empathizing vs. systemizing (all *p’s* > .148). Trait empathizing vs. systemizing moderated the relationship between condition and team selections (*β* = .23, *p* < .001; Figure B.22); however, a regions of significance analyses demonstrated an effect of condition across all levels of trait empathizing vs. systemizing (i.e., the regions of significance were outside the bounds of the model). A predicted mean analysis revealed that participants selected significantly more systemizing than empathizing at the highest levels of relative empathizing traits (*d* = .42), and at the highest levels of relative systemizing traits (*d* = .69). In sum, even though participants’ response to affordances depended on their trait-level empathizing vs. systemizing, their constrained affordances across conditions led women to consistently select more systemizing than empathizing teams in the counter-gendered (vs. gendered) condition.

![Figure B.16. Relationship between condition and team selections, moderated by trait empathizing vs. systemizing (Study 6)](image)
B.5 Studies 4-6

Studies 4-6 – Participant Suspicion Coding Scheme and Descriptive Statistics

To examine participant suspicion across Studies 4-6, we coded participants’ open-ended response to the question: “What did you think the purpose of today's study was?” Codes ranged from 0 to 3 as follows:

- **0** = Participants’ response explicitly stated they did not know or was completely irrelevant to the true purpose of the study
- **1** = Participants’ response mentioned the topic of study (e.g., “gender”) but did not clearly identify a hypothesis
- **2** = Participants’ response identified the independent or dependent variable, detected deception, or approached the hypothesis, but did not clearly indicate they knew the true purpose of the study
- **3** = Participants’ response correctly identified the independent and dependent variables, or otherwise suggested they understood the true purpose of the study
As shown in the table and figure below, across all three studies, only 7.99% of participants identified the exact purpose of the study. This was largely driven by participants in Study 5 (the article manipulation study). As reported in the supplement for each respective study, controlling for participant suspicion or removing suspicious participants (i.e., those coded as a “3”) does not change key effects.

Table B.28. Participant suspicion rates (Studies 4-6)

| Suspicion Code | Study 4 | | Study 5 | | Study 6 | | Total N |
|----------------|---------|-----------------|---------|-----------------|---------|
| 0 N % | 90 31.58 | 92 24.27 | 162 54.00 | 344 35.68 |
| 1 N % | 43 15.09 | 99 26.12 | 62 20.67 | 204 21.16 |
| 2 N % | 112 39.30 | 117 30.87 | 51 17.00 | 280 29.05 |
| 3 N % | 7 2.46 | 68 17.94 | 2 0.67 | 77 7.99 |
| Total N | 285 | 379 | 300 | 964 |

Figure B.17. Participant suspicion rates (Studies 4-6)