

Information Inequality

Gender Gaps in Knowledge

Molly King

Stanford University

March 31, 2019

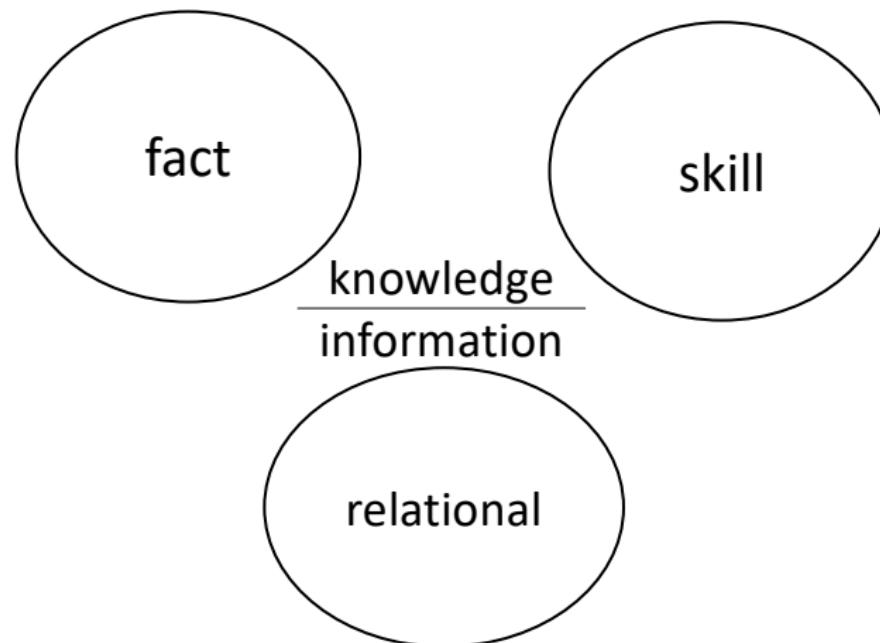
1. Good morning, everyone! Thank you for being here.
2. My name is Molly King.
3. I am a PhD candidate at Stanford University
4. Today I am going to be presenting on 1 of the papers from my dissertation about gender disparities in factual knowledge.
5. I will be exploring these gaps using analyses of quantitative survey data
6. I will show that women's disadvantage in most areas of factual knowledge closely follows patterns of social norms, suggesting the power of socialization and gendered expectation in creating knowledge inequalities.
7. I will discuss how this reflects what knowledge is valued in our society and sheds light on important sites of the reproduction of inequality.



1. Never before has such a wealth of information been so immediately accessible to so many and yet the filtering demands so high. Topics many believe to be objective and resolved after years of scientific consensus are now resurfacing as topics of factual debate. What people know, what we think we know, and what we think others know all have dramatic consequences for the future of the U.S. and the world. Information truly is power; and who possesses it and wields it most effectively has profound consequences for inequality and human welfare.
2. People disagree about facts, and this is shaping our public narrative more than ever.
3. But not only are we dealing with disagreements about facts. People also have differential access to the facts.

(Related theory: Hidalgo 2015, Simon 1971 | Cartoon credit: Loren Fishman | cartoonstock.com)

Types of Knowledge



1. Epistemology commonly distinguishes between three types of knowledge:
2. The category of “knowledge-that” is what interests me here.
3. Knowledge-that can roughly be thought of as declarative knowledge, propositional knowledge, or “explicit knowledge of a fact.”
4. My study focuses on inequalities in this type of knowledge.

Research Questions

Is there a gender gap in knowledge?

Does the gender gap in knowledge vary by domain?

What social process(es) explain(s) this gap?

1. So I wanted to perform a wide scan analysis of knowledge inequality, first finding out the answer to the most fundamental question: is there a gender gap in knowledge in different domains?
2. Then I wanted to look at who has and does not have knowledge in different domains, and how those inequalities might compare to each other.
3. I also sought to explore what social processes might explain the variance in knowledge by domain.

Data

Data Source	No. Datasets	No. Questions
General Social Survey	5	40
Pew Research Center	20	222
USC's Understanding America Study	3	48
Rand American Life Panel	2	24
National Financial Capability Studies	3	16
Kaiser Family Foundation	1	7
Health Information National Trends Survey	7	125
Integrated Health Interview Series	1	12
21st Century Americanism survey	1	4
Global Views American Public Opinion	1	2
Outlook on Life Survey	1	6
National Politics Study	1	5
Chicago Survey of Amer. Public Opinion	1	2
Total	48	513

1. in order to explore these questions, I collected data from 48 nationally representative data sets from between the years 2005 and 2015, each including at least one knowledge question.
2. I collected over 500 questions data from main public opinion survey repositories, like the General Social Survey, Pew Research Center, and ICPSR.
3. A question was included if it asked respondents about factual knowledge - a question with a generally agreed-upon answer.
4. These are true/false or multiple-choice questions that asked things like:
5. -“Chest pain is a symptom that someone may be having a heart attack - true or false.

Domains

physical science
biological science
technology
math
literature
social science
natural world
geography
domestic politics
foreign politics
economics
finance
health
religion
war
current events

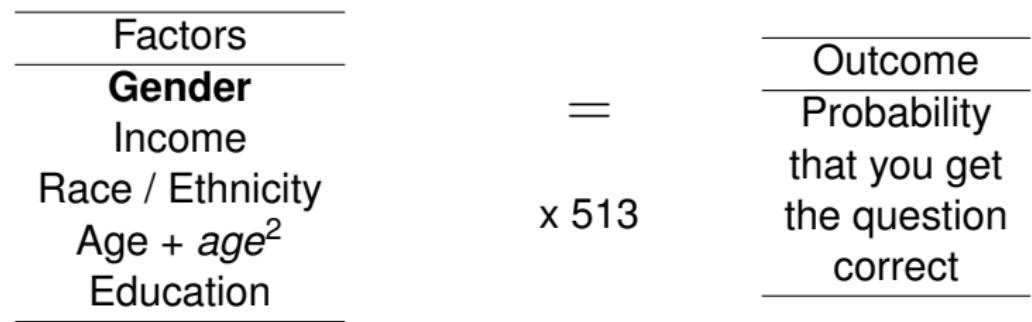
1. For each question, I marked for each individual whether they got the question correct or incorrect.
2. I curated these data and categorized them by domain.
3. This resulted in 16 topical domains.
4. People have studied knowledge gaps in many of these domains before. What is unique to my study is gathering these domains all together in one comparative framework, allowing me to look at the structured acquisition of knowledge.

What's Missing?

human development
child care
elder care
household maintenance
cooking
car maintenance
art
music

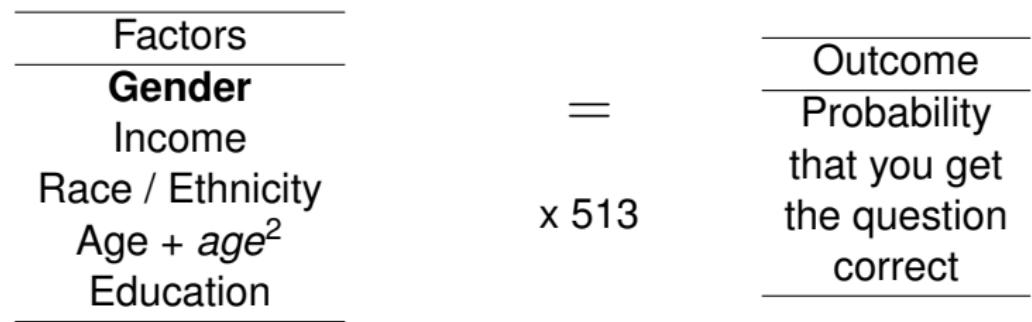
1. So that's what's here. It's also interesting to reflect on what is not here.
2. Across all of the surveys between 2005 and 2015, I was unable to find any questions related to facts about child development or elder care, household maintenance, cleaning, cooking, car maintenance, or art or music.
3. While many of these might be considered more skills based domains, there are also ways to ask questions about them that are more factually grounded.
4. Even if items around this type of knowledge are being included in nationally representative surveys, they are not being framed as knowledge.
5. The fact that these domains are missing is important commentary on what knowledge our culture considers important, a point I will return to in my hypotheses.

Model



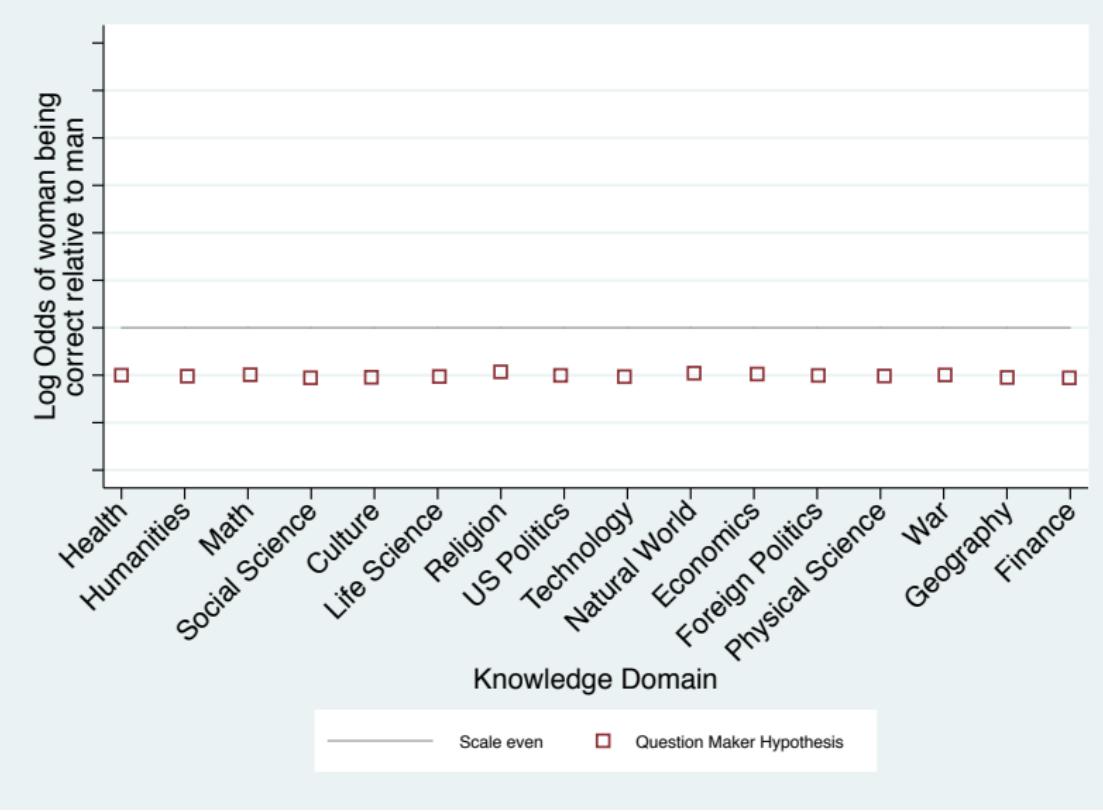
1. I also gathered many demographic characteristics about the individuals answering these factual knowledge questions.
2. For each question, I then use logistic regression to predict the probability of getting the question correct.
3. I run this analysis for each of my 513 questions.
4. I regressed the outcome of correct answer on the independent variables gender and control variables.
5. I also estimated linear probability models, and the results are much the same.
6. Because I am dealing with over 500 separate logistic regressions, I also adjust for multiple comparisons.
7. I mention this because the implication is that I can be extremely confident that these results would replicate using a different sample.
8. Finally, I created 95% confidence intervals using simulation methods, resampling from the standard error of each coefficient and averaging those across each domain.
9. I'm happy to discuss any details in the Q&A.

Model



1. I also gathered many demographic characteristics about the individuals answering these factual knowledge questions.
2. For each question, I then use logistic regression to predict the probability of getting the question correct.
3. I run this analysis for each of my 513 questions.
4. I regressed the outcome of correct answer on the independent variables gender and control variables.
5. I also estimated linear probability models, and the results are much the same.
6. Because I am dealing with over 500 separate logistic regressions, I also adjust for multiple comparisons.
7. I mention this because the implication is that I can be extremely confident that these results would replicate using a different sample.
8. Finally, I created 95% confidence intervals using simulation methods, resampling from the standard error of each coefficient and averaging those across each domain.
9. I'm happy to discuss any details in the Q&A.

Hypothesis: Question Maker Bias



1. This possibility leads me to my first hypothesis: the question maker bias hypothesis.
2. Under this hypothesis, questions in all domains are biased to be more likely to be answered correctly by men - or to be answered by men at all - because they are more likely to be written by men in the first place.
3. This is indicated here with predicted log odds in favor of men knowing more. The log odds of a woman knowing the correct answer relative to men is presented along the y axis. Under this hypothesis, men would be expected to know more in all domains.

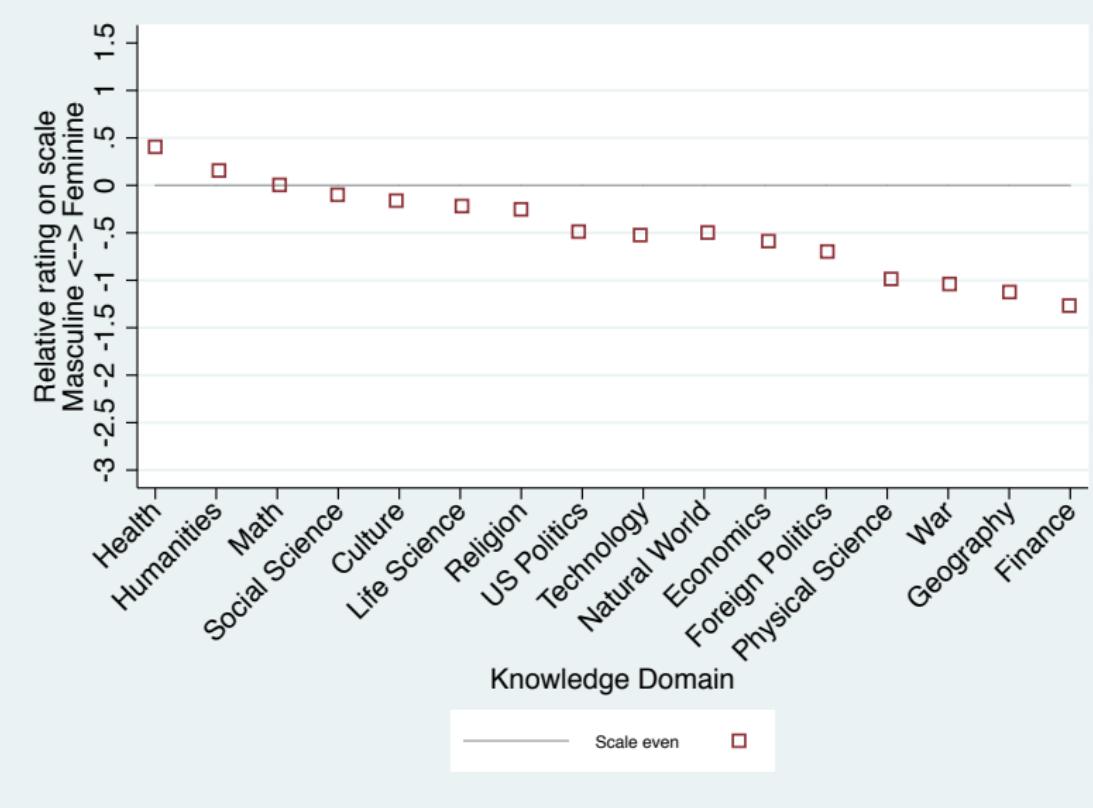
Hypothesis: Socialized Essentialism



(Related theory: Ridgeway 2006, West 1987, Cech 2013, Charles 2009 | Photo credit:

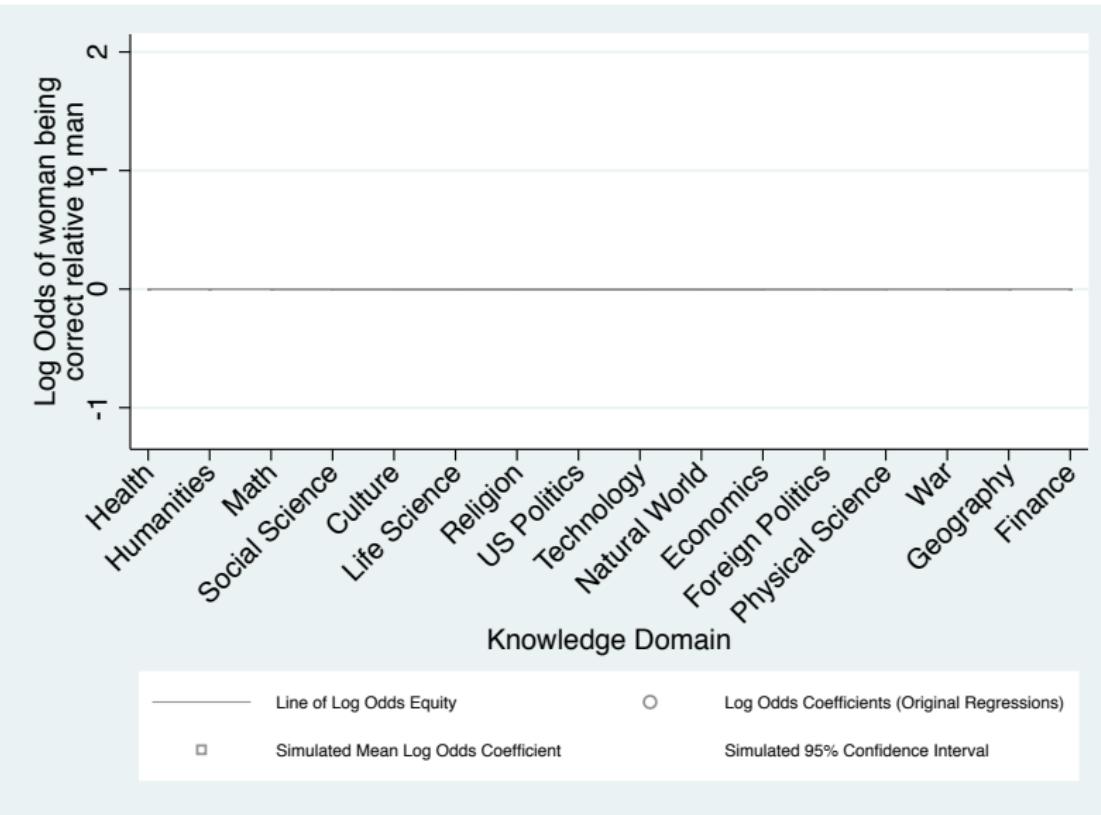
1. Alternatively, socialization may cause specialized knowledge by gender.
2. gender may have a direct impact on the types of knowledge acquired via education, occupations, and social networks.
3. Gender may also have an indirect impact via social norms, gaps in leisure time, or other socially shaped expectations of gendered knowledge interests.

Hypothesis: Socialized Essentialism



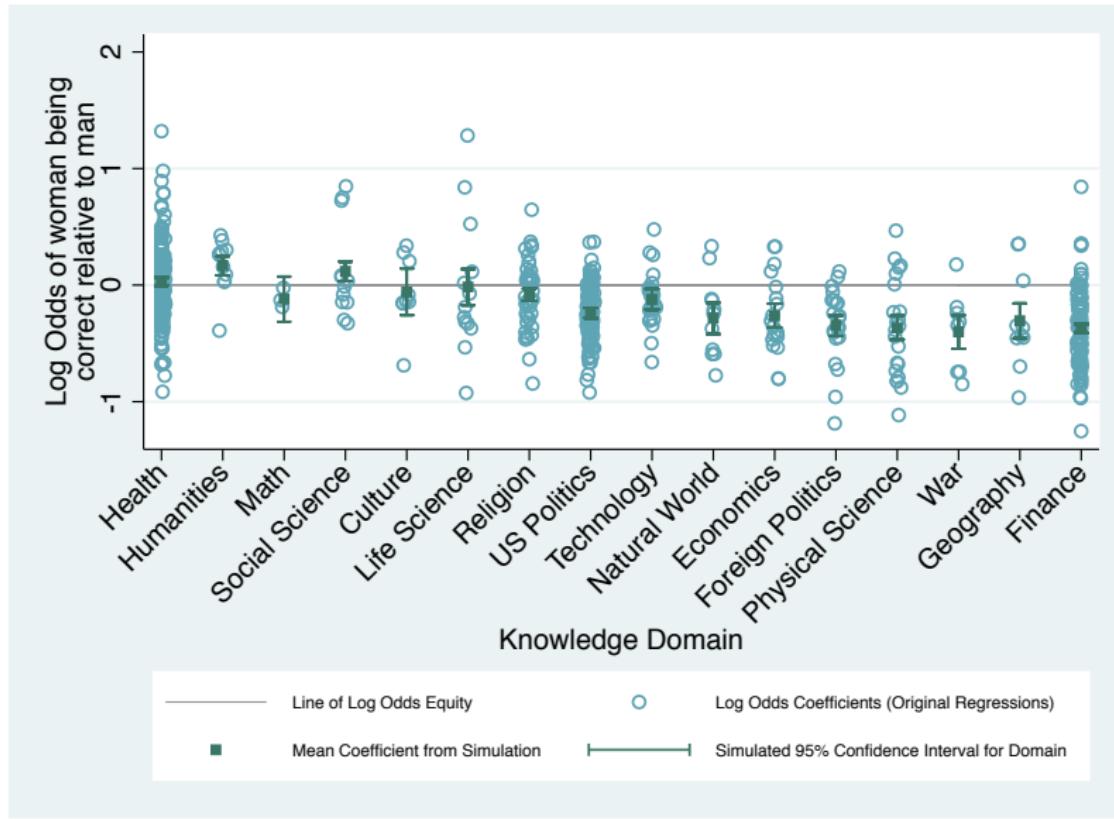
1. These interrelated mechanisms lead me to the Socialized Essentialism hypothesis.
2. Under this hypothesis, men and women specialize in knowledge domains understood to be stereotypically associated with their respective gender.
3. My estimates for the relative specialization of men and women in different domains in this hypothesis come from a survey I fielded asking about respondents' perceptions of each knowledge question.
4. I don't have time to discuss the details of this survey here, but I am happy to talk more about it in Q&A.

Domains Across X-Axis



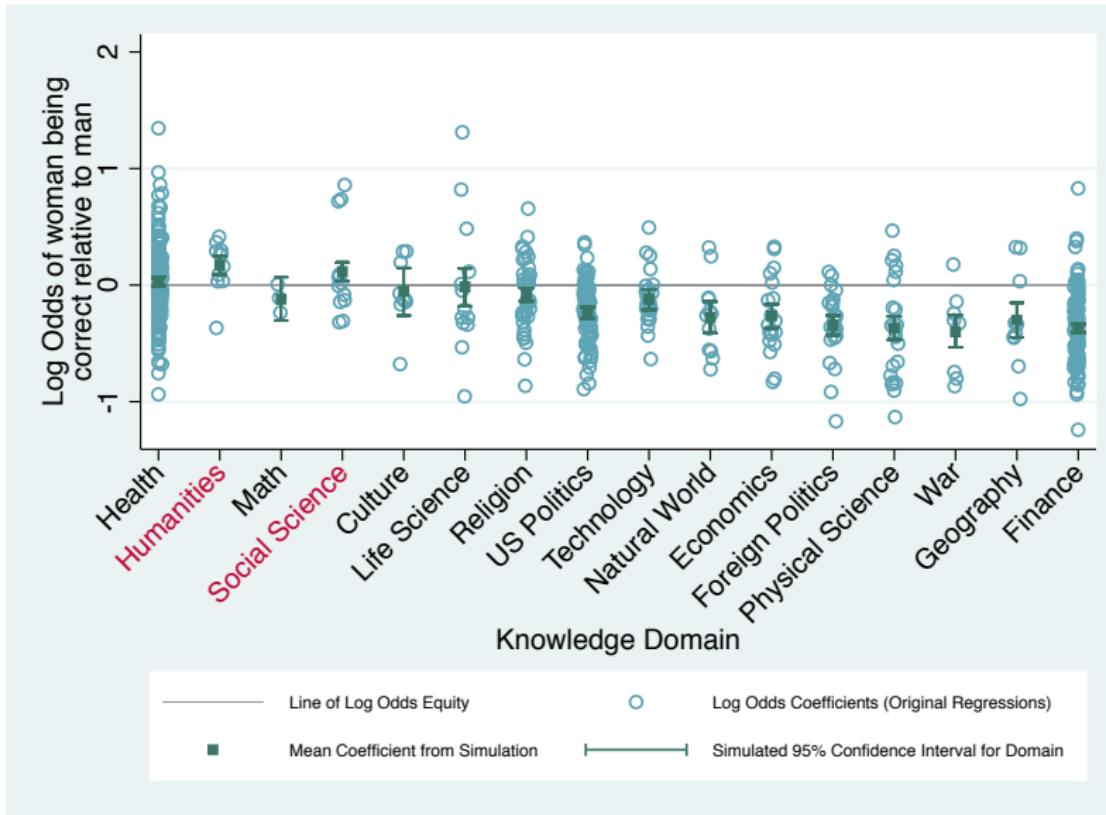
1. So here I take the 500 questions and divide them into these 16 domains across the x-axis.
2. I tested whether gender had a significant effect on knowledge within each entire domain.
3. If there were no difference in the average difference between men and women, we would see the confidence interval bar across the 0 line.

Each Model as a Data Point



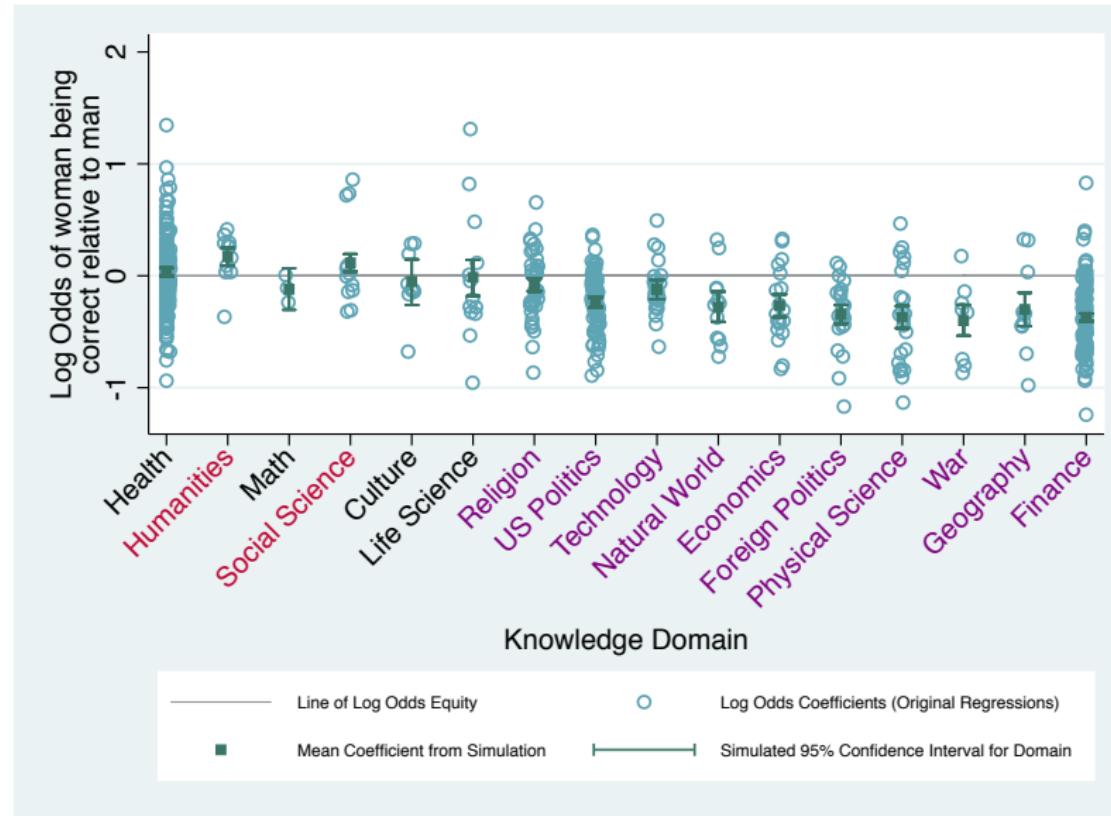
1. Each blue dot represents an individual question.
2. The green square is the mean knowledge level within each domain, and the confidence interval represents 95% simulated certainty around that mean for each domain.
3. For each domain, the simulated mean and confidence intervals show whether there is a significant difference between the 2 gender groups and the direction of that difference.

No mean gender difference in 4/16 domains



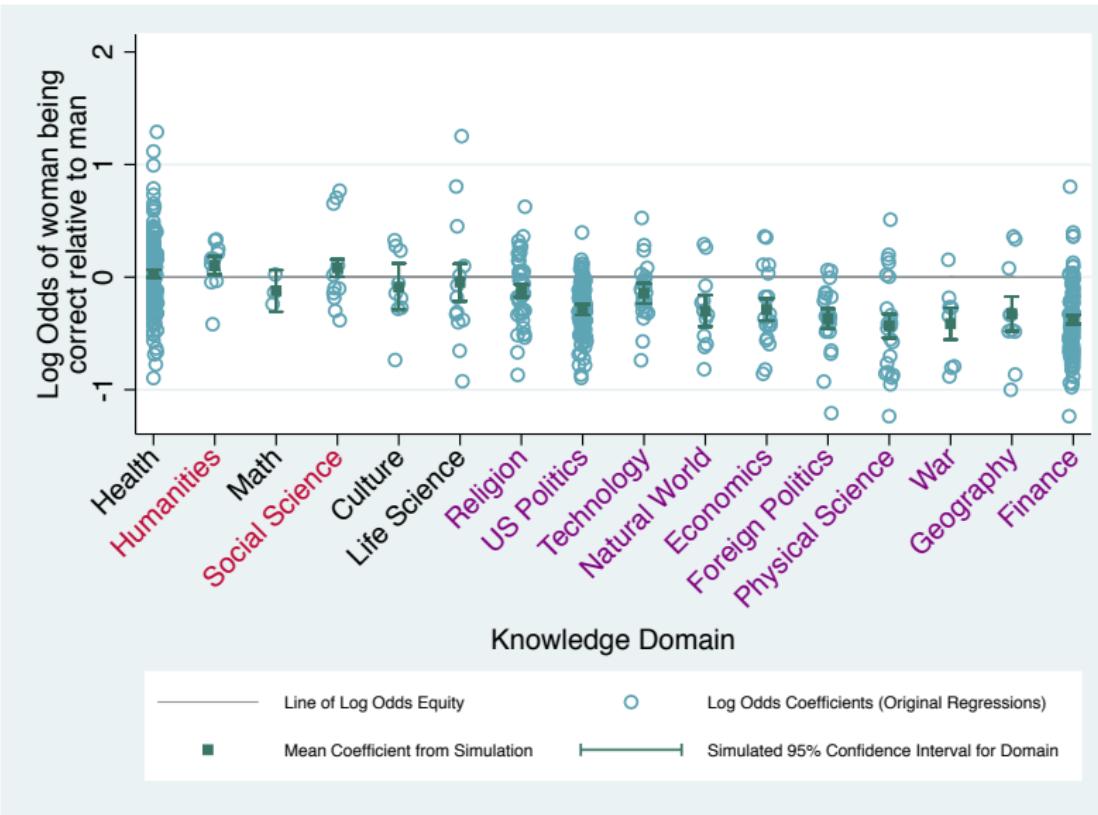
1. Women have greater average knowledge in the domains of social science and humanities.
2. There is no average gender difference in 4 of the 16 domains, because the confidence interval crosses 0.

Men have greater knowledge in 10/16 domains



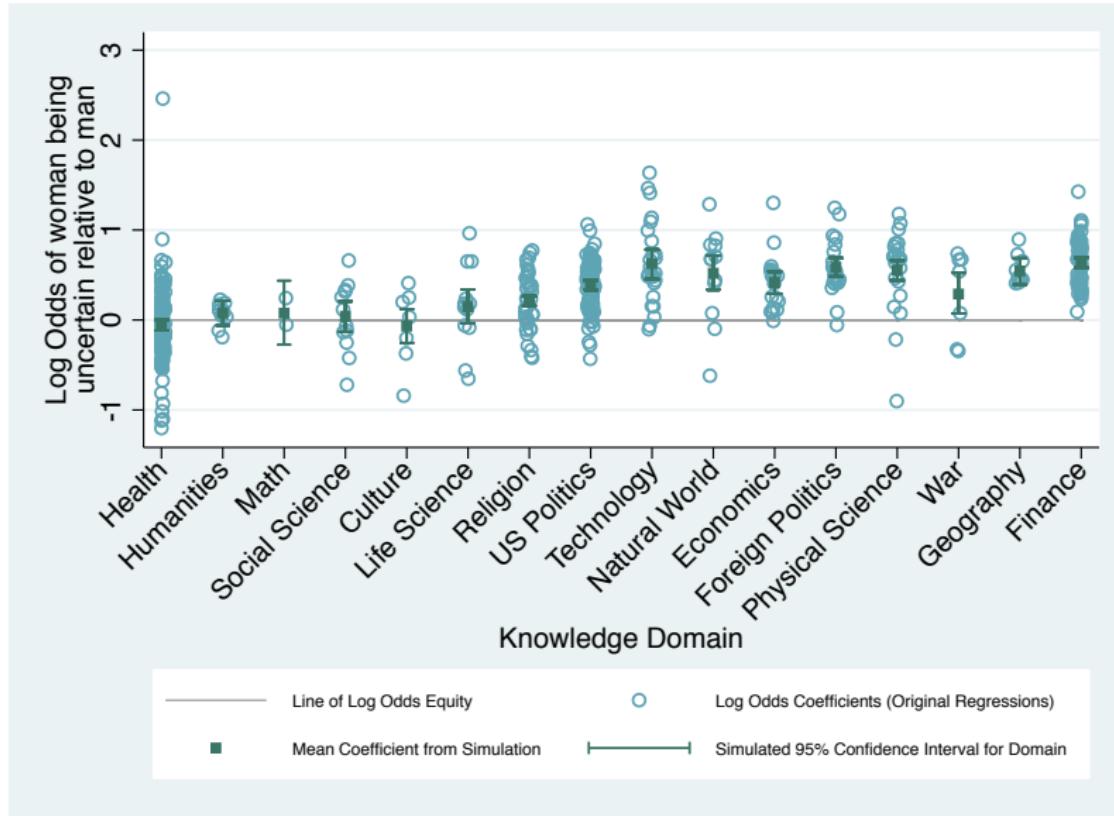
1. And men have greater average knowledge in 10 domains.
2. These are all domains where the questions were rated more masculine by respondents in my survey.

Result holds when controlling for education



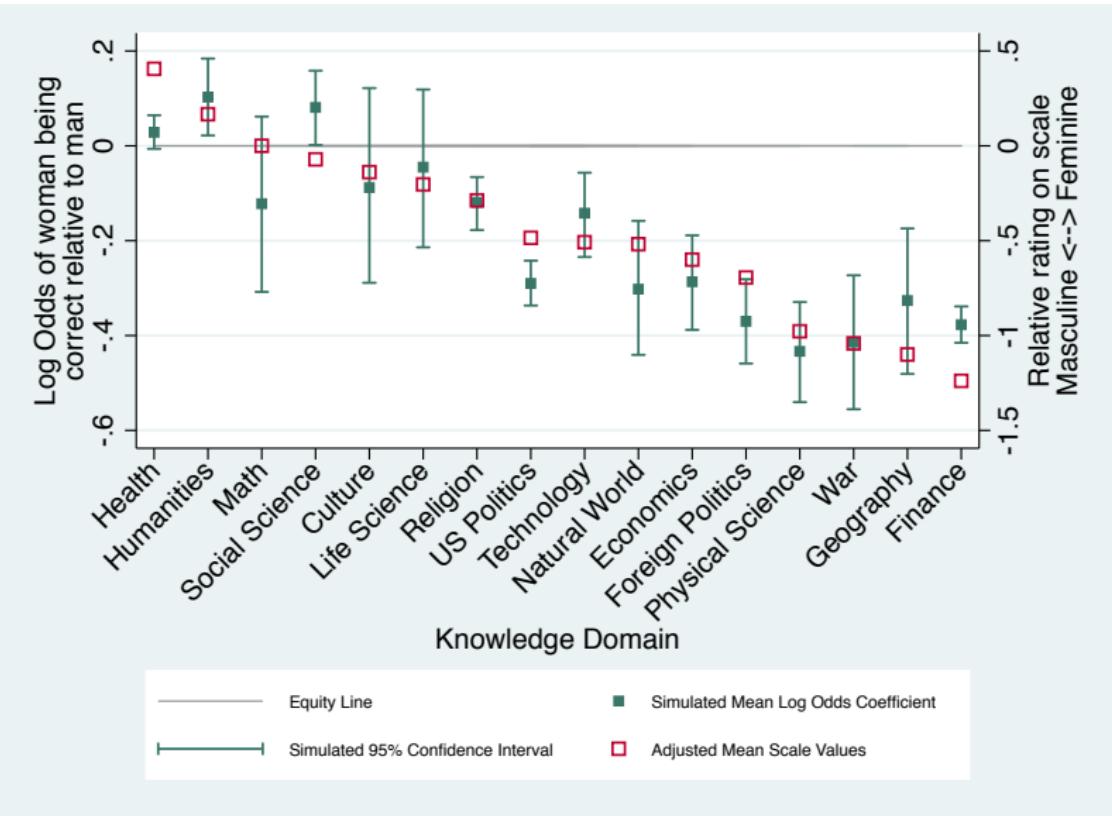
1. You might think this is driven by differences in education.
2. However, controlling for education increases the number of questions that men answer correctly significantly more often and decreases the number of questions that women answer correctly. This is because women are getting more of their knowledge advantage from their educational advantage relative to men - especially in the sciences.

Women answer more questions “don’t know”



1. I see a similar pattern when I model answers of “don’t know” to questions where that was available.
2. The same knowledge domains where men are more likely to answer correct on average are the domains where women are more likely to answer “don’t know”.
3. I’m still exploring whether this represents a true gap in knowledge or a difference in the certainty at which men and women are comfortable guessing, both of which could lead to an observed gap in knowledge.
4. In other words, I do not yet know if this is because women truly are less likely to know the information or are less willing to guess.

Results: Socialized Essentialism



1. Revisiting my original theory, I found my results most closely match the socialized essentialism hypothesis.
2. I do find high degree of correlation between what people think women and men are likely to know more about and the actual factual knowledge respondents report on nationally representative surveys.
3. Remember, my classifications into domains had nothing to do with the gendered ratings of the questions themselves. So the respondents' performance on the knowledge items and the ratings of how gendered each item is perceived to be are as independent as possible.
4. This is strong evidence that gendered socialization mechanisms are at work here in determining the domains of knowledge men and women specialize in.
5. In this way, our expectations shape our reality. Social expectations and socialization shape the knowledge acquired by men and women, with dramatic consequences for inequality.

Molly M. King

kingmo@stanford.edu



Theory Model Of Knowledge Status Differences Gender Segregated Networks Leisure

Time Socialization Next Steps: Exploring Mechanisms Methods Data Search &

Selection Data Structure My Survey on Gendered Perception of Knowledge
Questions Logit Details Multiple Comparisons Confidence Interval Simulations

Results Proportion of Don't Know Responses by Gender Other Research Gender & Self

Citation: Proportions Gender & Self Citation: Ratio

Information & Status Differences

1. Weber's three interrelated bases for inequalities in modern industrial societies are resources, power and status. Status differences between individuals may be created by virtue of information differences. Social status is defined as inequality grounded in differences in respect, esteem, and honor. "People use culture to make connections with one another"; information that is important to one's social contacts may lead those important others to publicly acknowledge that person and contribute to their sense of being valued. When aggregated at the group level, this control of information may be "transformed into more essentialized differences among 'types' of people, status beliefs that fuel social perceptions of difference." At the macro-level, I argue these differences in information, originally created on the basis of gender, give a force to status beliefs that can reproduce material inequalities independent of initial differences in power or resources.

● (*Back*)

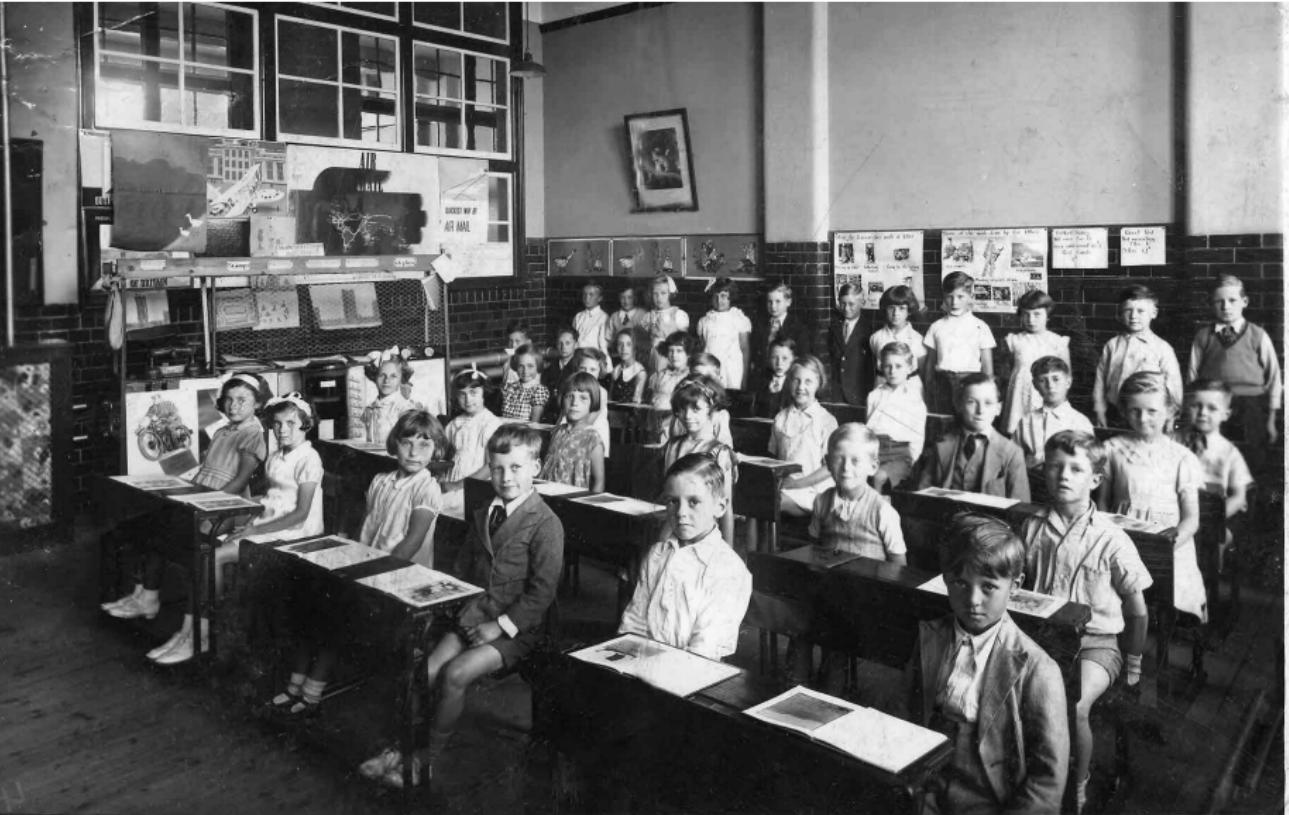
Leisure Time



(Related theory: Parker 2011 | Photo credit: East Riding Archives)

1. A number of different mechanisms may be at work behind this finding of gender essentialism in knowledge.
2. Notably, as adults, gender and class structure lives differently. Men and people of the middle and upper classes have more time for leisure, and we might imagine that leisure time is one of the key times when adults acquire facts about non-occupational related domains.
3. Depending on family structure, women have 3 to 5 fewer hours of leisure time per week than men.
4. This is in part due to different social roles and expectations, which may have an independent effect on knowledge gaps by funneling women and men toward specialized knowledge consistent with these duties and norms. Furthermore, ideas about what are socially acceptable and desirable interests for men and women in the United States may shape pursuit of knowledge in different domains.

Socialization



(Related theory: Ridgeway2006, West1987, Cech 2013, Charles 2009 | Photo credit:

1. Related, men and women are socialized differently from a very early age.
2. Children are socialized in gendered ways.
3. Additionally, gender and class differences in college major and occupation choice funnel individuals into different streams of knowledge exposure, which leads to reinforcing cycles of knowledge gaps and the resulting resources.
4. Gender differences in college major may be a result of self-expressive career choices, and occupational sorting may further reinforce knowledge differences throughout the life course.
5. After college, expectations of discrimination shape entry into jobs. Hostile environments in training and everyday discrimination in the workplace provide the structural underpinnings for demographic differences in knowledge.

Gender Segregated Networks



(Related theory: Lutter 2015, Marin 2012, DiMaggio & Garip 2012 | Photo credit:

<https://www.pbs.org/wgbh/>

Molly King (Stanford University)

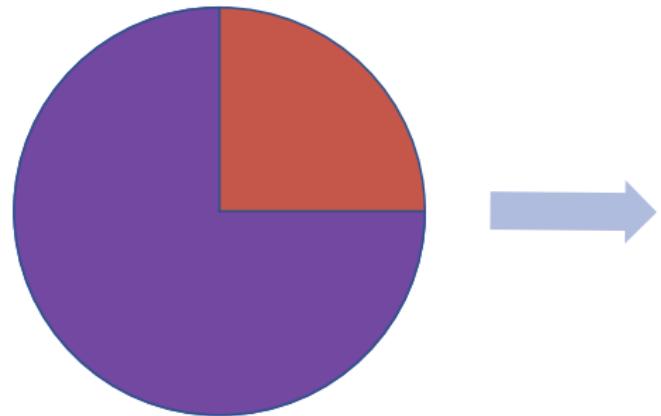
Information Inequality

March 31, 2019

5 / 24

1. Finally, while men's and women's networks are similar until they have children, they diverge dramatically thereafter. Social networks are important source of all types of information - from information about job to skill sharing - so there is no reason to believe that men and women would not be receiving different types of factual information through the networks as well.
2. Gender-segregated networks may also be an important vehicle for maintaining prior socialization.
3. I have studied the importance of networks and occupations in who gets credit for the production of NEW knowledge in a different set of projects by looking at the effect of gender on author order and self citation practices in academic publishing. I would be happy to talk about this work during the Q & A portion of the talk.

Next: Exploring Mechanisms



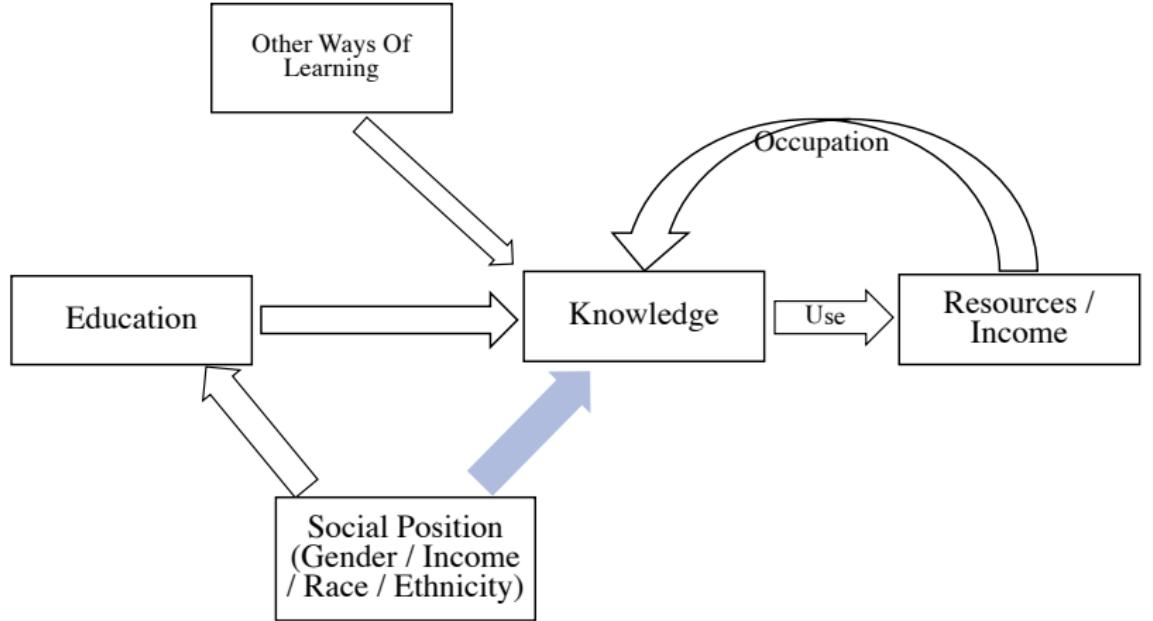
Assert Gender?

or

Adapt to Majority Knowledge?

1. My next projects will explore these potential mechanisms in more detail.
2. One project I have planned will investigate whether gender performance (or “doing gender”) contributes to knowledge gaps. Using a subset of my data where I can control for occupation, I will explore whether men and women in the same occupation differ in their knowledge of different domains. For example, do men in majority-female occupations use knowledge as a way to compensate and assert their masculinity, or do they adapt by acquiring gendered knowledge? I will compare these hypotheses and occupational-specific knowledge hypotheses with a null hypothesis of increasing gender equality over time. This will provide one test of possible theoretical mechanisms underlying the gender gap in knowledge among different domains.

Acquisition and use of knowledge



1. Again, revisiting my model of social influences on knowledge acquisition, I find that indeed gender has a significant effect.
2. Findings are consistent with a model that implies:
 3. — demographic characteristics affect the knowledge an individual has, and
 4. — using knowledge to access resources.
5. Again, my broad survey allows comparisons across knowledge domains and demographic status groups to understand how the socially structured acquisition of knowledge leads to stratified gaps in information resources.
6. All these different social forces filter toward concrete differences in knowledge that have implications for citizenship, occupational chances, and health, as well as many other opportunities to access resources.
7. This analysis is part of a broader research project that seeks to build out the study of knowledge inequality.
8. Understanding the demographic patterns of knowledge inequality can help sociology move toward a better understanding of the mechanisms underlying inequality more broadly.

Data Search & Selection

Databases searched:

- ICPSR – 4,581 surveys reviewed
- Data.gov – 1,117 surveys reviewed

Inclusion criteria:

- survey search includes the term “knowledge”
- years 2005 – 2015
- U.S. nationally representative on race, gender, age

(Back)

1. Additional surveys were identified by searching ICPSR using the term “knowledge,” limiting the search to the time period 2005 to 2015 in the United States. All (4,581) surveys returned from the search were assessed for relevance and inclusion in the study. I performed the same search for the term “knowledge” on the Data.gov database. Criteria for selection included whether the surveys were generally nationally representative and were limited to the time period 2005 to 2015. I reviewed the resultant 1,117 surveys from Data.gov for relevance.
2. This returned several additional surveys:
3. Annenberg National Health Communication Survey;
4. the National Financial Capability Studies
5. the 21st Century Americanism survey
6. the Global Views American Public Opinion and Foreign Policy Survey;
7. the Outlook on Life Survey;
8. the State of the First Amendment surveys;
9. and the Chicago Council Survey of American Public Opinion and U.S. Foreign Policy.

Data Structure

sc_earthsun	Linearized			
	Odds Ratio	Std. Err.	t	P> t
female	.4518979	.0450369	-7.97	0.000
race_hisp	.7261311	.150449	-1.17	0.243
race_black	.3949817	.0374355	-9.80	0.000
race_asian	1.91988	.0824735	+17	0.000
race_other	.7465925	.0794548	-2.75	0.007
dV_fincG_1k	1.001214	.0005461	2.22	0.031
edu_HS	1.673856	.000793	9.58	0.000
edu_someCol	2.300979	.0567248	33.80	0.000
edu_colPlus	6.263558	.5777689	19.89	0.000
_cons	3.218947	.2726742	13.80	0.000

DOMAIN	Literature		Science		
STATUS	L1	L2	S1	S2	S3
Gender (G)	GL1	GL2	GS1	GS2	GS3
Class (C)	CL1	CL2	CS1	CS2	CS3
Race (R)	RL1	RL2	RS1	RS2	RS3

(Back)

1. I regressed the outcome of correct answer on the independent variables income, gender, race and ethnicity categories, and controls education categories, and age and age squared.
2. Although the ideal measure of class here would have been parental occupation, I had to use respondents' family or household income because parental education was not available in most surveys.

(Back)

Adjusting for Multiple Comparisons

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	
0.01	
0.02	HIGHLY SIGNIFICANT
0.03	
0.04	
0.049	SIGNIFICANT
0.050	OH CRAP. REDO CALCULATIONS.
0.051	
0.06	ON THE EDGE OF SIGNIFICANCE
0.07	
0.08	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE
0.09	$p < 0.10$ LEVEL
0.099	HEY, LOOK AT
≥ 0.1	THIS INTERESTING SUBGROUP ANALYSIS

1. Because I am dealing with over 400 separate regressions, I also adjust for multiple comparisons using the Holland method.
2. I mention this because the implication is that I can be extremely confident that these results would replicate using a different sample.

Confidence Interval Simulations

Draw 1000 repetitions from a standard normal distribution, generate a distribution of 1000 possible coefficients for each knowledge question:

$$\hat{\beta}_i = \beta + (SE * x_i),$$

where $x_i \sim N(0, 1)$.

Order all $\hat{\beta}_i$ such that

$$\hat{\beta}_i \leq \hat{\beta}_{i+1}$$

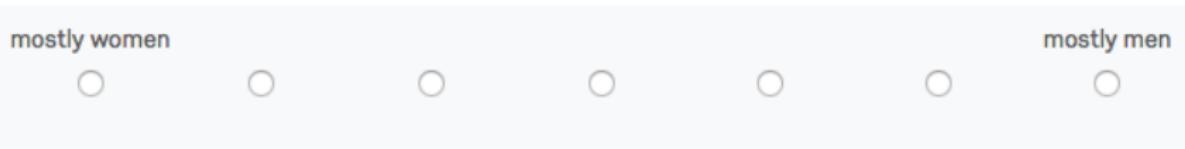
Find value of $\hat{\beta}_i$ at the 2.5th percentile and the 97.5th percentile.

1. I take each coefficient from each of my logistic regressions, and "resample" from a possible distribution of these coefficients adjusted by each coefficient's standard error multiplied by a random draw from a normal distribution.
2. Drawing 1000 repetitions from a standard normal distribution, I generate a distribution of 1000 possible coefficients for each knowledge question. In other words, I repeat this 1000 times for each of the knowledge questions (each of the 513 logistic regressions).
3. Then, within each knowledge domain, I line these resampled simulated coefficients up in order and use the 2.5th percentile and the 97.5th percentile as my upper and lower confidence bounds for that particular knowledge domain. These values are the lower and upper bounds of the 95 percent confidence interval for each information domain within each demographic subgroup.

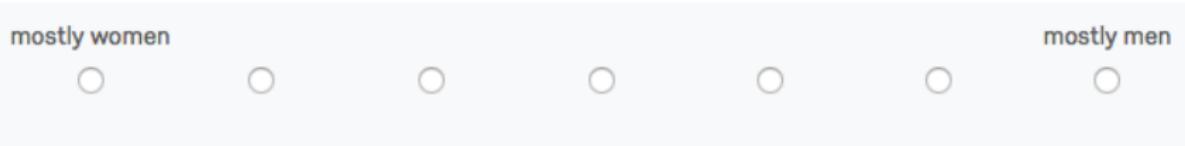
(Back)

My Survey: Gendered Perception of Knowledge Questions

- Who mostly watches news about X?



- Who mostly talks to their friends about X?

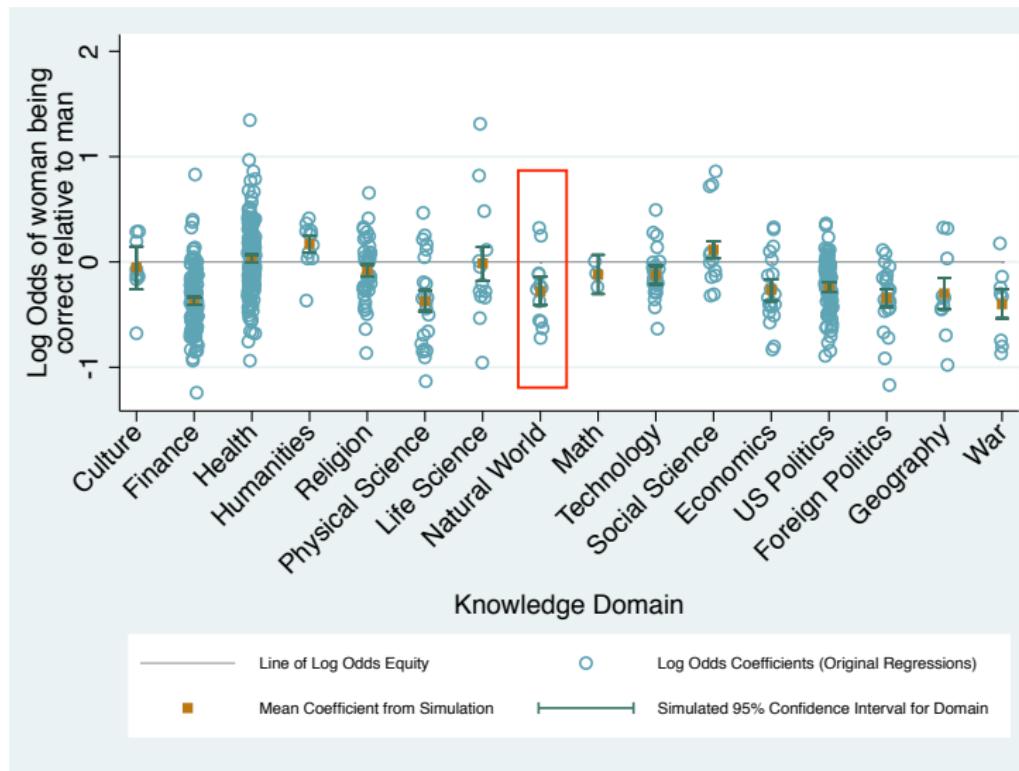


- Who do most people in our society think has more knowledge about X?



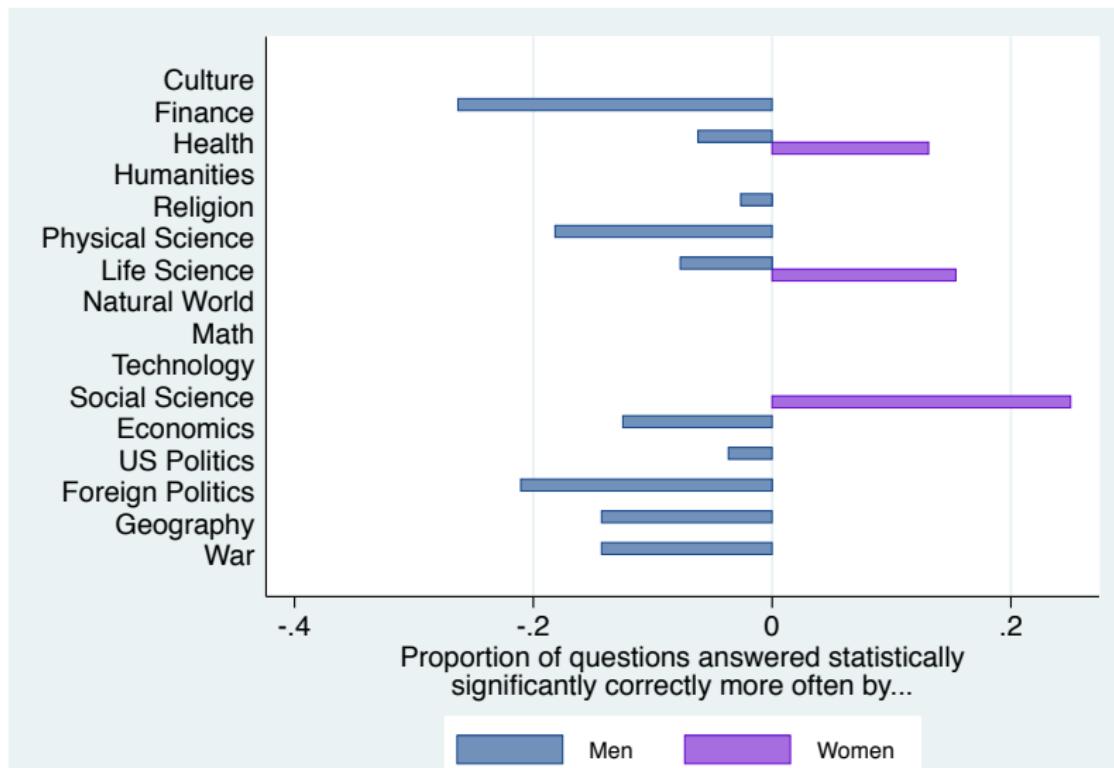
- I fielded a survey using these 3 items [read items]
- I chose these 3 items out of 6 original items based on which showed the highest inter item reliability scores.
- I used these questions to create a scale representing the socially perceived masculine or feminine nature of a given knowledge question.
- I then use this perceived gender of each item to develop an overall scale for my domains.
- The socialized essentialism hypothesis predicts that the respondents on the nationally representative surveys with the factual knowledge questions will perform in correlation to the gender each item is perceived to be. In other words, women answer more questions incorrectly in those domains that are perceived to be more "masculine," and vice versa.

Mean significance vs. Question significance



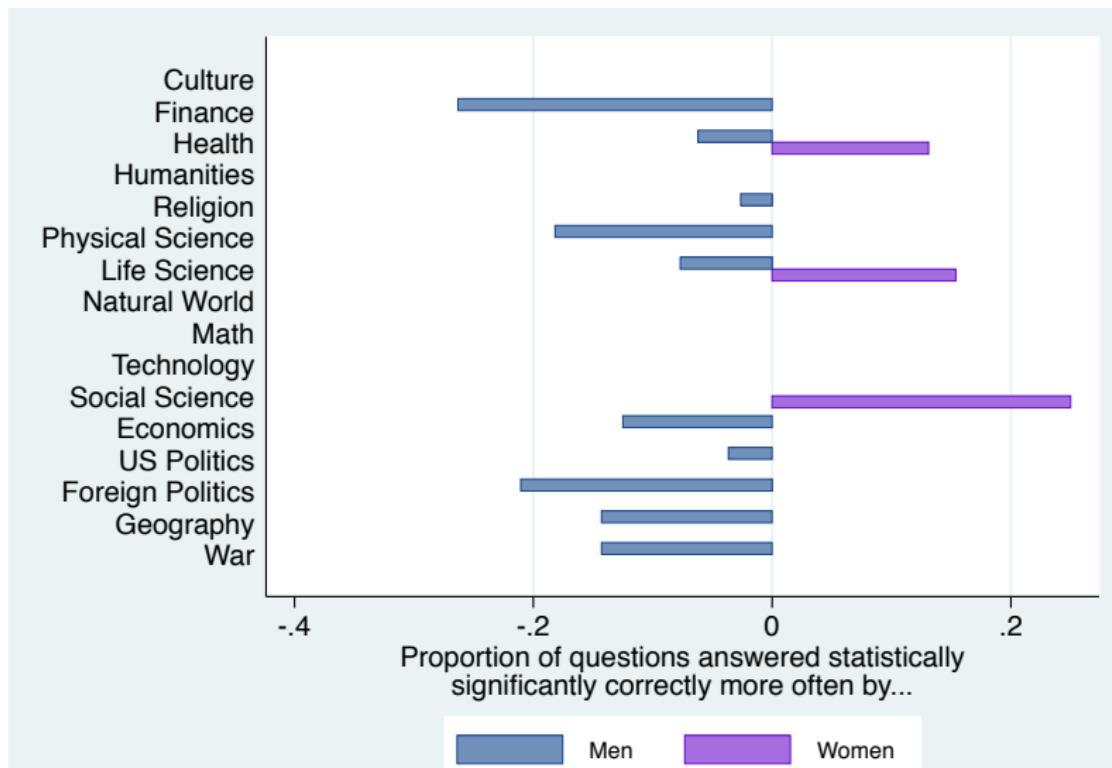
1. For example, note here how in the natural world domain women have .8 odds smaller of knowing the correct answer in the domain, on average.
2. But this first figure does not indicate how many of these questions show knowledge differences that are statistically significant.
3. So for each of these 500 questions, I tested whether the gender difference in knowledge was significant.

Men answer greater proportion of questions correctly in 65% of domains



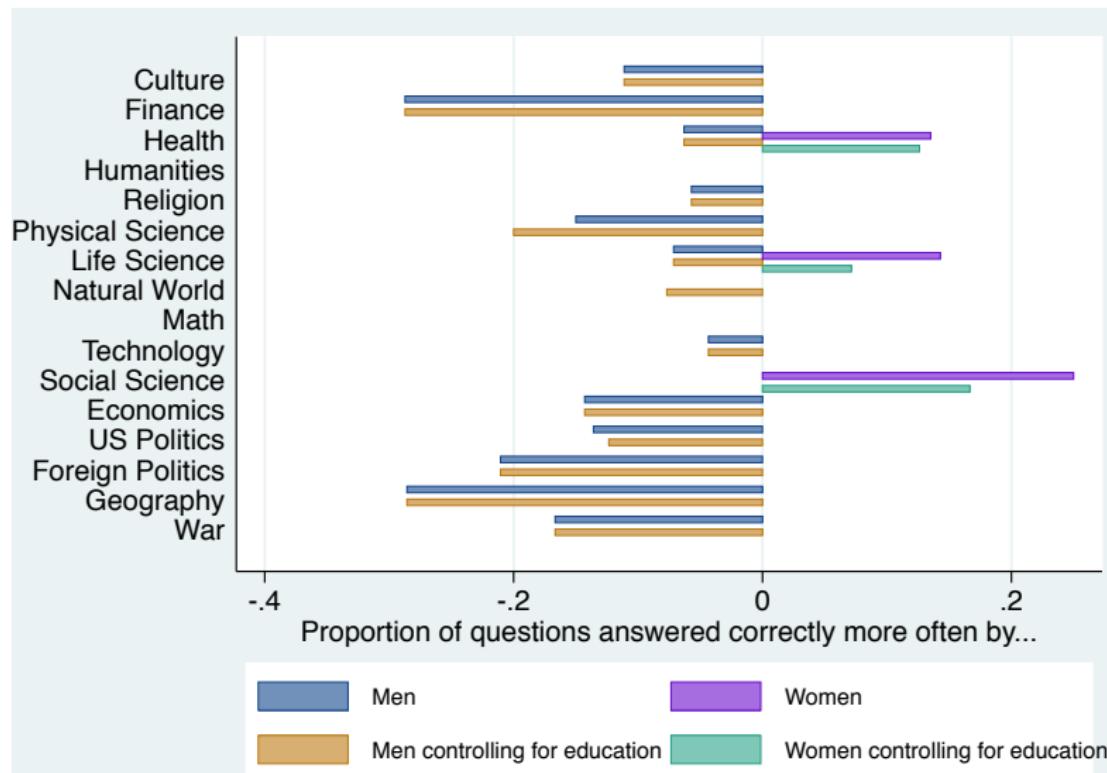
1. However, not all of those coefficients are statistically significant. In this figure, each bar represents the proportion of all questions in a given domain that men or women were more likely to answer correctly such that these differences were statistically significant.
2. So in the example of the natural world, the average gap between men and women across all items in the natural world domain is significantly different from 0, but there was no single question where men or women were more likely to be correct.
3. This can also go the other way, in theory, where the average gap is not significantly different from 0 but individual questions are. Though there are no examples of this here, social science does come close.

Men answer greater proportion of questions correctly in 65% of domains



1. I also find that women answer a greater proportion of questions correctly in the domain of health, though this is a bit more complicated. Both men and women have some questions that they are significantly more likely than the other group to answer correctly, so there are proportions in both directions on the graph.
2. One reason we might consider this to be notable is that men's poor health outcomes are typically explained by behavioral differences. Here differences in health outcomes might also be explained by a disparity in health knowledge.
3. You might assume this is driven by gender differences and education.

Controlling for education, men answer greater proportion of questions correctly in 63% domains



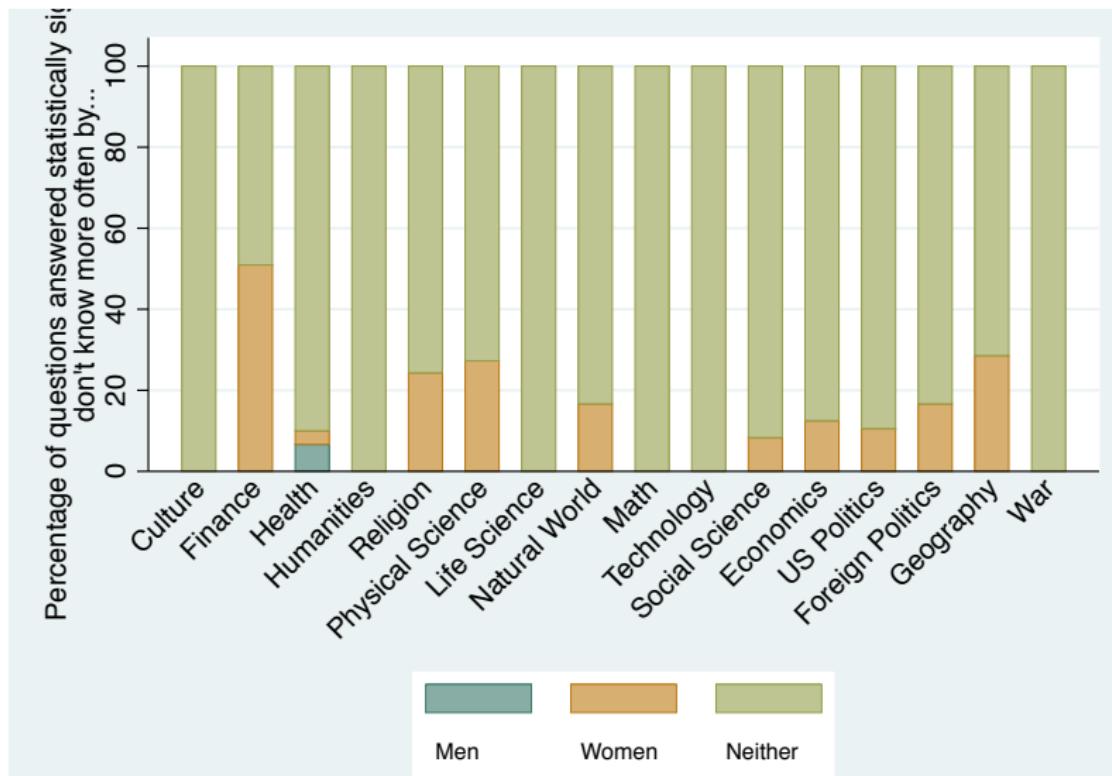
1. But overall the trend holds.
2. Men are particularly advantaged in the domains of finance, foreign politics, and geography.
3. Overall, Men answer a greater proportion of questions statistically significantly correctly in 3/5ths of the domains. Women answer a greater proportion of questions correctly in 1/5th of the domains.
4. Essentially, this residual effect of gender cannot be explained by differences in education level or income.

“Don’t Know” Model



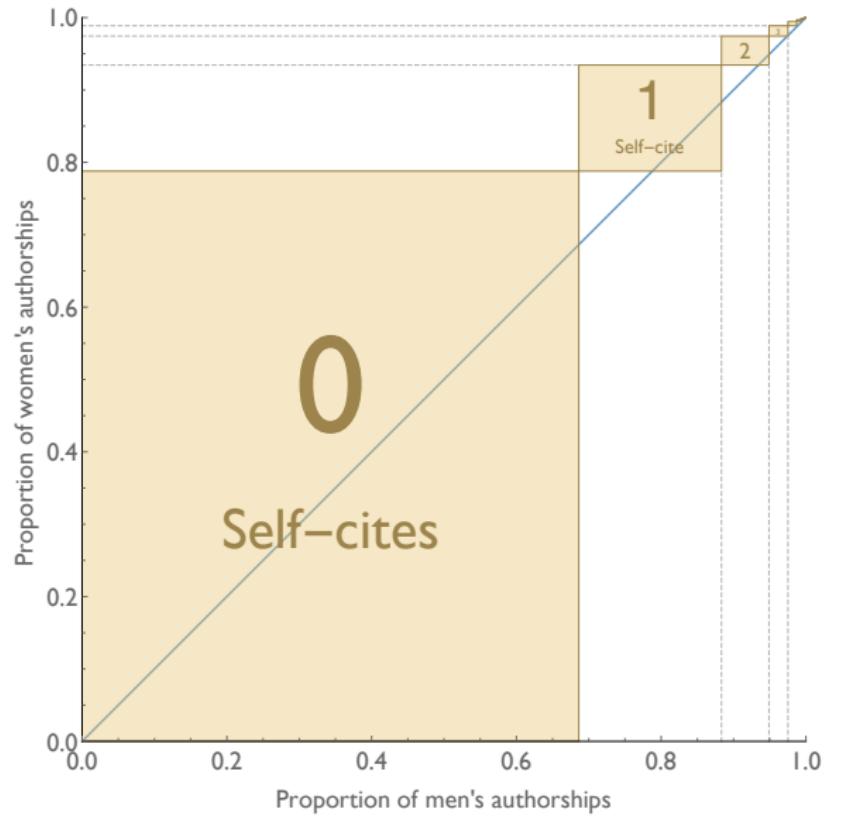
1. In a slightly different model, for each question, I then use logistic regression to predict the probability of answering ‘don’t know’ for all questions where that was an option.
2. I regressed the outcome of uncertainty on the independent variables gender and control variables.
3. I have also estimated linear probability models and the results are much the same.

Controlling for education, women answer greater proportion of questions “don’t know”



1. In all but 16 domains, women are more likely, on average, to answer don't know across all questions as well, even after controlling for education.
2. However, I do not yet know if this is because women truly are less likely to know the information or are less willing to guess.
3. I am currently working on a process to estimate each group's certainty threshold. I can talk this in q&a if you are interested.

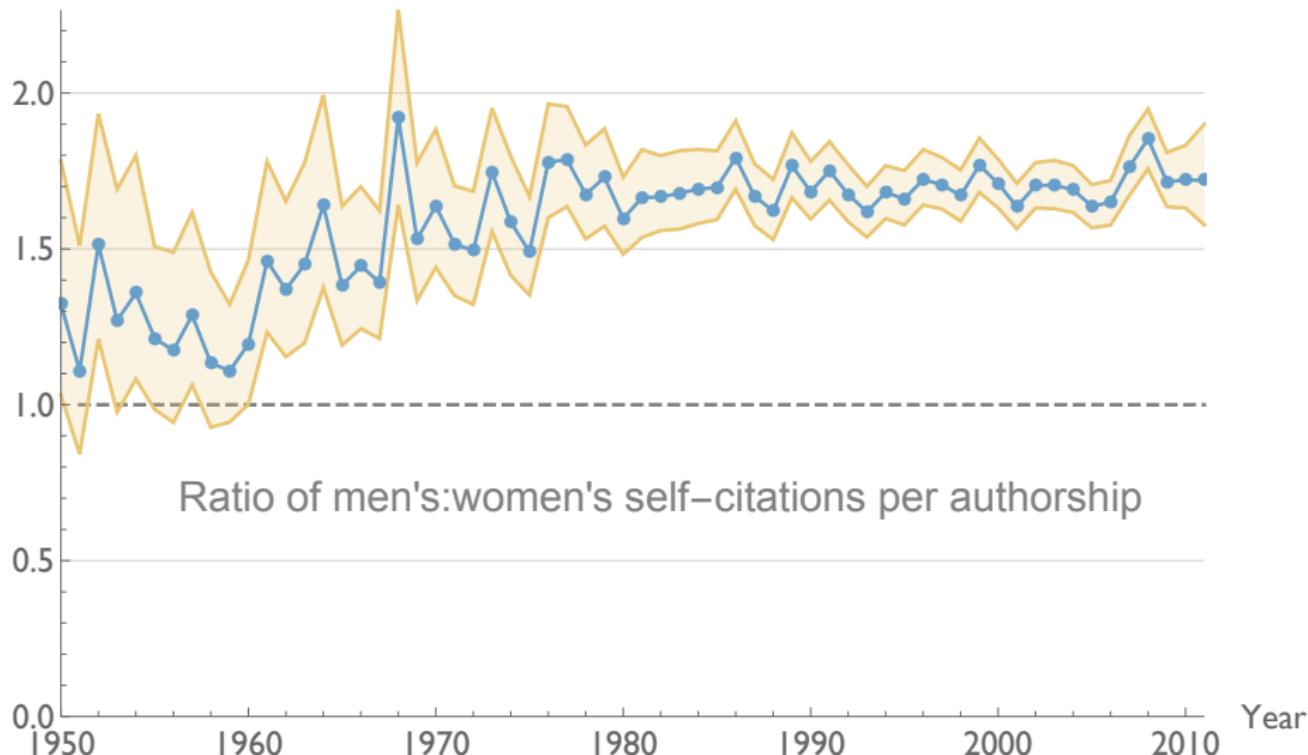
Gender & Self Citation: Proportion with Self Citation



1. shows members of self-citations grouped by proportions of menâŽs and womenâŽs authorships. We show the proportion of men with a certain number of self-citations on the x-axis and the corresponding proportion of women on the y-axis. If men and woman behaved similarly in their approaches to self-citation, the corners of the boxes would trace the x-y diagonal. Instead, wherever there is a difference in the proportion of men and women citing themselves a certain number of times, the corners of the boxes deviate from the diagonal.
2. relative to menâŽs authorships, womenâŽs authorships are more likely to feature zero self-citations. Women cite themselves one or more times in their papers less often than men do. In other words, compared with men, women are overrepresented in the zero self-citations category and underrepresented in terms of citing their papers at all. For example if in a paper you never cite another paper of your own, you are among the majority of men (68.6 percent) and women (78.8 percent) who do not cite themselves.
3. whenever a box is wider than it is tall, there is a greater proportion of men authorships in that category of self-citations. If you have one self-citation, you are in the 68th to 88th percentile range for men (representing 20 percent of menâŽs authorships) but the 78th to 93rd percentile for women (representing only 15 percent of womenâŽs authorships). With four self-citations in a single paper, a woman is in the

Gender & Self Citation: Ratio

Ratio M:W

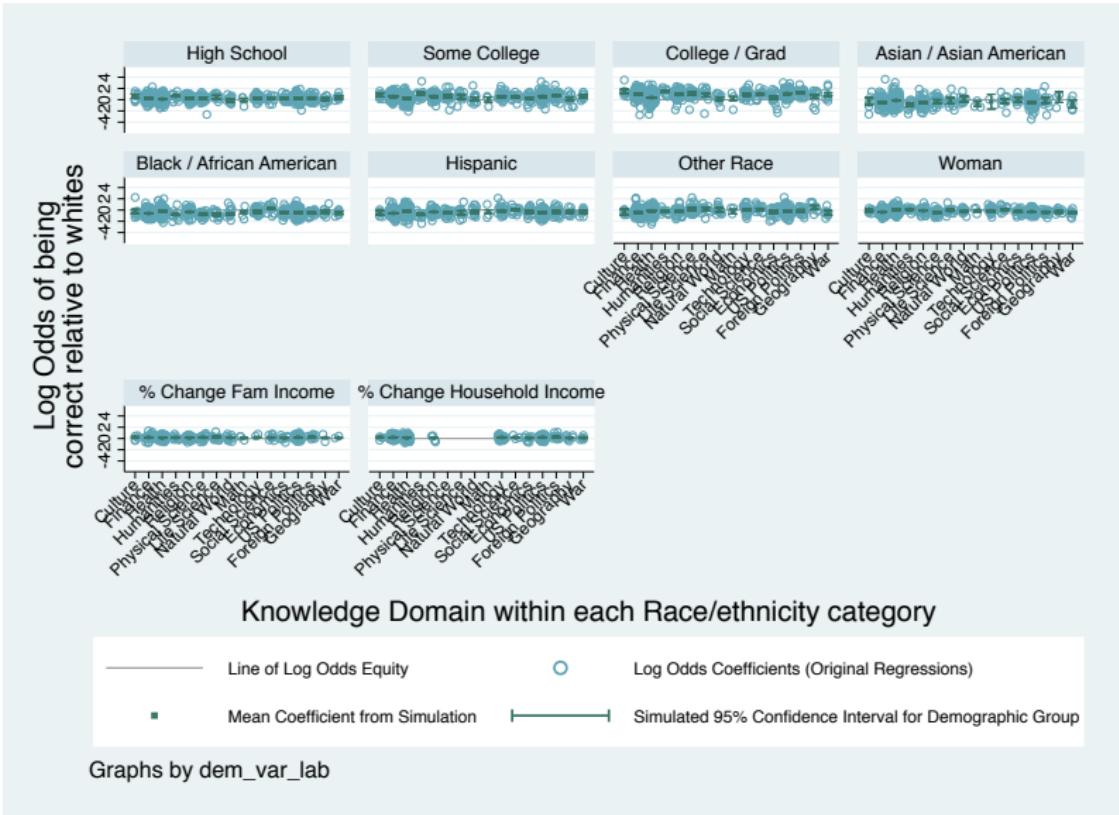


1. shows the self-citation ratio for each year. If men and women cited themselves at equal rates, the ratio shown would be 1.0. A value of 1.5 means that men cite themselves 50 percent more than women in papers published during that year. Shaded intervals represent 95 percent bootstrap confidence limits.
2. In the 1950s, the relative rate¹⁵ of menâŽs self-citations relative to womenâŽs self-citations was 1.23. However, during the 1950s, the bootstrapped 95 percent confidence intervals of the annual ratios overlap with an equality ratio of 1.0, indicating that we cannot reject the null hypothesis of gender equality in self-citation rate during this decade. However, beginning in the 1960s, the ratio of menâŽs to womenâŽs self-citations per authorship remains steadily significantly above 1.0. In the 2000s, the relative rate was 1.71. There is no evidence that that the gender gap is decreasing over time.

(Back)

Correct Knowledge by Race

1. .



Results with Linear Probability Models

1. I have also estimated linear probability models and the results are much the same.

(Back)