

# Information Inequality

## the Class, Gender, and Race of Knowledge Domains

Molly King

Stanford University

April 19, 2018

1. Today I am going to be presenting on 1 of the papers from my dissertation:
2. Information Inequality - the class, gender, and race of knowledge domains

1. Today I am going to be presenting on 1 of the papers from my dissertation:
2. Information Inequality - the class, gender, and race of knowledge domains

# Outline

## 1 Research Question

## 2 Methods

- Data
- Domains
- Model

## 3 Results

- Means across domains
- Proportions correct by group

## 4 Theory

- Factual Reality
- Model of knowledge
- Motivation

1. Just to give you a roadmap of where we're going:
2. I will start by outlining my research question.
3. Next, I'll tell you how I collected my data and the models I used to analyze it.
4. Then, we'll dive into some results from my study around class and gender.
5. Finally, I want to discuss the theoretical frame and motivation I am considering right now and focus your feedback on that.

# Research Question

1. So I wanted to perform a wide scan analysis of knowledge inequality, looking at who has and does not have knowledge in different domains, and how those inequalities might compare to each other.
2. So I wanted to ask, How does the status gap in knowledge vary by domain?
3. And here I mean status in the sense of status characteristics, like gender, class, and race.

*How does the status gap in knowledge vary by domain?*

---

General Social Survey  
Pew Research Center (21)  
Kaiser Family Foundation  
Health Information National Trends Survey (8)  
Integrated Health Interview Series  
Annenberg National Health Communication Survey  
USC's Understanding America Study (3)  
Rand American Life Panel (2)  
National Financial Capability Studies (3)  
21st Century Americanism survey  
Global Views American Public Opinion and Foreign Policy  
Outlook on Life Survey  
State of the First Amendment surveys  
Chicago Survey of Amer. Public Opinion and U.S. Foreign Policy

---

1. My data include 48 nationally representative data sets from between the years 2005 and 2015, each including at least one knowledge question.
2. I collected these data from places like main public opinion survey repositories, the General Social Survey and Pew Research Center.
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. — “True or false: A laser is a concentrated soundwave. The answer is false - lasers are concentrated light waves.”
6. — “Who is the vice president?”

# Domains

---

history  
natural world  
physical science  
biological science  
technology  
math  
culture  
geography  
domestic politics  
foreign politics  
economics  
finance  
health  
religion  
pop culture  
war

---

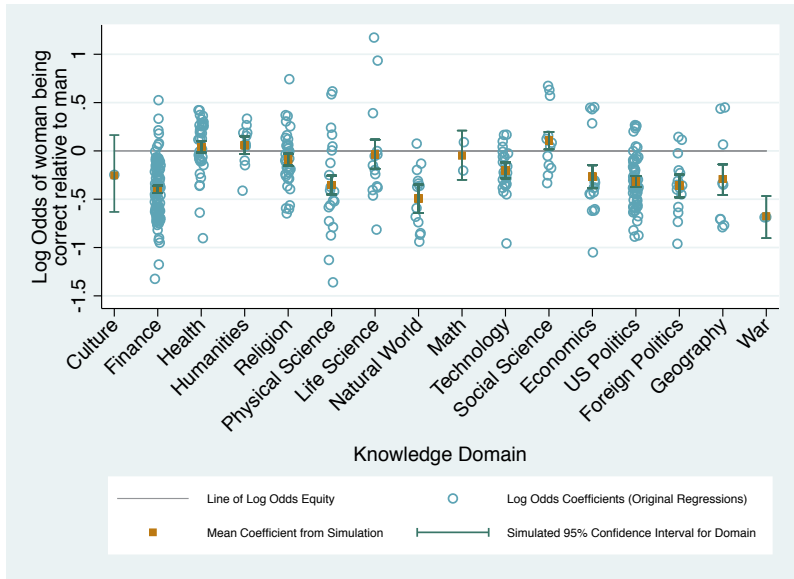
1. For each question, I mark for each individual whether they got the question correct or incorrect.
2. I curated these data and categorized them by domain.
3. This resulted in 16 topical domains.
4. People have studied knowledge gaps in many of these domains before. What is unique to my study is gathering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

# Model

Outcome	Factors
Probability that you get the question correct	Income Gender Race / Ethnicity Education Age + $age^2$

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 that an individual will get the question correct.
3. 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.
4. 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.

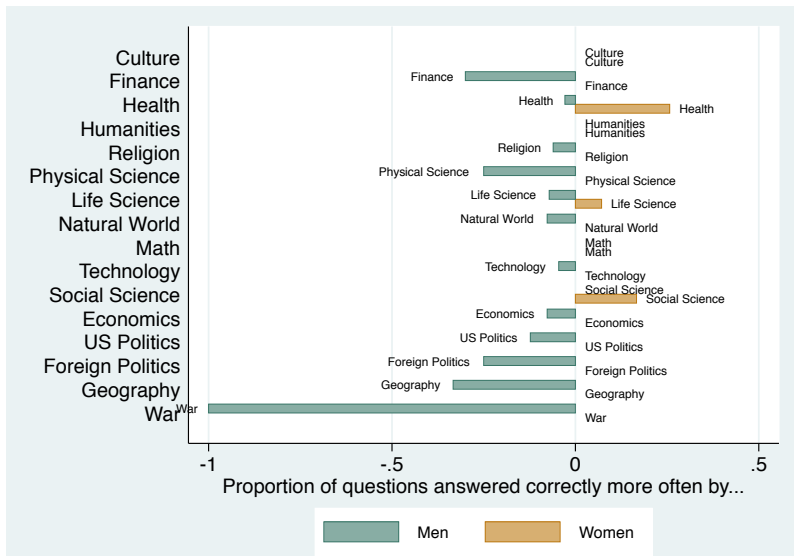
## No mean gender difference in 5/16 domains



1. So here we see that each blue dot represents an individual question. I have taken about 400 questions and divided them into these 16 domains across the x-axis.
2. The yellow square is the mean knowledge level within each domain, and the confidence interval represents 95% simulated certainty around that mean for each 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.
4. Here I tested whether gender had a significant effect on knowledge within each entire domain. For each domain, the simulated mean and confidence intervals allow us to see whether there is a significant difference between the 2 gender groups and the direction of that difference.
5. While there is no average gender difference in 5 of the 16 domains, men have greater average knowledge in 10 domains.
6. Women have greater average knowledge in the domain of social science.
7. One interesting difference I would like to focus your attention on is that men have greater average knowledge than women in the domain of religion. This is a surprising result given that U.S. women are much more likely to report that religion is “very important” in their lives, and women are largely responsible for the religious education of children.

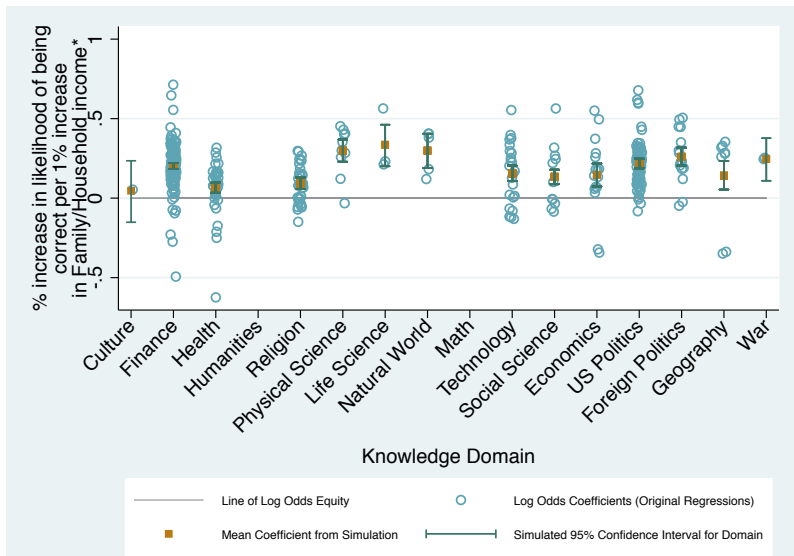


## Men answer greater proportion of questions correctly in 65% of domains



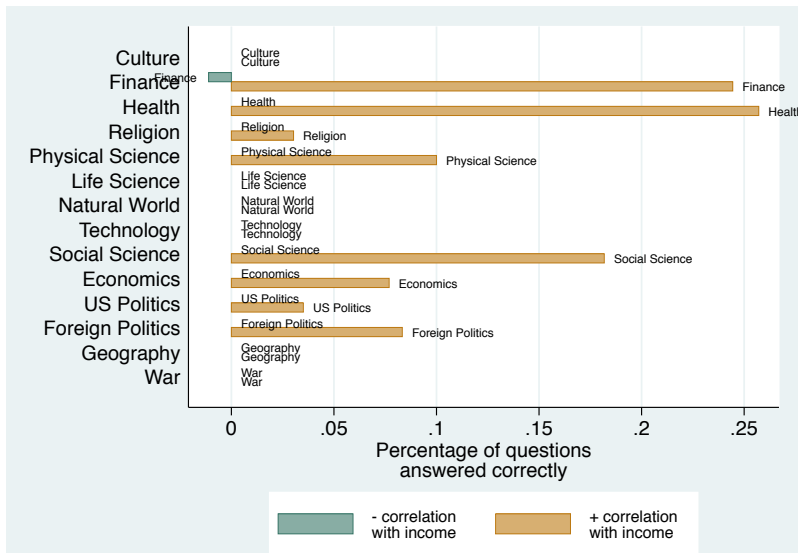
1. For each of these 400 questions, I tested whether the gender difference in knowledge with significant.
2. For health, I found that women answered 30% of the questions correctly more often than men.
3. We see that women answer greater proportion questions correctly in the domains of health and social science.
4. One reason we might consider this to be notable is that men's poor health outcomes are typically explained by behavioral differences. Here we see that differences in health outcomes might also be explained by a relatively large gender disparity in health knowledge.
5. Men are particularly advantaged in the domains of finance, physical science, foreign politics, geography, and war - though the last may be an artifact of sample size.
6. Overall, Men answer a greater proportion of questions correctly in 65% of the domains, while women answer a greater proportion of questions correctly in 12.5% of the domains.

## Income correlates with mean knowledge advantage in 13/14 domains



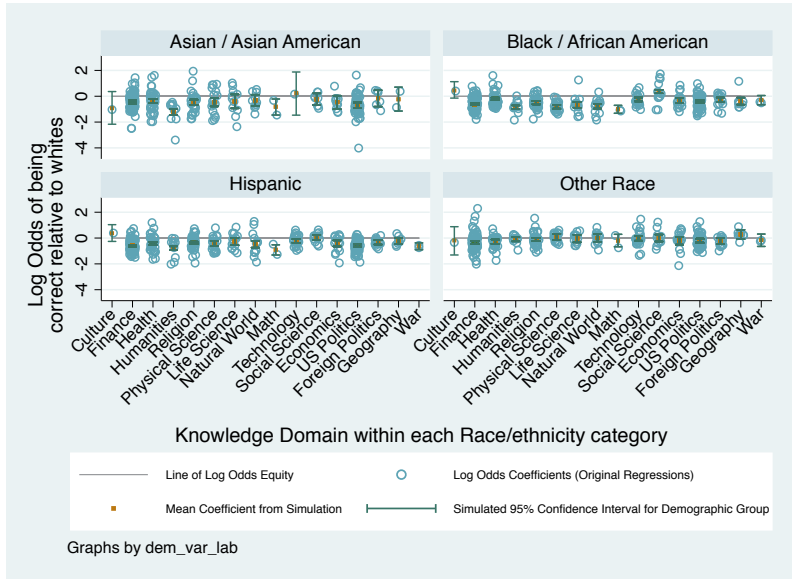
1. Here we see this model using income to predict the likelihood that respondents get the factual knowledge question correct, controlling for all other factors. This is essentially the income elasticity of knowledge.
2. As a reminder, this is after controlling for education - making this a conservative test for the effect of income (since in effect we are controlling twice for class).
3. So in this slide, I tested whether the effect of income on the average knowledge in the whole domain was significant. For each domain, we decide whether there is a significant effect of income, and then see that that effect largely favors those with higher incomes having more knowledge.

## Those with higher incomes answer a greater proportion of questions correctly



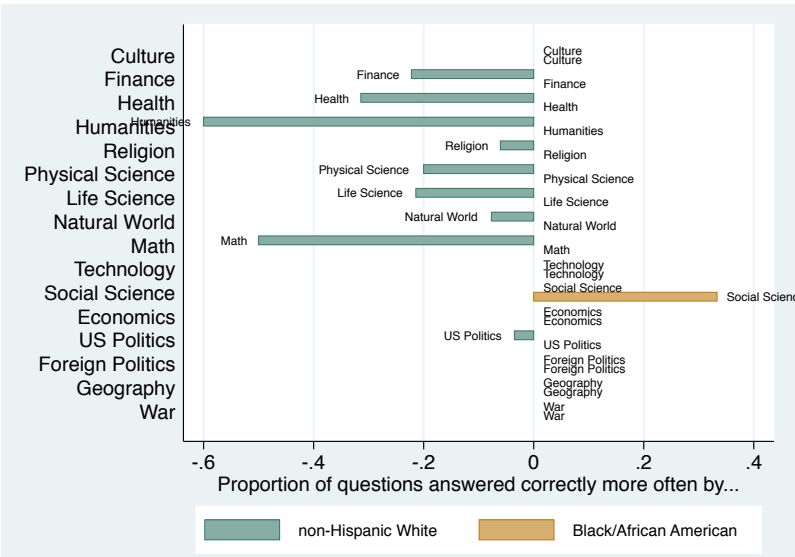
1. Again, I tested to see whether each question had a significant effect of income on knowledge.
2. We can also look at whether each question within a given domain was significantly different.
3. This way of viewing the data shows us that those with higher incomes answer a greater proportion of questions correctly in half of the domains.
4. Those with higher incomes are particularly advantaged in the domains of health, finance, and social science.

1. .



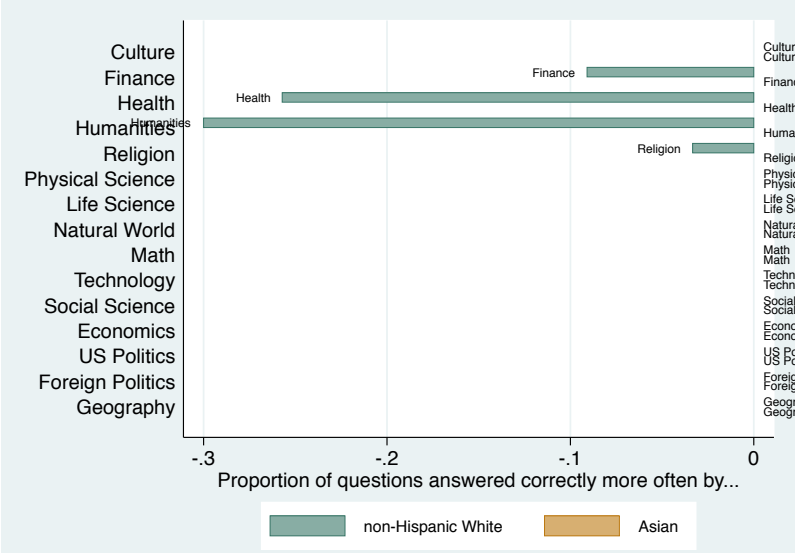
# Proportion of questions answered by non-Hispanic Whites vs Blacks

1. .



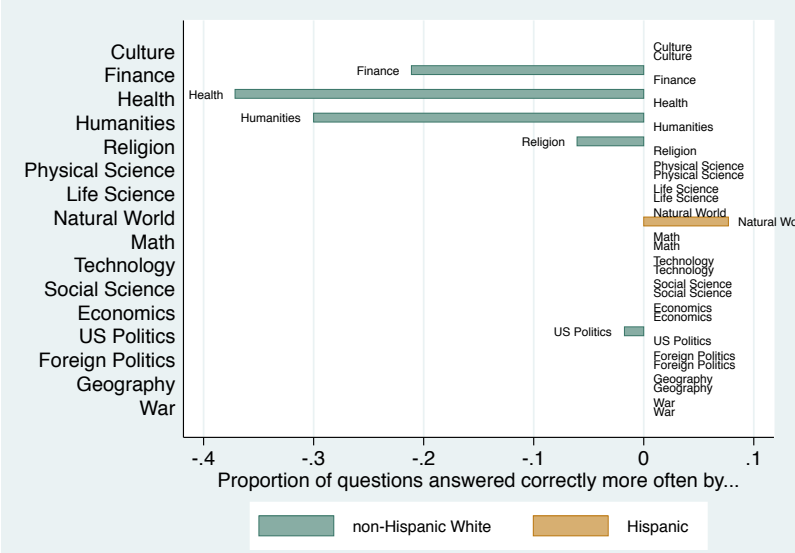
# Proportion of questions answered by non-Hispanic Whites vs Asians

1. .



# Proportion of questions answered by non-Hispanic Whites vs Hispanics

1. .



As we move wholesale into an information society, there is increasing evidence that command over information matters for a great many outcomes, yet rather surprisingly we really have no idea who knows what. To be sure, there is of course a large literature on IQ tests, but these are presumed to measure skill rather than content or knowledge. Insofar as we do study content, it is in a few stylized domains, like math and reading. The purpose, then, of my dissertation is to provide the first comprehensive assessment of which types of groups know what types of content (e.g., health, religion, sports, history, politics, science). Are some groups systematically advantaged across all types of content? Or is there a knowledge division of labor in which some groups specialize in some types of content? Here I provide concrete evidence on the size of these knowledge disparities.



1. At no time in history has understanding how people perceive correct and incorrect knowledge been more important than right now. Information is the currency of the new economy, and as such, it is becoming more important than ever in our modern world. Never before has such a wealth of information been so immediately accessible to so many through the internet and yet the filtering demands so high. Correct and incorrect knowledge dramatically shaped the most recent U.S. election cycle. 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. After all, reality may not appear identical to all of us.

***“Everyday life presents itself as a reality interpreted by men and subjectively meaningful to them as a coherent world”***

- Berger and Luckman, 1967

1. Rather, humans create and maintain an understanding of social reality through social interactions with other people and institutions. Some facts are “brute facts” – those which are physical, biological, and not dependent on language – and others are “social facts” – constructed out of language and institutional agreements.
2. But as Berger and Luckman note, none of these seem anything but real. “Everyday life presents itself as a reality interpreted by men and subjectively meaningful to them as a coherent world”
3. I argue that our different social statuses structure different understandings of this reality. In other words, individuals face the world with different sets of knowledge, shaped by their social contexts. Some of this knowledge may be correct, and some of it may be incorrect, as we see in the results of my analysis. People may be aware that some of their incorrect knowledge represents a gap in their understanding, but in other domains, they may falsely believe what they know is the truth.

- **confabulations** (Miller 1956)

“our heads are full of images, opinions, and information, untagged as to truth value, to which we are inclined to attribute accuracy and plausibility”

- DiMaggio 1997

In psychiatry, these “honest lies” are referred to as confabulations. Once thought to be the realm of only the memory-impaired, confabulations are increasingly understood to be a common human experience. We all have imperfect memories. Many studies have supported the finding that we are only able to remember and transmit about  $7(\pm 2)$  pieces of discrete information in our immediate memories at any one time. This has important implications for learning, particularly in adulthood when repetition may be less frequent. It also means that “our heads are full of images, opinions, and information, untagged as to truth value, to which we are inclined to attribute accuracy and plausibility”. This can lead to false conclusions about our memories of “facts”.

- confabulations (Miller 1956)
- ***schemata* / *heuristics*** (Forster 1999; Gigerenzer et al. 2011, Gigerenzer et al. 2011a, DiMaggio 1997)

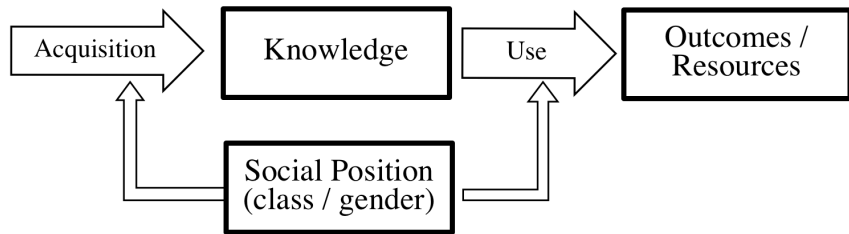
Cognitive science research tells us that information is organized and processed using schemata (or heuristics). These schemata/heuristics group similar concepts together and provide rules for information processing . When using these, people are more likely to perceive information that fits with existing schemata and biases. They also recall information more quickly and accurately when it is embedded in schemata . Perception of reality is shaped by existing schemata and heuristics. Understanding exactly which facts groups already have correct knowledge of can help us better understand what they are likely to perceive correctly in the future.

# Factual Reality

- confabulations (Miller 1956)
- schemata / heuristics (Forster 1999; Gigerenzer et al. 2011, Gigerenzer et al. 2011a, DiMaggio 1997)
- ***perfect information (canonical cite?)***

Furthermore, my study engages with the perfect information assumption of rational choice theory. Under rational choice theory, people have access to perfect information about a market choice game; under a slightly weaker assumption, people have at least some knowledge and are acting on it. These models also assume that the information people have access to is correct, or that it is true knowledge based on accurate perceptions of objective reality. Given the modern information landscape, these are significant assumptions to leave untested. Here I evaluate one important piece of the perfect information assumption: I argue that there are differential demographic patterns in knowledge. Considering race, class, and gender differences in factual information contributes to theory on the social constructions of reality and rational choice.

# Knowledge Acquisition as a Social Process



1. Figure depicts relationships among information, outcomes, and social position: social position influences the likelihood of acquiring information (the first arrow) and of using information to acquire resources or achieve outcomes (the second arrow). In this simplified illustration, I am interested in the first piece of this flow: knowledge as the dependent variable. The information acquisition process is influenced by group social position (as well as other factors). Knowledge, in this conception, is the outcome of interest.
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. 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.

# Why care about information inequality?

- Differences in information capacity itself are, by definition, a dimension of ‘inequality’;

1. First, knowledge is important for its inherent value.
2. Much like many people care about health or education as goods in and of themselves, not just for the ends they can help achieve, we may also think that knowledge is of value for its own sake.
3. Many philosophers consider learning and knowledge part of the “good life.”
4. Here I want to note that I do NOT mean to equate knowledge or information capacity with intelligence. Knowledge of facts is only one aspect of knowledge, after all, and IQ measures something more akin to analytic abilities.

# Why care about information inequality?

- Differences in information capacity itself are, by definition, a dimension of 'inequality';
- Differences in the amount of information people have are influenced by unequal social positions in our society;

1. I argue that a concept I am calling information inequality - or knowledge inequality - is important as both an outcome and cause of social inequality.

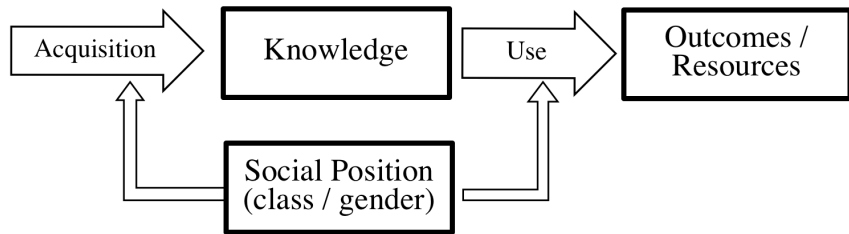


# Why care about information inequality?

- Differences in information capacity itself are, by definition, a dimension of ‘inequality’;
- Differences in the amount of information people have are influenced by unequal social positions in our society; and
- Information is a potential cause of later inequality in outcomes and access to resources.

1. Finally, information discrepancies enable differential access to resources and institutional positions, thereby causing later inequality as well.
2. So, while I argue that knowledge inequality is important from both ends of the causal arrow, in this research I focus on the idea that social status causes knowledge inequality.
3. The field of sociology has long studied the production of knowledge in science inequalities in knowledge careers; and information diffusion and its consequences. Many studies have evaluated information seeking behaviors and needs. But the tendency has either been to study knowledge in one specific domain (e.g., health) or to reduce knowledge across all domains to a single test score – and hence we know shockingly little about the everyday knowledge stock of Americans.

# Knowledge Acquisition as a Social Process



1. The field of sociology has long studied the production of knowledge in science; inequalities in knowledge careers; and information diffusion and its consequences. Many studies have evaluated information seeking behaviors and needs. But the tendency has either been to study knowledge in one specific domain (e.g., health) or to reduce knowledge across all domains to a single test score – and hence we know shockingly little about the everyday knowledge stock of Americans.
2. My contribution brings together these fields at a breadth never before undertaken and with an unusual twist on the dependent variable, focusing on information as the outcome of demographic and social forces.
3. Never before have so information domains been evaluated so comprehensively across demographic groups. This comprehensive study looks at information disparities from a broader scope than any other to date. This is important for understanding information both descriptively.