UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Conceptual Change About People

Permalink
https://escholarship.org/uc/item/7593t5kz

Author
Young, Neil

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Conceptual Change about People

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Neil S. Young

Dissertation Committee:
Professor Bill Tomlinson, Chair
Associate Professor Susanne M. Jaeggi
Associate Professor Andrew Shtulman

2018

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGMENTS

CURRICULUM VITAE

Neil S. Young

2018   Ph. D. In Education,
              University of California, Irvine

2014   M. A. in Education,
              University of California, Irvine

2006   B. A. in Psychology,
              Knox College

FIELD OF STUDY

Conceptual Change

ABSTRACT OF THE DISSERTATION

Conceptual Change about People

By

Neil S. Young

Doctor of Philosophy in Education

University of California, Irvine, 2018

Professor Bill Tomlinson, Chair

Students often fail to develop a sophisticated understanding of scientific topics despite years of formal education designed to help them do so. Through studying how students learn the physical sciences, conceptual change researchers have amassed a large body of evidence that people, rather than being empty vessels who passively accept scientific theories, already possess rich intuitive theories about the world. More recently, researchers have argued that these intuitive theories shape how people understand the social world as well. In this project, we explored these intuitive theories about people and how these intuitive theories differ from how experts in the social sciences think about people. To do this, we used the Delphi method, a method of surveying experts that uses iterative rounds of quantitative judgments and comments. We created seven Delphi panels, each comprised of professors from a specific field of the social sciences (Anthropology, Economics, Geography, History, Political Science, Psychology, and Sociology) and asked them to explain the most consequential ways that studying their field changes how one understands and explains the human world. Three of the panels were highly successful, one was moderately successful, and three were largely unsuccessful due to low participation.

Findings indicate that social scientists can describe several consequential ways they think about the human world differently from those who have not studied the social sciences. Similarities and differences across fields of the social sciences are discussed, as well as differences within the fields as described by panelists. However, these disagreements within and across fields, and especially disagreements over the correct way to interpret terms, make it difficult to generalize across the social sciences as a whole.

CHAPTER 1

Conceptual Change and Consilience

Conceptual Change is an interdisciplinary field, primarily drawing from the fields of developmental psychology and science education, which focuses on understanding how students learn scientific concepts and why they often find it so challenging to develop an expert-like understanding of science. Conceptual change began by looking at how students learned the scientific explanation for physical phenomena, particularly within the context of formal physics education, but has since expanded far beyond that (Vosniadou, 2008).

Conceptual change researchers argue that students failing to understand science is not a simple matter of them not yet having learned the correct concept, but a much more complex and difficult process in which "Students must not only learn unfamiliar concepts absent from everyday discourse, but they must also un-learn concepts acquired earlier in development for making sense of those same phenomena" (Shtulman & Harrington, 2016, p. 119).

A brief summary of conceptual change, though it lacks some of the nuance we will discuss later, would contain the following elements: People possess a wide variety of useful-but-wrong intuitive theories about how the world works and rely on these theories to understand, explain, and predict the world around them. These theories exist largely below the level of conscious awareness – often people are unaware they are even modeling the world and assume they are simply perceiving reality. These intuitive theories are not deterministic (not everybody will possess any particular theory), but nor are they merely idiosyncratic, and many intuitive theories can be found across individuals and even cross-

nationally. Most importantly for educators, while these intuitive theories are a good approximation of our everyday environment, they are nonetheless at odds with the modern scientific understanding of these domains and interfere with learning the correct scientific understanding.

These intuitive theories (and our reliance on them rather than scientific theories) are consequential. According to Bloom and Weisberg (2007), intuitive theories are a major reason people reject scientific theories, particularly when alternative theories are promulgated by trusted sources. For example, people's intuitive theories about animals interfere with understanding evolution. According to Gelman (2000), "people are deep-down essentialists" (p. 57) and assume that organisms have an innate and unchanging essence which determines their later development. While this intuitive theory is useful for predicting many aspects of organisms, it also makes it difficult to appreciate the within-species variation on which natural selection operates (Shtulman & Schulz, 2008; Shtulman & Calabi, 2012).

This rejection of scientific findings because they clash with intuitive theories can lead to needless human suffering; for example, vaccines are rejected by people in part because of their unintuitive nature (Miton & Mercier, 2015).[1] One likely reason humanity

---

[1] Interestingly, they frame this argument not in terms of conceptual change, but in terms of Cultural Attraction Theory (Claidière, Scott-Phillips, & Sperber, 2014), which argues that culture evolves towards cultural attractors non-randomly due to cognitive mechanisms and the broader cultural ecology.

has responded to climate change so sluggishly, despite the serious worldwide threat it poses, is that it clashes with our intuitive theories of how the world works. The actual processes through which the changing composition of the atmosphere affects the energy budget of the Earth can charitably be described as byzantine, requiring significant time and effort to understand. Given that many aspects of this process are far outside of our everyday experiences, people hold many misconceptions about the process through which invisible gases cause the Earth to heat up (Shepardson, Choi, Niyogi, & Charusombat, 2011). These misconceptions make understanding and caring about climate change difficult. Ungar (2000) noted that, compared to climate change, the public had been far more concerned about the ozone layer and attributes this to climate change lacking easily understood and evocative metaphors (it is easy to understand and be concerned about having a shield with a hole in it, but metaphors about having a slightly thicker blanket lack the same emotional impact). More broadly, we would argue that climate change is difficult to solve because it contradicts our intuitive theories about the world and requires us to think about the world (and ourselves) in new and complex ways.

History of Conceptual Change Research

      Kuhn and scientific revolutions.

      Before conceptual change emerged as a field, several theorists and empirical studies laid the intellectual foundations. Researchers, particularly in the field of science education, were inspired by the work of Thomas Kuhn (Vosniadou, 2008). In *The Structure of Scientific Revolutions,* first published in 1962, Kuhn argued that science operates using paradigms, grand collections of concepts and methods that provide a field with its basic understanding

of the world. These paradigms give scientists a common language and framework for studying phenomena, allowing them to work fruitfully together and push the boundaries of scientific understanding, and Kuhn argues that most scientists spend their careers working within a paradigm.

These paradigms, while they are useful, must eventually grapple with anomalies, places where current paradigms cannot explain findings. When these anomalies become too numerous and too problematic to ignore, a science enters a state of crisis, eventually replacing their old paradigm with a new one that resolves these anomalies, which Kuhn refers to as a paradigm shift. Afterwards, a new period of work within this new framework begins, until it too faces too many anomalies, and the cycle repeats (Kuhn, 1996).

Researchers in science education later applied Kuhn's insights to the process of learning for individual students. In this formulation, learners have a theory about some phenomenon which works well enough, but which begins to have anomalies when pressed. Eventually, when these anomalies become too problematic, the student rejects their current theory and adopts a new one (hopefully the scientifically correct theory). The act of learning, rather than simply being a process of knowledge acquisition, represents a fundamental paradigm shift within the minds of students. Unsurprisingly, given the influence of Kuhn's work on conceptual change, many of the questions and challenges to Kuhn's approach have echoes in the field of conceptual change (Vosniadou, 2008).

Piaget and the role children play in their own development.

Jean Piaget is a second crucial figure in the development of conceptual change. Both Piaget's theories, as well as later reactions to his theories, have helped shape the modern

4

fields of cognitive science and developmental psychology, and his understanding of children's development as the step-by-step growth of cognitive structures continues to undergird these fields even today (Flavell, 1996). The field of conceptual change, which inherits much from these fields, is therefore indebted to Piaget as well (for a large selection of Piaget's writings translated into English, see Gruber and Vonèche, 1977).

Piaget, with his long-time collaborator Bärbel Inhelder, took child development more seriously than earlier researchers had, and saw cognitive development in childhood as a critical part of gaining adult cognitive abilities. Rather than being passive accumulators of knowledge, Piaget argued that children played a major and active role in their own development. As children tried to understand the world, they inevitably learned information that contradicted their current beliefs, and Piaget argued that this conflict (and their attempts to resolve it) was what spurred their cognitive development. To Piaget, it was the children themselves, not just adults, who helped children develop sophisticated ways of thinking about the world.

Piaget utilized careful interviews of children's reasoning, designed a multitude of ingenious tasks to probe children's understanding of the world, and made voluminous notes about the development of his own children (though Carey, 1985, notes that his wife Valentine, herself a psychologist, was responsible for much of this work). Based on this work, he argued that children's thinking operated in fundamentally different ways from that of adults. He posited that children progressed through four domain-general stages of cognitive development (the sensorimotor stage, the pre-operational stage, the concrete

operational stage, and the formal operational stage, each with many substages), gaining new cognitive abilities for the first time in each stage.

According to Piaget, young children (that is, before they entered the concrete operational stage) were incapable of appreciating that properties of matter (mass, weight, volume, etc.) could be conserved despite apparent transformations. In one of his most famous tasks, Piaget showed children a container with water inside. He then lifted this container and poured the water into a taller, thinner container (or into several smaller containers). Young children responded as though the water changing appearance meant that there was now more water present. Older children, however, understood that the amount of water was conserved despite appearances (Piaget, 1952).

Before these abilities developed, argued Piaget, children were incapable of understanding the world in an adult-like way. Conservation was one such change, but Piaget argued that there many other such changes as well. For example, he argued that before they reached the concrete operational stage, children were only capable of reasoning egocentrically (that is, they are unable to take the perspective of another person). Likewise, he argued that children were incapable of reasoning counterfactually until they reached the formal operational stage.

In addition to his influence on the study of children's cognitive development, Piaget helped pave the way for the field of conceptual change philosophically as well. White and Gunstone (2008) argue that Piaget's work helped drive a reaction against Behaviorism, the dominant paradigm at the time. Behaviorism focused on behavior and was generally insensitive, if not actively hostile, to understanding mental processes. Piaget's work helped

spark a movement away from experiments that merely contrasted the effects of different educational interventions with little interest in the workings of the mind to a focus on how children understood and thought about the world. Once such a shift was achieved, researchers could begin to see that students were not blank slates but brought their own theories of how the world worked into the classroom.

Despite his immense influence on these fields, several aspects of Piaget's theories have been since been challenged. Some cognitive scientists characterize themselves as neo-Piagetians to reflect an agreement with his theories in general but with modifications of the details (Pascual-Leone & Smith, 1969; de Ribaupierre, 2015). More broadly, many cognitive scientists would say that while the field owes a great debt to Piaget, the field has since advanced (for example, see Carey, 1985). For our purposes, it is important to understand how cognitive science has since moved away from Piaget in some respects, as the field of conceptual change has inherited some of these challenges as well.

In particular, while Piaget's argument that children gained new cognitive abilities over development remains immensely influential, and his tasks used to demonstrate cognitive abilities of children remain well-regarded, cognitive scientists found problems with the idea of sharp boundaries between four distinct stages of development. For example, children could reason in a more sophisticated way in one domain (such as physics) than they could in another (such as biology), which should be impossible if stages are truly domain-general. In addition, children could be trained to perform tasks that should have been theoretically impossible for them given their Piagetian stage (Gelman & Baillargeon, 1983). In general, cognitive scientists now characterize the cognitive

development of children by many smaller domain-specific changes rather than four domain-general stages, while still allowing for the possibility of some domain-general changes (Carey, 1986).

This domain-specific understanding of cognitive changes permeates conceptual change. Vosniadou (2007) writes:

Most theories of learning and development, such as piagetian and vygotskian approaches, information processing, or sociocultural theories are *domain general*. They focus on principles, stages, mechanisms, or strategies that are meant to characterize all aspects of development and learning. In contrast, the conceptual change approach is a domain-specific approach. It examines distinct domains of thought and attempts to describe the processes of learning and development within these domains. (p. 48)

However, this is not to say that conceptual change researchers reject domain-general changes in cognitive development over the lifespan. Rather, conceptual change researchers argue that such domain-general approaches, by themselves, are not enough to capture the process of children learning about the world.

For our purposes, it is also important to note that cognitive scientists have also criticized Piaget's rejection of innate knowledge (though, as we will discuss later, conceptual change researchers have not, as a field, taken a stand on this issue). To Piaget, the stages of cognitive development were necessary and sufficient for adult cognitive competencies. But modern cognitive scientists, thanks in part to research on newborns, argue that people come into the world with certain expectations and predispositions

seemingly built in. In 1975, Piaget famously debated Chomsky (and the philosophers Hilary

Putnam and Stephen Toulmin) on this, and cognitive scientists have since largely, though

not entirely, sided with Chomsky (Piattelli-Palmarini, 1994).

Interestingly, while Piaget is now overwhelmingly remembered for his theories on

child development, he was not originally interested in psychology, calling himself a "genetic

epistemologist" (Boden, 1979), and seeing his research into children's development as a

temporary detour (de Ribaupierre, 2015). Trained initially as a biologist (his thesis was on

mollusks), Piaget saw cognition as crucial to how organisms adapted to their environments

and saw the study of children as a promising route for investigating this process. As

Gopnik, Meltzoff, and Kuhl (1999) note: "Piaget wanted to find a link between Kant and the

mollusks, between epistemology and biology. His great insight was that studying the

development of human children was one way to do this" (p. 15). Thus, Piaget's body of

work presages more recent work that attempts to understand human cognition as the

byproduct of adaptation to the physical and social environment.

Physics education meets cognitive science.

Starting in the 1970s, cognitive scientists and physics educators alike found it

worthwhile to investigate how students understood the principles of physics. Looking back

on this research, McDermott (1984) states: "some investigators have used physics as a

context for examining cognitive processes and approaches to problem-solving. For others,

the primary emphasis has been on conceptual understanding in a particular area of physics

such as mechanics, electricity, heat, or optics" (p. 24).

Researchers found that students used odd and unscientific frameworks for virtually every aspect of physics. For example, students often treated gravity as a property of an object (the object's tendency to go downwards), rather than as a force acting upon an object (Minstrell, 1982). Many students also had incorrect models of the world, such as believing that a ball dropped from an airplane would fall directly down rather than tracing a parabolic arc (McCloskey, 1983). Likewise, students often believed that an object leaving a curved tube would continue to travel in a curved motion until its tendency to curve was exhausted (McCloskey, Caramazza, & Green, 1980). These misconceptions have since been supplemented by many others (for a review of this work, see Brown & Hammer, 2008).

Researchers noted that these wrong answers bore a striking resemblance to earlier, now discarded theories of motion. For example, assigning a tendency to fall downwards to objects was how Aristotle thought about gravity. Other, long since discarded theories of motion showed up in student reasoning as well. Clement (1982) showed engineering students a drawing of a coin tossed into the air and asked them to draw and explain the forces acting on the coin when it was halfway between being tossed and reaching the apex of its flight. To a physicist, the correct answer is that the only force acting on the coin is the force of gravity accelerating the coin downwards. But when students diagrammed this scenario they often added a force on the coin pushing it upwards. When asked to justify their diagram, students argued that this force was imparted by the toss and would continue to push the coin upwards until it ran out, at which point the coin would begin falling downwards. As Clement noted, this is essentially the same reasoning that Galileo employed when he argued that objects possessed an impetus that kept them moving upwards until

10

exhausted. Clement described students as holding a "motion implies a force" preconception. McCloskey (1983) describes such mistakes more broadly as students possessing a "naïve impetus theory." Modern conceptual change researchers would argue that both the students and Galileo were drawing from the same intuitive theory of motion.

Researchers in physics education also noted that students often described their intuitive theories using misapplied scientific terminology. That is, they would use the term momentum, but mean impetus (McCloskey, 1983). This insight demonstrates why a student's unscientific conceptions often go unnoticed by educators and why careful interviews and well-structured tasks designed to elicit misconceptions are often necessary to elucidate them. Simply assuming that high performing students had a good conceptual understanding because they scored well on tests is not enough, as standard physics exams were poor predictors of conceptual understanding in these studies. Students would explain a concept well on an exam and then fail to apply the concept in the real world (McDermott, 1984).

The challenge these findings posed to science education were obvious to researchers from the beginning. Viennot (1979) writes:

It is commonly assumed that we think as we have been taught to think. The purpose of this paper is to show that even in physics where most people would imagine that they know nothing they have not been taught, we all share a common explanatory scheme of 'intuitive physics' which, although we were not taught it at school, represents a common and self-consistent stock of concepts and which, however wrong it may be, resists attempts to change or modify it. This 'intuitive physics'

presents, at the very least, a considerable challenge to teaching. Also, it makes much of our teaching less effective than we usually assume it to be. (p. 205)

It is fascinating to note how much of the current field of conceptual change was presaged by these early researchers. For example, in the same paper, Viennot even notes, without much comment, that intuitive theories are more likely to appear in experts when they are under time pressure, which later researchers have since confirmed. For example, Shtulman and Harrington (2016) demonstrated that adults take longer to affirm scientific statements that conflict with intuitive theories than ones that do not, and this was even true for professors of science.

While this research helped form the field of conceptual change, the fact that this research started in physics education may be a quirk of history (there were studies on children's conceptions in biology predating the field's emergence as well, so it is easy to imagine the field beginning there). Yet physics education has several factors that make it especially fertile ground for such research. Many aspects of physics, such as mechanics, are well-solved problems. This makes it easy to define the correct, expert-like understanding and also makes the contrast between the understanding of experts and novices sharp and easy to delineate. One can also easily create tasks to probe a student's understanding of physics, both for paper-and-pencil tasks and even create demonstrations where students interact with some physical system, which allows researchers to demonstrate that these misconceptions are more than merely theoretical but will actually influence how a student behaves when trying to achieve their goals. Likewise, researchers can probe student

intuitions by asking students to program their physics intuitions into a computerized system that will run them (Nancy & Ogborn, 1994).

In addition, since much of the teaching of physics occurs at the college level, this allowed the use of undergraduate students as research participants. While there are certainly problems with using undergraduate students as research participants (particularly if one relies solely on them), they do have their advantages. For one, they are relatively easy and inexpensive to recruit and manage (especially compared to studying children), they tend to be above average in their ability to be trained to perform tasks and ability to explain their reasoning (even when it is erroneous). And given that college students are usually assumed to be intelligent and to have received a high quality of education, their holding of basic misconceptions is harder to dismiss as idiosyncratic.

A particularly famous example of student misconceptions can be found in the short film *A Private Universe* (Sadler & Schneps, 1989). Researchers filmed several students who were graduating from Harvard and asked them to explain why the Earth had seasons. Many of the students gave the scientifically incorrect but intuitive response that the Earth was closer to the sun in the summer and farther in the winter (in actual fact, the northern hemisphere's summer occurs when the Earth is furthest from the sun), rather than the correct but unintuitive response that the seasons were caused by the tilt of the Earth on its axis. This demonstration of misconceptions surviving the educational process unscathed has since become a classic in the field of science education.

Though the empirical work on student misconceptions first flourished in physics education, there is ample reason to think that many of these findings on how and why

13

people make mistakes when understanding the world are not just restricted to this domain. Later researchers, working across domains, have since shown that many of the findings discussed above are reflections about how people come to understand the world, rather than being specific to physics.

Beyond physics education: Conceptual change as a field of inquiry.

In one of the most influential papers in conceptual change, Posner, Strike, Hewson, and Gertzog (1982) brought together Kuhn, Piaget, and the findings of physics education into a general theory of conceptual change. While they were impressed by the research on misconceptions coming out of physics education at this time, they noted that this research focused almost entirely on the mistakes students made, rather than understanding how students went about replacing their incorrect theories with the correct ones. As they argued: "there has been no well-articulated theory explaining or describing the substantive dimensions of the process by which people's central, organizing concepts change from one set of concepts to another set, incompatible with the first" (p. 211).

According to this theory, four factors led to conceptual change: a dissatisfaction with existing conceptions, an intelligible new conception, that this new conception appeared plausible, and that this theory should open up new areas of inquiry. These four aspects have since been challenged and refined, but this model sparked a tremendous amount of research aimed at understanding and systematizing people's misconceptions.

Generally speaking, the research which followed this model can be split into three distinct streams of research (for an overview of the field, see Vosniadou, 2008). The first focused on finding student misconceptions across scientific topics. The second has debated

when and how conceptual change occurs, building ever more sophisticated models to explain this process. And the third has tried to apply the insights of conceptual change to the classroom, creating educational interventions to help students understand and accept scientific conceptions. For example, Mills, Tomas, and Lewthwaite (2016) identified fifty-two studies intended to cause conceptual change on Earth and space science topics.

Probably the majority of the research that followed this paper is part of this first stream of research, applying the insights of conceptual change to find new domains in which intuitive theories come into conflict with scientifically correct theories. Indeed, finding student misconceptions has become something of a cottage industry, with many thousands of studies performed. You can now find studies on how people naïvely understand the world for a huge number of scientific topic, such as sound (Mazens & Lautrey, 2003), the greenhouse effect (Shepardson, Choi, Niyogi, & Charusombat, 2011), air pressure (Séré, 1992), and so on. In addition, the logic of conceptual change has also been used in applied fields, such as engineering (Streveler, Litzinger, Miller, & Steif, 2008) and the development of clinical knowledge (Kaufman, Keselman, & Patel, 2008).

Though the overwhelming amount of work so far has been on the physical sciences, some research has tried to apply the principles of conceptual change to the social sciences, generally in the sense of helping students think like topic experts when working with the problems characteristic of that field. For example, researchers have applied the insights of conceptual change to studying History (Halldén, 1997; Leinhardt & Ravi, 2008), finding that students find it difficult to think like Historians, and tend to think of historical events as a narrative instead of as interacting structures and actors.

This research has confirmed many of the initial findings from research in physics education across the fields of the physical sciences. Researchers found that students held misconceptions of how the world works, that these misconceptions were similar across people, and that these misconceptions could be found in children and adults alike.

Physics students often combined scientific models and intuitive theories in their explanations. Conceptual change researchers argue that this is a general process, and that people often create so-called synthetic models, which blend scientifically correct conceptions about the world with misconceptions. When Vosniadou and Brewer (1992) looked at children's conceptions of the shape of the Earth, they found that few gave either the intuitive response based on everyday experience (a flat plane) and only a third gave the scientifically correct response (a sphere). Nearly half of the children seemed to possess a model that attempted to harmonize the two sets of information, such as envisioning the world as shaped like a pancake so that it was both flat and round at the same time, or children thought that people lived on flat ground in the middle of a hollow sphere. This shows that people will combine conceptions and misconceptions together into a single model and may not notice that this synthetic model diverges from the scientifically correct one (nor, without asking the right questions, will teachers).

Across domains, many children's theories are similar to medieval theories about the world, for example, children seem to have a preformation theory about the origin of babies (Bernstein & Cowan, 1975), and such theories were popular from the time of Aristotle until the 19th century. Of course, this should not be taken as some sort of cognitive ontogeny recapitulating cognitive phylogeny – it is not that children will inevitably pass through

earlier, discarded theories on their way to the scientifically correct one. Rather, the way the brain processes information and the things it pays attention to makes certain theories more attractive than they would otherwise be. For example, Miton, Claidière, and Mercier (2015) argue that bloodletting (the act of removing "bad blood") is cognitively appealing, explaining its appearance in earlier medical science (as well as cross-culturally) despite its dangers and lack of efficacy. As children draw (in a less sophisticated fashion) from the same intuitive theories of how the world works as scientists of old, it should not surprise us to see such echoes across time and culture.

This research also showed (as suggested by early researchers in physics education) that intuitive theories – though they can be suppressed by formal education – never completely disappear. For example, children often have the misconception that plants are not alive (likely because they do not seem to move). While adults theoretically know better, when they are put under pressure to answer quickly, they become slower and less accurate at judging whether plants are alive. This applies not only to undergraduate students, but to biology professors as well (Goldberg & Thompson-Schill, 2009). Later researchers have since demonstrated that this is indeed a general phenomenon in human cognition, finding that people were slower to assess the validity of statements across several domains (including biology, physics, and mathematics) when intuitive theories conflict with scientific ones compared to when they align with them (Shtulman & Valcarcel, 2012).

Further supporting the idea that intuitive theories never completely vanish, we can also see intuitive theories common in childhood re-emerging in adults during cognitive decline. For example, children have a tendency to theorize about the world in a teleological

17

fashion, that is, they assume that objects exist in order to fulfill some function, such as arguing that mountains exist for climbing or that trees exist to provide shade. Adults sometimes reason in this way when it is inappropriate to do so, but children broadly reason in this way (Kelemen, 1999). Later research demonstrated that patients with Alzheimer's also show a preference for teleological reasoning compared to neurotypical adults (Lombrozo, Kelemen, & Zaitchik, 2007). According to these researchers, this supports the idea that the human preference for teleological reasoning is suppressed but not outgrown.

These findings are similar to modern dual-process accounts of thinking in cognitive science (Evans, 2003). In this view, an evolutionarily ancient and implicit system of reasoning operates alongside an evolutionarily recent system that allows for careful and logical reasoning. This first system contains many subsystems for understanding the world and is shared with many animals (the more closely related they are, the more similar their version of this system is). Such dual systems can explain why adults will behave as though magic is real despite their assertions that it is not, as can be seen when Subbotsky (1997) showed adults a box that appeared to sympathetically slice a postage stamp in half when the experimenter sliced a similar stamp in half. Though adults rejected the possibility that slicing one thing in half could magically cause another object to split, several participants nonetheless refused to allow the experimenter to slice a piece of paper in half when their driver's license was placed into the box.

The existence of such dual systems, and the importance of suppressing the intuitive system when trying to reason rationally, has been shown by Toplak, West, and Stanovich

(2011), who found that a person's tendency to carefully reflect on answers rather than answering intuitively is a more important predictor than cognitive ability for several reasoning tasks. Likewise, this tendency to suppress intuitive reasoning is a particularly important predictor of whether undergraduates will possess a scientifically accurate understanding in several scientific domains (Shtulman & McCallum, 2014).

Updated models of conceptual change and remaining theoretical debates.

The second stream of research inspired by Posner, Strike, Hewson, and Gertzog's (1982) model of conceptual change has attempted to characterize the process by which students replace old concepts with new. Any complete theory of conceptual change would be, by necessity, a complete theory of how humans learn, a proposition which is best characterized as nontrivially complex. Thus, while many theories have been proffered for how conceptual change works, drawing on metaphors of processes that share many similarities with the process of learning, none are complete or without problems. This stream of research has led to several new models of conceptual change, which have tended to increase in complexity over time, and has also led to much debate over how we should best characterize conceptual change.

The first question these models debate focuses on *when* conceptual change occurs. The original model of Posner, Strike, Hewson, and Gertzog's (1982) discussed above was focused on this question. A decade after the initial model was released, Strike and Posner (1992) updated their original model of conceptual change to clarify their position and take into account subsequent criticisms. They now argued that the process of conceptual change is less rational than they initially asserted. That is, while conceptual change drew from

ideas in the philosophy of science, actual students are not as interested in being rational or in understanding what it means to be rational as philosophers of science have been. As a consequence, students may react to incoming evidence that would disconfirm their current theories not by changing their theories, but by ignoring the incoming information, focusing on their performance in class rather than caring about conceptual understanding, or reacting with self-esteem-protecting relativism (i.e. dismissing the validity or importance of Physics).

Another question these models have tackled can be thought of as *who* undergoes conceptual change. Pintrich, Marx, and Boyle (1993) emphasized the importance of looking beyond a "cold" model of conceptual change and considering student motivation as well as the classroom context. Dole and Sinatra (1998) later combined this focus on student motivation, the classic models of conceptual change, and the Elaboration Likelihood Model (Petty & Cacioppo, 1986), a model of persuasion from social psychology, to create the Cognitive Reconstruction of Knowledge Model. This model posits that students undergo more conceptual change the more engaged they are with the material, and this engagement depends on aspects of the student, the incoming message, and the social context. Subsequent research has shown that other aspects of the learner, such as their thinking dispositions and epistemologies, are also important (see Sinatra & Mason, 2008, for a review).

One of the most long-running debates in the field turns on *how* people represent the world when they are trying to explain it. The framework theory approach (Vosniadou, Vamvakoussi, & Skopeliti, 2008), argues that people build relatively coherent theories they

use to explain phenomena in the world, and use these theories consistently. This approach emphasizes that children try to synthesize information about the world to understand it and do so by creating theories about the world that are domain-specific.

In contrast, the knowledge-in-pieces approach (diSessa, 1988; diSessa, 2008) argues that these coherent theories are an artifact. According to this approach, people use phenomenological primitives (basic ideas such as "more effort begets more results" which are abstractions from experience and require no explanation) and assemble them when making explanations. People will then use these primitives across domains, which diSessa (2008) likens to using the same code for different functions when writing a computer program.

We argue that both perspectives contain important insights into human cognition. From our perspective, this debate (aside from being a debate about the word theory) is also a matter of what scale you are looking at. We clearly do use coherent theories at times to explain and predict the world. Yet much of the processing the brain performs goes on below the level we are consciously aware of. If so, then concepts which we can call to mind are likely built, at some level, by ones we cannot. At some point, there must exist some very basic assumptions about the world that the brain makes that we cannot consciously access (lest we be confronted by the problem of infinite regress of underlying concepts).

Our preferred perspective looks at conceptual change from a complex systems approach, in which structure can emerge out of chaos (Brown & Hammer, 2008). This view harmonizes the framework theory approach and the knowledge-in-pieces approach, viewing them as descriptors of different levels of complex and recursive networks. In this

21

view, misconceptions are predictable but not deterministic and it is not the case that everybody will hold a given misconception. Nonetheless, certain ideas will act as powerful attractors for people, and we will find only a small subset of possible misconceptions about the world being held by any significant number of people. Seeing certain ideas as attractors also explains, for instance, why people's misconceptions respond non-linearly to educational interventions. Somewhat ironically, conceptual change researchers have shown that people struggle to understand emergent processes and tend to think in terms of direct processes (Chi, 2005), so while we expect that this perspective (along with Bayesian network modeling work coming out of cognitive science) will eventually eclipse other perspectives, it may not do so anytime soon.

Earlier perspectives on conceptual change, though they do not explicitly embody a systems approach, nonetheless contain some of its insights. For example, Strike and Posner (1992), when outlining their updated model, emphasize that misconceptions may not exist as hard rules that are always found in the mind of learners, but rather as "factors in a conceptual ecology" that people predictably gravitate towards when trying to explain phenomena. This is similar to the principles of dynamic systems, in which outcomes can be non-deterministic yet display predictable patterns. Likewise, Vosniadou and Brewer (1987) describe a mental model as a "dynamic structure which is created on the spot for the purpose of answering questions, solving problems, or dealing with other situations. Mental models are generated from and constrained by underlying conceptual structures" (p. 543).

Another theoretical debate that researchers in conceptual change have tackled can be thought of as focusing on *what* is changing when conceptual change occurs. The classic approach looks much like Kuhn's argument, arguing that a theory is replaced by another, more supported theory. Another view holds that conceptual change is best understood as the restructuring of knowledge, such that a concept is assigned from one ontological category to another. For instance, students coming to reassign electricity from the category of "thing" to the category of "process" (Chi, Slotta, & De Leeuw, 1994). Another theory generalizes this beyond ontological categories, arguing that conceptual change occurs when learners resubsume a domain of knowledge built for one purpose to explain some new domain (Ohlsson, 2009), although Shtulman (2009) argues that this theory explains why conceptual change is hard to complete, but not why it is hard to initiate.

In contrast to these perspectives, some researchers in conceptual change look at conceptual change from a sociocultural perspective. This view focuses on social interactions rather than cognition and emphasizes participation rather than learning (see Kelly & Green, 1998; Hynd, 1998). This perspective has been praised for drawing attention to important factors of students and their environments, but has also been criticized, in its extreme formulations, for denying the possibility of objective knowledge (Chinn, 1998; Vosniadou, Vamvakoussi, & Skopeliti, 2008) as well as minimizing the importance of cognition (Brown & Hammer, 2008).

Along with these theoretical debates, there is much debate over the meaning of specific terms, such as mental models and misconceptions (for working definitions of key terms in the field, see Murphy and Alexander, 2008). For example, many treat any non-

scientific conception as a misconception, whereas Strike and Posner (1992) argued that misconceptions were only those incorrect beliefs that operated as important planks for students' incorrect paradigms of a domain. We believe that until a foundational theory undergirds our understanding of human learning, these debates are unlikely to be resolved.

Despite these disagreements, there is broad agreement within the field that intuitive theories represent a profound challenge for education. Much of the work on how to overcome intuitive theories focuses on engendering scientific modeling, scientific argumentation, and helping students develop more sophisticated epistemological stances (for example, see Nussbaum, Sinatra, & Poliquin, 2008). Many tools have been developed as well, for example, researchers use Evidence-Model diagrams or Model-Evidence Link diagrams (Pluta, Buckland, Chinn, Duncan, & Duschi, 2008; Buckland & Chinn, 2010) to scaffold scientific modeling by making the process explicit. Other researchers have looked at refutational texts, which emphasize the clash between naïve conceptions and scientific conceptions (Sinatra, Kardash, Taasoobshirazi, & Lombardi (2012). Though these methods are often more effective in getting students to change their understanding of the world than traditional pedagogy, lasting conceptual change represents a difficult and often unachieved goal.

Intuitive theories and innate knowledge.

Conceptual change researchers have yet to take a strong stance on where intuitive theories come from. In contrast, cognitive scientists (and evolutionary psychologists) have largely come to the conclusion that people come into the world with expectations about how it works. This is sometimes referred to as "innate knowledge," though this knowledge

24

should be understood as a series of assumptions about the world and preferences for approaching some stimuli and avoiding others based on the way the brain self-assembles, rather than as a list of facts hard-coded into the brain (for an overview of the nature of innate knowledge, see Spelke, 1994).

The idea of innate knowledge helps solve otherwise intractable problems in cognition, such as how animals that are not raised by parents can nonetheless engage in complex adaptive behaviors they could not have learned, or how organisms can reduce the infinitely wide search space of potential explanators of phenomena to a manageable number. In addition, this view explains how babies seem to have expectations about the world from a very young age. For example, two-month-old infants will react with alarm and raised arms to an object growing larger in such a way that it is seemingly headed towards them, but not to an object that would seemingly miss them (Ball & Tronick, 1971). But how do infants know that looming objects are dangerous? Or that one should try to avoid impacts?

To the proponents of this view, this knowledge comes from systems that are innate to human cognition. According to Carey and Spelke (1996):

> Some of the cognitive achievements of children and adults do not result from processes of theory change, we believe, because they do not result from changes of any kind: they depend on core cognitive systems that emerge early in development and remain constant thereafter. (p. 516)

Furthermore, they argue that "there are at least four core conceptual systems encompassing knowledge of objects, agents, number, and space." (p. 517) and that "each

25

knowledge system is organized around a distinct set of basic principles which enable

infants to identify the entities in the domain and constrain reasoning about those entities"

(p. 517). From this view, certain theories in physics are intuitive to many different students

(as well as earlier generations of scientists) because everybody draws on the same innate

knowledge of how the physical world works. According to Spelke (1994) "If initial

conceptions are constant over the spontaneous development of common-sense knowledge,

then they also are universal across human cultures and historical times. They are a body of

knowledge that all humans share, whatever the diversity of our elaborated belief systems"

(p. 441).

Innate knowledge enables and shapes the development of more complex theories.

These initial conceptions about how the world works, in conjunction with experience, allow

people to "bootstrap" the creation of new and more complex concepts (Carey, 2004).

According to Spelke, Katz, Purcell, Ehrlich, and Breinlinger (1993), people share a "core

knowledge" of physics and:

> This knowledge remains central to common sense reasoning throughout
>
> development and constitutes the core of adults' physical conceptions. New beliefs
>
> emerge with development, amplifying and extending human reasoning and
>
> surrounding core physical conceptions with a multitude of further notions. As
>
> physical knowledge grows, however, initial conceptions are not abolished,
>
> transformed, or displaced. (p. 132)

This core of innate knowledge is likely extremely small compared to the intuitive theories people will generate. But from such relatively simple assumptions, we can see the rapid growth of complex intuitive theories about the world.

Compared to cognitive scientists however, it is difficult to say exactly where conceptual change researchers fall on the question of innate knowledge. As far as we know, there is not currently any empirical work on where researchers in the field would say such theories come from. Conceptual change researchers are mostly interested in the educational implications of conceptual change, that is, how to change people's intuitive theories into scientifically accurate ones.

However, we believe that this recent understanding of innate knowledge provides a better understanding of the origins of intuitive theories than is usually given in the conceptual change literature. We believe that innate knowledge explains why intuitive theories are so intuitive in the first place, as well as explaining many odd features of intuitive theories, such as their consistency across people and their longevity over the lifespan, though we note that distinguishing between innate knowledge and intuitive theories is difficult and contentious work (present at birth implies innate, but something, like adult teeth, can be innate without being present at birth). We also note that it is not necessarily the case that every intuitive theory has a core knowledge about a domain at its base (we probably do not have innate knowledge specifically about the shape of the Earth for example, because we do not need it to survive). But if all humans share the same innate knowledge, we will tend to see the same intuitive theories again and again.

Just as researchers have argued that humans possess an innate knowledge of physics, so too have researchers argued that people possess an innate knowledge of people, the roots of which can be seen in infancy. Research on infants has demonstrated that even newborns have expectations about how the world works and preferences for certain stimuli (such as faces) before they could have learned about such things (Gopnik, Meltzoff, & Kuhl, 1999). Newborns do not know that a face is a face, but they somehow know that faces are important and worth paying special attention to. Impressively, children seem to interpret the actions of others in terms of goals by only 12 months (Gergely, Nádasdy, Csibra, & Bíró, 1995). This also explains why "seeing" the actions of others as goal-directed does not take any effort, and indeed, it is difficult to *not* see agency. Heider and Simmel (1944), in a famous example of this, showed participants a short film of moving geometrical shapes, which participants interpreted in terms of the desires of the shapes, and even assigned personality traits to the shapes.

Researchers argue that in addition to possessing a naïve physics and naïve biology, people possess a naïve psychology (often called folk psychology) which leads them to explain the actions of others in terms of beliefs and wants (Wellman & Gelman, 1992). This naïve psychology is driven by people's "theory of mind" (Premack & Woodruff, 1978) that allows them to understand others in terms of their beliefs and wants. However, researchers have disagreed on how the theory of mind works in the brain. Evolutionary psychologists argue that humans have evolved "core cognitive architecture" in several specific domains to help us make sense of the world, including specific modules for understanding the physical world and the social world (Leslie, 1994, Cosmides & Tooby, 1994). In contrast,

Gopnik and Wellman (1994) accept innate knowledge about people but reject a specific theory of mind module, instead arguing for the "theory theory" perspective, in which "our everyday conception of the mind is an implicit naïve theory; children's early conceptions of the mind are also implicit theories, and changes in those conceptions are theory changes" (p. 257).

The idea of folk psychology ignited a contentious debate in the 1980s and 90s, particularly among Philosophers, with much of this debate centering around the existence and consequences of such systems. According to Stich and Nichols (2003), philosophers were interested in the idea of folk psychology because it spoke to the mind-body problem in philosophy and also because of the possibility that folk psychology is radically wrong as a way to understand the world. Indeed, some philosophers argued that the existence of folk psychology means that our perceptions of others having wants and desires is literally false and thus that such concepts will have no place in a sufficiently advanced neuroscientific framework (Churchland, 1981). This perspective, usually called eliminative materialism or eliminativism, argues from an analogy to intuitive physics: just as students intuitively reason about physics by referring to entities that are non-existent from the perspective of modern physics, they reason about human actions by reference to non-existent entities (i.e. beliefs are logically equivalent to phlogiston).

Researchers have since gathered evidence for the existence of a folk psychology system from multiple sources. For example, Call and Tomasello (2008) review research on whether Chimpanzees possess a theory of mind and conclude that Chimpanzees also understand the actions of others in terms of goals and intentions, though they lack an

appreciation that others can believe false things. Other researchers have shown the relative independence of the naïve physics and naïve psychological systems by looking at deficits in reasoning characteristic of specific syndromes. For example, Baron-Cohen, Wheelwright, Spong, Scahill, and Lawson (2001) showed that children with Asperger syndrome, compared to neurotypical children, had impaired performance on a folk psychology test but superior performance on a folk physics test. Contrariwise, Kamps, Julian, Battaglia, Landau, Kanwisher, and Dilks (2017) showed that adults with Williams syndrome performed worse than mental-age matched controls on a folk physics task, but not on a folk psychology task.

This understanding of folk psychology has also been taken up by researchers outside of psychology and cognitive science. For example, Rubin (2003) argues that there is a "folk economics," largely based on the same arguments as evolutionary psychologists, and that this folk economics drives how people understand the world. Leiser and Krill (2016) expand on this, emphasizing that people think about the world teleologically and in terms of intentions and also rely on metaphors that obscure accurate economic understandings. This idea has also been applied outside of the social sciences as well. For example, Plantinga (2011) argues that filmmakers and audiences alike rely on folk psychology to make and make sense of movies.

The folk psychology and conceptual change perspectives overlap, though imperfectly. Vosniadou, Vamvakoussi, and Skopeliti (2008) argue for four domain-specific framework theories for physics, psychology, mathematics, and language, and describes the change in children's concept of mind as similar to other forms of conceptual change.

However, they also argue that "these framework theories are constructed early on and are based on children's interpretations of their common everyday experiences in the context of lay culture" (p. 15), which implies an emphasis on developmental processes rather than innate knowledge.

Likewise, conceptual change and folk psychology overlap with other fields as well, although in complicated ways. Psychologists have long been interested in misconceptions about psychology (for example, see Nixon, 1925), though many of the misconceptions in this literature are factoids (people only use 10% of their brain, hypnosis works, etc.) rather than disagreements with key psychological principles. Furthermore, "although much work has focused on the prevalence of psychological misconceptions, less attention has been devoted to understanding how people come to acquire or develop them" (Hughes, Lyddy & Lambe, 2013). When researchers have tried to understand where such misconceptions come from, they often disagree. For example, Bensley and Lilienfeld (2017) emphasize a failure to think critically as the key cause. However, Amsel, Baird, and Ashley (2011), argue that misconceptions are caused in part by folk psychology and that overcoming these misconceptions is a process of conceptual change.

The misconception literature tends to focus on what students get wrong. In contrast, the idea of "threshold concepts" (Meyer & Land, 2003; Land, Cousin, Meyer & Davies, 2005) can be conceptualized as what students need to get right:

A threshold concept can be considered as akin to a portal, opening up a new and previously inaccessible way of thinking about something. It represents a transformed way of understanding, or interpreting, or viewing something without

31

which the learner cannot progress. As a consequence of comprehending a threshold

concept there may thus be a transformed internal view of subject matter, subject

landscape, or even world view. (Meyer & Land, 2003, p. 1)

Researchers have since used this perspective to investigate threshold concepts in

Economics (Davies & Mangan, 2007) as well as using the Delphi Method (described later)

to understand threshold concepts for informational literacy (Townsend, Hofer, Hanick &

Brunetti, 2016). However, while this perspective lists knowing being counterintuitive as a

reason that students struggle to learn accurate conceptions, it does not explain what makes

knowledge counterintuitive. This perspective did not originally reference conceptual

change, however, Davies and Mangan (2007) have since described threshold concepts as

being "within the 'conceptual change' rather than the 'enrichment' tradition" (p. 713).

More recently, researchers have argued for a much more important position for

intuitive theories in how people understand the world. In this view, intuitive theories are

the key way that people understand and make sense of the world for both the physical and

social world. According to Gerstenberg and Tenenbaum (2017): "understanding common-

sense reasoning requires at minimum two key insights: (i) human knowledge is organized

in terms of intuitive theories, and (ii) much of human cognition can be understood in terms

of causal inferences operating over these intuitive theories."

Previous work on intuitive theories also struggled to explain how these intuitive

theories worked or were instantiated in the brain. However, Gopnik and Wellman (2012)

have since updated the "theory theory" perspective, arguing that Causal Bayesian Nets can

generally explain learning and the development of intuitive theories, as these nets explain

how people make inferences about causation as well as update their models with new information. Researchers have also emphasized the importance of these intuitive theories in how people understand the social world. For example, Rhodes (2013) argued that intuitive theories underly the development of social categorization, particularly the theories that social categories are natural kinds (Hirschfield, 1996) and that social categories mark social obligations. Researchers have also argued that the process of moral development reflects a change in children's intuitive theories about the world rather than a progression through distinct stages (Rhodes & Wellman, 2017).

Relevance to the current project

If studying a social science causes conceptual change in the same way that studying a physical science does, then the process of becoming formally educated in a field should cause one to replace their initial conceptions (intuitive theories, folk theories, misconceptions, and so on) with scientifically correct conceptions. But if this is true for the social sciences, then what actually changes? That is, how does one differ in how they understand people after studying a social science compared to how they understood people before they began?

In this project, we sought to understand the ways that formally studying a social science changes how one understands and explains the human world. As in undergoing conceptual change in the physical sciences, this should be more than simply learning new terminology, but rather a profound reshaping of how one models and explains the world. Though studying a social science likely changes one's understanding of people in innumerable ways, here, we are interested in the ways that are the most consequential.

However, before we can answer this, two concerns present themselves. For one thing, can it be said that there is a scientific understanding of people that truly differs from the intuitive understanding of them, or is the difference merely one of complexity and terminology? Second, even if we can establish that experts do consistently differ in how they understand people from how non-experts do, how do we know that the experts are correct? Philosophically, it is probably impossible to demonstrate beyond any doubt that any perspective is the correct one. However, there are principles we can use to give greater weight to the likelihood of a perspective being correct, and one such principle is the principle of consilience.

Consilience

Consilience as a Logical Method

Consilience, broadly speaking, is the principle that we can put more trust in a conclusion when multiple, independent lines of evidence converge to the same answer. The more lines of evidence that converge, the better. The more independent the lines of evidence, the better still.

The term consilience was coined by Rev. William Whewell, an English polymath who, among other things, invented the word scientist. In his book on the philosophy of science (Whewell, 1840), he laid out a series of aphorisms on how science works. These aphorisms contain both his own ideas and ideas which predate him, and many of these aphorisms are still held, implicitly or explicitly, by scientists today. For example, these aphorisms emphasize the importance of rigorous and clear concepts, precision in measurement, and the testing of hypotheses.

In his fourteenth aphorism concerning science, Whewell introduces the idea of consilience: "The Consilience of Inductions takes place when an Induction, obtained from one class of facts, coincides with an induction, obtained from another class. This consilience is a test of the truth of the Theory in which it occurs" (Whewell, 1840, p. 36, capitalization as in original).

Though Whewell refers to this as a "consilience of induction" his use of the term induction is archaic and requires some explanation. According to Laudan (1971), Whewell's use of "induction" is closer to how we use the term "abduction" today. Both induction and abduction are forms of logical reasoning, but there are important differences between them. Induction refers to taking some set of observations and asserting the generalizability of this observation to as-of-yet unobserved cases. For instance, given the observation that the sun has risen every day in the past, we might conclude that the sun will rise tomorrow. Abduction, in contrast, is when we try to explain observations by appealing to the most likely theory which explains them. For example, given the observation that a fish has disappeared from our fish tank while another fish in the tank has grown noticeably fatter, we might conclude that the larger fish has eaten the smaller one. This type of reasoning, most famously employed by Sherlock Holmes, is commonly used in both everyday life as well as by scientists when they argue that some phenomenon is best explained by a particular theory. Thus, Whewell's aphorism about consilience, rephrased to reflect this, states that when a set of facts leads us to privilege a particular theory, and another set of facts leads us to privilege that same theory again, this makes it more reasonable to think that the theory is true.

Consilience works because it helps us ignore noise when trying to discern a signal. In any judgment, a person may deviate from the correct answer due to error. However, it is unlikely that one person's errors are exactly the same as another person's errors, or that the errors of those two people would be the same as those of a third. For example, two eyewitnesses may both incorrectly identify an innocent person as a criminal due to the vicissitudes of memory, but as the number of eyewitnesses increase, the chance of them all being mistaken in the same way diminishes. Even if many eyewitnesses were to incorrectly identify innocent people, because individuals make different errors, we would expect them to incorrectly identify *different* innocent people. Thus, if many witnesses instead identified the same person, this should make us more confident that their identification reflects reality, as whatever errors they individually made were not enough to cause their answers to diverge.

This is not the same as saying that agreement is the same as proof (even with multiple witnesses, incorrect identifications are inevitable), and this is particularly true when the judgments are not independent. If the eye-witnesses, for example, discussed their recollections with each other, then their judgments would converge on the same person. Yet this would not be consilience but rather more like getting the same opinion multiple times. This is why we are more impressed when multiple laboratories replicate a study compared to when a single laboratory replicates the same study repeatedly. But all else being equal, agreement among many sources of evidence should lead us to give greater weight to a theory than if a theory is supported by only one source of evidence.

Though it is rarely called by name, consilience is a deep part of the entire scientific enterprise, and a crucial part of creating a scientific consensus despite scientists necessarily possessing imperfect methods. Usefully, consilience allows us to test theories to an ever-higher standard of proof, even though no method of testing can ever be perfectly free of error. By aggregating noisy methods, we can nonetheless discern a strong signal, and consilience has recently been used for this purpose in some of the largest and most important debates within science. For example, McInerney, O'Connell, and Pisani (2014) argue that Eukaryotes (organisms who have cells with a nucleus inside of a membrane, which includes all plants, animals, and fungi) are derived from an ancient fusion of bacteria and archaea (bacteria-like organisms which are metabolically distinct from bacteria) based on a consilience of evidence from "molecular phylogenetics, palaeontology, bioenergetics, and modern cell biology and biochemistry – each of which has contributed important and surprisingly congruent insights relating to this argument" (p. 1). Likewise, Oreskes (2007) argues that there is consilience of evidence for climate change and that "Instrumental records, tree rings, ice cores, borehole data, and coral reefs all point to the same conclusion: things are getting warmer overall" (p. 90).

Consilience Among the Sciences

More recently, biologist E. O. Wilson has argued for an expansion of the term consilience. While he considers consilience, in the sense discussed earlier, as one of the most important aspects of science and a crucial aspect that distinguishes true science from pseudoscience, he also argues that consilience is not merely a method of science but something that the physical sciences *have already achieved*. In his formulation, consilience

37

refers not only to a scientific practice, but also a description of the trajectory of science itself towards an ever-greater unity of knowledge across fields (Wilson, 1998).

Wilson argues that the physical sciences have converged and will continue to do so in the future. That is, while chemists and physicists look at the world using different methods and at different scales, they have come to share a fundamental understanding of the world and how it functions. A chemist and a physicist may emphasize different properties of atoms in their projects and model them differently depending on their needs, but one would be hard-pressed to find any aspect of atoms on which they truly disagree. This agreement extends throughout the physical sciences as well as across scales. For example, a malacologist (one who, like Piaget, studies mollusks), would wholeheartedly accept the knowledge of physics and chemistry as true and relevant even if they were primarily interested in higher-order aspects of mollusks such as behavior. This shared understanding allows scientists from different disciplines to work together fruitfully at the intersection of their respective fields, and consequently, many interdisciplinary fields have formed at the frontiers of the physical sciences.

Wilson argues that, unlike the physical sciences, the social sciences have so far failed to reach consilience. He argues that this is for many reasons, including dissatisfaction with the enlightenment ethos, the rise of postmodernism, and because "social scientists by and large spurn the idea of the hierarchical ordering of knowledge that unites and drives the natural science" (p. 182). Wilson compares the medical sciences to the social sciences, arguing that they are both asked to solve complex problems. But while Wilson argues that the medical sciences continue to make dramatic progress, he argues:

There is also progress in the social sciences, but it is much slower, and not at all

animated by the same information flow and optimistic spirit. Cooperation is sluggish

at best; even genuine discoveries are often obscured by bitter ideological disputes.

For the most part, anthropologists, economists, sociologists, and political scientists

fail to understand and encourage one another. (p. 182)

He blames this lack of progress on the refusal of many social scientists to accept that

humans have evolved from other organisms and the consequences thereof, arguing that

this drives a major and bitter rift within the social sciences. He argues that eventually, the

social sciences will be consilient with each other as well as with the physical sciences, but

until then the social sciences "will continue to split within each of its disciplines, a process

already rancorously begun, with one part folding into or becoming continuous with

biology, the other fusing with the humanities" (p. 12).

Relevance to the Current Project

This project is interested in both senses of consilience.

Firstly, we are interested in whether there are any ways of understanding and

predicting people that span the social sciences. If so, we should feel more confident that

these ways represent a true scientific understanding of people, even though no social

science alone could prove this.

Secondly, the degree to which the fields of the social sciences are consilient (in the

sense of reaching a unity of knowledge in how they understand people) is an empirical

question. Though we do not believe that the physical sciences have reached complete

consilience, nor that the social sciences utterly lack it, it does seem likely that the hard sciences are closer than the social sciences are at this point in time to reaching consilience. By comparing how fields understand *people*, we can beyond differences in focus and terminology to understand where they are approaching consilience and where they are not.

However, these are difficult questions to investigate. As Wilson argues "The social sciences are hypercomplex. They are inherently far more difficult than physics and chemistry, and as a result, they, not physics and chemistry, should be called the hard sciences" (p. 183). Unfortunately, this makes it very difficult to separate what any field knows from the convoluted history, theoretical approaches, and jargon that surrounds it.

Fortunately, we are not the first to want to summarize a complex and disputatious body of knowledge. As we will show, methods for distilling such knowledge into a useful, though imperfect form, already exist. If we accept that experts in the social sciences know things, then all we have to do is ask the experts.

CHAPTER 2

The Delphi Method

The Delphi method is structured communication method that aggregates the judgments of experts on any topic of interest. Developed at RAND in the 1950s as "Project DELPHI", the Delphi method was initially used to study the vulnerability of industrial infrastructure to atomic warfare. Due to the sensitive nature of this research, the Delphi method was not described to the public until the 1960's (Dalkey & Helmer, 1963). After this, the Delphi method was quickly applied to forecasting in other areas where expertise exists but is difficult to consolidate, such as predicting future political and technological trends (Gordon & Helmer, 1964).

The developers of the Delphi method wanted a way to consolidate knowledge in the fields for which it was difficult to do so. They were skeptical of what they saw as a long-standing division between the "exact sciences" which are logical and mathematical and the "inexact sciences" which are vague and intuitive. Instead, they argued that the difference between the two was a matter of degree rather than kind, and that what really made something be scientific was objectivity. As they described it, they sought to describe the "quasi-laws" of the social sciences and argued that expert opinion could be an objective way to do this (Helmer & Rescher, 1959). The creators of the Delphi method were aware of, and likely inspired by, earlier work on expert consensus by Kaplan, Skogstad, and Girshick (1950), particularly their conclusion that "predictions made by groups of people are more likely to be right than predictions made by the same individuals working alone" (p. 93).

41

Since its creation, the Delphi method has been used in numerous domains where it would otherwise be difficult to consolidate expert knowledge. For example, Delphis have been deployed to design effective policy for protecting wetlands (Bardecki, 1984), predicting future applications of computer-mediated communications at colleges and universities (Holden & Wedman, 1993), predicting the future of electronic journals and the effects of such journals on libraries (Keller, 2001), studying the effects of five different forest management regimes on Nordic boreal forest biodiversity and habitat preservation (Filyushkina, Strange, Löf, Ezebilo, & Boman, 2018). and calculating how much people in various countries would be willing to pay in order to protect the Amazon rainforest (Strand, Carson, Navrud, Ortiz-Bobea, & Vincent, 2017).

Delphis also seem particularly popular in the medical field, with Delphis used to study issues such as the factors that help or hinder the ability of nurses to perform physical exams (McElhinney, 2010) and the process through which pathologists diagnose complex melanocytic lesions and why they sometimes disagree (Carney et al., 2016).

More recently, several papers have attempted to expand the use of the Delphi method across several fields, arguing that it is a useful method for problems that characterize those fields, reviewing uses of the Delphi method in that field so far, and recommending best practices for future research. Such papers have been written for mental health research (Jorm, 2015), educational research (Green, 2014), and for the design of clinical trials (Sinha, Smyth, & Williamson, 2011).

While most Delphis are intended to consolidate knowledge about a topic, several variants of the Delphi have also been developed. For example, the policy Delphi (Turoff,

1970), brings together views of stakeholders to understand the range of outcomes preferred by stakeholders without seeking consensus. Likewise, the decision Delphi (Rauch, 1979) uses the decision makers themselves as the participants, in order to structure the decision-making process and create the future rather than predicting it. Researchers continue to innovate new variants of the Delphi Method. For example, the Morphological Delphi Method (Mozuni & Jonas, 2018) combines the Delphi Method and Morphological Analysis (a method for studying irreducibly complex problems) in order to predict outcomes for complex systems and their evolutionary trajectory.

Delphis have also found wide use in educational research. For example, there have been Delphis to understand engineering topics that should be included in high school curricula in order to prepare students to study engineering in college (Childress & Rhodes, 2008), Delphis to understand teachers' pedagogical content knowledge in mathematics (Manizade & Mason, 2011), and Delphis to identify important as well as neglected research areas in distance education (Zawacki-Richer, 2009). In one particularly impressive example of a Delphi, researchers were able to get several hundred faculty members, across several geographically distributed campuses, to agree on a new curriculum (Dailey & Holmberg, 1990).

Of particular relevance to the current project, researchers studying misconceptions have used the Delphi method in order to better understand common student misconceptions within a scientific domain as well as what the scientifically correct concepts are that students should learn. For example, the Delphi method has been applied to understand concepts that students struggle to master in thermodynamics (Streveler,

Olds, Miller, & Nelson, 2003), the scientific concepts that students must know in order to understand climate change (Jarrett, Takacs, & Ferry, 2011; Jarrett, Ferry, & Takacs, 2012), and difficult topics in introductory computing courses (Goldman et al., 2008).

The Delphi method has been used as the basis for numerous dissertations, having seen use in over a thousand (Landeta, 2006), and Skulmoski, Hartman, and Krahn (2007) argue that the Delphi method is an attractive method for dissertations due to its methodological flexibility and applicability to a wide range of topics. The Delphi method has become popular enough as a method for dissertations that there has even been a Delphi on the use of the Delphi method in dissertations (Davidson, 2013).

How the Delphi Method Works

Despite its unusual origins and portentous name, the Delphi method is straightforward in terms of the actual methodology. In brief, the Delphi method asks experts to make a judgment on some topic of interest. Following this, the experts are provided with feedback from other experts and are then asked to make their judgments again. This process continues until stability (though not necessarily consensus) is reached.

This process allows expert opinions to be aggregated while avoiding the pitfalls that would naturally occur if one was to simply put experts in a room together. As Gordon (2009) describes it:

> In a sense, the Delphi method is a controlled debate. The reasons for extreme opinions are made explicit, fed back coolly and without anger or rancor. More often than not, groups of experts move toward consensus; but even when this does not occur, the reasons for disparate positions become crystal clear. Planners reviewing

this material can make judgments based on these reasons and their own knowledge and goals. (p. 4)

The Delphi method straddles both quantitative and qualitative methods. Unusually, it does so within a single method rather than in the mixed-method sense of separate quantitative and qualitative methods supporting each other. Qualitative and quantitative research perspectives are often thought of as contradictory, but according to Bryman (1984) much of this debate about apparently technical issues is really a debate about philosophical issues (i.e. whether objective knowledge exists and can be "collected" by scientists.) As Bryman points out, method and epistemology are not intrinsically linked and:

> some survey researchers seem to exhibit a commitment to the epistemology of qualitative research, in particular its emphasis upon seeing through the respondents' eyes yet use the technical paraphernalia of the survey. In contrast, qualitative researchers frequently make quasi-quantitative assertions, such as 'many', 'frequently', or 'some of the time.' (p. 88)

In this sense, while the Delphi uses both quantitative and qualitative techniques, it shares more in common philosophically with quantitative methods, in that it assumes that knowledge exists and can be gathered fruitfully.

Martino (1993) states that there are three features that distinguish a Delphi from face-to-face interactions: "anonymity, iteration with controlled feedback, and statistical response" (p. 17).[2]

Anonymity of the panelists was important as it would reduce the influences of social factors that inevitably occur when bringing people together into a room. As Dalkey and Helmer (1963) explain:

> The method employed in the experiment appears to be more conducive to independent thought on the part of the experts and to aid them in the gradual formation of a considered opinion. Direct confrontation, on the other hand, all too often induces the hasty formulation of preconceived notions, an inclination to close one's mind to novel ideas, a tendency to defend a stand once taken or, alternatively and sometimes alternately, a predisposition to be swayed by persuasively stated opinions of others. (p. 459)

In addition, the unequal social positions of experts may cause some experts to avoid stating their true opinions, to avoid offense or career-damaging conflicts.

---

[2] Not all researchers divide the necessary features of a Delphi in this way. For example, Rowe and Wright (1999) state that the four necessary features of a Delphi are: "anonymity, iteration, controlled feedback, and the statistical aggregation of group response" (p. 354). However, the differences between how various researchers conceive of the Delphi seem minimal compared to the similarities.

Iteration with controlled feedback is important because it allows panelists to take their colleagues views into account and reminds them of any aspects they may not have originally considered. Using controlled feedback, as opposed to simply giving all feedback directly back to panelists, is important because it "permits the group to concentrate on its original objectives, without being distracted by self-chosen goals such as winning an argument or reaching agreement for the sake of agreement" (Martino, 1993, p. 18). More practically, providing controlled feedback as opposed to unfiltered feedback helps keeps panelists focused on the original purpose of the Delphi rather than getting sidetracked.

Lastly, representing the responses of the group with statistics is important because panelists are typically asked to give numerical responses to the questions posed by Delphi facilitators, and seeing the ratings of their fellow colleagues is often more informative than comments alone would be when making such judgments. In addition (depending on the nature of the statistical feedback presented to panelists) the breadth of opinions for the entire group, not just the majority opinion, can be shown to panelists.

Steps of a Delphi

The first step in performing a Delphi is identifying experts in a domain of interest (often by asking experts to nominate other experts). The purpose and scope of the Delphi are explained to the experts, and those interested in participating become panelists for the Delphi. In the first round of the Delphi, these panelists are given a largely unstructured questionnaire and asked to highlight the issues that they believe are important in the domain of interest. Many Delphis skip this unstructured round entirely (for example, Childress & Rhodes, 2008), particularly when the researchers can draw on existing

47

literature and are relying on expert guidance primarily to rank the importance of various factors provided by the experimenters.

The responses from the first round (if employed) are then compiled by the facilitator, who codes the items for topics and removes duplicate items. Once completed, this new list becomes the basis for the second round. The second round, unlike the first, is structured: panelists are given this list and asked to make quantitative judgments about the items. For example, they might be asked to predict the year by which some technological or social advance will occur. For Delphis that are not based around forecasting, the panelists are instead typically asked to use Likert items to rate the degree to which they agree with given statement or the degree to which they believe that statement is important, and so on. For example, Goldman et al. (2008) used a Delphi to understand concepts that students struggle with in introductory computer science classes. After obtaining potential items from an unstructured first round, they had panelists rate these items using three ten-point Likert items: how important a concept was, how difficult a concept was, and how likely a student was to master the concept in an introductory course. After rating such items in this round, panelists are typically allowed to explain the reasoning for their ratings. Some guides to performing Delphi methods recommend that panelists must present their reasoning if their prediction is an outlier (Martino, 1993) though others say that this should be optional. According to Clayton (1997), some Delphis also give the panelists an opportunity to add additional items at this stage.

At the end of the second round, the responses are then compiled once more by the facilitator. In the third round, the panelists receive not only the list of items again, but also

statistical feedback of the responses of the panelists (such as means, medians, quartiles, etc.) as well as the comments of the other members. In many Delphis, they are also provided with their own previous predictions. Panelists are then asked, based on all this information, to again rate the items.

This process of rating and feedback continues until the chosen stopping criterion is reached. Originally, the Delphi continued until there was consensus among experts. However, according to Landeta (2006) "[l]ater applications of the technique have eliminated the restriction of the obligatory search for consensus, so that today it might be defined as a social research technique whose aim is to obtain a reliable group opinion using a group of experts" (p.468). As consensus is no longer seen as the goal, its use has largely been supplanted by "stability" of the panel's responses as the stopping criteria. Generally, this stability of items is reached by round 3 (for Delphis without an unstructured first round) or round 4 (for Delphis with one).

Originally, panelists participated in Delphis through the mail, but such pen-and-paper Delphi have since largely been supplanted by e-Delphis, typically administered via e-mail. Some studies have used multiple methods of gathering responses simultaneously, including e-mails, faxes, and web surveys, ensuring that these methods were as equivalent as possible but leaving it up to the respondents to choose which to use (Okoli & Pawlowski, 2004). Even more recently, Gordon and Pease (2006) have introduced the real-time Delphi, where rounds are obviated in favor of "roundless" computer-mediated participation with up-to-date feedback of the group's opinions. Follow-up studies have found that these

Delphis make the process quicker and reduce drop-out rates without affecting robustness of results (Gnatzy, Warth, von der Gracht, & Darkow, 2011). [3]

Validity and Reliability of the Delphi Method

The Delphi method is not without its critics. This criticism can be split into two types: criticism of the methods employed during a Delphi, and criticism of the Delphi itself as a methodology. Sackman (1974) is critical of the Delphi's use of expert opinion, pointing out such issues as biased or incorrect experts, a lack of research on what expertise is and how to identify it, and skepticism about the superior judgments of experts (though see Clayton, 1997, for a rebuttal). Much of the Sackman's criticism, however, focuses on methodology and how the Delphi method often departs from best practices in the social sciences, a criticism echoed by proponents of the Delphi method as well (Rieger, 1986). It is not surprising that this is the case given that the Delphi method originated outside of the social sciences and is often used by researchers for whom survey methods are not a core aspect of their field.

Some studies have compared the Delphi Method to similar methods. For example, Van De Ven and Delbecq (1974) compared the Nominal Group Technique (a structured communication method based on face-to-face interactions they had developed previously) to the Delphi Method as well as to traditional unstructured interacting (discussion) groups by randomly assigning 420 individuals to 60 different panels (20 for each of the three techniques). They found that Delphi and the Nominal Group Technique outperformed

---

[3] Unfortunately, real-time Delphis require specialized, often proprietary, software to run.

discussion groups for quantity of unique ideas generated, though they also found that panelists rated the NGT as highest in perceived group satisfaction.

Rowe and Wright (1999) reviewed several studies and looked at within-subjects improvement of accuracy for the Delphi (that is, how accurate the Delphi's results were compared to the results at the end of the first round, before panelists saw any feedback or comments), finding that "Evidence for Delphi effectiveness is equivocal, but results generally support its advantage over first round/ staticized group aggregates by a tally of 12 studies to two" (Rowe & Wright, 1999, p. 364).

More recently, Graefe and Armstrong (2011) compared the accuracy of face-to-face interactions, the Delphi Method, Nominal Groups, and prediction markets on estimation tasks whose answers were known but which individuals were unlikely to know (for example, the percentage of Americans who had completed at least 4 years of college in 2006). All four forms of group interaction significantly outperformed individual estimates and the average of individual's estimates, with the Delphi the most accurate of the four group methods (but not to a statistically significant level).

Hill and Fowles (1975) argue that the Delphi lacks evidence for its reliability, in part because of the difficulty of comparing forecasts from different studies. Furthermore, they argue that the reliability of the Delphi is likely threatened by its lack of standardization in its instantiation. Woudenberg (1991) is more critical still, arguing:

> The only justified conclusion seems to be that factors other than the specific method used (capability of the group leader, motivation of the participants, quality of the instructions, etc.) to a large extent determine the accuracy of an application of a

judgment method. In accordance with this, one of the most consistent findings is that the method which was the primary focus of an article, and which can be expected to be preferred by the author(s), was almost always found to rank highest in accuracy. (p. 139)

We chose the Delphi Method for this project as we could not find any superior method for aggregating expertise. In addition, we felt that some form of back-and-forth with experts, which the Delphi method provides, was essential to the success of this project, given that we would be traversing into fields which are both complex and which use specialist terms. Without this back-and-forth, it was too likely that we would fail to interpret statements in the manner that panelists originally intended. This need meant that other methods, such as single-round surveys, would not be suitable for this project. While performing interviews or surveying the literature would likely attenuate this issue somewhat, it would have come at the cost of increasing the amount of the researcher present in the results rather than the experts.[4]

Both Delphi proponents and critics alike argue that the quality of a Delphi depends on the quality of the methods employed. There are several papers on how to create high-quality Delphis, and in our project, we draw from the guidelines provided by Chaffin and Talley (1980), Clayton (1997), Gordon (2009), Martino (1993), Rieger (1986), and Ziglio (1996). Such guidelines include recommendations concerning the structure of the Delphi,

---

[4] Unfortunately, our hope that using the Delphi would largely replace our judgment with the judgment of experts was only partially fulfilled.

such as explaining the purpose and structure of the Delphi method to the panelists before beginning (to reduce dropout and confusion), piloting questions for clarity, letting panelists shape the list as much as possible rather than the moderator, and ending the process only when there is statistical stability of responses between rounds.

Many of the recommendations given by Delphi proponents are quite familiar to social scientists, being similar to recommendations for high-quality quantitative and qualitative methods generally. These include practices such as ensuring clarity of questions and potential answer choices, avoiding double-barreled questions, and making surveys of a reasonable length. (for example, see Krosnick and Presser, 2010 for guidelines on best practices on survey design). However, while the guidelines for Delphis would be familiar to social scientists, they are not as expansive or as detailed as the guidelines that social scientists typically use, and we used the more expansive guidelines typical of social science when creating survey questions when possible.

In addition, while social scientists have yet to create a method to improve on the Delphi, some researchers have suggested ways to improve the Delphi method. For example, Brady (2015) discusses adapting the Delphi method for qualitative research (though focusing on theory building) and in their own work supplemented the Delphi method by borrowing "from constructivist inquiry and grounded theory by using a methodological journal during the course of the study to document major research decisions and rules, such as how consensus was defined" (Brady, 2015, p. 5).

More generally, while we agree that the Delphi is imperfect, we believe that this method can provide a first-order approximation of the knowledge of a field, a task which may otherwise be impossible given the hypercomplex nature of the social sciences. When confronted with the problem that we cannot formalize this knowledge perfectly, we agree with Helmer (1967), who states:

> We can either throw up our hands in despair and wait until we have an adequate theory enabling us to deal with socioeconomic and political problems as confidently as we do with problems in physics and chemistry, or we can make the most of an admittedly unsatisfactory situation and try to obtain the relevant intuitive insights of experts and then use their judgments as systematically as possible. (p. 4)

CHAPTER 3

Methods of the Current Project

Selection of Fields

One of the first decisions in this project was one of the trickiest: which fields are social sciences, and which are not? Many academic fields straddle the physical sciences, the social sciences, and the humanities. Consequently, (and also for other institutional reasons) different schools position the various fields in complex and contradictory ways.

Often, even for fields generally considered a social science (such as psychology) subfields within each field may be closer to the physical sciences or the humanities. This messy picture gets even messier when including interdisciplinary fields such as cognitive neuroscience that combines psychology and neuroscience (which itself is an interdisciplinary field that draws heavily from biology).

Given this, it is likely impossible to make a definitive list of the social sciences that would please everybody, nor did we seek to do so. Instead, as this project was designed to understand the intuitive theories that people have about humanity in contrast to the scientifically accepted theories, we chose the fields that we felt most clearly possessed the following features:

- Have as a goal the creation of accurate theories about people and why they do the things they do.
- Scientifically progressive, in the sense that these theories are tested against evidence and improved or rejected accordingly.
- Interested in all human contexts, not just one.

- Interested in all scales, from individuals to the transnational.

- Interested in all humans, from people living in small hunter-gatherer tribes to people living in vast megaregions of tens of millions.

Seven fields, in our opinion, met these criteria: Anthropology, Economics, Geography, History, Political Science, Psychology, and Sociology. These seven fields thus formed the seven panels for this project.

Four of these fields (Anthropology, Political Science, Psychology, and Sociology) are generally considered social sciences. Economics, especially broadly construed as the study of how people allocate scarce resources, also fits within the social science.[5]

Geography is unusual in that it can be split into Physical Geography and Human Geography, and while Human Geography is clearly a social science, Physical Geography (the study of the atmosphere, hydrosphere, etc.), is often not.[6] However, given that Geography as a field has a strongly interdisciplinary character, self-described Physical Geographers often study the interaction between the physical environment and human populations, such as how people modify the environment for their uses or how populations respond to environmental changes. For this reason, we did not insist that panelists had to label themselves as Human Geographers, instead, Panelists were asked to participate only if they felt that they could thoroughly explain how Geographers understood people. In practice,

---

[5] A panelist in the Economics panel made this point as well, arguing that people often erroneously think that Economics is only about money.

[6] Geography is also unusual in that it is far more prominent in Europe than it is in America.

this meant that nearly all Geographers who participated did consider themselves Human

Geographers, but two Geographers who primarily considered themselves Physical

Geographers and another who primarily considered themselves a Geographic Information

Systems researcher also chose to participate (though it was not a requirement, all three did

study people in their own research).

History is probably the most contentious field to include, as many Historians do not

consider History a science at all but instead consider History to be part of the Humanities.

In our conversations with Historians during recruitment, it seemed to us that some

Historians did consider themselves social scientists, not only in name, but in practice. That

is, they considered it their job to broadly understand people, create theories for why people

work as they do, and test these theories against the available historical evidence to see

which theories best match the evidence. However, several of the Historians in the panel

profoundly disagreed with the classification of History as a social science. Unfortunately,

while there seems to be a philosophical division within the field over the nature of History

as a field of study, unlike in other fields, there does not seem to be relatively clear and

labeled division between the science-focused and the humanities-focused part of the field.

There are subfields that are more interested in History as a science, such as Big History (for

an overview, see Christian, 2017)[7], but these are subfields rather than broad organizing

labels used by most researchers.

---

[7] Christian (2017) explicitly describes consilience as a goal of Big History, arguing that not

only can Historians learn from the natural sciences, but that the natural sciences can learn

Many fields of the social sciences were not included in this project, for both theoretical and practical reasons. Fields that focused on only one group of people were excluded, as they are not focused on creating understandings that are applicable to all people. Likewise, fields that were mainly about a specific human context, such as Law, Business, or Criminology, were also excluded. Finally, fields that focused on one specific aspect of people, such as Communication or Education, were also excluded because developing increasingly accurate and broad understandings of people are not the primary focus of these fields.

Note that this does not mean that these fields lack important insights into people – indeed, we would argue that they almost certainly do. It would be hard to imagine that one could study any of these fields (or a field in the Humanities) without learning many profound truths about people. However, since it is not the jobs of these fields to understand people broadly, it is unclear whether their insights would have been broad enough to compare across fields. Likewise, as it is not the job of these fields to understand people in a scientifically accurate way, it is likely that their understanding of people would combine deep insights into people and intuitive theories together, without the scientific testing necessary to disambiguate between the two.

from History when trying to reconstruct the History of the natural world from scattered clues.

Anonymity vs. Confidentiality

According to the proponents of the Delphi Method, one of the key features of the Delphi Method is the anonymity of panelists. This anonymity, in theory, allows panelists to share their opinions freely without concern of social censure or professional consequences.

However, from what we could tell, few Delphis are *anonymous* in the sense used by social scientists. Rather, they are what social scientists would call *confidential*. In confidential research, the experimenters can match a participant with their responses but choose to withhold this information. In anonymous research however, not even the experimenter can match a respondent to their responses. What Delphi proponents mean when they say that Delphis are anonymous is that the *panelists* do not know who the other panelists are.

We chose to make our panels truly anonymous, as our understanding is that researchers should always endeavor to do so unless it is not possible as a matter of protecting participants. We did not think that participating in the current research would cause negative professional consequences for panelists. However, by maintaining strict anonymity we could ensure this.

To maintain anonymity, we also chose not to collect demographic information about Panelists. Because we drew from a population that is both relatively small and whose information is publicly available, even basic demographic information would be enough to uniquely identify some of the panelists.

Identification Codes

When using the Delphi method, one tries to see how panelists change their responses as they take the feedback of their peers into account. For example, one panelist might critique an item, and this critique may cause other panelists to change their rating accordingly in the next round.

How does one know when the stop this process? According to Dajani, Sincoff, and Talley (1979), the key criterion is that panelists ratings have reached *stability,* that is, when panelists give similar ratings on an item across subsequent rounds. Stability is important as when panelists no longer change their ratings across rounds, this implies that panelists have already made up their mind on the item and that further rounds of ratings and commenting are unnecessary (or would at least provide diminishing returns in terms of the information gained). More technically, we would argue that instability demonstrates potential problems with an item more than stability demonstrates a lack thereof and stability is probably best thought of making us more confident that an item represents the view of experts, rather than representing that an item is free from error. Though removing items from consideration in this way may cause some nuances to be missed, it is important to remove items in order to keep the Delphi method manageable for panelists, as otherwise the Delphi would simply grow larger and larger with every round.

According to von der Gracht (2012), Delphi facilitators use many methods, of varying quality, to calculate stability (and some facilitators do not test for stability at all and instead only use consensus as their stopping criteria). Often, Delphi practitioners simply compare the average score on a question from one round to the next. However, Chaffin and

60

Talley (1980) argue that the most correct approach is to compare the ratings of *individual* panelists in one round to the ratings of those same panelists in the next, rather than comparing the average of all panelists together. In doing so, one gets the clearest picture of whether individual panelists are changing their minds. In contrast, when one only looks at group averages, one cannot tell whether individual panelists gave similar ratings across rounds or whether individual panelists did change their ratings but had this shift cancelled out by other panelists changing their ratings in the opposite direction. That is, using the group average, though easy to measure, may mask individual turmoil.

This presented a problem for the current research: to calculate the stability of an individual panelist's ratings across rounds, it is necessary to first link a panelist's responses in one round to that same panelist's responses in the next. However, the panelists in this project were anonymous. Thus, in order to link responses across rounds, while at the same time maintaining anonymity, we asked the panelists to create an anonymous identification code.

Panelists were given two options for this: the first option was to simply create a passcode of their choosing; the second option was to create a Self-Generated Identification Code.

If panelists choose to create a passcode of their own choosing, they were asked to create a code of between five and fifteen characters, made of any combination of letters and numbers. They were asked not to put their name or any other identifying information in their code. Lastly, they were instructed to write this passcode down somewhere (with a note to themselves to remind them of the purpose of this code) as they would need the

code in subsequent rounds and we would not be able to remind them of their code were they to lose it.

About half of the panelists chose this option. This system for creating codes worked adequately, but not perfectly. Unfortunately, panelists who chose this option sometimes forgot their code. One panelist apologized for doing so, noting that they themselves did longitudinal research and it annoyed them when their own participants forgot their codes.

When matching codes, we did not require codes to be an exact match across rounds. For example, some panelists transposed elements of their code, or they repeated a specific sequence of numbers a different number of times across rounds. In these cases, the chance that two similar codes were from different panelists was considered acceptably small.

The second option for panelists to create an anonymous code was to create a Self-Generated Identification Code (Kearney, Hopkins, Mauss, & Weisheit, 1984). The Self-Generated Identification Code (SGIC) is a way, commonly used in anonymous longitudinal research, that participants can create a stable code without needing to remember their code. Rather, they algorithmically generate this code each time they take the survey based off information about themselves that they find easy to call to mind. The Self-Generated Identification Code is similar to an earlier method, the Context-Determined, Rule-Generated, Pseudonym technique (Carifio & Biron, 1978; Carifio & Biron, 1982), though the SGIC includes more items (seven instead of five), and was more forgiving, in that participants would be considered matched on subsequent surveys as long as six out of seven items were the same across rounds rather than needing all items to match.

Good Self-Generated Identification Codes are surprisingly tricky to create, as it is difficult to create questions that remain stable over time[8]. For example, the Context-Determined, Rule-Generated Pseudonym technique asked five questions: first letter of middle name, first letter of month of birth, first letter of sex, first letter of the name of their street, and first letter of mother's name. Though some of these are obviously less stable than others over time, one can imagine scenarios that could change each of these (especially street name) between rounds of a survey. Likewise, the original SGIC dropped the question on the first letter of their street and added in questions on the first letter of the father's name, the student's racial/ethnic category, and the number of siblings. But one can imagine scenarios where participants would answer these in one way at one point in time and in a different way at another. Fortunately, an advantage of using a panel of people who are similar in some way (in this case, all professors in a social science) is that one can make more assumptions about the kinds of things they might have strong preferences about.

Another issue when creating SGICs is that most stable information about participants (birth month, first letter of name, etc.) also tends to be the information about a

---

[8] Schnell, Bachteler, and Reiher (2010) created scripts to match participants across rounds using the Levenshtein string distance function (which measures the transformations needed to turn one string into another) and a linear sum assignment problem algorithm to avoid the problem of multiple assignments. Unfortunately, the links to their scripts were non-functional at the time of this research.

person that is the most publicly available. This can undermine anonymity, especially for samples drawing from small pools. Furthermore, when creating a Self-Generated Identification Code, one must weigh how fault-tolerant to be. As one reduces the amount of information needed to connect a participant across rounds, one also makes it more likely that two different panelists will be linked incorrectly. In the case of the current research, we only required four out of six codes to match, as we considered false positives unlikely given the small samples and because if two panelists did happen to be falsely linked this would tend to bias the statistical results away from our desired outcome (as linking two different panelists incorrectly would tend to make items look more unstable).

Lastly, as weak preferences are inherently less stable than strong preferences, we also added an option to put "no strong preference" for each item. We argue that when somebody lacks a strong preference, the most likely future outcome is that they will continue to lack a strong preference. Whereas if somebody lacks a strong preference, asking them to enumerate their "favorite" is likely to cause spontaneous guessing, which would cause unstable responses over time.

For our version of an SGIC, we asked the following six questions:

- First letter of your favorite wild animal

- First letter of your favorite color

- First letter of your favorite non-Earth planet

- First letter of the last name of your favorite Philosopher

- First letter of the last name of the scientist you most respect

- Number of siblings

We briefly piloted these questions before deploying them in the study. One issue we found was that asking people their favorite planet would occasionally cause people to choose one planet but then immediately backtrack as they reasoned that, technically speaking, Earth was their favorite planet. We modified the questions for our SGIC based on this process and dropped any questions that caused confusion.

This system worked adequately, but not perfectly. Codes were largely stable, but not completely so. Partially this is due to the nature of such codes, as panelists circumstances or preferences may change between rounds. However, there was some evidence that several panelists did not take this task seriously despite its importance to the project. For example, some panelists did not read these directions carefully and would answer the questions with complete words rather than only the first letter, and one panelist put that they had "no strong preference" for every question, including their number of siblings.

Recruitment

Human subjects protection.

The Institutional Review Board (IRB) process had an unusually large influence on the methods of this project, enough so that failing to discuss this influence would leave many of the methodological decisions in this project otherwise inexplicable.

For this research, we sought to recruit panelists from the entire University of California and California State University systems. However, after our local IRB approved our proposal, they also informed us of an additional requirement we would need to fulfill:

at each campus we wished to recruit at, we would also need local IRB approval or the equivalent.[9]

Gaining this approval was not a straightforward process. Unfortunately, while all IRBs agree that it is their job to regulate research originating at their home institution, it is currently a point of contention among IRBs whether they also protect the population of the institution *itself* from outside research. Given this, many institutions refuse to consider outside research at all and those who do have very different processes and requirements for considering such research.

This need to get IRB approval from each campus impacted the current research in three ways. First, getting this approval was a time-consuming process. Getting in contact with the relevant personnel was not always possible. Websites sometimes listed outdated information, key personnel were unavailable, and emails and calls often went unanswered. Likewise, schools have idiosyncratic requirements for outside research. For example, one school requested signed letter of approvals from the Deans of each of the department we wanted to recruit from.

Second, this process halved the pool of professors we could recruit from and likely reduced the size of all panels. Several schools insisted on criteria we could not meet. For

---

[9] At the time of writing, the common rule (the federal regulations governing human subjects protection) is being updated to streamline IRB procedures. This update would likely obviate much of the process discussed here, though it is unclear when these new rules will take effect as their rollout has been delayed multiple times.

example, several schools required that at least one member of the research team was a researcher at that institution. This left a smaller pool to recruit from and the recruiting process itself was less efficient than it otherwise would have been. These factors reduced the number of panelists for all panels but was particularly problematic for the Geography panel. As not every school has a Geography department, losing access to several schools that did have one proved consequential.

The main way that the IRB process shaped this project, however, was that it made it impossible to change any procedures after we began recruiting at other campuses, even when it would have been optimal from a research perspective to do so. While IRBs do have expedited processes for handling changes to a research project, in the case of the current research, any change to the process would have then required reapproval at all IRBs. This process would have delayed rounds, which would likely cause some panelists to drop out of the study. Not being able to modify procedures meant, for example, that we could not supplement panels with other forms of recruitment. This meant we could not increase the size of the Geography panel even when it became clear that there were simply not enough Geographers around for us to reach our desired panel size. Likewise, we could not add additional rounds to the project, and while many comparisons successfully resolved by the end of this project (either by meeting our criteria for success or by meeting our criteria to be dropped), not all comparisons did.

Recruiting of panelists.

The participants for this study were recruited from the University of California (UC) and California State University (CSU) systems. Although we had secured permission from

the relevant IRBs to recruit at nineteen campuses, two campuses proved logistically impractical to travel to as we were not able to secure approval at any nearby campus. This left seventeen campuses that we recruited at, all of which provided at least one panelist to the project.

To recruit panelists, campuses were visited in person. As the campuses were spread out over a large geographical area, the recruitment process was largely dictated by logistical and financial considerations. As a consequence, while most schools were visited at least twice, we were able to visit some schools only once.

We chose to use this method of recruitment, even though recruiting at conferences would have potentially been more expedient, for two reasons. First, we could not ensure that simply recruiting at the major national conference for each field would give us a good cross-section of the field, as different parts of the field may not attend such conferences in proportion to their prevalence. Recruiting at conferences therefore risked creating biased sample (which might, or might not, be mitigated by attending multiple conferences per field). In contrast, we argue that aggregating all the professors within a specific field across all the campuses of the UCs and CSUs likely represents a reasonable cross-section of a field. Even though individual UCs or CSUs may emphasize one aspect of a field, this would tend to be compensated for by an opposing focus at another school.

Second, even if one did recruit professors at a conference, it is unlikely that professors recruited in this way would feel compelled to complete the somewhat lengthy Delphi process. By recruiting professors in their offices rather than in passing, we were able to meet the professors face-to-face and answer any questions the professors had about

68

the research. We believe that doing so likely increased engagement with the project, which is critical to the success of a Delphi Panel. In addition, several panelists commented that they often received requests to participate in research projects (usually via E-mail) and that they would have rejected our request, as they typically do, had we not taken the effort to recruit them in person.

We did not recruit experts by asking them to nominate other experts, as is usually done in the Delphi method, for two reasons. First, we wanted to avoid the risk that snowball sampling would cause us to reach deeply into subfields rather than broadly across fields. Second, we did not think it was necessary to ensure that panelists were experts, as they had already been verified as such by others in their field not once, but twice, and possibly even a third time; once when receiving a doctorate, again when gaining an academic position, and once again (if applicable) when obtaining tenure.

Though this recruiting process was, technically, a convenience sample, we did take steps to minimize the inevitable bias this sampling method presents. When possible, when we would return to a campus, we would do so on different days of the week and at different times of day, to avoid biasing the sample towards professors who might only be in on campus in the morning or in the afternoon. Nonetheless, an unavoidable consequence of recruiting professors in person is that the panelists tend to be professors who spend more time in their offices.

At each campus, we walked door-to-door, asking professors in the relevant fields for a few minutes of their time. If the professors agreed, we explained to them the purpose of the current research project and asked them if they were interested in participating. Often,

the recruitment discussion was brief. At other times, it led to lengthy discussions about the nature of the research and the particulars of its methods. Professors from more quantitative fields were skeptical about the value of surveys as a method of generating knowledge. In addition, professors from several fields doubted the premise of the current research, either arguing that their field did not actually know things, or at the very least, that their field would be unable to reach consensus on what it knew. Other professors argued that, philosophically, it was impossible to scientifically know things about the world at all. Unsurprisingly, given the nature of this project, the panels are biased towards professors who believe that science is epistemologically capable of understanding the world. However, there were a handful of professors who agreed to participate despite disagreeing with, or at least questioning, the idea of scientifically objective knowledge.

Before traveling to a campus, we first created spreadsheets and maps for each campus. The spreadsheets contained the information for the professors within a specific field at that school, and the maps contained the location of the various departments. Doing so proved necessary for several reasons. Firstly, this helped prevent us from attempting to recruit the same professor twice. Secondly, departments are usually, but not always grouped together, and sometimes professors from different departments are confusingly comingled. In addition, as only professors were eligible to participate in the current research, the spreadsheets helped ensure we were recruiting professors, instead of lecturers, researchers, or staff. This system was crucial for efficient recruiting but did not always work perfectly: websites were not always up-to-date, and professors, on occasion, were in the wrong offices.

Most professors (approximately three out of four) agreed to hear about the

research. Of the professors we talked to about the research, approximately one out of every

three professors agreed to participate. Most professors who declined to speak to the

researcher or to participate in the research cited their limited time as the reason. When a

professor did agree to participate, they were sent a link to the first round of the project as

soon as possible, often on the same day. While the recruitment process generally went

smoothly, this was not always the case. Several professors expressed skepticism about

whether the project was real or whether the researcher was an actual graduate student.

In total, the recruitment process took around three months. Clayton (1997)

recommended between fifteen and thirty panelists when using experts from within one

discipline, and we were able to meet this for all panels except for Geography (for overall

participation in this project, as well as the participation across rounds, see Table 1).

Table 3.1

Effective Participation for each Panel

		Structured Delphi Rounds				
Field	Recruited	Round 1	Round 2	Round 3	Round 4	Round 5
Anthropology	25	16	12	12	12	10
Economics	22	10	12	7	7	5
Geography	13	7	7	5	4	3
History	21	11	11	8	6	5
Political Science	21	8	8	8	6	5
Psychology	29	18	12	12	10	10
Sociology	20	16	12	9	10	7

Note: Effective participation refers only to panelists who rated at least one item or proposed at least one comparison during a round.

Due to the anonymous nature of the project, coupled with panelists forgetting or changing their codes, it is not possible to know precisely how many of the professors who were recruited chose to participate at some point during the study. However, we can give a lower bound of the number of panelists who participated per panel (for example, no fewer than eleven Historians participated, as at least eleven participated in a single round), and this lower bound is close to or within expectations for the Delphi method. According to Gordon (2009) "Our experience indicates that a response rate from 40 to 75 percent of the participants can be anticipated" (p. 8).

Panelists Withdrawing from the Study

Over the course of the study, around twenty panelists asked to be dropped from their respective panel. Most panelists who did so cited time as the reason for their leaving the panel. Given the decreasing size of the panels over time, it is likely that other panelists chose to leave the study but did not choose to explicitly inform us of this.

Contacting Panelists

For each round of this project, panelists were sent an email letting them know that the round was open. This email also contained a link they could use to participate. This link was quasi-random. The random elements were added so that panelists could not easily deduce the URL for other panels and peek at their comparisons, but some consistent aspects were added to each field's URLs so that we could easily check that panelists were getting the correct link for their field.

During each round, panelists would also receive reminders to participate via email. Usually, only one or two reminders were sent for each round, but in cases of low

participation for a round, the round was extended and an additional reminder was sent. In addition, whenever a panelist asked for more time on a round, the round was extended and an email was sent to let panelists know this.

Using email to communicate with panelists proved trickier than originally expected. Unfortunately, when one sends out emails, there is no way to check whether panelists ever receive them, as email servers will sometimes block emails (or send them to spam folders) without notifying the sender. One panelist reported that this was indeed the case for them, and that they did not receive an email letting them know that a round was open. While we cannot quantify the percentage of emails that went through successfully, we do know that many did go through due to some panelists successfully participating, and also because we received many replies (as well as automated responses) from panelists in response to our emails.

To compensate for emails potentially not going through, we sent out three to four emails to panelists per round and also sent out emails in such a way as to maximize the chances that professors would receive notifications about the study. For each round we would both send out group emails and at least one set of emails sent to each professor individually. We also staggered the emails when possible to avoid sending more than 100 emails in any single day, as sending too many emails is thought to increase the chance of emails being diverted by spam filters.

The anonymous nature of the survey also added additional issues with using email to communicate with panelists. Since we could not know who had participated in a particular round, we had to send email reminders to all panelists. This meant that some

panelists would email us to inform us that they had already taken the round. Sometimes, the panelists seemed concerned that their results had not been recorded, other times, panelists seemed annoyed that they had received reminders even though they had already participated in a round.

Timeline of Surveys

Ideally, each round of the project was scheduled to last one month: two weeks for the survey, plus two additional weeks to analyze the results and create the next round of the survey. In practice, the rounds often took five to six weeks. Surveys were often extended by request as panelists asked for more time due to having other obligations, such as papers or teaching. Other times, we extended the rounds (typically for one additional week) due to insufficient participation. The time between rounds was often extended as well, due to the logistical demands of analyzing seven panels and creating and testing seven new surveys. Often, this meant that three weeks were needed between rounds.

Each panel was made aware of the existence of the other panels, though not of the specifics of their operation, such as the comparisons that had been suggested in other panels. The panels were run in parallel with some staggering. The psychology survey was the first to be sent out, with other surveys sent to their respective fields as they became available.

<div align="center">

Rounds of the Project

</div>

For each of the seven fields, there were five rounds of surveys, for a total of thirty-five surveys. Panelists participated in this project via the SurveyMonkey online survey

software. This software allowed the professors to take the survey via a web browser at a time and place of their convenience.

Procedure for Round 1

Given the enormity of the literature of any particular field, much less seven of them, the first round in the current project was unstructured. That is, we asked the experts themselves to provide the concepts that later rounds would refine, rather than providing such a list of concepts ourselves (which would have also tended to bias the panels towards our understanding of their field rather than their own).

Round 1 thus acted as the seed round for later rounds, and most of the comparisons that panelists debated throughout the study were generated in this round. Sometimes these suggested comparisons made it through the entire study unaltered, but more often, these comparisons were modified in subsequent rounds until they were either dropped from the study or refined into their final form.

In this round, panelists were first directed to think about ways that experts in their field thought about people differently than non-experts did. They were directed to focus on theories about people that are both representative (held by many non-experts / experts), and important (theories that are central to how non-experts / experts understand people), as opposed to idiosyncratic or inconsequential theories. They were also directed to use plain language such that people lacking expertise in their field could nonetheless understand these comparisons. For the task directions as panelists saw them, see Appendix A.

To make these comparisons, Panelists were asked to complete two statements. For example, Anthropologists were asked to complete the following two statements:

Many non-experts think that people:

But Anthropologists think that people:

We asked panelists to complete these statements, rather than simply asking them to describe a difference between non-experts and experts in their particular field for several reasons. The primary reason was to help focus these comparisons on completing *theories about people*, rather than directly comparing non-experts and experts themselves. Second, we hoped that this way of phrasing the question would help maintain parallel language in these comparisons. This was not always successful, as some suggested comparisons simply ignored the prompts or were not parallel.

We used "non-expert" as the contrasting phrase, due to the need for brevity. The ideal phrase would have been something like "people who have never formally studied (your field) or another field of the social sciences" but the need for compactness in the survey prevented such a statement. We explained to panelists during recruitment that the goal of the study was to understand what experts in their field learned about people in the course of becoming an expert, so that panelists could draw both on their own experiences studying the field as well as their experiences teaching their field to students, and we also reiterated this understanding in early rounds. Nonetheless, panelists sometimes understood "non-expert" in a way we did not intend, especially as the study progressed, and assumed that non-expert meant unsophisticated or uneducated.

Panelists were asked to complete as many comparisons as they liked (up to seven) and could take this survey as many times as they liked before the round closed (so that they could add comparisons as they occurred to them, rather than all at once). We also asked panelists if this was their first time taking this survey or if they were returning to add additional items. Only one panelist, out of all seven panels, said that they were returning to do so.

Data Analysis for Round 1

Items sent to round 2 for rating by panelists.

As Round 1 was the seed round for later rounds, there were fewer analyses performed for this round than in subsequent rounds. The only analysis performed after Round 1 was determining which comparisons would be passed onto Round 2 and, if so, which required modifications or clarifications to fit within the study. Most of the comparisons suggested by panelists in Round 1 were passed unaltered into Round 2, but many were not.

Some proposed comparisons were not suitable as written and required minor editing. When comparisons had obvious spelling or grammatical errors, or when then comparisons were incomplete sentences as written, such comparisons were fixed before the items were sent to panelists in Round 2.

Likewise, when comparisons were conceptually clear but were phrased in such a way that they did not fit into the study format as written, they were edited to do so. Most commonly, panelists forgot that they were completing phrases and instead simply stated how non-experts and experts thought about people differently. For example: "Many non-

experts think that people: Humanity is evolving, or moving towards one goal or purpose. But Sociologists think that people: Humanity, as a species, is neither moving forwards nor backwards." In these cases, the items were rephrased to fit within the study format (in this case, by removing "humanity" but retaining "as a species" to make it clear what sense of the word people was intended).

The most complicated alterations performed on comparisons were when two or more suggested comparisons appeared to be about the same underlying concept. As the Delphi Method is already a lengthy process that requires two or more rounds of ratings for each item, it is important to reduce redundancy whenever possible. In these cases, the suggested comparisons were combined into a single comparison. Sometimes these new comparisons used the language of the clearest suggested comparison to represent all similar comparisons. Others had the non-expert statement drawn from one comparison and the expert statement drawn from a different comparison. And lastly, some comparisons combined the language of several similar comparisons to create a new comparison (for an example of such a combined comparison, see Table 2).

Table 3.2

Example of Combined Comparison

	Many non-experts think that people:	But Anthropologists think that people:
Suggested Comparison:	Were unsophisticated in the past (often times attributed to "cave people")	Experts widely recognize that peoples of the past were very intelligent and readily able to cope with a broad range of climatic and social environments
Suggested Comparison:	were more primitive in the past.	have been "modern" in their behaviors and thought patterns for at least 30 thousand years, and perhaps longer.
Combined Comparison:	were unsophisticated and primitive in the past.	have been modern in their behaviors and thoughts for at least 30 thousand years, and have always been very intelligent and readily able to cope with a broad range of climactic and social environments.

To double-check that these newly created comparisons were reasonable combinations of the original comparisons proposed by Panelists, all such comparisons were run past another member of the research team for verification. In most cases, the proposed combinations were considered acceptable combinations of the original items, and any disagreements were resolved via discussion.

Items sent to panelists for additional clarification.

Some comparisons were more problematic and would have required significant alterations for clarity or to fit within the study format. However, as the researchers are not experts in these fields, we did not feel confident that we could accurately capture the intended meaning of a proposed comparison. For this reason, these comparisons were neither altered nor carried forward to be rated in their current form in Round 2. Instead, they were sent back to panelists with a request for additional clarification.

The following types of comparisons were sent back to Panelists for clarification:

- Comparisons that used specialist or ambiguous terms.

- Comparisons that seemed to be incomplete thoughts.

- Comparisons that were contextual (about a particular time, place, or society) rather than general statements about how people work.

- Comparisons that were normative rather than descriptive (statements about how people should act instead of how they do act).

- Comparisons that compared several different ways that non-experts and experts thought about people differently rather than comparing a single way.

- Comparisons where the relationship between the two parts of the comparison was unclear.

- Comparisons that seemed similar conceptually to other comparisons, but which were too unclear to be combined as outlined above.

- Comparisons that directly compared non-experts to experts, rather than comparing their theories about people.

- Comparisons about concepts rather than competing theories of people (e.g. prices)

Contradictory comparisons.

When two suggested comparisons seemed to contradict one another, we sought further clarification on these comparisons before sending them to be rated. These were sent back to panelists for clarification as well. These comparisons were sent back with a complex series of questions designed to help us understand whether they truly contradicted or merely appeared to do so.

80

Items dropped from the study

Some comparisons were not carried forward at all and were instead dropped from the study completely. These comparisons were not about people's theories about how people work, but rather, about people's perceptions of the fields themselves. For example: "Many non-experts think that people: Economics only apples[*sic*] to financial (or money) issues. But Economists think that people: Economics can be applied to a lot of areas where individuals or groups make decisions on how to allocate scarce resources."

While people's theories about the fields themselves are interesting (and likely related to people's theories about people), these kinds of comparisons were not sent forward for ratings and comments as they were outside the scope of the study.

Procedure for Round 2

Round 2 was the first round in which the Delphi Method was fully enacted, and all subsequent rounds follow Round 2 in design.

Panelists were first asked if they had participated in Round 1, and if they had not done so, asked to go through the consent procedure before beginning. All panelists were then asked to create an anonymous identification code through one of the two procedures described previously.

Panelists then read the directions for Round 2. Panelists were reminded that they would be judging suggested comparisons on how non-experts and experts in their field view the human world differently. They were also reminded that these comparisons were, by and large, those originally suggested by their colleagues. Lastly, panelists were instructed that they both could rephrase the comparisons (to make them more

representative and/or important) or comment on the comparisons so that their colleagues could take their reasoning into account in future rounds.

Primary task for round 2.

The main task for Panelists in Round 2 was to rate the comparisons suggested in Round 1. The field with the least comparisons ready to be rated (Geography) had only four at this stage. In contrast, the field with the most comparisons ready to be rated (Psychology) had thirty-four.

Comparisons were displayed to panelists in a random order. For each comparison, Panelists were presented with a proposed comparison, and then asked to make three different ratings. The first rating was the proportion of non-experts that thought in the proposed manner. The second rating was the proportion of experts in their field that thought in the proposed manner. These two ratings describe how *representative* the comparison is. The third rating was the degree to which shifting from the first perspective to the second changes how one understands and explains the human world. This rating describes how *consequential* the comparison is.

Panelists were asked about the proportion of their field, rather than their personal opinion on the statement, because we are interested in how the field thinks rather than how this sample does. While we attempted to recruit panelists broadly within fields, the panels are not randomized samples of the field, which means that asking them for their direct opinion runs the risk of misrepresenting a field. In contrast, even if the sample was composed entirely of researchers in a field who held some unpopular position, as long as

they were aware that their position was the minority position, they could accurately rate the perspective of the field.

All three ratings were on a seven-point scale. After rating the comparisons, panelists were given the chance to improve the item by rephrasing it. Panelists were also given the chance to explain their ratings about the item to their fellow panelists. An example of the questions in Round 2, including the exact wordings of the prompts and rating scales, can be seen in Appendix B.

The questions, as written, were an attempt to juggle clarity and brevity, as we sought to have each comparison take up as little space as possible, and ideally, only a single computer screen. Not doing so would mean that each comparison would have required multiple pages or would require panelists to scroll down the webpage to read and answer all the questions for a comparison. In our opinion, needing to do so would have caused the survey to feel longer, more effortful, and more tedious, and would have also likely increased the chance that panelists would miss questions. In addition, spreading questions over multiple pages would have meant that comparisons would have to be displayed in a set order, rather than randomized. Unfortunately, the web interface we used only gives partial control over how questions are presented, and this meant that both question labels and answer choices had to be as brief as possible to fit in the allotted space. Several different variants of the survey interface were designed and piloted; the one used was the version we judged to best fit our criteria.

Item clarification.

As mentioned previously, many items were not yet ready to be rated as they first required additional clarification. Panelists were instructed that we sought their assistance clarifying ambiguous or contradictory items, but that this section was optional.

For problematic comparisons, the comparison (or group of comparisons) were presented to panelists with a brief description of why they did not, as written, fit within the study format. Panelists were then asked to help resolve the relevant issue, either by explaining the problem to us, or preferably, by creating a better-phrased version of the comparison. When comparisons were about concepts other than people (such as about trade or voting), we asked them to explain what this told us about people or suggest a new comparison to that effect.

Clarification of contradictory comparisons.

When a proposed comparison appeared to contradict another proposed comparison, these comparisons were subjected to a special, multi-part clarification process. This was done as it is difficult to tell statements that merely *appear* contradictory (due to issues of phrasing or definitions) from statements that truly represent a disagreement within the field, particularly if one is not situated within that field.

To clarify these items, panelists were informed that some proposed comparisons appeared to contradict other proposed comparisons. They were then asked whether they would like to help us interpret these comparisons or whether they would prefer to skip this section.

On these items, panelists were first asked whether these comparisons did not actually contradict, whether they were only contradictory as written (but could be combined conceptually), or whether they were truly contradictory. If panelists said that the items were not contradictory, they were asked whether the comparisons were actually about different concepts, or whether they were about the same concept but phrased differently. If panelists instead said the items were contradictory only as written, they were asked to explain how to reconcile these items. Lastly, if panelists said the items were truly contradictory, they were asked whether this contradiction came from a profound divide within the field on how to understand people, or for some other reason. If they said it was driven by a profound divide, they were asked to explain the nature of the divide, as well as describe the groups on either side of the divide. If Panelists instead said the contradiction was caused by some other reason, they were asked to describe this.

Divisions within fields.

To help ensure that all perspectives within a field were represented, we asked Panelists to nominate their group, subfield, or theoretical perspective if they felt that it had an understanding of people that differed profoundly from that of others in their field.

Secondly, panelists were asked the degree to which groups of researchers within their field agreed or disagreed on how best to understand people. This was rated on a seven-point scale, from almost complete disagreement to almost complete agreement. All fields were asked this question Round 2, except for the Psychology panel which was asked this question in Round 4.

Lastly, panelists were asked to explain any important and consistent way that researchers in their field differed in their understanding of people.

Final questions.

Now that panelists had seen many of the comparisons suggested by their colleagues, they were given a chance to suggest additional comparisons. They were also given a chance to make comments or ask questions about the survey itself.

Data Analysis for Round 2

Just as the survey for Round 2 forms the basis of later surveys, the analysis for Round 2 is likewise the basis for the analysis of later rounds. In brief, each comparison that panelists rated was analyzed both quantitatively and qualitatively. Those comparisons that met our criteria were then passed forward for further rating, those that did not were sent back for clarification or dropped from the study.

Proposed comparisons were analyzed both quantitively and qualitatively in parallel, and the fate of comparisons (whether they would be retained as-is, modified, sent for clarification, or dropped) depended on both analyses taken together. This was not a formalized process, but instead was a judgment call about whether ratings or comments called the comparison's fitness (as a representative and consequential comparison between experts in a field and non-experts) into question.

For a diagram of the analysis process for comparisons, see Figure 3. Panelists spent most of the study either rating or clarifying comparisons. In theory, this process was to continue until all comparisons either successfully passed all tests and were considered high

quality items or until they were dropped from the study. In practice, the study ended before this occurred, leaving many comparisons in an unresolved state.

Analysis of suggested comparisons.

Quantitative analysis.

Given that this was the first time that comparisons were rated by panelists, the quantitative analysis in Round 2 was simpler than in subsequent rounds. Each comparison had been rated three times by panelists (two ratings on the representativeness of the comparison, one on the consequentiality). For an item to pass the quantitative analysis successfully, all three ratings needed to have an average of 3.5 or higher (on a seven-point scale). This was a more forgiving target than later rounds (which required four out of seven), to avoid dropping items that were strong conceptually, but which were poorly phrased when originally proposed, as panelists could (and did) rephrase items to make them stronger over the course of the study.

The logic of this analysis is that, were these comparisons to have poor scores on these ratings, it would imply that the proposed comparison failed to capture how non-experts thought, how experts thought, or that the comparison only contrasted non-experts and experts on an inconsequential matter. For example, the comparison: "Many non-experts think that people use only a small percentage of their brain at any one time. But Psychologists think that people use as many brain areas as are needed to complete a task, which can be all or some" was dropped from the study, as while both non-expert and expert statements were rated as highly representative, the shift from one perspective to

the other was rated as largely inconsequential (2.2 out of seven). That is, changing one's

perspective on this issue would have only a minor impact on how one understands people.

Comparisons that met these criteria were eligible to pass onto Round 3 to be rated

by panelists again, but not all comparisons that met these criteria were sent on in this

matter. If panelists raised serious concerns about the item in their comments, these

comparisons would be sent back to Panelists for further clarification. In addition, when

comparisons had poor ratings, but also had comments that implied the comparison

touched on an important concept but was poorly phrased, these comparisons were sent

back to panelists to clarify or rephrase the comparisons if they wished.

Qualitative analysis.

The Delphi Method has reasonably well-enumerated principles for recruiting

participants and guiding their deliberation, many of which would be familiar to qualitative

researchers unfamiliar with the Delphi Method. However, when it comes to analyzing the

qualitative data produced by panelists, it is a different matter. Even though the Delphi

Method relies on the statements of panelists, it does not seem to have any widely agreed-

upon system for analyzing the statements of panelists. Indeed, many guides for performing

the Delphi hardly mention the topic.

For example, Martino (1993, p. 18), in regard to analyzing the initial seed round,

states that "The panel moderator extracts from the questionnaires only those pieces of

information that are relevant to the issue and presents these to the group." without

describing how the moderator decides what is relevant and what is not. Likewise, for

analyzing comments on items in later rounds, he states that moderators should provide

88

statistical feedback, as well as "a consolidated summary of the panel's reasons for advancing or delaying the forecasts. Similar arguments are combined, and lengthy arguments are summarized." (Martino, 1993, p. 20) without explaining when or how one should combine or summarize.

In fairness to Delphi proponents, it is hard to see how this process could be specified more precisely. When performing a Delphi, one must make almost philosophical judgments. Is this statement clear or ambiguous? Are these statements about the same concept or different ones? It is hard to see how one could specify how to weigh abstract concepts in advance, which means these are inescapably judgment calls, to be made when they are encountered. This is an inescapable fact for qualitative research in general, and one reason that qualitative researchers talk about the importance of a skilled analyzer.

However, while the methods of qualitative analysis for Delphi panels are not well specified, Delphi proponents and qualitative methods researchers are on agreement on the most important part of qualitative research: that the final result should come from the data, not the analyzer. Of course, this is the aspirational goal and is probably impossible in practice. Even if the analyzer does not blatantly insert their own opinion into the analysis, the analyzer still must make thousands of tiny decisions when combing through the results. As these decisions are inevitably guided by the biases of the analyzer, through this process, some amount of the analyzer is inevitably present in the result.

This was a major source of concern for this project. But one major reason we chose the Delphi method, rather than the much simpler process of simply surveying the fields a single time (or performing a literature review), was so that if we did misrepresent a

difference between how experts and non-experts thought about people, panelists had the opportunity to correct this with their ratings and comments.

In addition, just as we applied our training in quantitative methods to the survey construction and quantitative analysis, we attempted to apply our training in qualitative methods to the qualitative analysis, although the squishier nature of qualitative analysis made this more difficult. We approached this project much as one approaches the analysis of interview data, with a systematic analysis of themes that occur in the statements of participants, followed by grouping utterances that shared similar themes. However, unlike most qualitative research, where this step is followed by the creation of a codebook to systematize the labeling of themes and their application to the data, here, we sought to maintain the utterances of panelists close to their original form, including maintaining much or all of their original language (as the more you transform the data, the less the participants' original intentions are conveyed). Thus, while we thematically analyzed the statements of panelists, this was mainly done in order to ensure proposed comparisons were relevant to the study and to reduce conceptual redundancy.

The primary aspect of the qualitative analysis was answering the following questions: did a proposed comparison (or proposed improvement to a comparison) produce a clear contrast between how experts and non-experts thought about people? If not, was this because the comparison was poorly phrased or was the problem more fundamental? Secondly, we often needed to judge whether proposed comparisons were different instantiations of the same underlying concept, and if so, how best to combine them. This was often tricky as we would sometimes need to combine five or more original

comparisons into one comparison that represented the key insight. Answering these two questions made up, by far, the bulk of the qualitative analysis.

Each round was analyzed in a manner informed by, but mostly independent from the previous rounds. As panelists rephrased comparisons, they would sometimes begin to overlap conceptually with other comparisons. The fact that these comparisons did not overlap previously was irrelevant, and these now overlapping comparisons might be combined or sent back to panelists to clarify the overlap.

As is done with most qualitative research (but not seemingly in most Delphi panels) we sought to have a second rater confirm that our analysis was a reasonable summation of the underlying data. Unfortunately, given the large volume of items to be processed and the short turnaround time between rounds, we were not able to have a second rater perform all the same steps independently. However, we did think it was important that major decisions were considered from another viewpoint. To this end, we asked another member of the research team to double-check all items that were significantly altered between rounds. When doing so, they were presented both with the original items and with our proposed modifications. For example, when we combined items, we would show both the original suggested comparisons and our new proposed comparison, to ensure that the original items really were conceptually similar and, if so, that the new proposed comparison captured the most important elements of the originals. During this process, most proposed alterations were agreed upon, and disagreements were resolved by discussion.

Given the large number of comparisons in each round, not all comparisons were double-checked by another member of the research team. In particular, comparisons that needed only minor alterations (such as spelling or grammar) were not double-checked. Nor were comparisons checked that were being sent back to panelists for clarification, as the panelists themselves would be double-checking the comparisons (this is similar to member checking, a technique in which qualitative researchers ask participants to verify the reasonableness and accuracy of their results).

Analyzing comparisons sent back for clarification.

Many panelists chose to rephrase or comment on at least one comparison sent back for clarification. Some comparisons, however, did not receive clarification. Comparisons were analyzed in much the same way as the original proposed comparisons were analyzed in Round 1.

Comparisons adequately clarified.

When panelists provided an adequate rephrasing of a problematic comparison (that is, a rephrasing that appeared to solve the apparent issues with the comparison as originally phrased) the comparison would be considered rehabilitated, and this new version would be sent to the panelists to be rated. Likewise, if panelists did not rephrase the comparison, but clearly explained the issues with the comparison and how to resolve them, the comparisons would be rephrased accordingly and sent back to panelists.

Sometimes, different panelists would suggest different ways of rephrasing a comparison. If these rephrasings were conceptually similar, we would go with the clearest suggested rephrasing or combine them. However, if these suggested rephrasings seemed to

be about different underlying concepts, we would then split the original comparison into two or more child comparisons, each of which would then be sent to panelists for rating or further clarification. Likewise, if panelists suggested a rephrasing of a comparison that seemed conceptually distinct from the comparison as originally written, we would split these into two or more child comparisons and send them forward for either rating or further clarification. This process was iterative, and some of the comparisons at the end of the study were the children of earlier child comparisons.

Comparisons sent back for verification.

Sometimes, it was not clear whether a proposed rephrasing adequately solved the issues with the comparison as originally phrased. In this case, we would send the original item and the suggested revision back to panelists, to ask for verification that the proposed revision truly improved the comparison or that the newly phrased comparison did in fact represent a true comparison between experts and non-experts.

Comparisons not clarified or inadequately clarified.

When comparisons did not receive clarification, or when these clarifications were unclear, these comparisons would be sent forward to Round 3, in the hopes of receiving additional clarification there. Some comparisons were only partially clarified (for example, when panelists had rephrased half of a comparison but not the other half). These were also sent back for further clarification.

Contradictory items.

By and large, our complicated multi-question system for clarifying contradictory comparisons did not work as intended. Few panelists chose to take this section (many

chose to skip it entirely), and when panelists did take this section, they frequently disagreed with each other. For example, some panelists would say that apparently contradictory comparisons were truly contradictory, while others would say that they were not.

When panelists did participate in this section, their statements were not always comprehensive enough to disentangle the contradictory comparisons (or panelists outright disagreed). Consequently, we abandoned our intended way of analyzing these items and instead treated these comparisons in a similar way to other comparisons, in that we tried to extract clear and self-contained comparisons out of the panelists' statements. If such comparisons were conceptually clear (either originally, or after panelists rephrased or commented on them during this process), they were sent to panelists to be rated. If such comparisons were not, they were sent forward to the next round to be clarified.

Though this system did not work as well as we hoped, it did produce a number of new comparisons, some of which progressed all the way through the study successfully. However, this did mean that some of the comparisons panelists saw in future rounds did contradict other comparisons. We left it up to panelists to rate and comment these comparisons accordingly, and through this, determine which comparisons better reflected their field.

This system was also useful in that some panelists explained why people in their field suggested contradictory comparisons, which helped limn some of the disagreements within and across fields.

Additional Comparisons

Any additional comparisons suggested by Panelists in Round 2 were treated much the same as the original suggested comparisons were treated in Round 1. Many were passed forward to be rated in Round 3, while others were either dropped or sent back to Panelists for further clarification.

Procedure for Round 3

Round 3 was very similar to Round 2. As in earlier rounds, Panelists were first asked to go through the consent process if they had not done so previously. Panelists were then asked whether they had created an anonymous code previously or not. If they had not, they were put through the same code creation process that was used in Round 2. If they had previously created a code, they were asked which of the two kinds they had created. If they had created their own code, they were asked to enter it again. If they had instead created a Self-Generated Identification Code, they were instead given the prompts for recreating their code algorithmically. Panelists who forgot which type they had created or their code itself were asked to create a new one instead.

Panelists then read the directions for Round 3. The directions for Round 3 briefly reminded panelists about the purpose of the research and the nature of their task. The directions also reminded panelists that they could add comments to any comparison, and that it was especially important to do so if their ratings differed significantly from those of their colleagues, so that their colleagues could understand their reasoning.

In previous rounds, panelists had raised a number of questions about the project, such as why some items were ambiguous or confusing. These questions were briefly answered here before panelists advanced to rate comparisons.

Primary task for round 3.

The main task of Round 3 was largely similar to that of Round 2. Once again, panelists were shown a suggested comparison about how non-experts and experts thought about people differently and asked to rate this comparison on its representativeness and consequentiality.

However, unlike in Round 2, in which panelists were rating comparisons for the first time, panelists were now rating items that had already gone through a round of ratings and comments. Therefore, when panelists saw the comparisons this time, they could also see the ratings of their colleagues, as well as the comments (if any) of their colleagues. For an example of these questions, including the feedback panelists received about their colleagues' answers, see Appendix C.

In the Delphi method, it is necessary to give panelists a sense of how their colleagues rated an item in previous rounds. Delphi practitioners have used a variety of methods to do this, including reporting means, medians, and interquartile or interquintile ranges.

In the case of the current research, we chose to present the ratings themselves to panelists in the form of histograms. Given that panelists were likely already familiar with reading and interpreting histograms, we felt that presenting the data directly gave them the most possible information to weigh when making their judgments. In addition, measures of central tendency can be misleading, particularly for bimodal or otherwise non-

normal distributions, and in the case of the current research, it was important that panelists be able to tell the difference between a low average due to people tepidly agreeing versus a low average because some panelists strongly disagreed. With the histograms, we did also include summary information, including the number of panelists who had made a rating on that item in the previous round, as well as the average and standard deviation of these ratings, in case panelists would find this information useful but did not want to have to calculate it themselves.

As there were three ratings for each comparison, Panelists needed to see three histograms. Given this, it proved difficult to display this information for panelists in a compact fashion. We sought to display this display this information in a way that allowed the panelists to see, at the same time, the suggested comparison, the ratings and comments of their colleagues, and the questions where panelists were asked to make their new ratings. We felt that allowing panelists to see as much of the information as possible at the same time was important, as panelists would have to remember all the aspects of the question they could not currently see.

We tested several ways of presenting this information compactly and clearly within the standard SurveyMonkey interface but found these unsatisfactory. Instead, we created scripts for SPSS that created the individual histograms and measures of central tendency for each rating, but then ran these histograms through an action in Adobe Photoshop that assembled the three histograms per comparison into a single image and then added appropriate titles and other UI elements. This allowed the data to be presented back to panelists as compactly and clearly as we could manage. Even so, once histograms and

previous comments were added, many panelists probably needed to scroll down a webpage to see all aspects of that comparison, particularly if panelists used small computer screens or there were lengthy comments on a comparison.

Creating and displaying these histograms proved logistically complex. Many histograms needed to be generated for each round, sometimes over fifty for a single panel. The automation of this process was a crucial aspect of making seven simultaneous panels logistically feasible, but even so, creating and double-checking that histograms were paired to the correct comparison proved to be a time-consuming task and was a major contributor to delays between rounds of the panel.

Clarifying items.

Panelists were again asked to rephrase or clarify confusing or ambiguous comparisons, in much the same way as they had in round 2. Several panels had over a dozen such comparisons in need of clarification.

Divisions within fields revisited.

Panelists were presented with the information gleaned from the previous round and asked to elaborate on the suggested philosophical divisions or subfields that differed from other subfields or the field as a whole. They were also asked to suggest any additional ways that researchers within the field disagreed on how they understood people.

Data Analysis for Round 3

Analysis of suggested comparisons.

Comparisons were analyzed similarly to how they were analyzed in Round 2. However, as items already had an opportunity to be improved by panelists, the criteria for

the ratings being considered successful was increased (from at least 3.5 out of 7 to at least 4 out of 7). The exception was the Economics panel: as every single comparison for the Economics panel that panelists had been asked to rate in this round had at least one rating below 4. Due to this, we held off on raising the cutoff for one additional round for the Economics panel, in case Economists, for some reason, found it generally more difficult to agree than other fields. As the Economics panel only had two comparisons reach the end of the study, and one of the comparisons had been sent back for clarification rather than ratings at this point, this decision ultimately only prevented one comparison from being dropped.

Calculating stability of comparisons.

Delphi panels, in theory, continue until the panelists' ratings have reached stability. In this project, we followed the recommendation of Chaffin and Talley (1980) and calculated the stability of individual panelist's responses across rounds rather than the testing the averages of the entire group. However, given that our panel was truly anonymous, doing this calculation proved far more complicated than we anticipated.

Essentially, statistical tests for the social sciences are generally designed to be used with data that is either independent (drawn from different samples) or repeated (the same sample responding twice). Often, there is a specific statistical test (i.e. independent samples t-test vs. paired-samples t-test) for each version. However, we realized that the standard tests were inappropriate for our data, as our data was neither purely independent nor purely repeated.

The problem of having a combination of independent and repeated data is not only a problem for the current research but has been recognized as a problem for longitudinal research generally. Recently, statisticians have begun tackling how handle data that is neither perfectly independent nor repeated, though there is still much disagreement about the best way to do so, and many different methods have been proposed and critiqued. For example, Looney and Jones (2003) propose a corrected z-test for this kind of data, whereas Samawi and Vogel (2014) recommend using either a weighted z-test or weighted t-test.[10]

Unfortunately, it was unclear which, if any, of these proposed methods were correct for our data, as our data is a combination of *three* kinds of data: independent data (from panelists taking one round but not the subsequent round or vice versa), repeated data (from panelists taking two rounds in a row), and repeated-but-unpairable data (from panelists who did participate in subsequent rounds but forgot their codes or were otherwise unpairable). The last type of data is particularly problematic, as it is not possible to pair such data, but nor would it be correct to treat such data as independent either. Additionally, the need to prepare rounds as quickly as possible meant we were not able to enact the far more complicated statistical procedures these tests require.

After discussion with colleagues with greater statistical expertise, we compared only those panelists who could be matched across rounds and ignored other panelists when calculating stability. This meant the comparison itself was the correct comparison

---

[10] As many Delphis test for group stability rather than individual stability, they avoid this problem, but are likely (for reasons enumerated earlier) testing the wrong thing.

given the nature of the data. However, doing so came with the downside of a reduced n for these comparisons.

The specific test used was the Wilcoxon signed-ranks test, a non-parametric repeated measures test similar to the paired-samples t-test, but which does not assume normality of the distributions. We used a relatively conservative criteria of $p < .1$ rather than $p < .05$ for this statistical test. This is the opposite of most social science research, in which $p < .1$ is the more forgiving criteria. This is because in most research, you are trying to demonstrate that two averages are significantly different from each other, but here, we were testing that they were not. That is, a greater distance between the two averages (and thus, more change across rounds) is needed to reach the .05 level of significance than the .1, which means that using the .05 level of significance makes comparisons appear more stable (or technically, less likely to appear unstable). We chose this more conservative criteria to make it more difficult for comparisons to pass through the study successfully, and thus, be more likely to represent the fields accurately when they nevertheless did so.

As there were three ratings per panel, all three ratings needed to be stable for the comparison to be considered stable. If any of the three ratings differed significantly across rounds, the comparison was considered unstable, and sent back to panelists for another round of rating. Despite this requirement, most of the comparisons were stable when tested.

Before comparisons would be tested for stability, we required that at least 5 panelists could be matched across rounds. Unfortunately, requiring at least five panelists meant that there were several comparisons where it was impossible to test stability across

rounds, especially in later rounds as the number of panelists shrunk. And even for the comparisons that could be tested, the n was often so low (due to low participation and difficulty with matching panelists across rounds) that these analyses should be interpreted with caution.

Creating comparisons with three components.

When analyzing panelist suggestions for how non-experts and experts think about people, we noticed a recurring pattern. In some cases, while experts within a field agreed on the *expert* perspective, they disagreed on the *non-expert* perspective. That is, while panelists agreed on the correct way to understand the world, they disagreed on how non-experts naively did so.

For example, Anthropologists, in their expert statement, argued that people share some universal moral drives but express these in varied ways across cultures. In contrast, some Anthropologists argued that non-experts believed in a universal morality, while other Anthropologists argued that non-experts believed that morality is entirely driven by culture.

In other cases, Anthropologists commented that non-experts commonly held one of two competing theories. For example, on a comparison about whether people are naturally peaceful, one Anthropologist commented: "There are two popular myths: the noble savage and, its opposite, the ignoble savage. The non-expert statement reflects the noble savage myth, but in my experience the ignoble savage myth is equally as common."

In our view, both perspectives are likely correct. That is, some non-experts think like the former, and some like the latter. In hindsight, it is obvious why this should be so:

just as our intuitions of physics leads people to have many different incorrect models of the physical world (i.e. potential shapes of the Earth), the same thing should happen for the social world. The underlying intuitions that give rise to our cognitive models can support many such models, depending on which intuitions are most heavily recruited. And just as with physical intuitions, we should expect that there would be some domains in which a single misconception dominates, and other domains in which there is a more complicated landscape with many misconceptions present but none clearly dominating. [11]

However, in this study, we did not ask experts to list the various perspectives non-experts had. Instead, this project, by the nature of the comparison prompt, forced panelists to highlight a single popular way that non-experts think about people. It is therefore not

---

[11] These kinds of conflicts will occur when non-experts tend to take extreme positions on a topic, whereas experts take a more nuanced view. Nonetheless, it would be a mistake to conclude that one can take competing non-expert positions and find truth by splitting the difference There are cases in which the expert view of the physical world is utterly orthogonal to all non-expert perspectives. For example, quantum mechanics is certainly not a compromise position between competing naïve physics. Nor has any group ever naively invented anything like modern cosmology or climatology or nuclear physics. There are physical realities that are so bizarre and unintuitive that humans never stumble upon them without generations of scientific research, and whether this is true for social realities as well remains to be seen.

surprising that some panelists would highlight one popular way that non-experts think, and another panelist would highlight a different one.

To handle cases like these, we combined similar items to create a comparison with three components instead of two. The first two parts of this comparison were common non-expert perspectives, and the last part was the expert perspective that contrasted with both.

These items were analyzed similarly to the normal items, with a few differences. There were now five, rather than three, components to rate: two on the representativeness of the non-expert perspectives, one on the representativeness of the expert perspective, and two on the consequentiality from changing from the non-expert perspectives to the expert one.

Secondly, the presence of multiple popular perspectives makes it more difficult for any particular perspective to dominate. Therefore, the criteria for dropping these items was reduced from four (half) to three (a large minority), but only on the proportion of non-experts who held that specific perspective. Not doing so would have caused all comparisons with three components to be dropped, as it is not logically possible that two mutually exclusive positions could be held by more than half of non-experts, or even by half if there are any other additional perspectives. The criteria for dropping these items was not changed for the other ratings (that is, an average of at least four out of seven was still required). We did this as it was still equally important that experts agreed on the correct perspective and that experts agreed that changing one's perspective from either of the non-

expert views to the expert one would significantly change how one understood the human world.

The criteria for stability was not changed for these items. All five of the components needed to be stable across rounds for the comparison to be considered stable.

Combining comparisons for political science.

When analyzing the data for the Political Science panel, it became clear that many comparisons suffered from serious conceptual redundancy. For example, several items gave some variation of the idea that non-experts think people use information in a sophisticated manner, but experts think people do not.

For example, one comparison about whether people use information in a sophisticated manner was about "evaluating campaign messages." Another was about "voting." And another was about "supporting candidates." That is, these comparisons agreed in *concept*, but differed in *topic*. But these topics – though obviously a focus of intense study for Political Scientists – were not the central purpose of this study.

Thus, we combined the comparisons that shared the same underlying concept and subordinated the topics (such as voting) to be an example of the conceptual difference rather than the main point of the comparison. As before, these new comparisons were created by using the original language of the comparisons as much as possible, and we generally created these items by starting with the best-phrased version of the comparison and modifying it as needed to cover ideas raised in other comparisons.

By combining these conceptually similar comparisons, we were able to boil eleven comparisons down to only two and reduce both redundancy and the amount of work

panelists needed to do, but at the loss of some breadth and accuracy. Like other combined comparisons, these new comparisons were run past another member of the research team to verify that the combination was a reasonable summation of the original comparisons.

Procedure for Round 4

Round 4 operated in much the same way as Round 3. Panelists were again asked whether they had participated earlier or not and again asked to create or re-enter their anonymous code.

Panelists were again asked to rate comparisons suggested by their peers. When these comparisons had been rated previously, panelists would see the previous ratings and comments of their fellow panelists. Comments were displayed cumulatively. That is, comments made in Round 2 on a comparison would still be shown in Round 4. The exception to this was when the item had been altered, as comments would no longer be discussing the present version of the comparison.

Panelists were not asked to clarify items in this round. As the Delphi method requires at least two rounds of ratings for an item to potentially reach stability, and since the next round would be the final round, there would have been no chance for any newly clarified items to be rated twice. For the same reason, we did not ask panelists to suggest new items in this round.

Anthropologists and Geographers were also introduced to comparisons with three components (as described above) and asked to rate and comment on these types of comparisons as well.

Data Analysis for Round 4

The data analysis for Round 4 was performed in much the same was as in Round 3. Items that were rated harshly by panelists were dropped. Items that had been previously rated well by panelists, and were rated well again in Round 4, were tested for stability across rounds. Items that reached stability were considered successful and removed from further rating, items that did not were sent back to panelists for further rating and commenting. As the next round would be the last, no items were sent back to panelists to go through the clarification process.

Rephrasing political scientist comparisons.

In the previous round, we had combined eleven different comparisons into two new ones, based on conceptual similarity, to reduce redundancy in the Political Science panel. However, while panelists were positive about one of the two comparisons, the comparison about whether people were rational immediately set off an argument about the meaning of rationality. Several panelists commented that their understanding of the comparison depended on which definition of rationality was being used as this was a topic of great disagreement for Political Scientists. In addition, this comparison was rated poorly by panelists.

However, as panelists thought the original comparisons were important enough to warrant mentioning, we did not want to toss this item entirely as it was likely that the fault lay with our particular combination. Thus, we went back to the original comparisons and combined them again, but this time, we did not group together all items that *we* considered to be about rationality but instead tried to get at the specific concepts the panelists raised.

One of these new comparisons focused on whether people made judgments carefully; the other focused on whether people are good judges of the quality of information. These versions proved more successful and less controversial.

Procedure for Round 5

Round 5 operated in much the same way as Round 3 and 4 had. Once again, panelists rated and commented on proposed comparisons. However, there were some differences from previous rounds.

Special questions.

A handful of items, despite receiving acceptable ratings, had comments from panelists that called their validity into question. For example, one panelist criticized a statement that non-experts think criminals act due to "amorality," arguing that "immorality" was the correct concept. Unfortunately, since this was the final round, there was no longer time to clarify the items or to modify them and submit them for additional rounds of ratings.

For these questions, we asked the panelists to rate these comparisons as normal, but also asked panelists questions about whether the comparisons would be improved if the suggested changes were made. The reasoning behind this was, essentially, the transitive property: if panelists liked a comparison, but panelists also thought changing it would improve it, it stands to reason that panelists would like the new version even more. Unfortunately, this process did not work as intended, as panelists did not unanimously agree on whether proposed changes would improve an item or which of two proposed versions they preferred. These items were consequently dropped from the study as

comparisons between non-experts and experts, though some of the debates they demonstrated within fields are discussed later.

Concepts that caused disagreement.

Several concepts seemed to cause disagreement both within and across fields (i.e. nature vs. nurture). To help clarify these conceptual disagreements (especially ones that involved researchers preferring different definitions for a shared term like rationality), we looked at concepts for which there seemed to be disagreements within or across field. Based on this, we created thirteen sets of contrasting statements intended to help us understand where fields sat on these issues. For these questions, we asked panelists what proportion of researchers in their field would argue or each perspective.

However, panelists often rejected the presented dichotomy, arguing that their field either did not consider the debate relevant to their field or that researchers in their field preferred a third perspective. For this reason (and due to low participation for several panels) we dropped these questions from the study, however, panelist comments on these questions did help elucidate the divisions across and within fields.

Final questions and comments.

As their last task, Panelists were asked a series of optional questions. Some questions asked panelists about their field's understanding of people, other questions asked panelists about their experience participating in this project.

Data Analysis for Round 5

Round 5 was analyzed much the same as Round 4. Once again, ratings and comments were used to judge the quality of items, and comparisons that had already been

rated previously were tested for stability across rounds. In addition, all additional

questions were analyzed by calculating standard descriptive statistics.
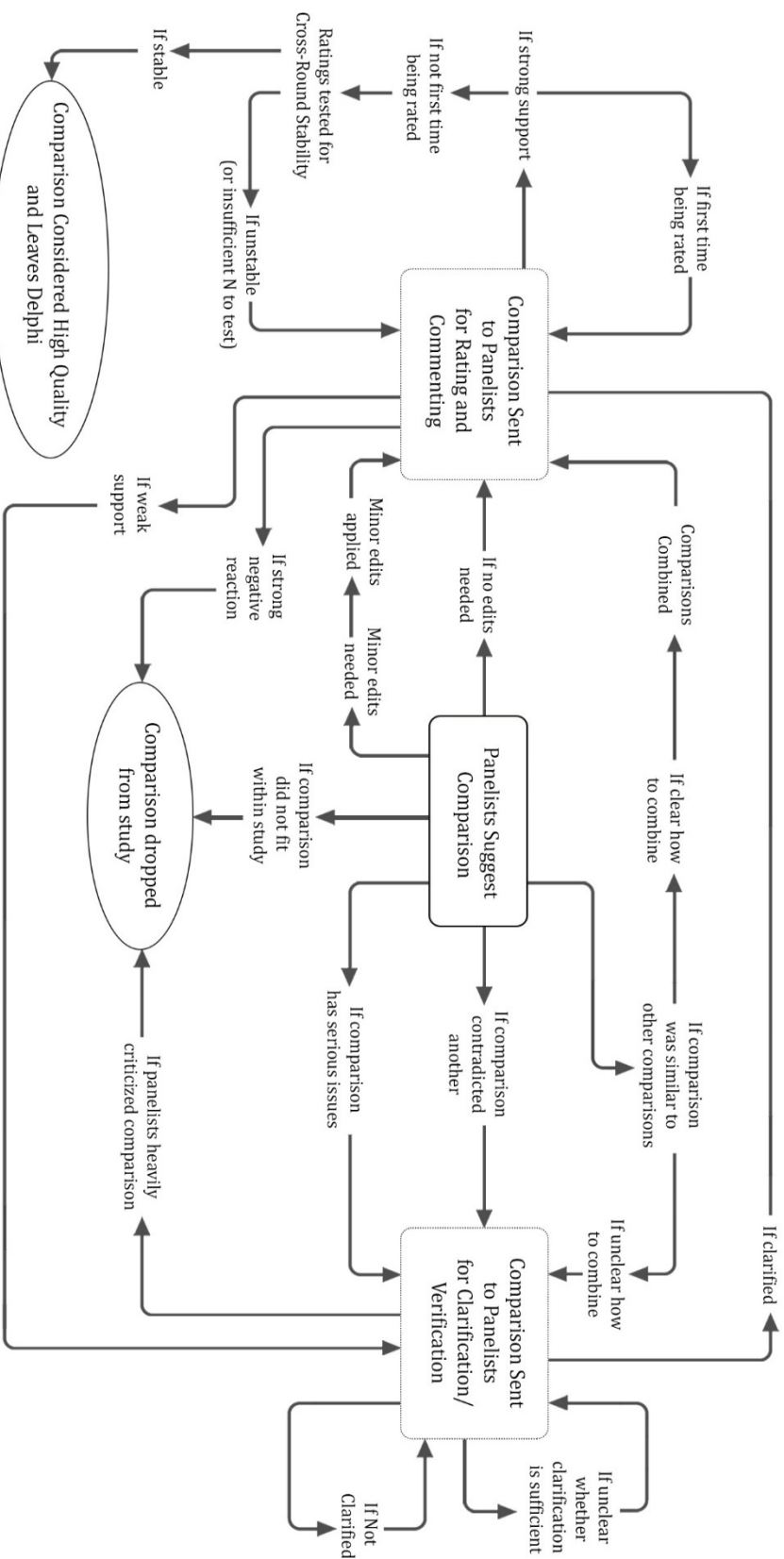
Figure 3: Pathway of comparisons through the study

111

CHAPTER 4

Overview of Results for all Panels

Broadly speaking, three of the panels (Anthropology, Psychology, and Sociology) were successful Delphi panels, producing several comparisons that were highly rated and which reached stability. One of the panels (History) was moderately successful, producing a handful of successful comparisons. Lastly, three of the panels (Economics, Geography, and Political Science) were largely unsuccessful. While these panels did produce comparisons that were rated highly by panelists, the low participation in these panels meant that it was never possible to check whether comparisons had reached stability. The results of all seven panels are presented, in alphabetical order, in subsequent chapters.

While some panels were not successful, this does not, in our judgment, mean that these panels were failures. Nor does a panel being successful mean that the comparisons suggested are without flaws. While a strict interpretation of the Delphi method might simply drop all comparisons that did not reach stability (and thus drop all panels that did not have stable comparisons) and present all successful comparisons without further comment, we have chosen not to do so for several reasons.

The primary reason is that the difference between a comparison being stable or not was often simply a matter of the number of panelists who could, despite their anonymity, be connected across rounds. For example, even if seven panelists participated in two subsequent rounds, it would only take three of them forgetting their code to make it impossible to test comparisons for stability. Often, there were enough panelists to test stability across rounds for many of the comparisons, but if any of the panelists skipped all

(or even just some) of a particular comparison in either round, that comparison could no longer be tested for stability. Because of this, while many comparisons failed to reach stability, it was not because they proved unstable. Rather, this failure to reach stability almost always demonstrated a failure not of the comparison, but of the process.

The second reason is that panelists, in their comments, made many critiques of the successful items. These critiques often argued that comparisons were suboptimal as phrased, and many of these comparisons, despite having high ratings and reaching stability, can probably be refined further, though doing so would require additional work from experts in the relevant fields.

We have grouped items by whether or not they reached stability, however, our view of stability is that it is a useful criterion but not an exhaustive one. While it is not necessarily the case that comparisons which reached stability are better comparisons than those comparisons which did not, stability does still demonstrate that panelists had largely made up their minds about an item. Thus, we would argue that we should interpret stability as a sign that we should have greater, though not complete, confidence that a comparison does truly represent a field.

Comparisons are displayed by field. Within each field, comparisons are displayed by their combined ratings, with comparisons that had higher average ratings displayed first (ties were settled by coin flip). We have included the actual ratings of panelists with each comparison, as the degree of agreement within fields is of theoretical interest.

We have also included many of the panelists' comments about the comparisons. In particular, we included any comments that expressed disagreement with the comparison

or which called aspects of the comparison into question. However, we did not include comments that expressed support for the comparisons (since the ratings already do that), comments about whether or not people in the field would agree with the expert statement (since the ratings already do that), or comments on earlier versions of the comparison (since the issues they discussed may no longer apply).

The amount of information presented here does make the results complicated, far more so than we originally envisioned. The original hope for this project was that we would be able to distill all the knowledge of the fields of the various social sciences neatly, perhaps into a single uncomplicated list for each field.

This proved hopelessly naïve. In the end, we overestimated the degree to which Delphi panels would return comparisons between experts and non-experts that are unimpeachable. The disagreements within fields, and even just differences in how individuals understand and frame things, meant that it was often impossible to boil down comparisons into simple contrasting statements that everybody agreed on. For this reason, while we have classified these comparisons as either successful or in progress, it is probably best to consider all comparisons, even the successful ones, as works in progress to varying degrees. Nonetheless, while we were unable to refine comparisons to the degree we had hoped for, panelist ratings and comments do make it clear that many of these comparisons represent strong, though not perfect, comparisons of how non-experts and experts differ in how they understand and explain the human world.

CHAPTER 5

Results of the Anthropology Panel

The Anthropology panel produced several successful items. While there were some disagreements over how Anthropologists understood the world (largely based on the divide between Biological and Cultural Anthropology), the major issue with suggested comparisons in the Anthropology panel was over the non-expert statements. All eight items dropped from this panel for low ratings were dropped because of their ratings on the representativeness of the non-expert statement. Panelists sometimes disagreed about the how non-experts understood the human world, with some arguing that non-experts held one view, and others arguing that non-experts held an entirely different view. As explained in the section on the data analysis for Round 3, we were able to handle some, but not all, of these cases by creating comparisons with three components. These combined comparisons proved successful – all had reasonably strong ratings and reached stability.

Below we have listed the most successful comparisons from this panel, as well as comparisons that were still in progress when the study ended. Some of the comparisons include ratings that are not whole numbers due to one panelist taking the last round twice. As there was no logical reason to prefer one run to the other, that panelist's scores are the average of their two runs.

Many non-experts think that people:

can be classified into different "races" based on biological variation between populations.

But Anthropologists think that people:

cannot be classified into different races because socially-constructed racial categories have an arbitrary and variable relationship to biology, given the lack of substantial biological variation between populations.

Mean =5.75
Std. Dev. =0.622
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =6.75
Std. Dev. =0.452
N =12

Proportion of Anthropologists
(Panelist Ratings)

Mean =6.42
Std. Dev. =0.669
N =12

Consequentiality of Shift
(Panelist Ratings)
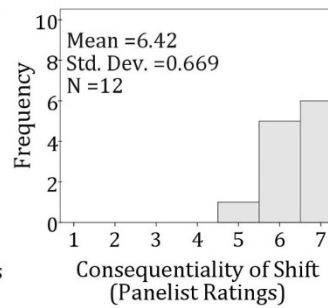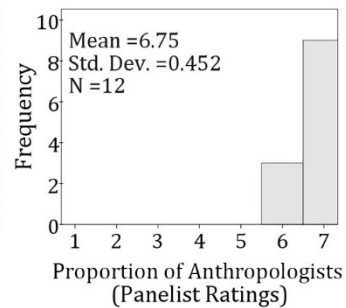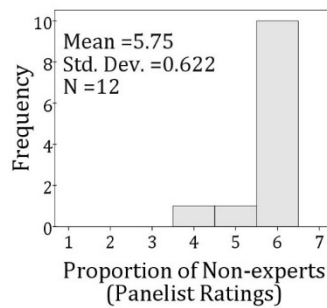
Stable: Yes
(Avg N: 5.66)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 5.1: Stable Anthropology Comparison 1

Many non-experts think that people:

are either male or female.

But Anthropologists think that people:

have variation in both sex (phenotypes) and gender (socially-defined roles).

Mean =5.5
Std. Dev. =0.674
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =6.67
Std. Dev. =0.492
N =12

Proportion of Anthropologists
(Panelist Ratings)

Mean =5.82
Std. Dev. =0.982
N =11

Consequentiality of Shift
(Panelist Ratings)
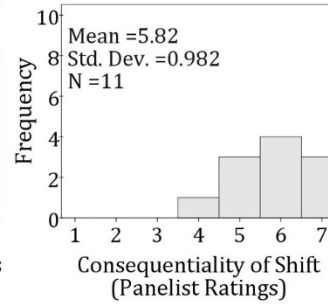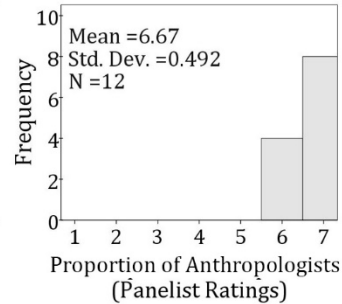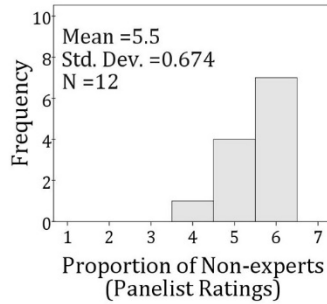
Stable: Yes
(Avg N: 5.67)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 5.2: Stable Anthropology Comparison 2

Many non-experts think that people:

in other cultures do crazy things.

But Anthropologists think that people:

in other cultures do things that make sense in the context of those cultures.

Mean =5.09 Std. Dev. =1.221 N =11	Mean =6.73 Std. Dev. =0.467 N =11	Mean =5.78 Std. Dev. =0.667 N =9
Proportion of Non-experts (Panelist Ratings)	Proportion of Anthropologists (Panelist Ratings)	Consequentiality of Shift (Panelist Ratings)
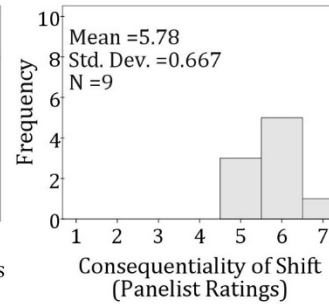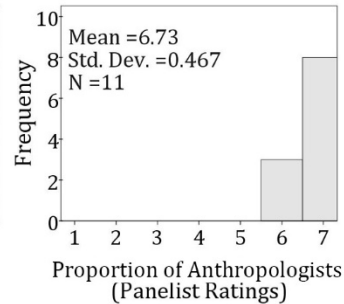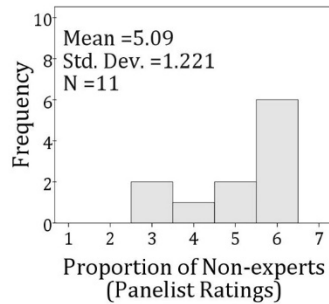
Stable: Yes
(Avg N: 6.67)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 5.3: Stable Anthropology Comparison 3

Many non-experts think that people:

were unsophisticated and primitive in the past.

But Anthropologists think that people:

have been modern in their behaviors and thoughts for at least 30 thousand years, and have always been very intelligent and readily able to cope with a broad range of climactic and social environments.

Mean =5.5
Std. Dev. =1
N =12

Mean =6.75
Std. Dev. =0.452
N =12

Mean =5.17
Std. Dev. =1.115
N =12

Stable: Yes
(Avg N: 6.00)

Proportion of Non-experts
(Panelist Ratings)

Proportion of Anthropologists
(Panelist Ratings)

Consequentiality of Shift
(Panelist Ratings)
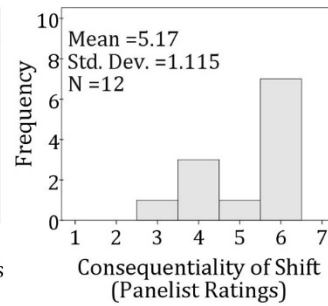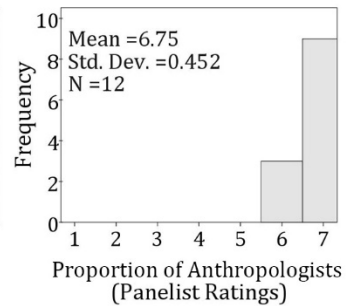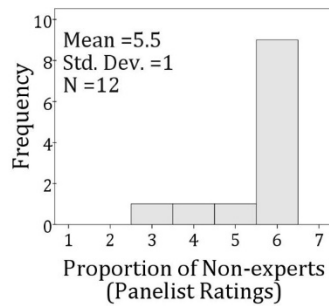
Clarifying Panelist Comments:
Panelists mentioned the importance of mentioning 30,000 years ago, as Anthropologists see it as a time of an important change in human behavior, though Anthropologists disagree over the cause.

Other Issues:
None for this item.

Figure 5.4: Stable Anthropology Comparison 4

Many non-experts think that people:

are completely different across cultures.

But Anthropologists think that people:

share many cultural features (such as language, religion, kinship, marriage, etc), although manifested in different ways across cultures.

Mean =4.85
Std. Dev. =1.334
N =10

Mean =6.7
Std. Dev. =0.483
N =10

Mean =5.5
Std. Dev. =0.972
N =10

Stable: Yes
(Avg N: 6.67)

Proportion of Non-experts
(Panelist Ratings)

Proportion of Anthropologists
(Panelist Ratings)

Consequentiality of Shift
(Panelist Ratings)
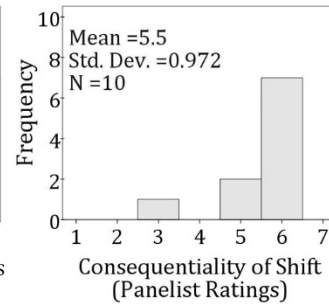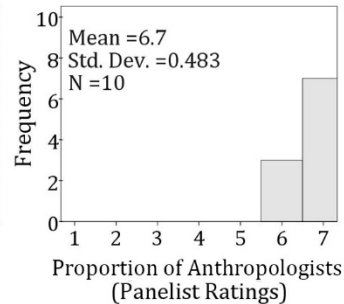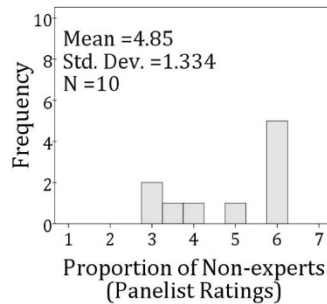
Clarifying Panelist Comments:
None for this item.

Other Issues:
The original formulation of this item was about "cultural traits," but this created a debate over whether "traits" should be understood as a biological term and was removed.

Figure 5.5: Stable Anthropology Comparison 5

Many non-experts think that people:

living in small groups [e.g. in the jungle] speak "primitive languages."

But Anthropologists think that people:

everywhere speak languages that are equally capable of abstract, complex thought - there is no such thing as a primitive human language.

Mean =4.42
Std. Dev. =1.782
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =6.92
Std. Dev. =0.289
N =12

Proportion of Anthropologists
(Panelist Ratings)

Mean =5.36
Std. Dev. =1.206
N =11

Consequentiality of Shift
(Panelist Ratings)
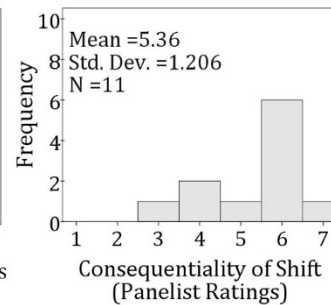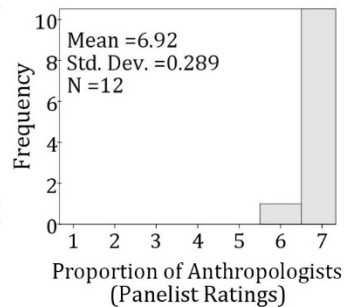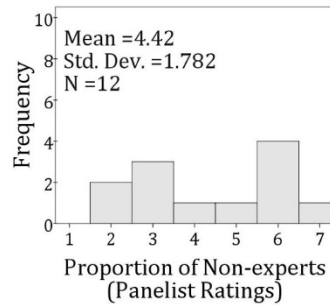
Stable: Yes
(Avg N: 5.67)

Clarifying Panelist Comments:
Panelist mentioned that few people have ever thought about the language of people living in small groups, but that those who do think about it do make that assumption.

Other Issues:
None for this item.

Figure 5.6: Stable Anthropology Comparison 6
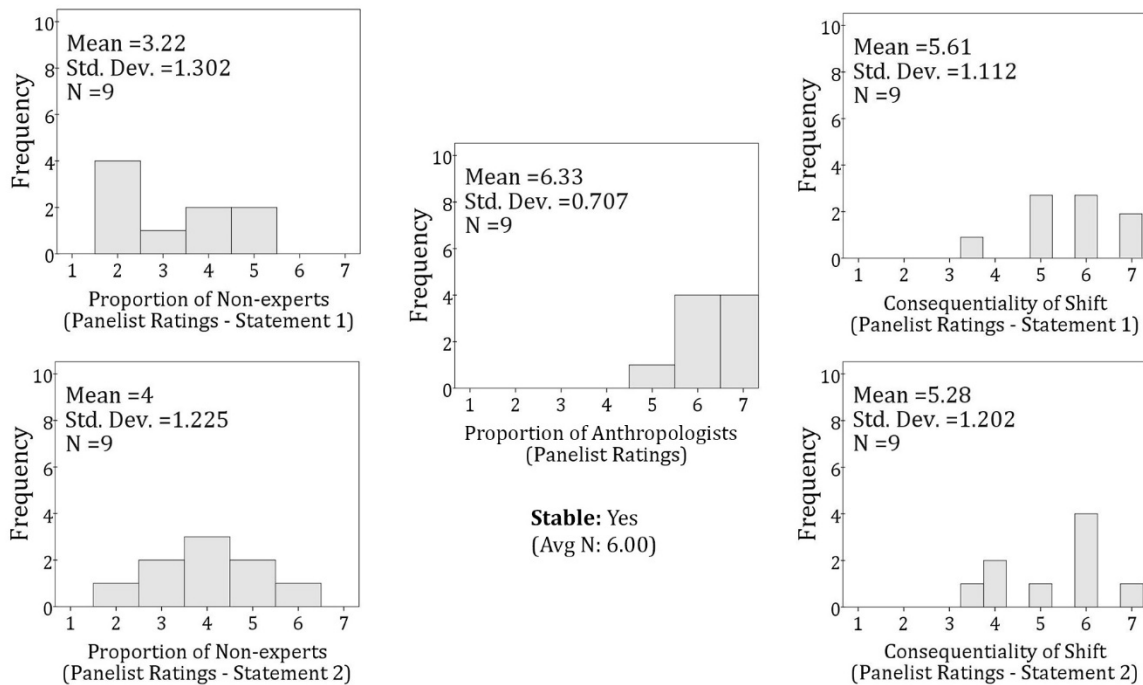
Many non-experts think that people:

behave in ways that can easily be explained.

But Anthropologists think that people:

are complex and that human behavior is not easily explained by any one factor.

Mean =4.33
Std. Dev. =1.67
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =6.67
Std. Dev. =0.651
N =12

Proportion of Anthropologists
(Panelist Ratings)

Mean =5.55
Std. Dev. =0.934
N =11

Consequentiality of Shift
(Panelist Ratings)
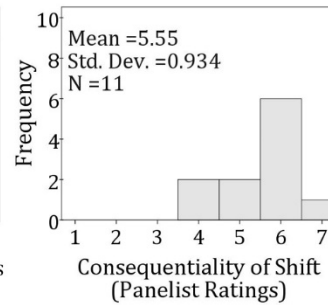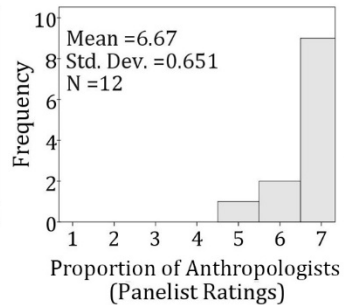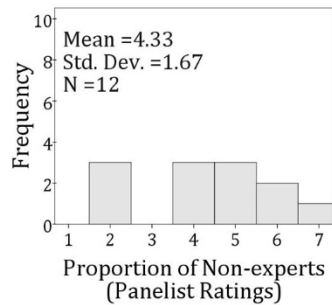
Stable: Yes
(Avg N: 6.33)

Clarifying Panelist Comments:
Panelist commented that the word "easily" made this hard to rate. Another commented that subfields are full of people who think their "factor" explains most ot it.

Other Issues:
None for this item.

Figure 5.7: Stable Anthropology Comparison 7

Many non-experts think that people:

naturally live in one particular family structure (e.g. nuclear families in Western countries).

But Anthropologists think that people:

live in family structures that vary cross-culturally.

Mean =4.95
Std. Dev. =1.117
N =10

Proportion of Non-experts
(Panelist Ratings)

Mean =6.7
Std. Dev. =0.483
N =10

Proportion of Anthropologists
(Panelist Ratings)

Mean =4.5
Std. Dev. =0.535
N =8

Consequentiality of Shift
(Panelist Ratings)
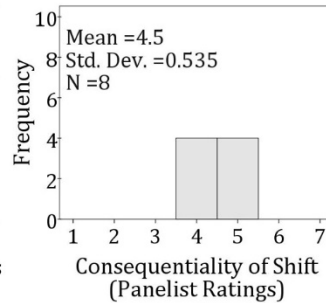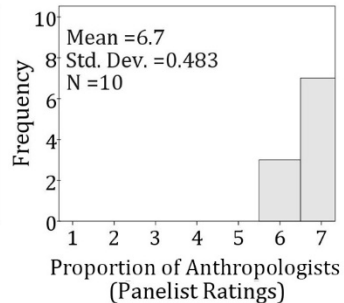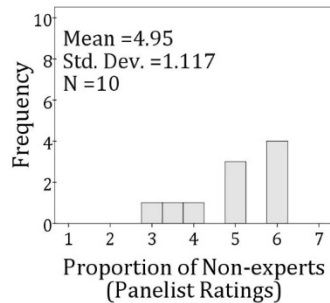
Stable: Yes
(Avg N: 6.33)

Clarifying Panelist Comments:
Panelist commented that there could be "more specificity about what consistitutes a natural family structure."

Other Issues:
The original version included the idea that family structures varied because of how people made a living, but this aspect caused disagreement and was removed.

Figure 5.8: Stable Anthropology Comparison 8

Many non-experts think that people:

make fully-informed decisions, in which they can predict the consequences of their actions, based on reason and logic.

But Anthropologists think that people:

make decisions based on partial and selective information, in which predictive ability is limited, based on emotion and cultural "rules of thumb" that they have learned.

Mean =4.56
Std. Dev. =0.882
N =9

Proportion of Non-experts
(Panelist Ratings)

Mean =5.67
Std. Dev. =0.866
N =9

Proportion of Anthropologists
(Panelist Ratings)

Mean =5.11
Std. Dev. =1.269
N =9

Consequentiality of Shift
(Panelist Ratings)
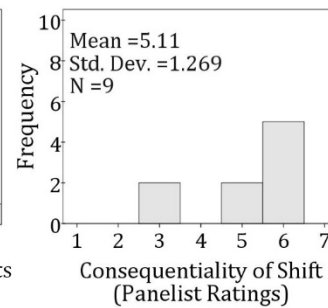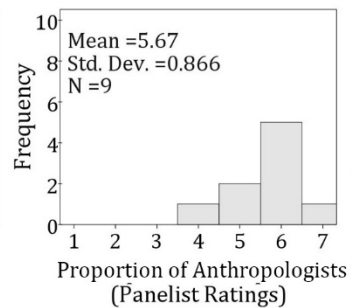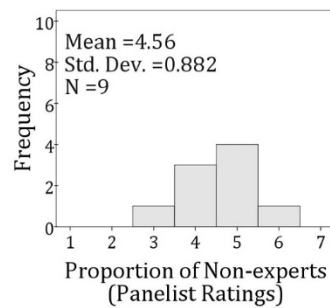
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
One panelist disagreed, arguing: "I think most people, including non-experts and anthropologists, understand that people 'do the best they can' when they make decisions, but rarely have full information or the ability to predict the consequences." Another panelist argued that they think the opposite, as research makes human decision-making more understandable.

Other Issues:
None for this item.

Figure 5.9: Stable Anthropology Comparison 9

Many non-experts think that people:

are fully conscious of their attitudes and behaviors, including biases, and know why they do what they do.

But Anthropologists think that people:

behave unconsciously and habitually, on the basis of acquired cultural patterns of thought and behavior.

Mean =4.36
Std. Dev. =1.502
N =11

Proportion of Non-experts
(Panelist Ratings)

Mean =5.18
Std. Dev. =1.401
N =11

Proportion of Anthropologists
(Panelist Ratings)

Mean =5.22
Std. Dev. =1.302
N =9

Consequentiality of Shift
(Panelist Ratings)
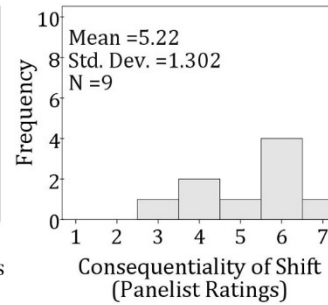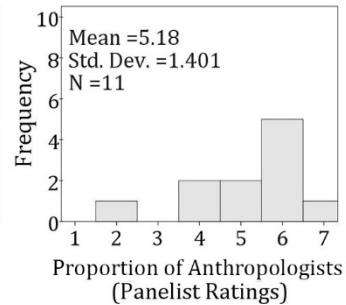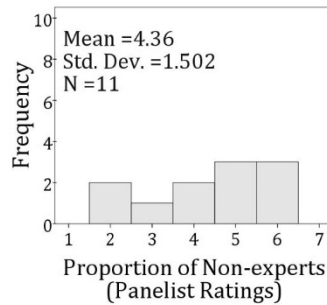
Stable: Yes
(Avg N: 6.33)

Clarifying Panelist Comments:
This item caused considerable debate for both statements. For the non-expert statement, one panelist commented that "few people believe that they or others are fully conscious of their attitudes and behaviors." For the expert statement, one panelist commented that it was environmentally deterministic and while "much of human behavior is result of cultural processes, there is much (unconscious) human behavior that is shaped by biological processes." Another commented that this ignored the concept of agency and suggests "we are culturally programmed and can't challenge learned ideas." In contrast, another panelist commented that this "is one of the fundamental things I teach in my undergraduate anthropology classes, and I have witnessed the effect of this change first hand on numerous occasions."

Other Issues:
None for this item.

Figure 5.10: Stable Anthropology Comparison 10

Many non-experts think that people:

are hard-wired from birth on how to behave.

While other non-experts think that people:

are separated from their biology, and learn all of their behaviors.

But Anthropologists think that people:

behave in ways that are reflections of complex interactions between culture and biology.

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 5.11: Stable Anthropology Triple Comparison 1

126

Many non-experts think that people:

are born as "blank slates" and shaped entirely by culture.

While other non-experts think that people:

are naturally and innately selfish, greedy, caring, altruisitic, etc. across time and geography.

But Anthropologists think that people:

have some innate and natural characteristics, but are also shaped by experience and historical, cultural, and social processes (though Anthropologists vary on the specifics and the relative importance of these factors).
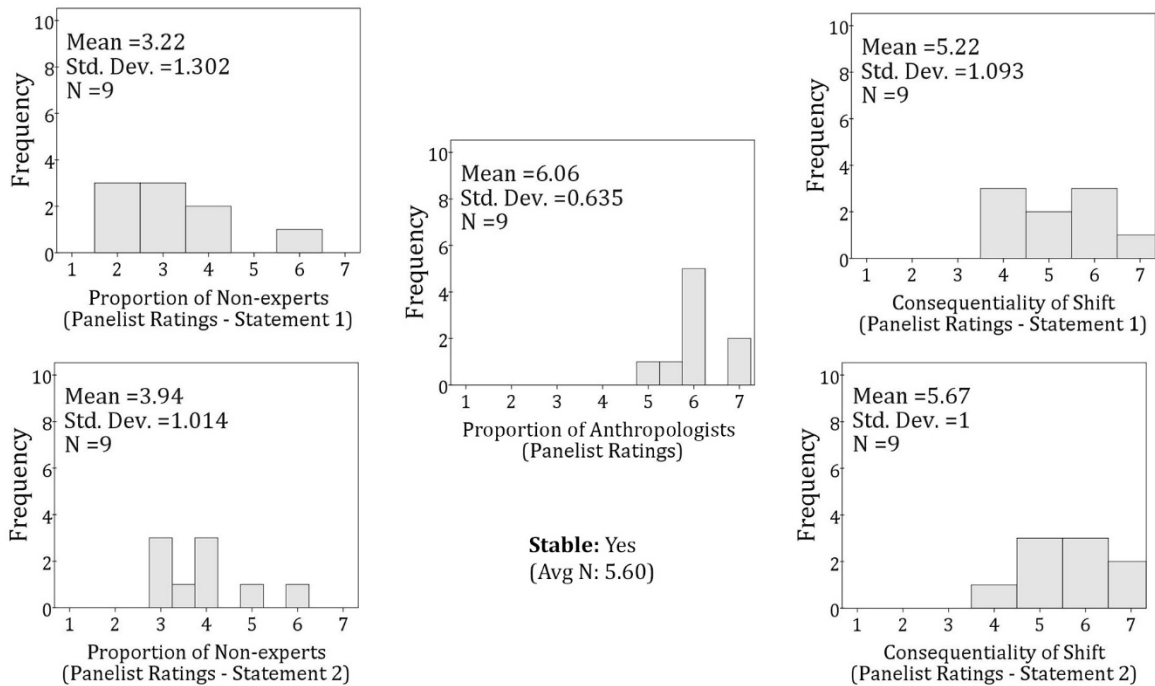
Clarifying Panelist Comments:
Panelist commented that "these statements are oversimplified and anachronistic, and most non-experts would recognize the two statements as a false dichotomy." Another panelist argued that the comparison ending with "behave in ways that are reflections of complex interactions between culture and biology" was a better worded version of the same idea.

Other Issues:
None for this item.

Figure 5.12: Stable Anthropology Triple Comparison 2

Many non-experts think that people:

are innately good, peaceful, and unbiased, until they learn (through experience and culture) to become xenophobic, racist, ethnocentric, or biased.

While other non-experts think that people:

are innately selfish and prone to violence, though this can be restrained.

But Anthropologists think that people:

are fundamentally driven to identify themselves as a member of one group and not others, which underlies our capacity for both conflict and cooperation, but which can also be reshaped through experience, culture, education, and public sentiment.

Stable: Yes
(Avg N: 5.80)
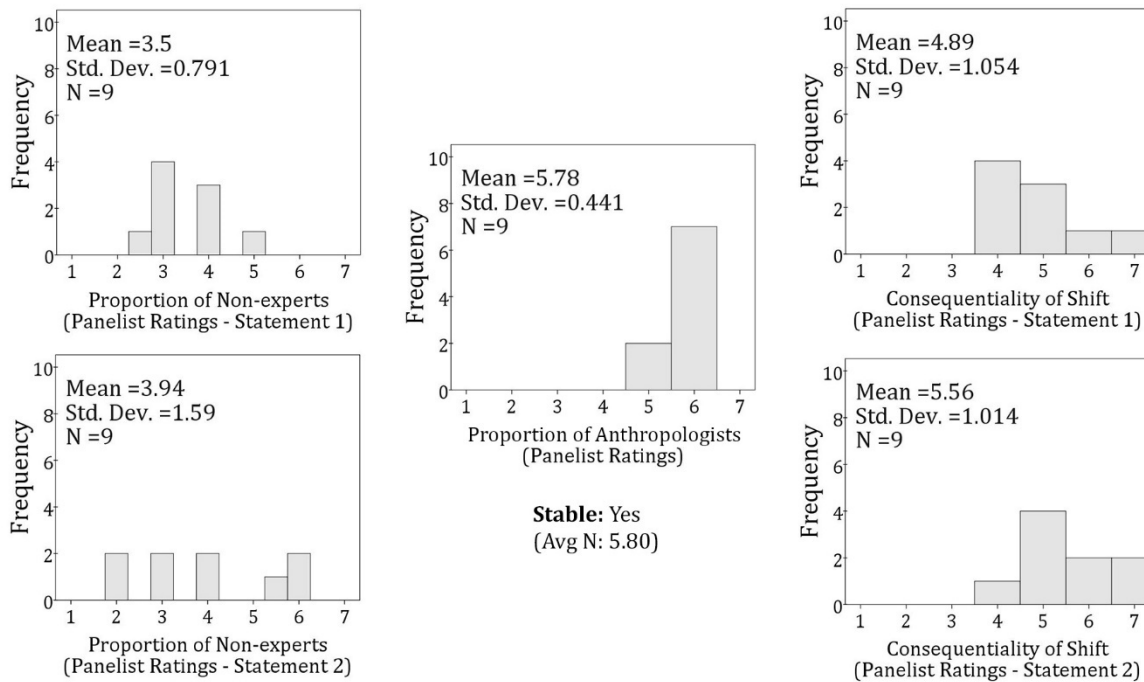
Clarifying Panelist Comments:
None for this item.

Other Issues:
This item was created by combining several earlier comparisons together. On one of those earlier version, a panelist commented on the existence of two competing non-expert perspectives: "There are two popular myths: the noble savage and, its opposite, the ignoble savage. The non-expert statement reflects the noble savage myth, but in my experience the ignoble savage myth is equally as common."

Figure 5.13: Stable Anthropology Triple Comparison 3

Many non-experts think that people:

have an innate morality that is self-evident.

While other non-experts think that people:

acquire morality from their culture, and there is tremendous variation between cultures in morality.

But Anthropologists think that people:

share basic moral drives, such as altruism, but the specific rules of behavior are shaped by culture and vary cross-culturally and situationally
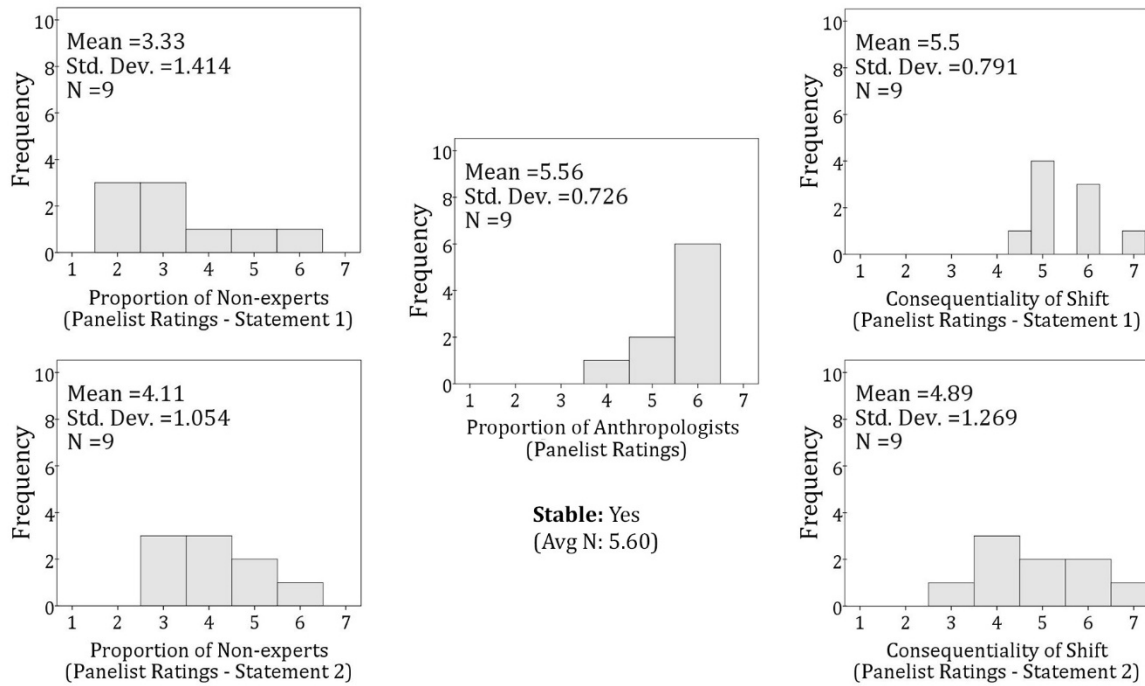
Clarifying Panelist Comments:
Panelist commented that these were compatible ideas and that most people would agree with both if asked. They also commented that the expert statement simply rephrased the ideas as compatible.

Other Issues:
None for this item.

Figure 5.14: Stable Anthropology Triple Comparison 4

129

Many non-experts think that people:

are born either male or female.

But Anthropologists think that people:

have sex and gender that are shaped by both biology and culture.

Chart 1 (left): X-axis "Proportion of Non-experts (Panelist Ratings)"; Y-axis "Frequency"; Mean =5.67, Std. Dev. =1, N =9

Chart 2 (middle): X-axis "Proportion of Anthropologists (Panelist Ratings)"; Y-axis "Frequency"; Mean =6.5, Std. Dev. =0.612, N =9

Chart 3 (right): X-axis "Consequentiality of Shift (Panelist Ratings)"; Y-axis "Frequency"; Mean =5.75, Std. Dev. =0.886, N =8
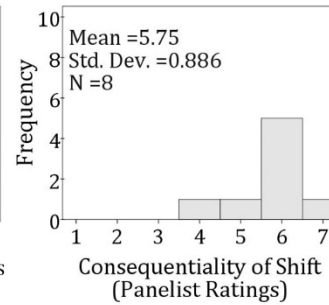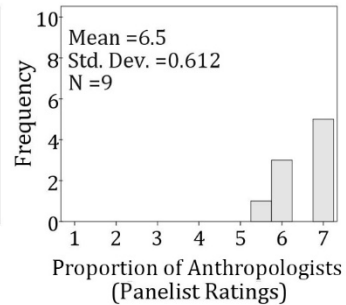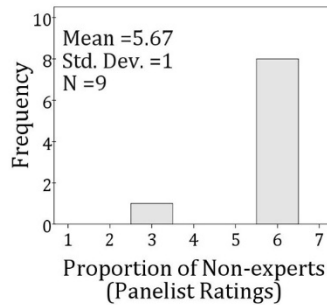
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
Panelists wanted it clarified that non-experts tend to think that sex is rooted in biology but fewer think gender is.

Other Issues:
This item caused debate on both statements. Some panelists argued that non-experts think sex and/or gender is entirely biological. Others argued that non-experts think sex and/or gender is entirely determined by culture. As such, this comparison would have made a good triple comparison, however, according to panelists, Anthropologists themselves disagree over whether sex and/or gender differences are partially innate (compare to 5.2, which was uncontroversial for Anthropologists).

Figure 5.15: In-Progress Anthropology Comparison 1

Many non-experts think that people:

from other cultures are fundamentally different, and
that some cultures are more "advanced" than others.

But Anthropologists think that people:

are biologically similar, but culturally different with no
culture considered more advanced than another.

Mean =5.11
Std. Dev. =0.928
N =9

Proportion of Non-experts
(Panelist Ratings)

Mean =6.39
Std. Dev. =0.601
N =9

Proportion of Anthropologists
(Panelist Ratings)

Mean =5.94
Std. Dev. =0.678
N =8

Consequentiality of Shift
(Panelist Ratings)
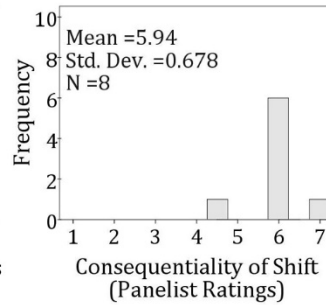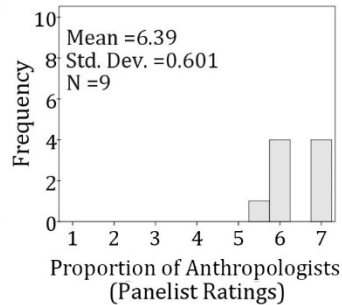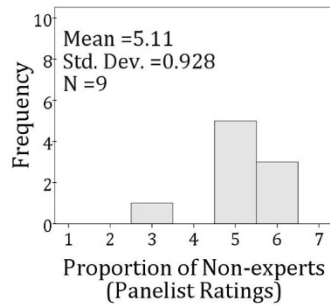
Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
Panelist comments: panelist wanted the word "advanced" clarified, to disambiguate more advanced technological
development from the idea of a culture being "better," as different cultures have "adapted according to its environment
and needs, and access to technologies."

Other Issues:
None for this item.

Figure 5.16: In-Progress Anthropology Comparison 2

Many non-experts think that people:

are separate from the natural world, especially in modern times.

But Anthropologists think that people:

are always engaged in a dynamic coupling with natural systems, despite considerable technological achievements.

Mean = 5.18
Std. Dev. = 1.471
N = 11

Proportion of Non-experts
(Panelist Ratings)

Mean = 6.27
Std. Dev. = 0.786
N = 11

Proportion of Anthropologists
(Panelist Ratings)

Mean = 5.9
Std. Dev. = 1.197
N = 10

Consequentiality of Shift
(Panelist Ratings)
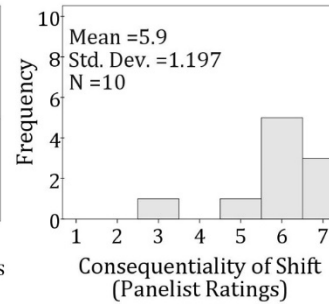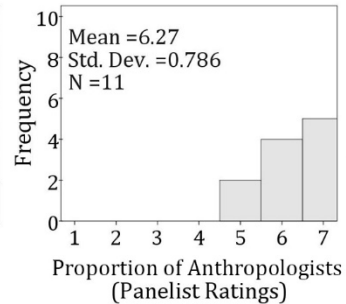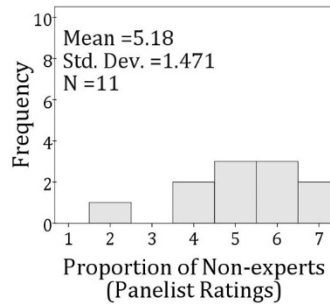
Stable:
Partial (see note below)

Clarifying Panelist Comments:
Panelists commented the numbers who do not think this way would go way up if we included non-Western people. Another panelist commented that most new Anthropologists think this, but some "functionalists" may not.

Other Issues:
This item was erroneously considered stable and not tested further. Both the non-expert and Anthropologist statements were stable (n = 5), however, there was insufficient n (n = 4) to determine if the consequentiality rating was stable.

Figure 5.17: In-Progress Anthropology Comparison 3

132

Many non-experts think that people:

have religious beliefs and religious practices based on Truth given by a supreme supernatural being.

But Anthropologists think that people:

create religion and religious practices.

Mean =4.8
Std. Dev. =1.135
N =10

Proportion of Non-experts
(Panelist Ratings)

Mean =6.55
Std. Dev. =0.599
N =10

Proportion of Anthropologists
(Panelist Ratings)

Mean =5.88
Std. Dev. =0.835
N =8

Consequentiality of Shift
(Panelist Ratings)
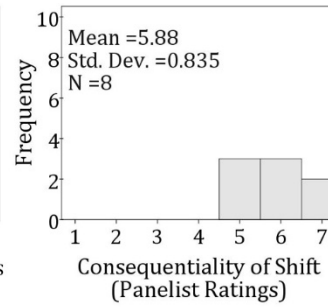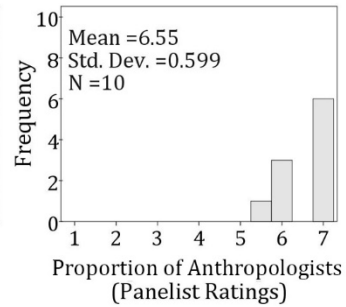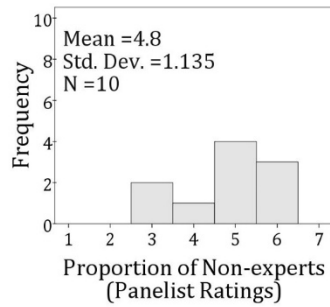
Stable:
Insufficient N to determine.

Clarifying Panelist Comments:
One panelist wanted this amended to "learn and create religious practice".

Other Issues:
None for this item.

Figure 5.18: In-Progress Anthropology Comparison 4

CHAPTER 6

Results of the Economics Panel

The Economics panel, from the perspective of the Delphi Method, was an unsuccessful panel. Due to low participation across rounds, no comparisons reached statistical stability for the Economics panel. However, panelists did suggest some interesting ways that Economists think about the human world differently than people who have never studied Economics.

The Economist panel had a fair amount of internal disagreement. Several suggested comparisons were strongly disliked by panelists, who commented that the proposed non-expert and expert statements should be switched. Such items included whether people were mainly influenced by prices or other factors (habits, customs, and social conditioning) when making decisions and whether people think before making decisions or rely on instincts. It is possible that these debates are based on the division within the field between Neoclassical and Behavioral Economists, but we cannot be certain of this.

The Economist panel was also unusual in that nearly all of the comparisons suggested by panelists were eventually dropped from the study due to low ratings. Approximately half were dropped due to the representativeness of the non-expert statement, the other half were dropped due to the comparison being considered inconsequential, with only two comparisons dropped in part because they were unrepresentative of Economists. Even though many items were dropped as being unrepresentative of non-experts, Economists did not generally propose an alternate non-

expert view. Likewise, though they rated shifts as inconsequential, they did not explain why they thought this.

Many of the proposed comparisons for this panel did not fit within the study format. Some proposed comparisons contrasted non-experts and experts directly, rather than their different theories about people. Some proposed comparisons were not really about people at all, but about concepts. And some proposed comparisons were not about how people do think, but how they should think. Sometimes, proposed comparisons had multiple issues. For example, one original suggested comparison was "Many non-experts think that people: room and board are a cost of attending university. But Economists think that people: room and board are not costs because those costs would be incurred anyway." Our attempts to get panelists to explain what this told us about competing theories of people often failed, though panelists did explain why these statements were correct (in this example, because you are really only paying the difference, if any, between what you would pay at university at what you would pay elsewhere rather than paying the full cost of university housing in a vacuum).

The Economics panel also touched on several concepts not discussed by any other panel, though these were usually dropped due to low ratings. For example, one comparison (dropped due to a consequentiality rating of 3.9) was "Many non-experts think people: are fundamentally unwilling to attach a dollar cost to a human life. But Economists think that people: implicitly attach a dollar cost to human life via the decisions and risks that they take." Another example (dropped due to a consequentiality rating of 3.0) was: "Many non-experts think that people: derive happiness from money. But Economists think that people:

derive happiness from the freedom that money represents." Another example (dropped due to a non-expert rating of 3.5) was: "Many non-experts think that people: pay limited attention to economic incentives. But Economists think that people: are highly influenced by economic incentives."

As these items were dropped from the panel, we cannot say that they represent how Economists think about the human world in contrast to how non-experts do. However, these comparisons do seem to get at important ideas about people that are not widely discussed outside of the field and which may have real-world consequences. For example, the degree to which people are motivated by economic incentives profoundly affects what kinds of policies would be effective at changing behaviors. In the end, it is not clear whether Economists have fewer ways they differ from non-experts than other fields, or whether determining what those ways are would simply require more sensitive methods than this study employed.

Many non-experts think that people:

in one nation only gain at the expense of people in other nations when free trade is introduced.

But Economists think that people:

in both countries could benefit from free trade.
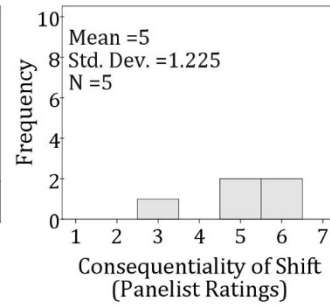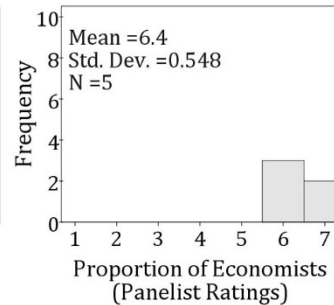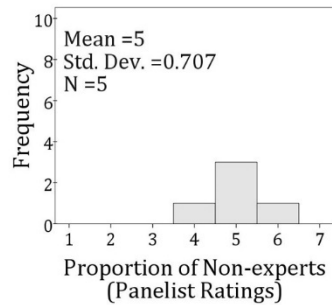
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
Panelists explained that specialization "allows people to consume more for the same cost."

Other Issues:
We asked panelists to explain what this told us about people generally (rather than trade). Panelists did not directly answer this, but one panelist commented that "the costs [of free trade] are concentrated and thus easier to observe," whereas "the benefits are diffuse and harder to observe."

Figure 6.1: In-Progress Economics Comparison 1

Many non-experts think that people:

often make decisions that are not in their own best interests.

But Economists think that people:

almost always act in their own best interests, based on the information and resources they have at the time of the decision.

Left histogram: Mean =4.4, Std. Dev. =0.548, N =5. X-axis: Proportion of Non-experts (Panelist Ratings)

Middle histogram: Mean =5.6, Std. Dev. =0.548, N =5. X-axis: Proportion of Economists (Panelist Ratings)

Right histogram: Mean =4.2, Std. Dev. =0.837, N =5. X-axis: Consequentiality of Shift (Panelist Ratings)
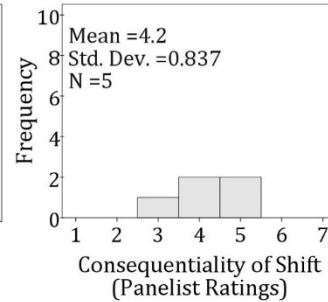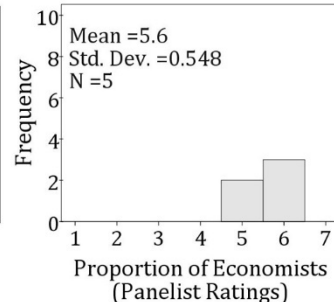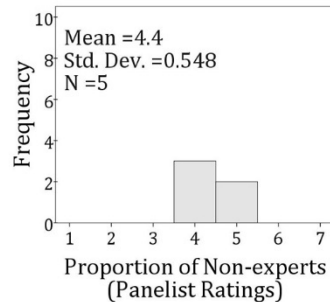
Stable: Insufficient N to determine

Clarifying Panelist Comments:
One panelist elaborated on this comparison, commenting that "people do make decisions in their own best interests as they see them but they may view them such that the decisions don't appear that way." In contrast, "Economists are more careful about framing the parameters of said decisions." Lastly, they commented that this change was important because "Careful framing makes choices and decisions more readily apparent to the social scientist studying the situation."

Other Issues:
None for this item.

Figure 6.2: In-Progress Economics Comparison 2

CHAPTER 7

Results of the Geography Panel

The Geography panel, from the perspective of the Delphi Method, was an unsuccessful panel. The Geography panel had, by far, the lowest participation of any panel. This was largely due to problems with our limited ability to recruit panelists but was also partially due to the nature of the field. Geographers tend to do research that focuses on one part of the globe, which may be far distant from their institution. As such, they, more than professors in many other fields, are likely in the field rather than in their offices.

This low participation meant that no items could possibly succeed, as it was never possible to test for stability across rounds. Secondly, this means that the estimates provided by panelists are based on very few respondents and are consequently less likely to represent the views of the field as a whole compared to the other panels (for example, while only three Geographers participated in the final round, ten Psychologists did). Despite this, panelists suggested several interesting ways that Geographers think about the human world differently than non-experts.

The Geography panel also has one comparison that has three components. This is because, as in the Anthropology panel, Geographers proposed opposing ways that non-experts thought about a concept that Geographers broadly agreed on.

Many non-experts think that people:

's behaviors are shaped by their identities (race, ethnicity, gender, age).

But Geographers think that people:

's behaviors are not only shaped by their identities, but also by space and place. In other words, peoples' environments play a significant role in shaping behaviors and identities.

Mean =5.67
Std. Dev. =0.577
N =3

Proportion of Non-experts (Panelist Ratings)

Mean =6.33
Std. Dev. =0.577
N =3

Proportion of Geographers (Panelist Ratings)

Mean =4.67
Std. Dev. =1.155
N =3

Consequentiality of Shift (Panelist Ratings)
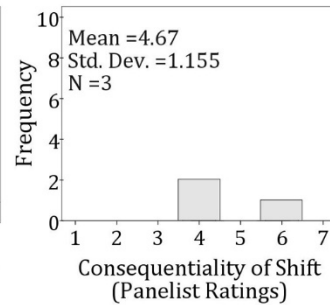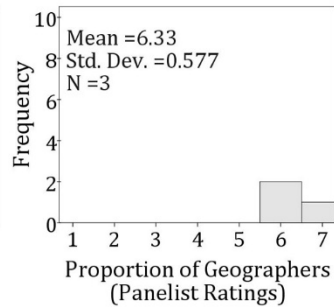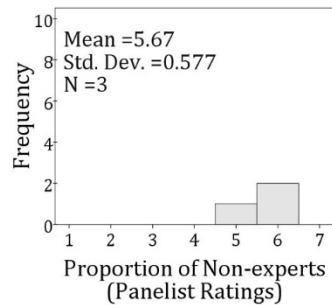
Stable: Insufficient N to determine

Clarifying Panelist Comments:
One panelist commented that "many identities are space and place dependent." Another commented that the Geographer's statement "is a bit close to environmental determinism for my taste. If you replaced this with 'influenced by', I think the responses you get would be a bit more accurate."

Other Issues:
None for this item.

Figure 7.1: In-Progress Geography Comparison 1

Many non-experts think that people:

's consumption decisions don't matter.

But Geographers think that people:

's consumption decisions have significant social, environmental and political impacts around the world.
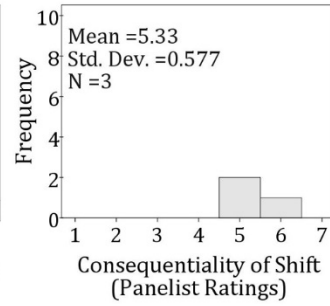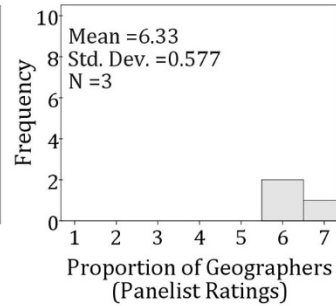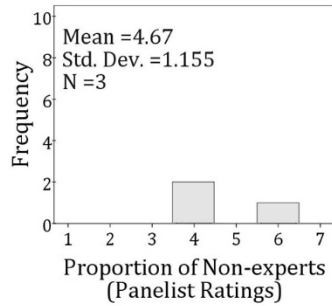
Mean =4.67
Std. Dev. =1.155
N =3

Proportion of Non-experts
(Panelist Ratings)

Mean =6.33
Std. Dev. =0.577
N =3

Proportion of Geographers
(Panelist Ratings)

Mean =5.33
Std. Dev. =0.577
N =3

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

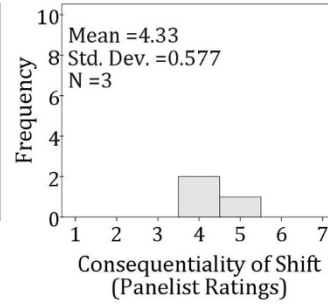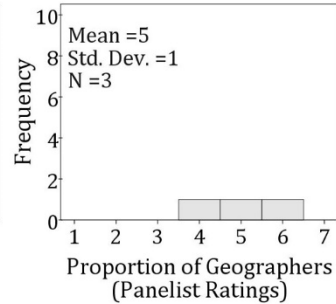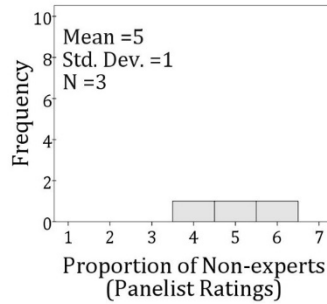*Figure 7.2:* In-Progress Geography Comparison 2

Many non-experts think that people:

are separate from the natural environments.

But Geographers think that people:

are deeply connected to and reliant upon the natural environment through their resource use and consumption habits.

Mean =4.67
Std. Dev. =1.155
N =3

Proportion of Non-experts
(Panelist Ratings)

Mean =6.33
Std. Dev. =0.577
N =3

Proportion of Geographers
(Panelist Ratings)

Mean =5
Std. Dev. =1
N =3

Consequentiality of Shift
(Panelist Ratings)
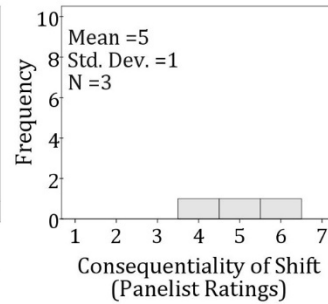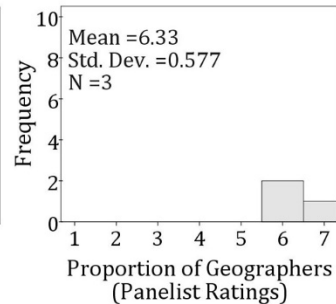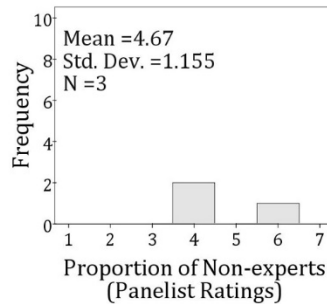
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
One panelist commented that "non-experts in urban settings are likely to have very different answers from non-experts in rural settings. "

Other Issues:
None for this item.

Figure 7.3: In-Progress Geography Comparison 3

Many non-experts think that people:

belong to different human races.

But Geographers think that people:

are genetically similar and therefore the concept of race is invalid.

Proportion of Non-experts (Panelist Ratings) — Mean =6, Std. Dev. =0, N =3

Proportion of Geographers (Panelist Ratings) — Mean =5, Std. Dev. =1.732, N =3

Consequentiality of Shift (Panelist Ratings) — Mean =5, Std. Dev. =1.732, N =3
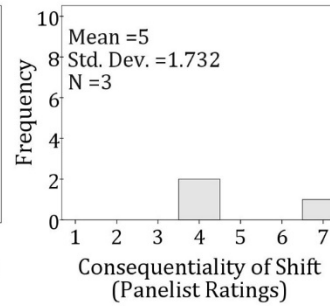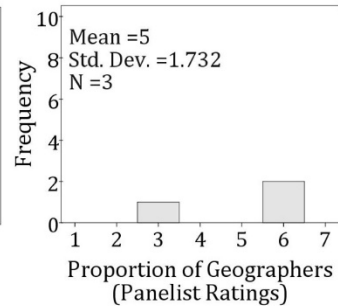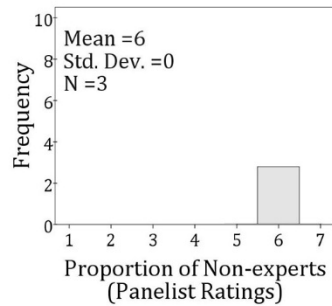
Stable: Insufficient N to determine

Clarifying Panelist Comments:
One panelist commented "I don't agree that the concept of race is invalid. It may biologically invalid but it operates as a powerful social category that has profound impacts on society."

Other Issues:
None for this item.

Figure 7.4: In-Progress Geography Comparison 4

Many non-experts think that people:

respond rationally to environmental threats and hazards (like flood risks).

But Geographers think that people:

often respond in seemingly irrational ways to environmental threats and hazards due to imperfect knowledge, conflicting goals, or larger structural constraints.

Mean =5
Std. Dev. =0
N =3

Proportion of Non-experts (Panelist Ratings)

Mean =6
Std. Dev. =0
N =3

Proportion of Geographers (Panelist Ratings)

Mean =4.33
Std. Dev. =0.577
N =3

Consequentiality of Shift (Panelist Ratings)
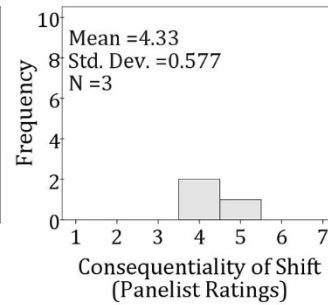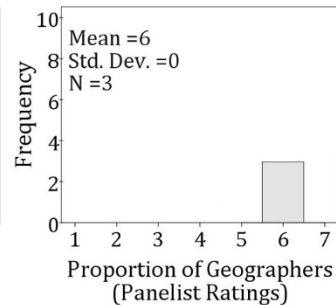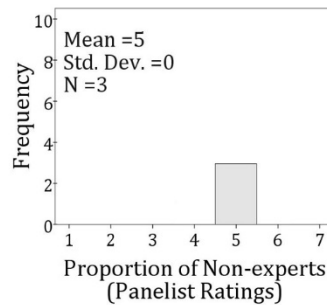
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
"Rationality" caused some disagreement here. Some panelists emphasized the role of knowledge in responding "rationally." Others argued that one could respond rationally even with poor information, and "Information may have an effect in how good the response is not in how rational it is." In addition, one panelist commented that this was old (potentially implying outdated).

Other Issues:
None for this item.

Figure 7.5: In-Progress Geography Comparison 5

Many non-experts think that people:

act as individuals.

But Geographers think that people:
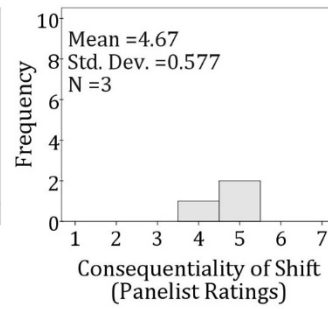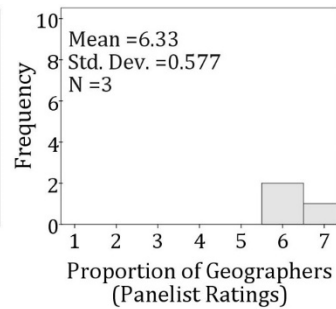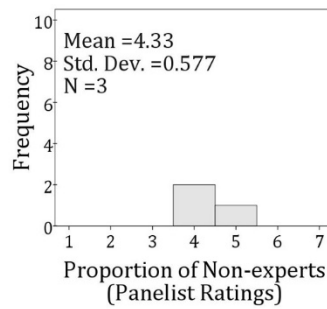
operate within complex and interconnected systems.

Mean =4.33
Std. Dev. =0.577
N =3

Proportion of Non-experts
(Panelist Ratings)

Mean =6.33
Std. Dev. =0.577
N =3

Proportion of Geographers
(Panelist Ratings)

Mean =4.67
Std. Dev. =0.577
N =3

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 7.6: In-Progress Geography Comparison 6

145

Many non-experts think that people:

change their opinion based on being presented with factual information.

But Geographers think that people:

have opinions based on a wide range of factors which may or may not have to do with factual information.

Mean =5
Std. Dev. =1
N =3

Frequency

Proportion of Non-experts
(Panelist Ratings)

Mean =5
Std. Dev. =1
N =3

Frequency

Proportion of Geographers
(Panelist Ratings)

Mean =4.33
Std. Dev. =0.577
N =3

Frequency

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
Panelists commented that these were not parallel, but did not suggest ways to make them so. One panelist commented "there is a lot of talk about this right now, so things change." Another mentioned that "no[t] everyone acts according to how they think." One panelist commented: "facts? factual? factors? [A]ll loaded terms."

Other Issues:
None for this item.

Figure 7.7: In-Progress Geography Comparison 7

Many non-experts think that people:

's behaviors aren't influenced by place.

While other non-experts think that people:

are controlled by their (physical or social) environment.

But Anthropologists think that people:

, while place matters and shapes their lives and experiences in profound ways, are influenced/limited, but not controlled, by their physical and social environment.

Stable:
Insufficient N to determine
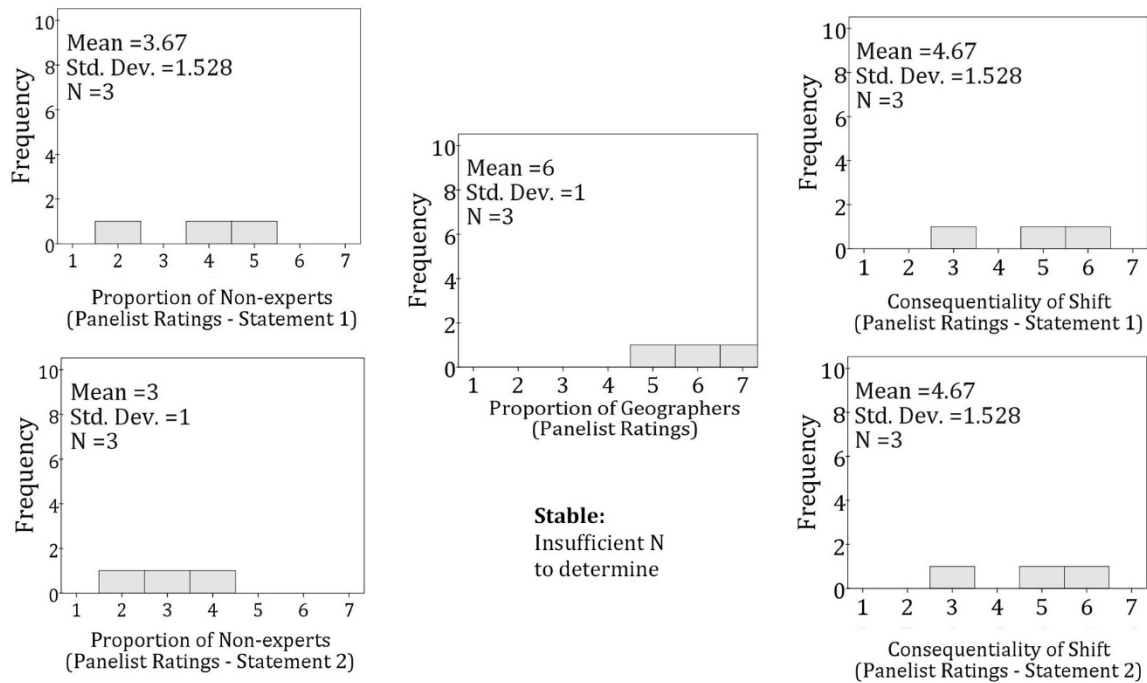
Clarifying Panelist Comments:
One panelist commented "The term controlled can be viewed as having a variety of meanings, esp. related to degree of control."

Other Issues:
None for this item.

Figure 7.8: In-Progress Geography Triple Comparison 1

147

CHAPTER 8

Results of the History Panel

The History panel was moderately successful, producing a handful of comparisons that were both well-rated and stable. After Round 3, participation was too low to test stability. Thus, all items that succeeded did so relatively early in the process.

One problem unique to the Historian panel was that several items were phrased only in terms of "people in the past" rather than being theories about people that would also apply to the past. We expected that many items would involve people in the past, but in order to fit into the study, they would need to contrast general understandings of people. For example, if non-experts thought people in the past were different from people today in some manner but Historians thought they were not (or vice versa), this would be fine. But if both statements are about people in the past, it is trickier to say what this tells us about people generally as opposed to telling us what people know about History. Several suggested comparisons were also based on divergent understandings of a single historical context, though panelists were eventually able to generalize the underlying ideas.

The ratings for the Historian panel were noticeably odder than in other panels. Like other panels, most of the comparisons dropped over the course of the project were only marginally below the threshold, however, several comparisons received extremely low scores (i.e. averages of two out of seven or lower) for how representative they were of Historians. Such ratings are surprising, given that these comparisons were suggested by Historians as ways of thinking important to Historians. Similarly, the distribution of ratings for the panel seemed larger than for other panels. Taken together, these imply that

148

Historians may agree less or be less aware how other Historians (or non-experts)

understand the human world compared to other fields.

Many non-experts think that people:

are divided into races.

But Historians think that people:

are very uniform biologically, and that visible "racial" differences have no intrinsic significance.
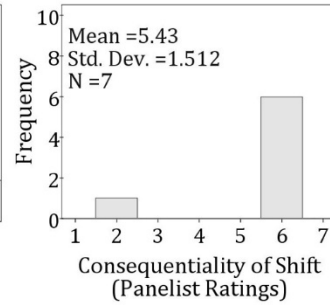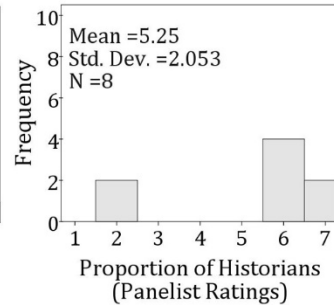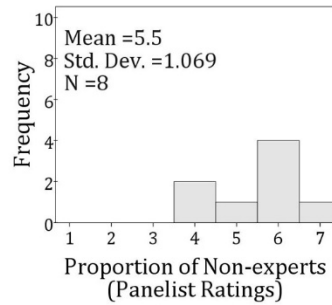
Stable: Yes
(Avg N: 5.00)

Clarifying Panelist Comments:
One panelist disagreed, commenting that "[t]he propositions confronted here are not in conflict. They talk about essentially different subjects." Another panelist commented that "Historians continue to use "race" as a necessary category of understanding. They know the biology, but have not yet embraced the implications of the idea that there is only one human race. The historical record does not reflect the biological insight."

Other Issues:
None for this item.

Figure 8.1: Stable History Comparison 1

Many non-experts think that people:

mistrust people from different racial/ethnic/religious traditions than their own because it is "human nature" to do so.

But Historians think that people:

acquire and express such mistrust in specific historical situations and for reasons that vary with historical circumstances.

Mean =4.13
Std. Dev. =1.246
N =8

Frequency

Proportion of Non-experts
(Panelist Ratings)

Mean =6
Std. Dev. =0.756
N =8

Frequency

Proportion of Historians
(Panelist Ratings)

Mean =5.88
Std. Dev. =0.641
N =8

Frequency

Proportion of Historians
(Panelist Ratings)
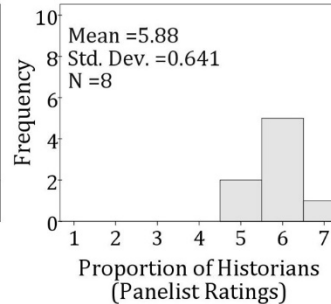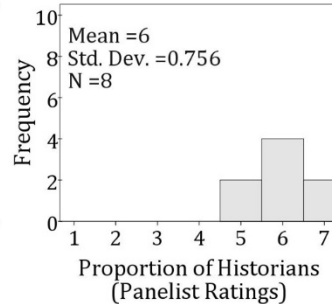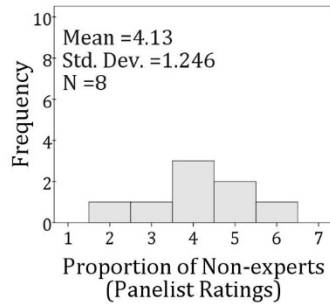
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
Panelist commented that the first statement was too restrictive, and that non-experts "tend to think this for reasons other than 'human nature.'"

Other Issues:
None for this item.

Figure 8.2: Stable History Comparison 2

Many non-experts think that people:

are either "good" or "evil."

But Historians think that people:

cannot be easily summarized and are complex.

Proportion of Non-experts (Panelist Ratings) — Mean =4.88, Std. Dev. =1.458, N =8

Proportion of Historians (Panelist Ratings) — Mean =5.63, Std. Dev. =1.598, N =8

Consequentiality of Shift (Panelist Ratings) — Mean =5.43, Std. Dev. =0.787, N =7
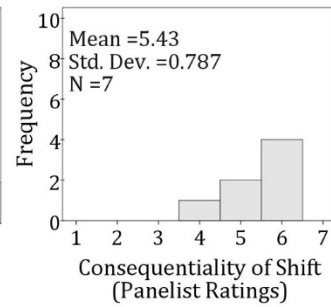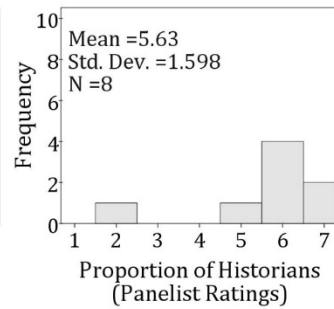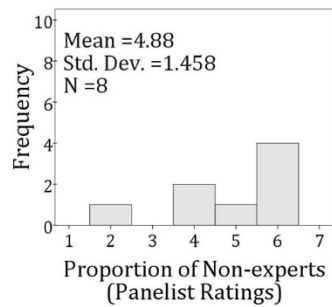
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
One panelist wanted the Historian statement to be more specific. Another panelist commented "It seems to me that what historians do in general is make things more complicated!"

Other Issues:
None for this item.

Figure 8.3: Stable History Comparison 3

Many non-experts think that people:

are always the same in their basic perceptions and motivations (i.e. human nature never changes).

But Historians think that people:

see the world differently and want different things from it in different cultures and periods (i.e. human nature is not always the same).

Mean =5.75
Std. Dev. =1.258
N =4

Frequency
Proportion of Non-experts
(Panelist Ratings)

Mean =5
Std. Dev. =0.816
N =4

Frequency
Proportion of Historians
(Panelist Ratings)

Mean =5.25
Std. Dev. =0.5
N =4

Frequency
Consequentiality of Shift
(Panelist Ratings)
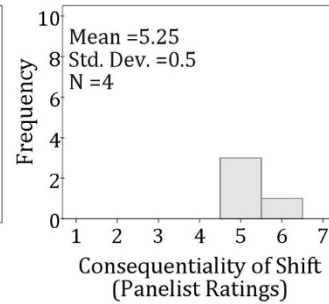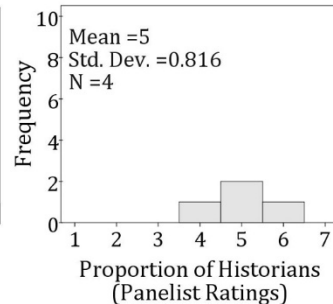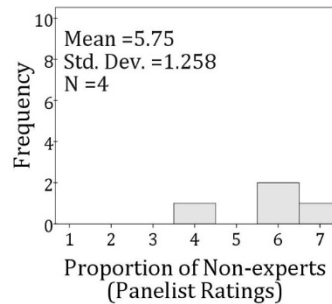
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
This item, despite getting strong ratings across rounds, caused a lot of debate. Several panelists disagreed with the expert statement, arguing "There do seem to be some traits that cross cultures." Another commented " I think there's a reason why human nature is called human nature. If there is nothing universal about being human, then we should discard the term." A third commented "I think I would say that human nature may stay the same -- we're all driven by Maslow's hierarchy of needs – but there is no constant in how we think those needs are best met. And different cultures prioritize different needs, so the relative importance of various desires and needs shifts." In addition, two panelists wanted the Historian statement to be made more specific (in terms of cultures and periods).

Other Issues:
None for this item.

Figure 8.4: In-Progress History Comparison 1

Many non-experts think that people:

have always thought (about the world) the way we do.

But Historians think that people:

think about the world in varied ways across time and place, shaped by their values, politics, scientific knowledge, and their historical context.
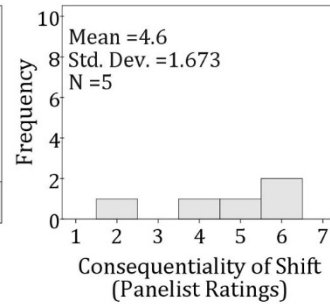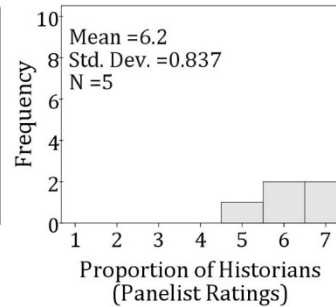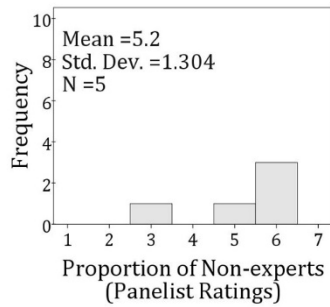
Mean =5.2
Std. Dev. =1.304
N =5

Frequency

Proportion of Non-experts
(Panelist Ratings)

Mean =6.2
Std. Dev. =0.837
N =5

Frequency

Proportion of Historians
(Panelist Ratings)

Mean =4.6
Std. Dev. =1.673
N =5

Frequency

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 8.5: In-Progress History Comparison 2

154

Many non-experts think that people:

who believe things we think of as superstitious are ignorant or stupid.

But Historians think that people:

who believe things we think of as superstitious have those beliefs as part of coherent ways of understanding the world.

Mean =4.6
Std. Dev. =1.817
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =5.8
Std. Dev. =0.447
N =5

Proportion of Historians
(Panelist Ratings)

Mean =5
Std. Dev. =0.707
N =5

Consequentiality of Shift
(Panelist Ratings)
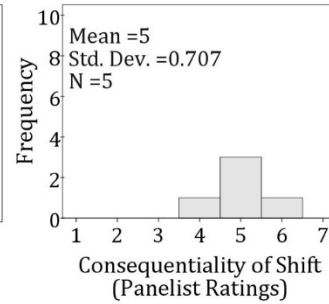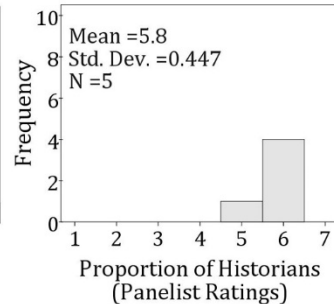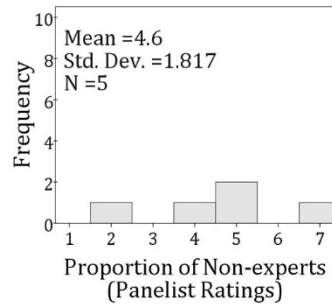
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

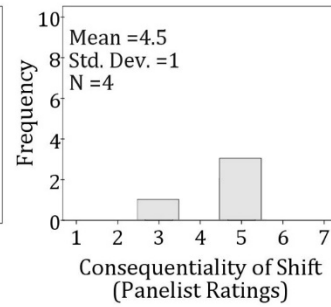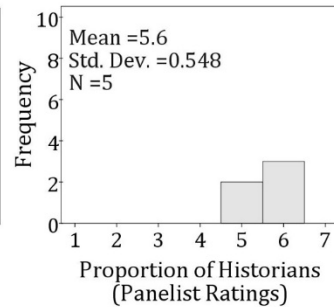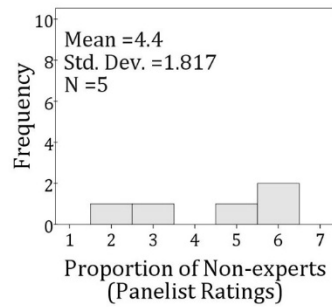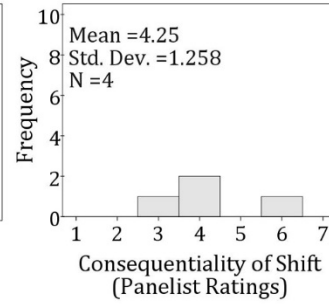*Figure 8.6*: In-Progress History Comparison 3

Many non-experts think that people:

who were more "civilized" naturally dominated those who lacked the markers of an advanced civilization.

But Historians think that people:

in the "civilized" world came to dominate everyone else for very specific reasons.

Mean =5.8
Std. Dev. =0.837
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =4.4
Std. Dev. =1.517
N =5

Proportion of Historians
(Panelist Ratings)

Mean =4.6
Std. Dev. =1.14
N =5

Consequentiality of Shift
(Panelist Ratings)
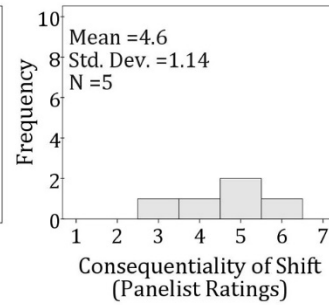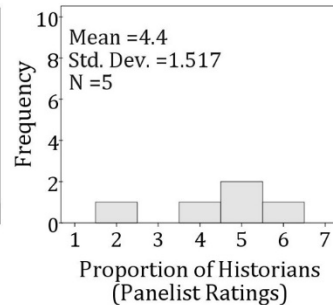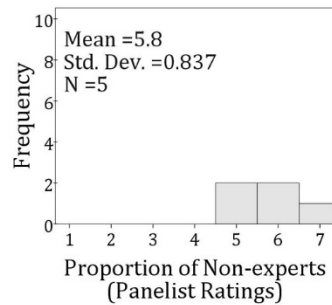
Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
One panelist did not think "many historians would talk about the "civilized" world!"

Other Issues:
The original suggested comparison was about competing explanations for European Colonization. We asked panelists to make a more general comparison about theories of people and the this was the most general version suggested.

Figure 8.7: In-Progress History Comparison 4

Many non-experts think that people:

can be understood by reference to a single explanatory principle.

But Historians think that people:

are best understood by reference to the multiple contexts within which they live and act.

Mean =4
Std. Dev. =1
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =6.2
Std. Dev. =0.837
N =5

Proportion of Historians
(Panelist Ratings)

Mean =4.2
Std. Dev. =1.304
N =5

Consequentiality of Shift
(Panelist Ratings)
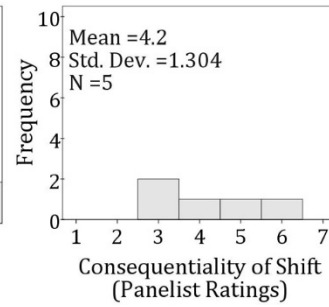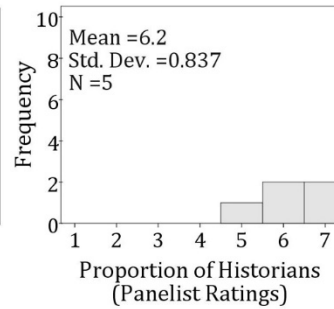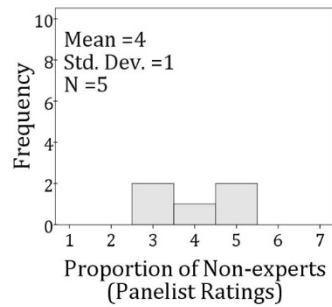
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 8.8: In-Progress History Comparison 5

Many non-experts think that people:

act on the basis of simple identities of class, race/ethnicity, and gender.

But Historians think that people:

act on the basis of intersecting and overlapping identities.

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
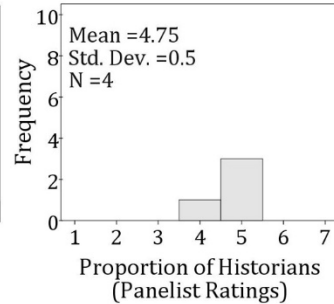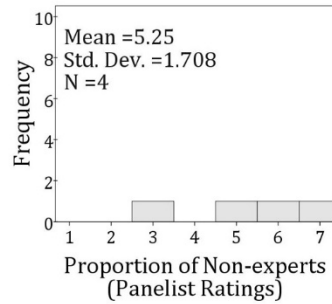None for this item.

Other Issues:
None for this item.

Figure 8.9: In-Progress History Comparison 6

Many non-experts think that people:

were much less tolerant of cultural and racial differences in the past.

But Historians think that people:

's tolerance of cultural and racial difference varied for many reasons over time.

Mean =5.25
Std. Dev. =1.708
N =4

Proportion of Non-experts
(Panelist Ratings)

Mean =4.75
Std. Dev. =0.5
N =4

Proportion of Historians
(Panelist Ratings)

Mean =4.25
Std. Dev. =1.258
N =4

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
One panelist was unsure how "tolerance" was intended. Another commented that they were not sure that "tolerance" was the right word, and "the issue is about how easily outsiders/those who are different are incorporated into communities and social groups."
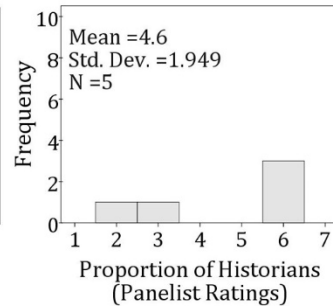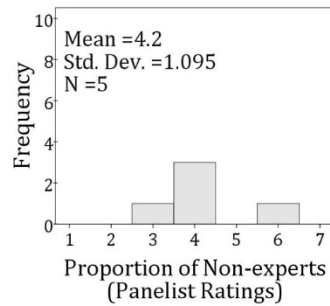
Other Issues:
None for this item.

Figure 8.10: In-Progress History Comparison 7

Many non-experts think that people:

are motivated primarily by self-interest.

But Historians think that people:

are motivated by both a variety of values (religious, political, economic) as well as by personal self-interest.
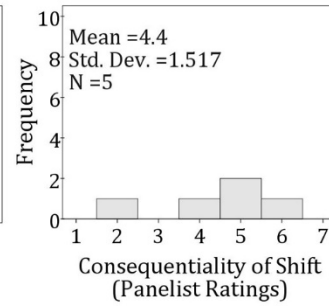
Mean =4.2
Std. Dev. =1.095
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =4.6
Std. Dev. =1.949
N =5

Proportion of Historians
(Panelist Ratings)

Mean =4.4
Std. Dev. =1.517
N =5

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
One panelist commented "Wouldn't the other categories also be self-interest?"

Other Issues:
None for this item.

Figure 8.11: In-Progress History Comparison 8

CHAPTER 9

Results of the Political Science Panel

The Political Science panel, from the perspective of the Delphi Method, was an unsuccessful panel. Due to panelists choosing not to participate across subsequent rounds or panelists forgetting their codes, it was never possible to match five panelists across rounds. This means that no comparisons could reach stability, and thus, no items were successful. Despite this, several comparisons were rated highly by panelists.

The major issue with the Political Science panel, as described earlier, was that many suggested comparisons were both contextual and conceptually redundant. That is, many items seemed to have the same insight about people but applied this insight to different contexts relevant to Political Scientists (voting, evaluating campaign messages, etc.). We attempted to reduce this redundancy by combining items, which eventually resulted in the creation of three items that were reasonably popular with panelists. Since the process for creating these items was unusual, we have labeled these items below.
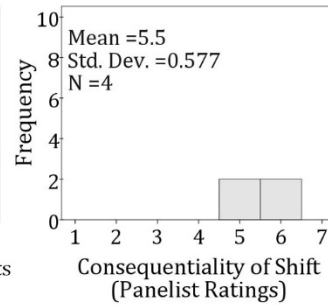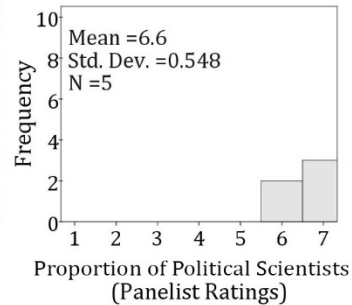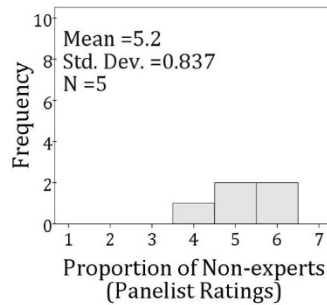
A common issue with this panel was that panelists disagreed about how to interpret the term "rationality." This is discussed in greater detail below, in the section on divisions within fields as described by panelists.

Many non-experts think that people:

understand government institutions, and don't need any expertise to run government.

But Political Scientists think that people:

don't understand the complexity of the systems of government.

Mean =5.2
Std. Dev. =0.837
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =6.6
Std. Dev. =0.548
N =5

Proportion of Political Scientists
(Panelist Ratings)

Mean =5.5
Std. Dev. =0.577
N =4

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
None for this item.

Other Issues:
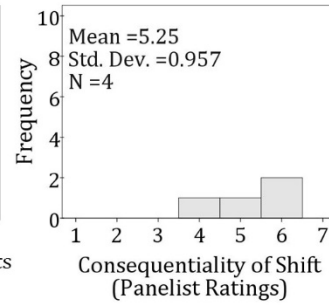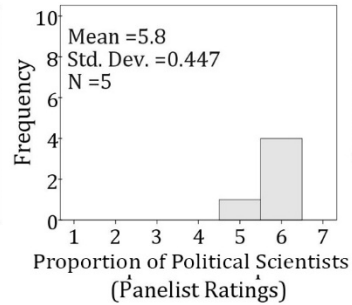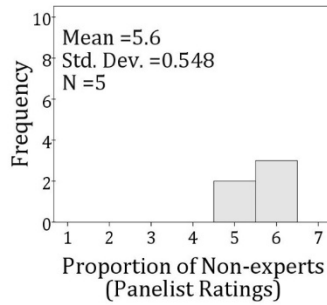None for this item.

Figure 9.1: In-Progress Political Science Comparison 1

Many non-experts think that people:

in leadership positions tend to exercise leadership based on their individual skills, views, and personal experiences before attaining office.

But Political Scientists think that people:

in leadership positions are constrained and/or guided by their institutional settings, and the existing policies that they encounter upon entering that position.

Mean =5.6
Std. Dev. =0.548
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =5.8
Std. Dev. =0.447
N =5

Proportion of Political Scientists
(Panelist Ratings)

Mean =5.25
Std. Dev. =0.957
N =4

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
Panelist commented that "capability to exercise leadership" would be more accurate for the first statement.

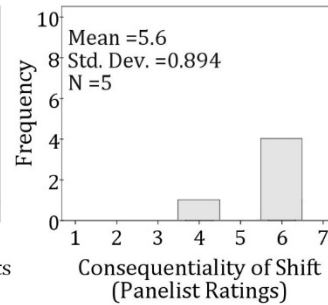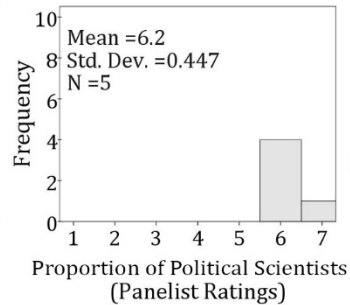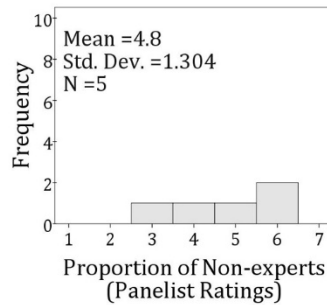**Other Issues:**
None for this item.

Figure 9.2: In-Progress Political Science Comparison 2

163

Many non-experts think that people:

make decisions (such as who to vote for) based on a careful and unbiased evaluation of information.

But Political Scientists think that people:

make decisions (such as who to vote for) by relying on heuristics, intuitions, and predispositions.

Mean =4.8
Std. Dev. =1.304
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =6.2
Std. Dev. =0.447
N =5

Proportion of Political Scientists
(Panelist Ratings)

Mean =5.6
Std. Dev. =0.894
N =5

Consequentiality of Shift
(Panelist Ratings)

Stable:
Not possible to test (see below)

Clarifying Panelist Comments:
Panelist commented that they did not "think there is a very big difference here between experts and non-experts; other than that non-experts might find 'uninformed' decisions less understandable and more 'stupid'."

Other Issues:
This item was created by combining several similar items together. It was not possible to test this comparison for stability, as it was first seen by panelists in Round 5.
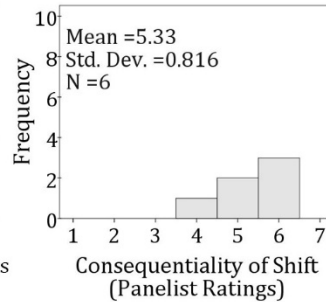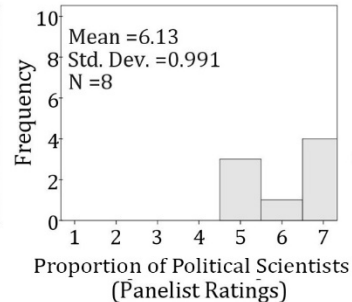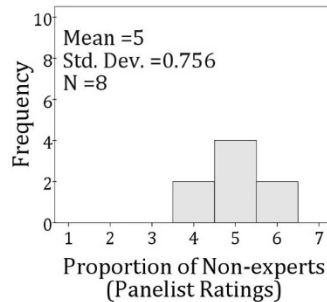
Figure 9.3: In-Progress Political Science Comparison 3

Many non-experts think that people:

can find the "best" solution if they just get together and work it out.

But Political Scientists think that people:

in general, cannot do this as collective choices are inherently more complicated and "best" alternatives may not exist for groups.

Mean =5
Std. Dev. =0.756
N =8

Proportion of Non-experts
(Panelist Ratings)

Mean =6.13
Std. Dev. =0.991
N =8

Proportion of Political Scientists
(Panelist Ratings)

Mean =5.33
Std. Dev. =0.816
N =6

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine (see below)

Clarifying Panelist Comments:
None for this item.

Other Issues:
This item was erroneously considered stable and not tested further. However, it was not possible for comparisons to reach stability for this panel, due to low participation.
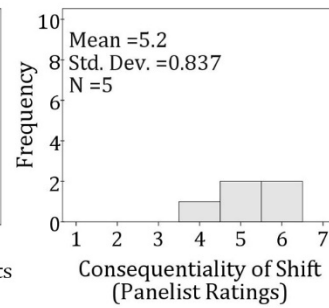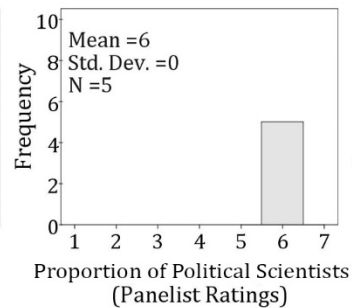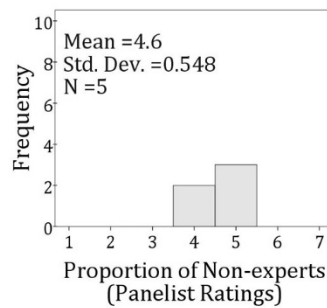
Figure 9.4: In-Progress Political Science Comparison 4

Many non-experts think that people:

understand social and political issues, and are competent at making informed judgments about them (such as what policies are worth supporting).

But Political Scientists think that people:

are poorly informed and are bounded/limited in their rationality, their time, and their ability to understand the complexities of the social and political world, and are therefore not very competent at making informed judgments.

Mean =4.6
Std. Dev. =0.548
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =6
Std. Dev. =0
N =5

Proportion of Political Scientists
(Panelist Ratings)

Mean =5.2
Std. Dev. =0.837
N =5

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
None for this item.

Other Issues:
This item was created by combining several similar items together.
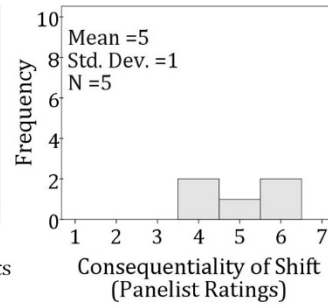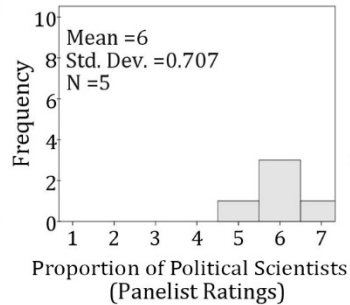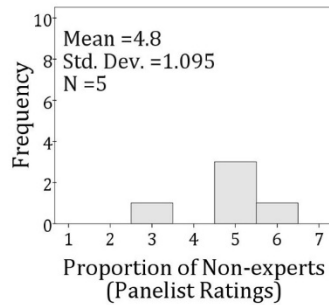
Figure 9.5: In-Progress Political Science Comparison 5

Many non-experts think that people:

have a single identity.

But Political Scientists think that people:

have multiple "identities" that manifest in different ways depending on context.

Left chart:
Mean =4.8
Std. Dev. =1.095
N =5
Frequency (y-axis: 0 to 10)
Proportion of Non-experts
(Panelist Ratings)

Middle chart:
Mean =6
Std. Dev. =0.707
N =5
Frequency (y-axis: 0 to 10)
Proportion of Political Scientists
(Panelist Ratings)

Right chart:
Mean =5
Std. Dev. =1
N =5
Frequency (y-axis: 0 to 10)
Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
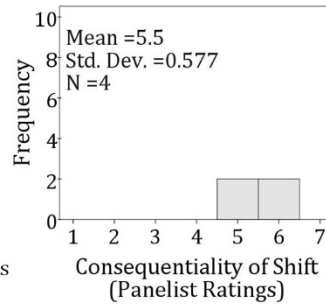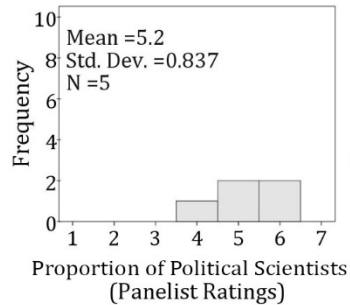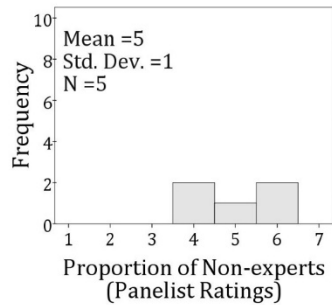Panelist commented that non-experts may not be familiar enough with the concept of identity to think in those terms.

Other Issues:
None for this item.

Figure 9.6: In-Progress Political Science Comparison 6

Many non-experts think that people:

are often political independents.

But Political Scientists think that people:

are mostly partisans, even if they claim otherwise.

Mean =5
Std. Dev. =1
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =5.2
Std. Dev. =0.837
N =5

Proportion of Political Scientists
(Panelist Ratings)

Mean =5.5
Std. Dev. =0.577
N =4

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
Panelists argued that the expert statement was characteristic of Political Scientists who study the U.S., but not other Political Scientists, who are less interested in party.
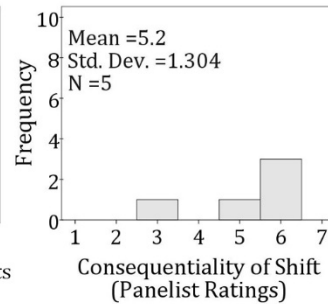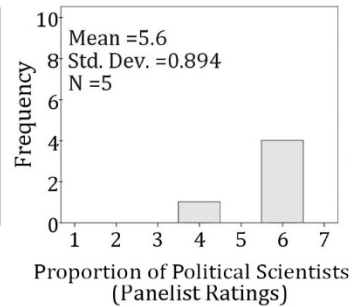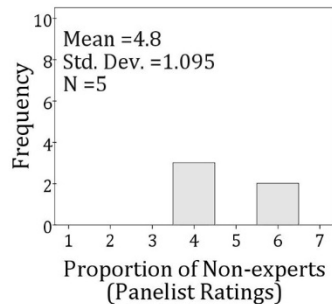
Other Issues:
None for this item.

Figure 9.7: In-Progress Political Science Comparison 7

Many non-experts think that people:

can judge when information is useful and when it is manipulative.

But Political Scientists think that people:

are swayed by flimsy arguments and influenced by how information is framed.

Mean =4.8
Std. Dev. =1.095
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =5.6
Std. Dev. =0.894
N =5

Proportion of Political Scientists
(Panelist Ratings)

Mean =5.2
Std. Dev. =1.304
N =5

Consequentiality of Shift
(Panelist Ratings)

Stable:
Not possible to test (see below)

Clarifying Panelist Comments:
None for this item.

Other Issues:
This item was created by combining several similar items together. It was not possible to test this comparison for stability, as it was first seen by panelists in Round 5.
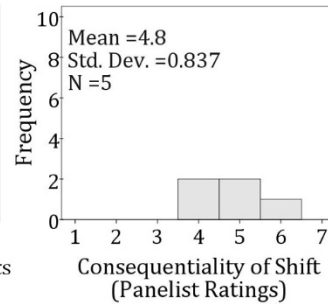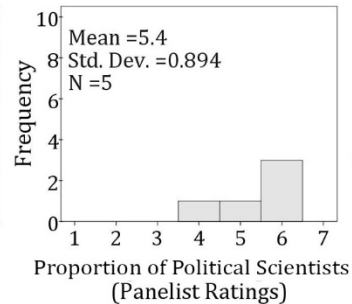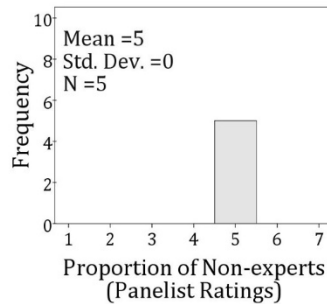
Figure 9.8: In-Progress Political Science Comparison 8

Many non-experts think that people:

hold consistent views/ideologies.

But Political Scientists think that people:

really don't hold consistent views/ideologies as they are influenced by context and how information is framed.

Mean =5 Std. Dev. =0 N =5	Mean =5.4 Std. Dev. =0.894 N =5	Mean =4.8 Std. Dev. =0.837 N =5
Proportion of Non-experts (Panelist Ratings)	Proportion of Political Scientists (Panelist Ratings)	Consequentiality of Shift (Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
Panelist commented: "While non-expert [*sic*] tend to think that people are either set in their views or are inconsistent, political scientists don't necessarily see having consistent views and being manipulated by information as being incompatible, i.e., many would argue that people have relatively fixed values but that the views with regard to policies, etc. might depend on information, etc. And many would draw a distinction between views and ideologies (which are closer to values)."

Other Issues:
None for this item.

Figure 9.9: In-Progress Political Science Comparison 9

Many non-experts think that people:

make informed judgments.

But Political Scientists think that people:

do not make informed judgments because of the lack of incentives to become informed.
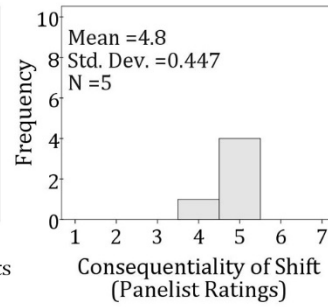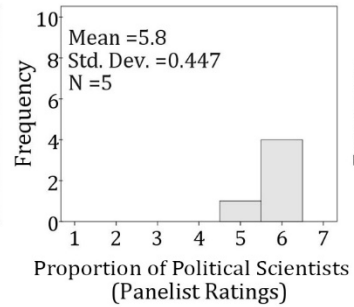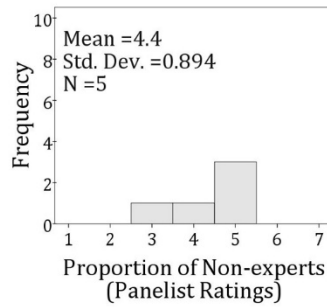
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
None for this item.

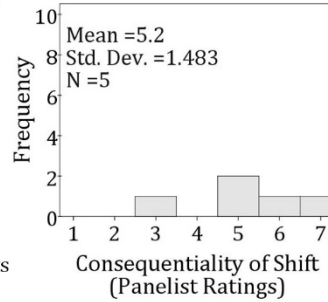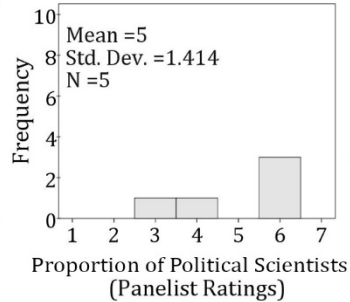**Other Issues:**
None for this item.

Figure 9.10: In-Progress Political Science Comparison 1*0*

Many non-experts think that people:

are instrumentally rational (choose actions in order to best achieve their goals).

But Political Scientists think that people:

often behave in ways contrary to their immediate self-interest due to social factors.

Mean =4.6
Std. Dev. =0.548
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =5
Std. Dev. =1.414
N =5

Proportion of Political Scientists
(Panelist Ratings)

Mean =5.2
Std. Dev. =1.483
N =5

Consequentiality of Shift
(Panelist Ratings)
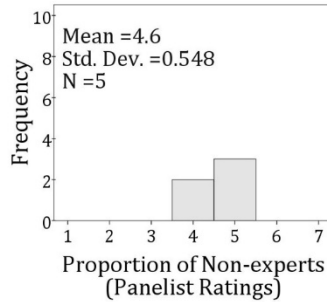
Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
Regarding the Political Scientist statement, panelist commented that "the challenge of things like war come into play." Another panelist disagreed with this item, commenting "the statements could be reversed. I would think the majority of political scientists see people as instrumentally rational but that in the presence of limited information people often rely on various cues and information shortcuts."

Other Issues:
None for this item.

Figure 9.11: In-Progress Political Science Comparison 1*1*

Many non-experts think that people:

prefer democratic over undemocratic forms of government.

But Political Scientists think that people:

prefer the form of government that they believe will benefit them the most.
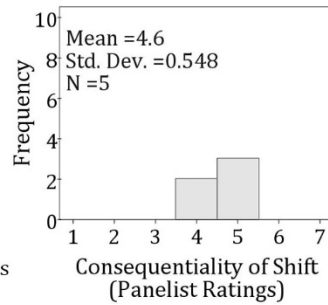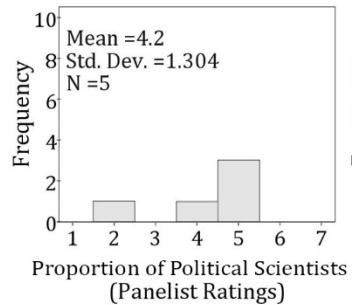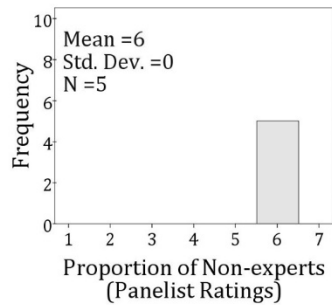
Mean =6
Std. Dev. =0
N =5

Proportion of Non-experts
(Panelist Ratings)

Mean =4.2
Std. Dev. =1.304
N =5

Proportion of Political Scientists
(Panelist Ratings)

Mean =4.6
Std. Dev. =0.548
N =5

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
Panelist commented that they answered the first statement as if it was constrained to those who live in a democracy.

Other Issues:
None for this item.

Figure 9.12: In-Progress Political Science Comparison 1*2*

Many non-experts think that people:

can be neatly divided into racists and non-racists.

But Political Scientists think that people:

fall on spectrums on a number of dimensions of racial resentment.
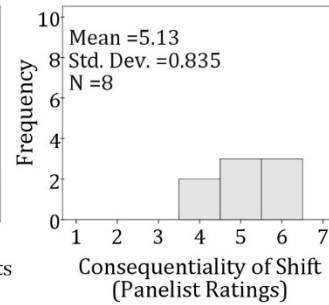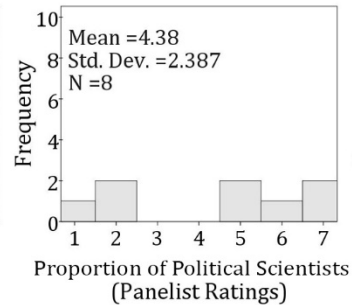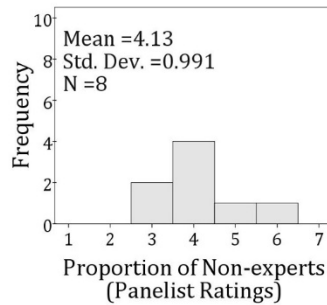
Left histogram:
Mean =4.13
Std. Dev. =0.991
N =8
Proportion of Non-experts
(Panelist Ratings)

Middle histogram:
Mean =4.38
Std. Dev. =2.387
N =8
Proportion of Political Scientists
(Panelist Ratings)

Right histogram:
Mean =5.13
Std. Dev. =0.835
N =8
Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine (see below)

Clarifying Panelist Comments:
Panelist commented that those Political scientists who do not study race would tend to think in terms of a single spectrum.

Other Issues:
This item was erroneously considered stable and not tested further. However, it was not possible for comparisons to reach stability for this panel, due to low participation.

Figure 9.13: In-Progress Political Science Comparison 1*3*

CHAPTER 10

Results of the Psychology Panel

The Psychology panel was probably the most successful panel. It not only generated an enormous number of comparisons (over fifty of them), but all of the comparisons, except for two, either reached stability or were dropped, and those two items would have likely reached stability had they not erroneously been removed from the process prematurely.

Nearly three dozen items suggested by panelists were later dropped due to low ratings. Unlike the Anthropology panel, which dropped items for not representing the non-expert perspective, most of the items dropped from the Psychologist panel were dropped because they were deemed inconsequential (a few were additionally considered unrepresentative of non-experts). Several of these comparisons were common psychological myths, for example, that opposites attract, that genes can "skip a generation," and that we only use 10% of our brain. Other items dropped in this manner were more typical of clinical psychology (such as why people seek out therapy) and typically were considered only moderately consequential (between three and four out of seven) rather than central to how people understand and explain the human world.

Many non-experts think that people:

are rational and logical in their decision-making and behavior.

But Psychologists think that people:

rely on emotions, cognitive shortcuts, and stereotypes.

Mean =5.42
Std. Dev. =0.669
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =5.75
Std. Dev. =0.754
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =4.92
Std. Dev. =1.24
N =12

Consequentiality of Shift
(Panelist Ratings)
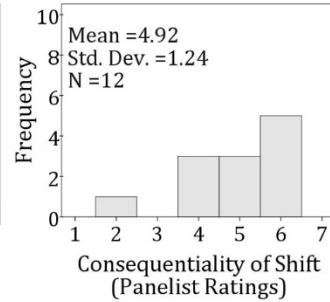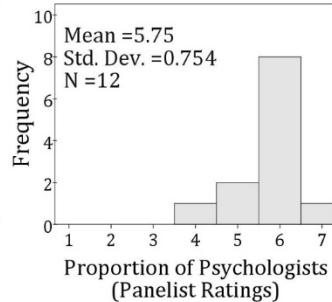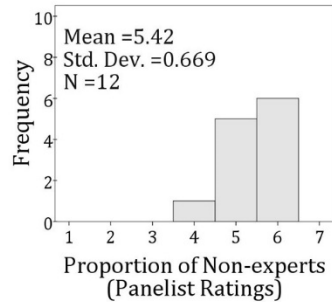
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
One panelist disagreed, arguing that it is not that "non-experts think people are rational so much as they think they are capable of rational thought and so when they don't act rationally it is intentional."

Other Issues:
None for this item.

Figure 10.1: Stable Psychology Comparison 1

Many non-experts think that people:

recall events as they happened, like a tape or video recorder.

But Psychologists think that people:

reconstruct memories from many sources, leading exact recall to be extremely rare if not impossible.
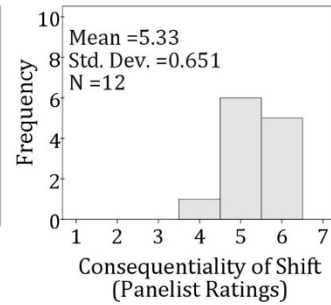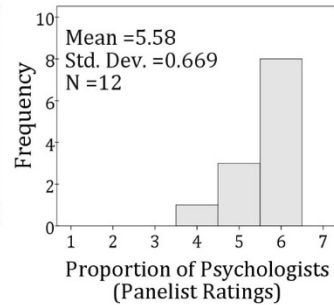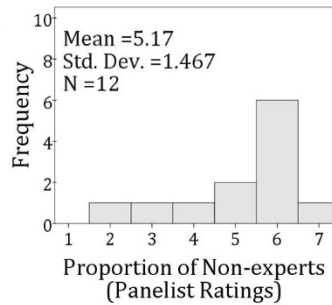
Mean =5.17
Std. Dev. =1.467
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =5.58
Std. Dev. =0.669
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =5.33
Std. Dev. =0.651
N =12

Consequentiality of Shift
(Panelist Ratings)

Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 10.2: Stable Psychology Comparison 2

Many non-experts think that people:

are good at recognizing their own biases.

But Psychologists think that people:

are not typically good at recognizing their biases.
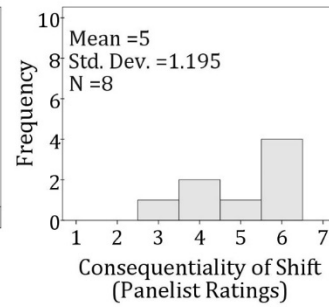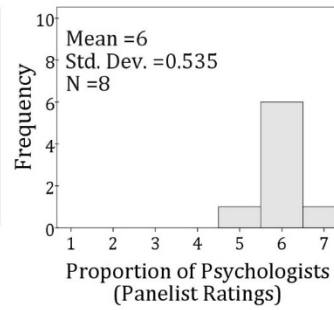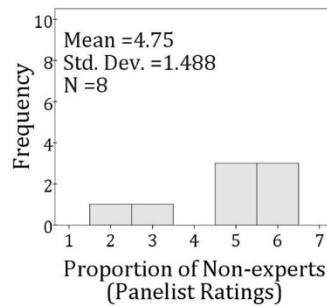
Mean =4.75
Std. Dev. =1.488
N =8

Proportion of Non-experts
(Panelist Ratings)

Mean =6
Std. Dev. =0.535
N =8

Proportion of Psychologists
(Panelist Ratings)

Mean =5
Std. Dev. =1.195
N =8

Consequentiality of Shift
(Panelist Ratings)

Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 10.3: Stable Psychology Comparison 3

Many non-experts think that people:

accurately and objectively perceive themselves and the world around them.

But Psychologists think that people:

are subject to countless biases and errors, some conscious and some entirely outside of their awareness.
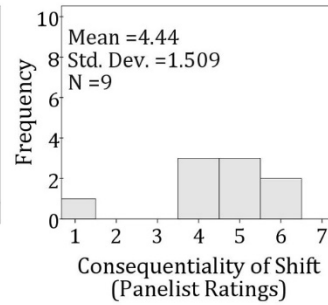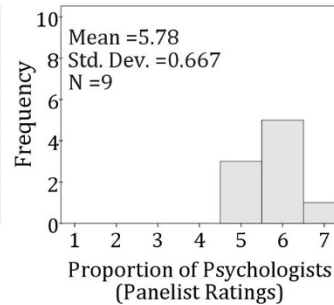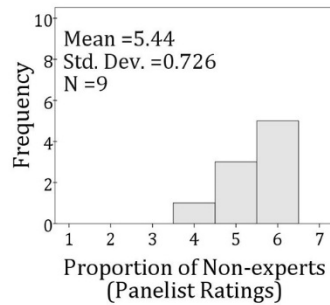
Mean =5.44
Std. Dev. =0.726
N =9

Proportion of Non-experts
(Panelist Ratings)

Mean =5.78
Std. Dev. =0.667
N =9

Proportion of Psychologists
(Panelist Ratings)

Mean =4.44
Std. Dev. =1.509
N =9

Consequentiality of Shift
(Panelist Ratings)

Stable: Yes
(Avg N: 6.67)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 10.4: Stable Psychology Comparison 4

179

Many non-experts think that people:

act deliberately and consistently across situations.

But Psychologists think that people:

are influenced by their surroundings more than they realize, and behave quite variably across situations.
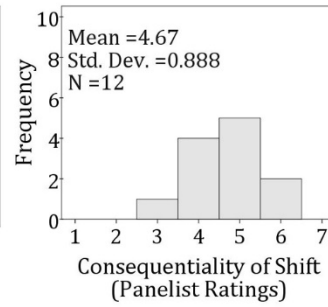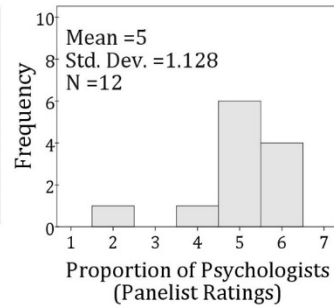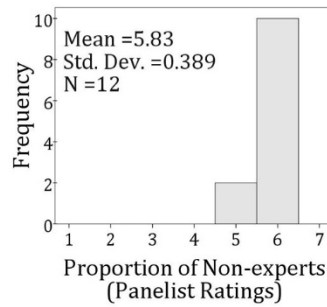
Mean =5.83
Std. Dev. =0.389
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =5
Std. Dev. =1.128
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =4.67
Std. Dev. =0.888
N =12

Consequentiality of Shift
(Panelist Ratings)

Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 10.5: Stable Psychology Comparison 5

Many non-experts think that people:

succeed through willpower.

But Psychologists think that people:

need lots of things (such as luck and circumstances) to go their way in order to succeed; willpower alone is insufficient.

Mean =5.4
Std. Dev. =0.699
N =10

Proportion of Non-experts
(Panelist Ratings)

Mean =5.7
Std. Dev. =0.675
N =10

Proportion of Psychologists
(Panelist Ratings)

Mean =4.33
Std. Dev. =1
N =9

Consequentiality of Shift
(Panelist Ratings)
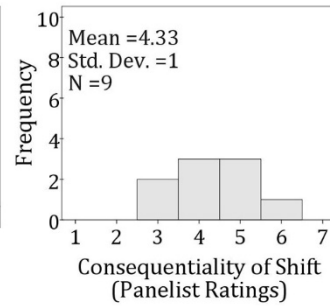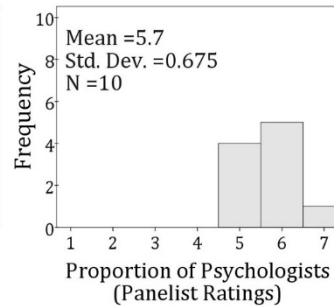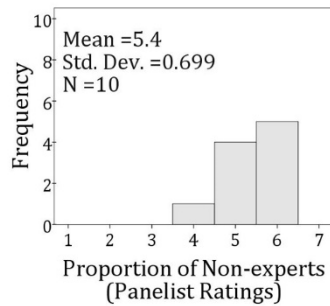
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
Panelist wanted "self management skill" added, as they considered it essential to success.

Other Issues:
None for this item.

Figure 10.6: Stable Psychology Comparison 6

Many non-experts think that people:

have accurate insight into the reasons for their thoughts, feelings, and behaviors.

But Psychologists think that people:

are quite poor at accurate introspection.
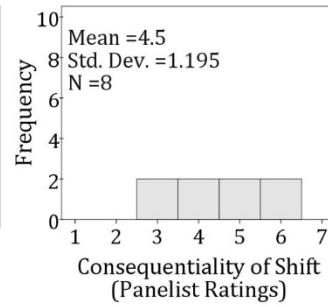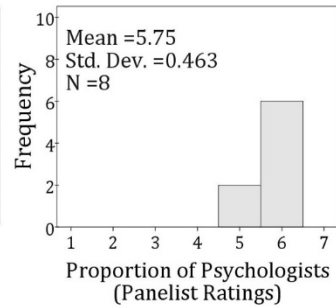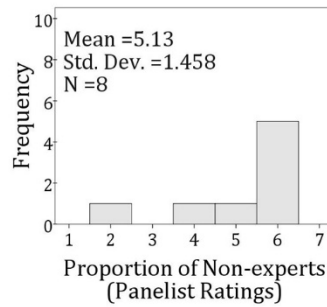
Clarifying Panelist Comments:
One panelist disagreed, saying that "many nonexperts are aware that people often don't know their true motives."

Other Issues:
None for this item.

Figure 10.7: Stable Psychology Comparison 7

Many non-experts think that people:

generally know what's good for them in terms of health, relationships, happiness, etc.

But Psychologists think that people:

are pretty bad at understanding themselves and knowing how best to get what they want.

Mean =5.58
Std. Dev. =0.793
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =5.08
Std. Dev. =1.311
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =4.67
Std. Dev. =1.303
N =12

Consequentiality of Shift
(Panelist Ratings)
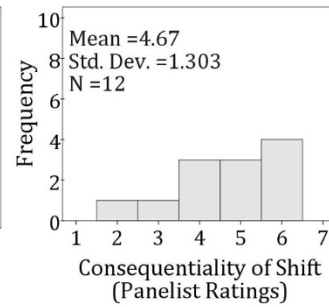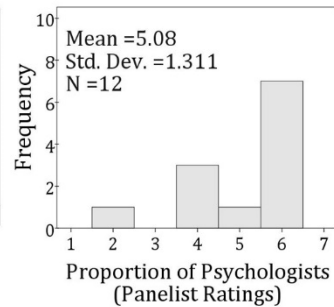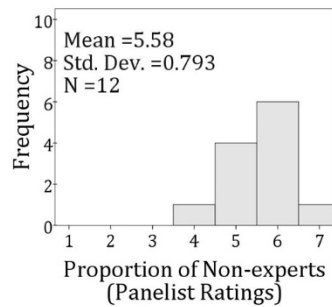
Stable: Yes
(Avg N: 5.66)

Clarifying Panelist Comments:
one panelist disagreed, arguing that "people believe people know what is good for them" and when people act non-optimally, non-experts assume it is intentional and look for explanations.

Other Issues:
None for this item.

Figure 10.8: Stable Psychology Comparison 8

Many non-experts think that people:

with mental illnesses are dangerous/violent.

But Psychologists think that people:

with mental illnesses are mostly nonviolent.
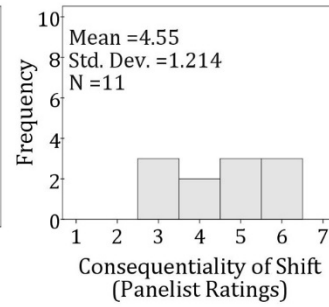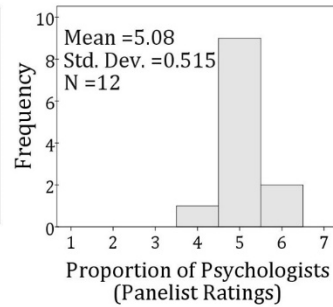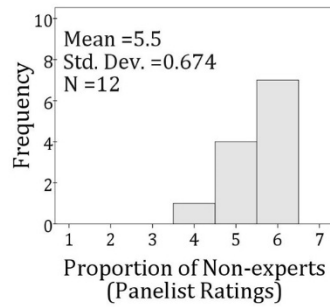
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
Panelist mentioned that people likely do not think this for all disorders, given the proportion of anti-depressants and anti-ADHD meds.

Other Issues:
None for this item.

Figure 10.9: Stable Psychology Comparison 9

Many non-experts think that people:

think like they do.

But Psychologists think that people:

overestimate the similarity between their own thoughts and others' thoughts.
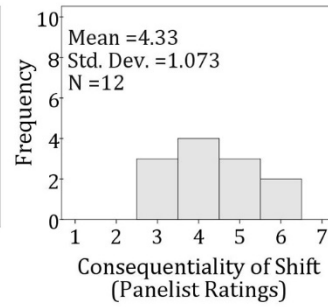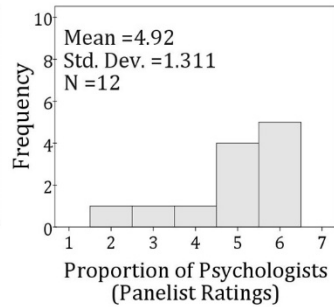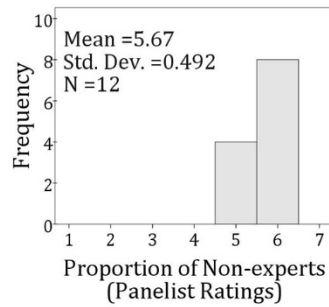
Mean =5.67
Std. Dev. =0.492
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =4.92
Std. Dev. =1.311
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =4.33
Std. Dev. =1.073
N =12

Consequentiality of Shift
(Panelist Ratings)

Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 10.10: Stable Psychology Comparison 1*0*

Many non-experts think that people:

are either one way or another (e.g. for gender, people are either male or female).

But Psychologists think that people:

have characteristics that vary along a continuum.
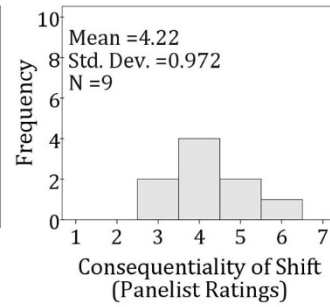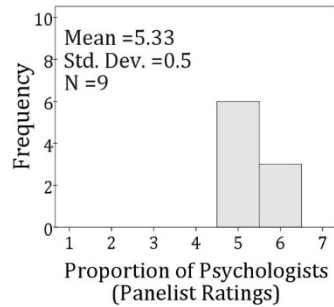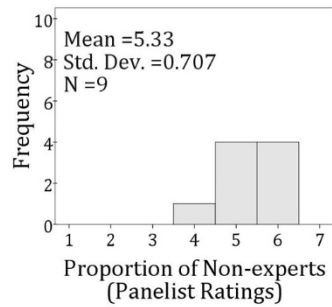
Clarifying Panelist Comments:
Panelist argued that it depends on the trait, with non-experts seeing some traits as dichotomous and others traits as being on a continuum.

Other Issues:
None for this item.

Figure 10.11: Stable Psychology Comparison 1*1*

Many non-experts think that people:

engage in behaviors and thoughts that are mostly conscious.

But Psychologists think that people:

engage in behaviors and thoughts that are mostly unconscious.

Mean =5.92
Std. Dev. =0.515
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =4.25
Std. Dev. =1.422
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =4.42
Std. Dev. =1.311
N =12

Consequentiality of Shift
(Panelist Ratings)
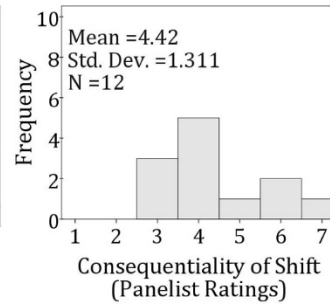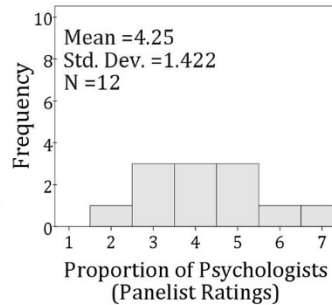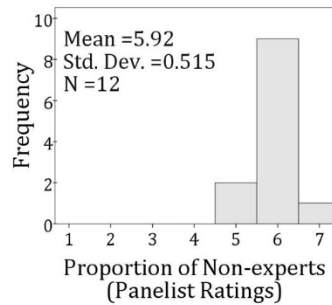
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
Panelist mentioned that people "likely think people do things unconsciously as well, but may think about it more like a freudian, motivated unconscious rather than the less intentional unconscious most experts think about."

Other Issues:
None for this item.

Figure 10.12: Stable Psychology Comparison 1*2*

Many non-experts think that people:

are mostly motivated by money.

But Psychologists think that people:

have many goals that are more important, including feeling that they are respected and doing something worthwhile.

Mean =5	Std. Dev. =1.128
N =12	

Proportion of Non-experts
(Panelist Ratings)

Mean =5
Std. Dev. =1.414
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =4
Std. Dev. =0.775
N =11

Consequentiality of Shift
(Panelist Ratings)
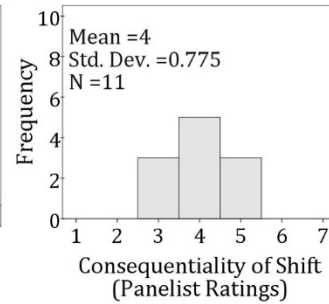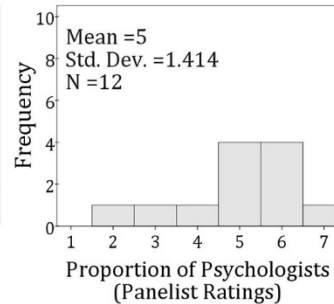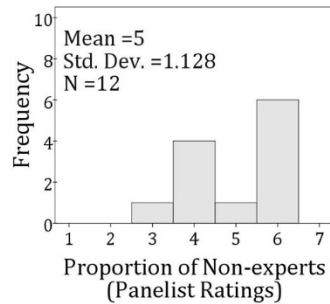
Stable: Yes
(Avg N: 5.66)

Clarifying Panelist Comments:
Panelist mentioned that this was tricky, because it "depends on context."

Other Issues:
None for this item.

Figure 10.13: Stable Psychology Comparison 1*3*

Many non-experts think that people:

are interested in discovering and disseminating truth.

But Psychologists think that people:

are mostly self-serving, taking positions and holding beliefs that benefit them.

Mean =4.75
Std. Dev. =1.215
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =5
Std. Dev. =0.853
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =4.17
Std. Dev. =1.03
N =12

Consequentiality of Shift
(Panelist Ratings)
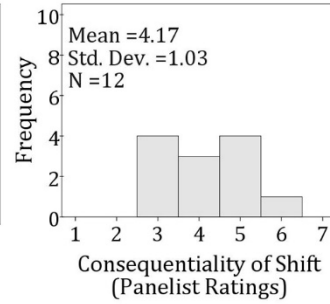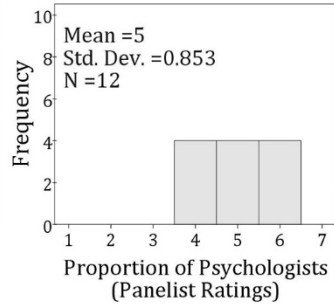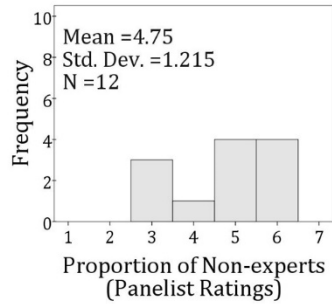
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
One panelist mentioned that they were "not sure this is psychology as opposed to education and logic."

Other Issues:
None for this item.

Figure 10.14: Stable Psychology Comparison 14

189

Many non-experts think that people:

are not that influenced by advertising.

But Psychologists think that people:

are significantly influenced by advertising/persuasive messaging.
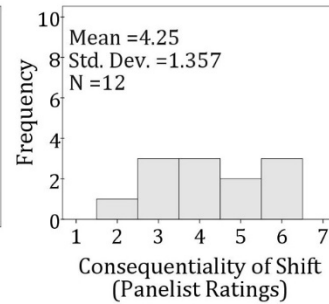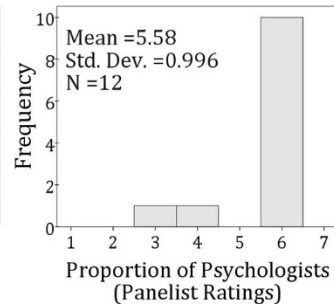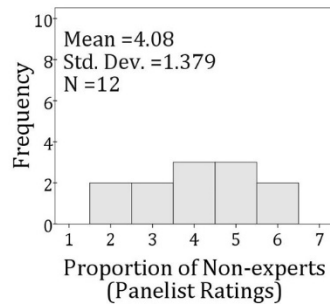
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
One panelist disagreed, arguing that "people know that marketing works – most companies have a marketing department for a reason."

Other Issues:
None for this item.

Figure 10.15: Stable Psychology Comparison 1*5*

Many non-experts think that people:

will change their minds when presented with logical evidence to do so.

But Psychologists think that people:

have many ways to defend their beliefs, therefore attitudes are highly stable and difficult to change.

Mean =5.33
Std. Dev. =0.778
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =5.42
Std. Dev. =1.311
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =4.89
Std. Dev. =0.928
N =9

Consequentiality of Shift
(Panelist Ratings)
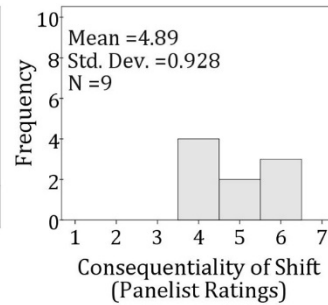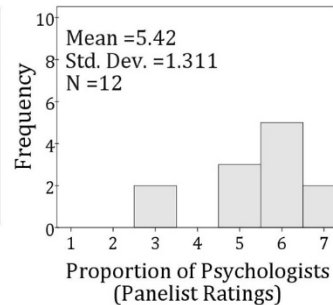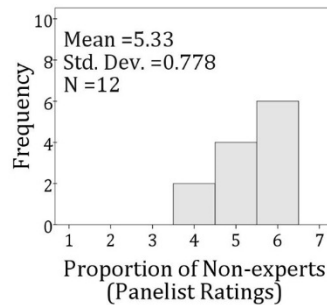
Stable:
Partial (see note below)

Clarifying Panelist Comments:
None for this item.

Other Issues:
This item was erroneously considered stable and not tested further. Both the non-expert and Psychologist statements were stable (n = 6), however, there was insufficient n (n = 4) to determine if the consequentiality rating was stable.

Figure 10.16: In-Progress Psychology Comparison 1

Many non-experts think that people:

think rationally and base their opinions on a range of evidence.

But Psychologists think that people:

seek out information that confirms their beliefs.
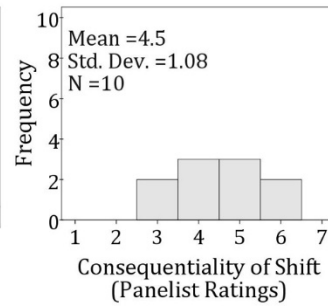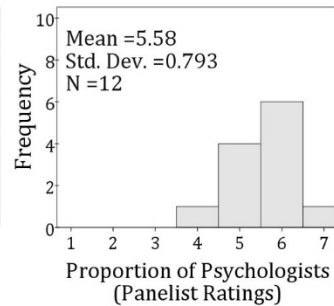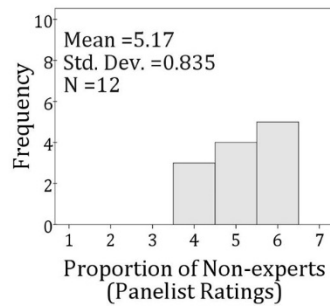
Mean =5.17
Std. Dev. =0.835
N =12

Proportion of Non-experts
(Panelist Ratings)

Mean =5.58
Std. Dev. =0.793
N =12

Proportion of Psychologists
(Panelist Ratings)

Mean =4.5
Std. Dev. =1.08
N =10

Consequentiality of Shift
(Panelist Ratings)

Stable:
Partial (see note below)

Clarifying Panelist Comments:
None for this item.

Other Issues:
This item was erroneously considered stable and not tested further. Both the non-expert and Psychologist statements were stable (n = 6), however, there was insufficient n (n = 4) to determine if the consequentiality rating was stable.

Figure 10.17: In-Progress Psychology Comparison 2

CHAPTER 11

Results of the Sociology Panel

The Sociology panel was successful, producing several items that reached stability. However, several highly rated items never reached stability, primarily because low participation in the final round made it impossible to test for stability across rounds.

One unusual aspect of the Sociology panel was that several of the suggested comparisons were proposed by many panelists simultaneously. For example, five different panelists suggested comparisons revolving around the idea that non-experts think people have complete control over their life course (while Sociologists do not). Usually, such overlapping suggested comparisons were difficult to disentangle and combine, especially as the number of overlapping comparisons increased. For the Sociology panel, however, the comparisons were so similar in spirit that they were relatively easy to combine. In addition, the combined comparisons were also rated very highly by panelists. These two facts together imply that Sociologists share a core set of ways of understanding the human world and also think that this perspective is not broadly shared by those outside of Sociology (or at least, outside of the Social Sciences).

Contrariwise, while the Sociology panel had many items prove immediately stable, it was also the panel with the most open debates at the end of the project. Many of these debates seemed to be over whether a comparison was too strongly worded. On the other hand, panelists sometimes commented that comparisons were not strongly worded enough. Panelists also suggested comparisons that were highly contextual. For example, one suggested comparison was about whether people misrepresent themselves online.

However, when we pointed out to panelists that an item was contextual, they often

successfully rephrased the comparisons into more general statements about people,

several of which were highly rated.

Many non-experts think that people:

are completely autonomous actors, able to make any choice they want.

But Sociologists think that people:

are constrained by inequalities and social structures, and have far less agency than they recognize.
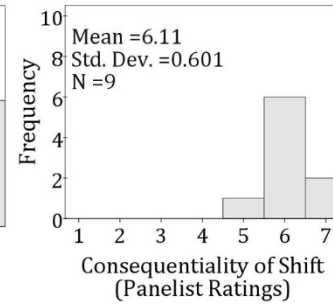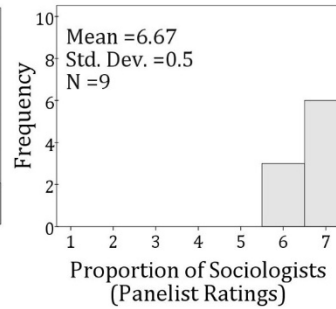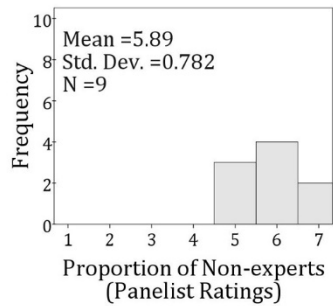
Mean =5.89
Std. Dev. =0.782
N =9

Proportion of Non-experts
(Panelist Ratings)

Mean =6.67
Std. Dev. =0.5
N =9

Proportion of Sociologists
(Panelist Ratings)

Mean =6.11
Std. Dev. =0.601
N =9

Consequentiality of Shift
(Panelist Ratings)

Stable: Yes
(Avg N: 5.66)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.1: Stable Sociology Comparison 1

Many non-experts think that people:

receive punishments and rewards in life because of their personal qualities (e.g. poor people are poor because they are lazy, rich people are rich because of merit).

But Sociologists think that people:

receive punishments and rewards in life because of social causes (e.g. poor people face structural inequalities and discrimination, rich people benefit from connections and inheriting wealth).
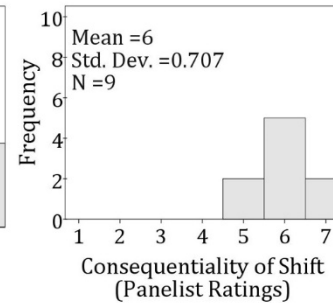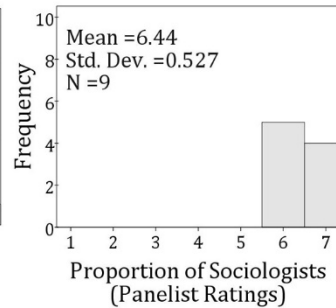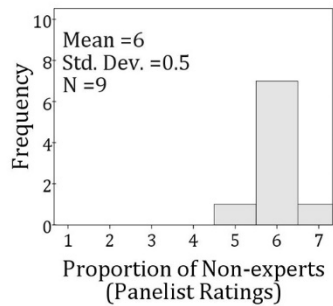
Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.2: Stable Sociology Comparison 2

196

Many non-experts think that people:

are born definitively female or male.

But Sociologists think that people:

become "female" or "male" through a combination of their hormonal, chromosomal, social, and environmental experiences.

Mean =6.56
Std. Dev. =0.527
N =9

Proportion of Non-experts
(Panelist Ratings)

Mean =5.78
Std. Dev. =1.093
N =9

Proportion of Sociologists
(Panelist Ratings)

Mean =5.89
Std. Dev. =0.333
N =9

Consequentiality of Shift
(Panelist Ratings)
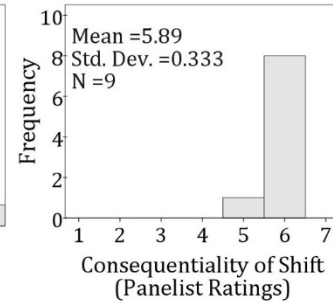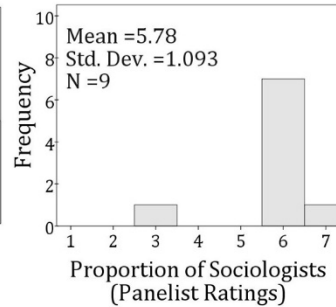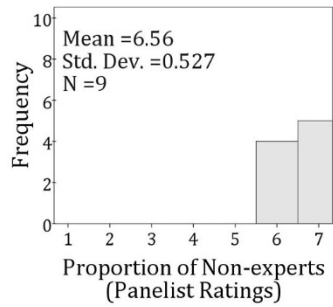
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.3: Stable Sociology Comparison 3

Many non-experts think that people:

can "pull themselves up by their bootstraps" and just need to work hard to be successful.

But Sociologists think that people:

of different backgrounds face different obstacles to success, many outside of their control.

Mean =5.78
Std. Dev. =1.093
N =9

Proportion of Non-experts
(Panelist Ratings)

Mean =6.56
Std. Dev. =0.527
N =9

Proportion of Sociologists
(Panelist Ratings)

Mean =5.89
Std. Dev. =0.601
N =9

Consequentiality of Shift
(Panelist Ratings)
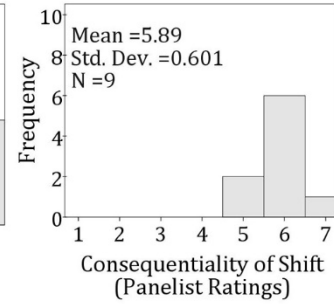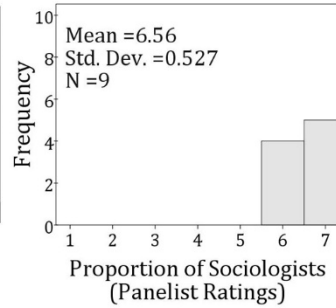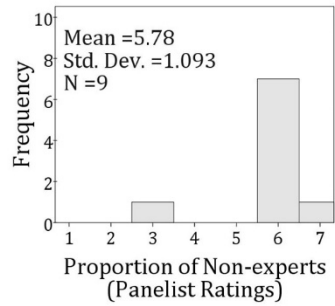
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.4: Stable Sociology Comparison 4

198

Many non-experts think that people:

are most likely to be harmed or killed by strangers.

But Sociologists think that people:

are much more likely to be killed by someone they know.
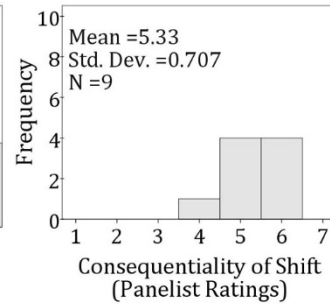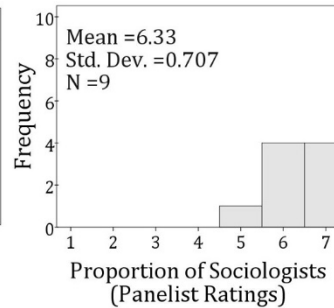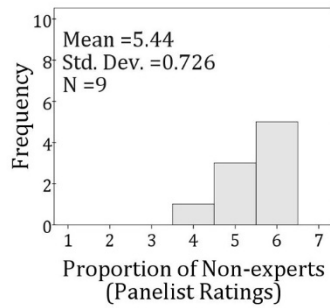
Mean =5.44
Std. Dev. =0.726
N =9

Frequency

Proportion of Non-experts
(Panelist Ratings)

Mean =6.33
Std. Dev. =0.707
N =9

Frequency

Proportion of Sociologists
(Panelist Ratings)

Mean =5.33
Std. Dev. =0.707
N =9

Frequency

Consequentiality of Shift
(Panelist Ratings)

Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.5: Stable Sociology Comparison 5

199

Many non-experts think that people:

can be fundamentally understood by their racial, gender, or sexual categories.

But Sociologists think that people:

who share a racial, gender, or sexual category are much more heterogeneous (dissimilar) than they might at first appear, and the categories themselves change.

Mean =5.44
Std. Dev. =0.726
N =9

Proportion of Non-experts
(Panelist Ratings)

Mean =6
Std. Dev. =0.5
N =9

Proportion of Sociologists
(Panelist Ratings)

Mean =5.38
Std. Dev. =0.744
N =8

Consequentiality of Shift
(Panelist Ratings)
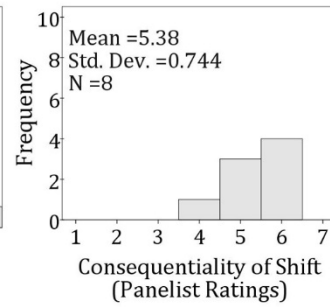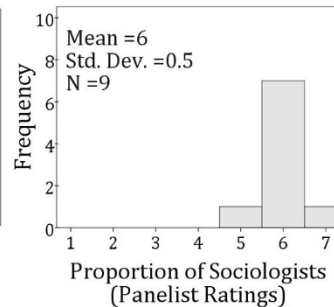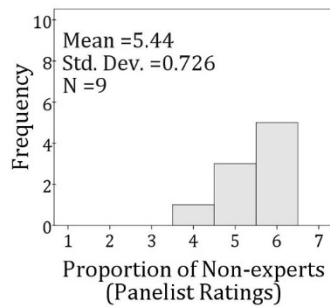
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.6: Stable Sociology Comparison 6

Many non-experts think that people:

are self-reliant.

But Sociologists think that people:
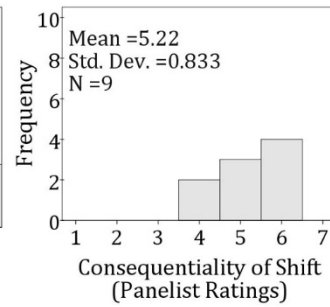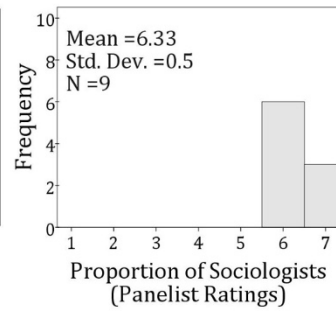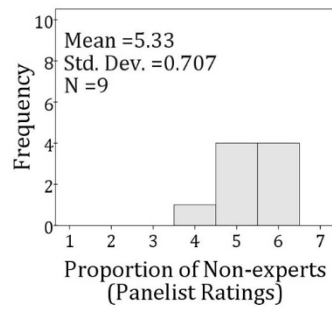
are interdependent.

Mean =5.33
Std. Dev. =0.707
N =9

Frequency

Proportion of Non-experts
(Panelist Ratings)

Mean =6.33
Std. Dev. =0.5
N =9

Frequency

Proportion of Sociologists
(Panelist Ratings)

Mean =5.22
Std. Dev. =0.833
N =9

Frequency

Consequentiality of Shift
(Panelist Ratings)

Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.7: Stable Sociology Comparison 7

Many non-experts think that people:

have essential qualities to their personalities that are immutable, innate, and unchanging.

But Sociologists think that people:

are dynamic in their personalities, experiencing changes in who they are based on life experiences and contact with other groups of people.

Mean =5.4
Std. Dev. =0.699
N =10

Proportion of Non-experts
(Panelist Ratings)

Mean =6.2
Std. Dev. =0.422
N =10

Proportion of Sociologists
(Panelist Ratings)

Mean =5
Std. Dev. =0.707
N =9

Consequentiality of Shift
(Panelist Ratings)
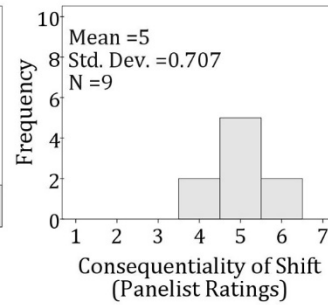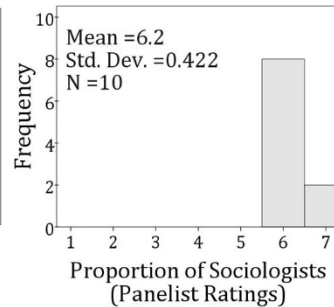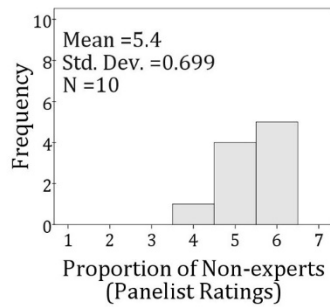
Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
One panelist thought the statement was too strongly worded, and most sociologists "think that there is a balance between immutable personal characteristics and dynamic ones (but that the balance across both differs across individuals)."

Other Issues:
None for this item.

Figure 11.8: Stable Sociology Comparison 8

Many non-experts think that people:

act strategically.

But Sociologists think that people:
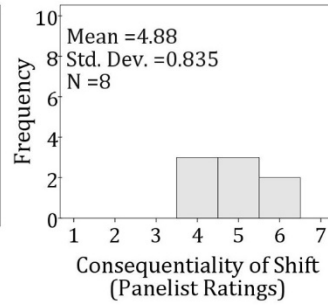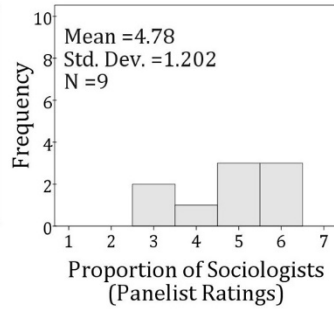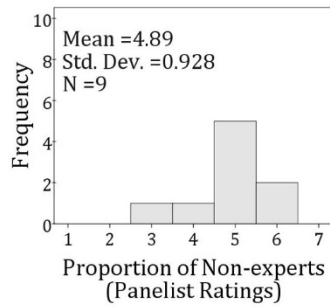
are mostly creatures of habit and emotion.

Stable: Yes
(Avg N: 6.00)

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.9: Stable Sociology Comparison 9

Many non-experts think that people:

have control over everything in their lives (such as how healthy they are), and are the masters of their own destiny.

But Sociologists think that people:

are impacted by social factors, and vary in their ability to control their life course.
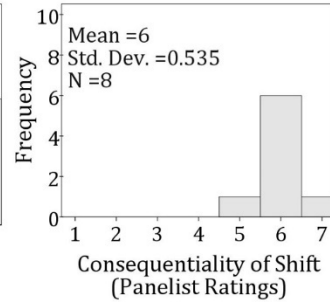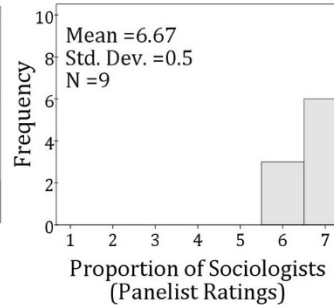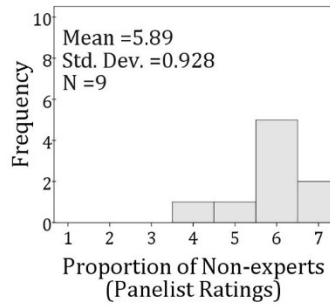
Stable:
Partial (see note below)

Clarifying Panelist Comments:
None for this item.

Other Issues:
This item was erroneously considered stable and not tested further. Both the non-expert and Sociologists statements were stable (N = 5), however, there was insufficient N (N = 4) to determine if the consequentiality rating was stable.

Figure 11.10: In-Progress Sociology Comparison 1

Many non-experts think that people:

who immigrate to one country from another are more prone to crime and delinquency than non-immigrants.

But Sociologists think that people:

who immigrate are no more, and often less, likely to likely to commit crimes or engage in delinquent behaviors than non-immigrants.
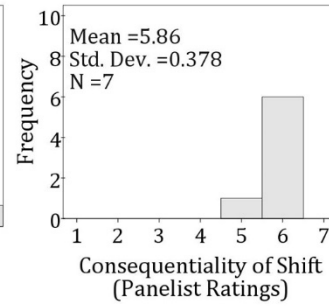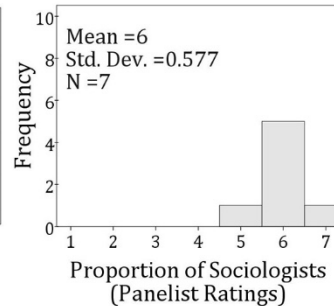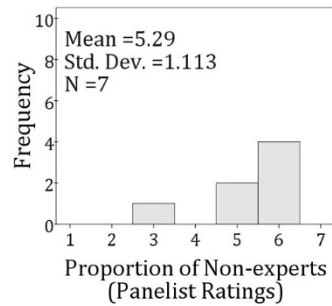
Mean =5.29
Std. Dev. =1.113
N =7

Frequency

Proportion of Non-experts
(Panelist Ratings)

Mean =6
Std. Dev. =0.577
N =7

Frequency

Proportion of Sociologists
(Panelist Ratings)

Mean =5.86
Std. Dev. =0.378
N =7

Frequency

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.11: In-Progress Sociology Comparison 2

Many non-experts think that people:

have moved beyond racism, embracing a post-racial world.

But Sociologists think that people:

, especially those from dominant groups, ignore real impacts of institutional racism, discrimination and privilege.

Mean =5.29
Std. Dev. =0.951
N =7

Proportion of Non-experts
(Panelist Ratings)

Mean =5.71
Std. Dev. =0.951
N =7

Proportion of Sociologists
(Panelist Ratings)

Mean =6
Std. Dev. =0.632
N =6

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N
to determine

Clarifying Panelist Comments:
One panelist did not think "ignore" was the right word as "it implies a willful denial." However, they did not suggest an alternative.

Other Issues:
None for this item.

Figure 11.12: In-Progress Sociology Comparison 3

Many non-experts think that people:

who immigrate take jobs away and lower wages for citizens.

But Sociologists think that people:

who immigrate are not at fault for taking jobs and lowering wages, employers who pay immigrant workers less reduce jobs for native workers.

Mean =5.86
Std. Dev. =0.69
N =7

Proportion of Non-experts
(Panelist Ratings)

Mean =5.43
Std. Dev. =0.787
N =7

Proportion of Sociologists
(Panelist Ratings)

Mean =5.17
Std. Dev. =1.329
N =6

Consequentiality of Shift
(Panelist Ratings)
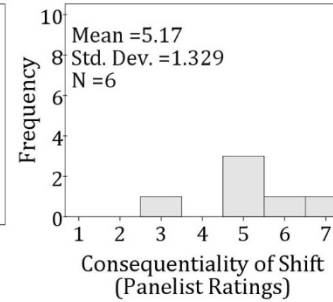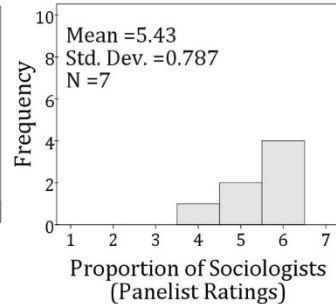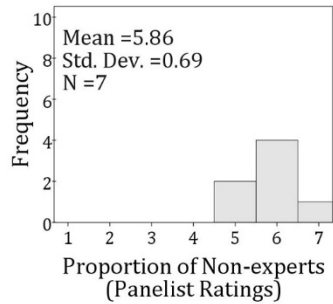
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
Several panelists were critical of the Sociologists statement. One commented that this was "a complicated empirical question that deal [*sic*] with a lot of economic issues". Another commented: "I am not sure if most sociologists would agree on this statement. I think there's more nuance and argument around issues of capitalism and market forces. Are there any sociologists who really would contend if the borders were closed and immigrants repatriated that employers would raise their wages and hire native workers? I'm wondering if a better way of stating it would be 'But Sociologists think that a capitalist system encourages employers to look for ways to reduce labor costs, including seeking out immigrant labor.' Another commented that immigrants "may have downward pressure on wages, but they are likely to create more jobs than take them away. When more people come to the US they need goods and services and more jobs are created supplying those goods and services. My point is that its not clear that immigrants actually reduce the number of jobs. The actual economics of this are more complex."

Other Issues:
None for this item.

Figure 11.13: In-Progress Sociology Comparison 4

207

Many non-experts think that people:

share universal understandings about many things, such as what is good or bad.

But Sociologists think that people:

vary greatly in their interpretations of things across cultures and societies.
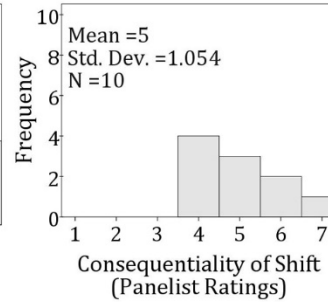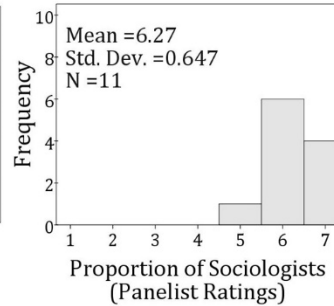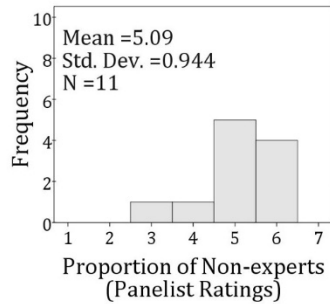
Mean =5.09
Std. Dev. =0.944
N =11

Proportion of Non-experts
(Panelist Ratings)

Mean =6.27
Std. Dev. =0.647
N =11

Proportion of Sociologists
(Panelist Ratings)

Mean =5
Std. Dev. =1.054
N =10

Consequentiality of Shift
(Panelist Ratings)

Stable:
Partial (see note below)

Clarifying Panelist Comments:
None for this item.

Other Issues:
This item was erroneously considered stable and not tested further. Both the non-expert and Sociologist statements were stable (N = 6), however, there was insufficient N (N = 4) to determine if the consequentiality rating was stable.

Figure 11.14: In-Progress Sociology Comparison 5

Many non-experts think that people:

on public assistance refuse to work, and are lazy.

But Sociologists think that people:

on public assistance are mostly the working poor, tend to only utilize public assistance as a last resort, and have structural barriers making it impossible to escape poverty.
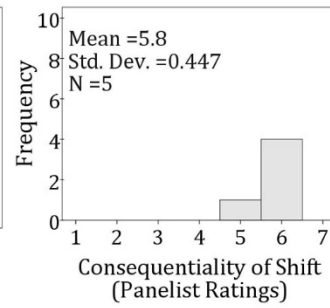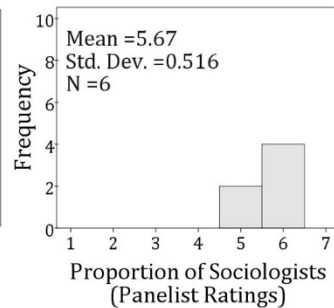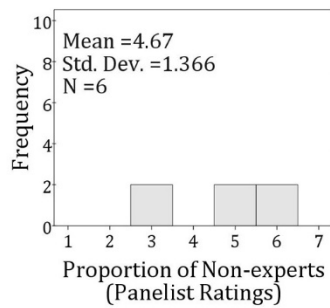
Mean =4.67
Std. Dev. =1.366
N =6

Proportion of Non-experts
(Panelist Ratings)

Mean =5.67
Std. Dev. =0.516
N =6

Proportion of Sociologists
(Panelist Ratings)

Mean =5.8
Std. Dev. =0.447
N =5

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
One panelist commented that they thought "it is merely improbably difficult, not impossible, to escape poverty."

Other Issues:
None for this item.

Figure 11.15: In-Progress Sociology Comparison 6

Many non-experts think that people:

are open to new ideas if they are good ones.

But Sociologists think that people:

are very unlikely to adopt new ideas even if they are really good ones unless there is mass pressure to do so and the new ideas do not challenge key identity markers.

Mean =5.57
Std. Dev. =0.976
N =7

Proportion of Non-experts
(Panelist Ratings)

Mean =5.29
Std. Dev. =0.756
N =7

Proportion of Sociologists
(Panelist Ratings)

Mean =5
Std. Dev. =0.816
N =7

Consequentiality of Shift
(Panelist Ratings)
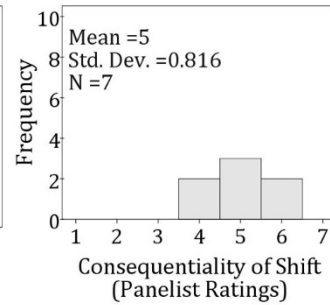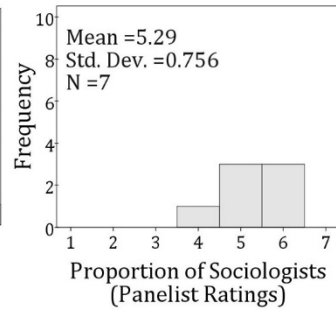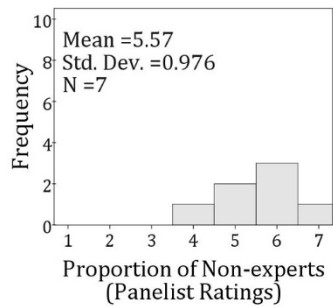
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
Panelist commented "what makes an idea good?" Another was unsure what was meant by "key identity markers."

Other Issues:
The Sociologist statement was created by combining an original suggested comparison (the mass pressure part) and a proposed rephrasing (the key identity markers part). In hindsight, this item probably should have been split into two items rather than combined.

Figure 11.16: In-Progress Sociology Comparison 7

Many non-experts think that people:

should maintain a singular understanding of the concept of family (the nuclear family) and should strive to achieve it.

But Sociologists think that people:

have different forms of family, based on social-structural factors, and that there is no "ideal" form of family.

Mean = 4.86 Std. Dev. = 0.9 N = 7	Mean = 5.71 Std. Dev. = 0.951 N = 7	Mean = 5.14 Std. Dev. = 0.69 N = 7
Proportion of Non-experts (Panelist Ratings)	Proportion of Sociologists (Panelist Ratings)	Consequentiality of Shift (Panelist Ratings)
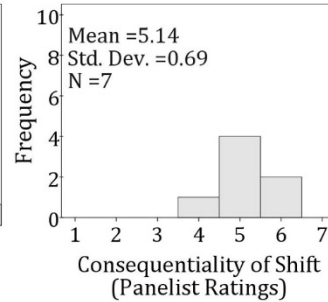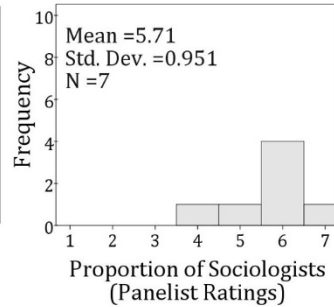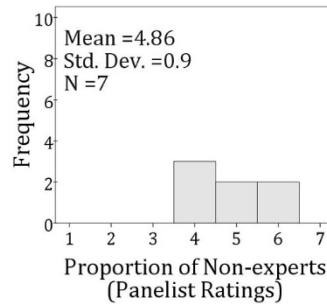
Stable: Insufficient N to determine

Clarifying Panelist Comments:
One panelist commented: "given how our world works, we do need normative commitments to caring about each other and our children." They also mentioned that "the terms above are slippery and unclear."

Other Issues:
None for this item.

Figure 11.17: In-Progress Sociology Comparison 8

Many non-experts think that people:

take action on the basis of psychological factors oriented around personal self-interest.

But Sociologists think that people:

take action on the basis of psychological and sociological factors, including rational and nonrational aspects of decision making.

Mean =5.14
Std. Dev. =1.215
N =7

Proportion of Non-experts
(Panelist Ratings)

Mean =5.71
Std. Dev. =0.756
N =7

Proportion of Sociologists
(Panelist Ratings)

Mean =5
Std. Dev. =1
N =7

Consequentiality of Shift
(Panelist Ratings)
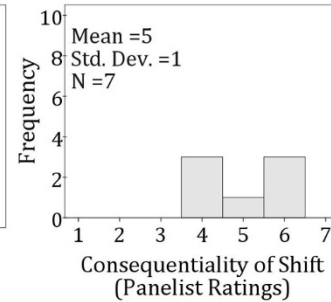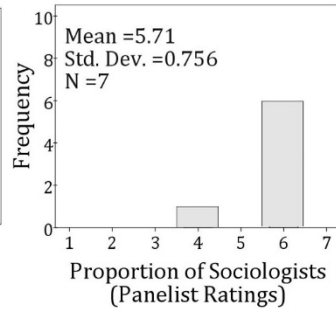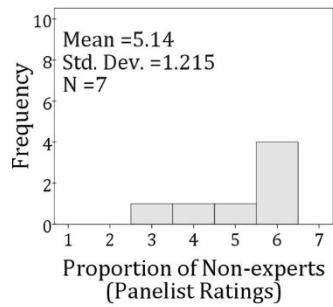
Stable:
Insufficient N to determine

Clarifying Panelist Comments:
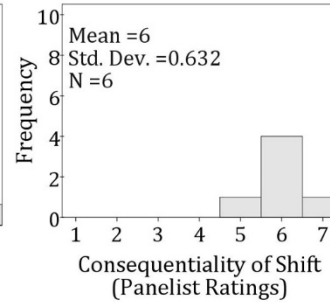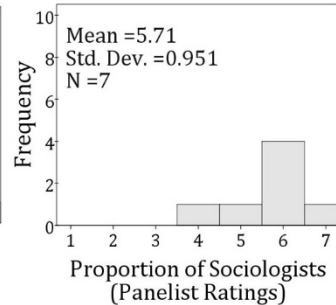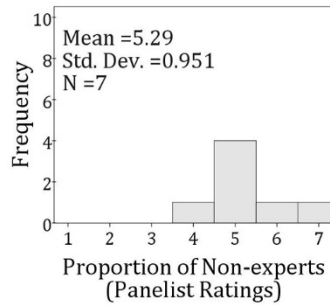None for this item.

Other Issues:
None for this item.

Figure 11.18: In-Progress Sociology Comparison 9

212

Many non-experts think that people:

who are religious are part of social institutions unlike another social institutions.

But Sociologists think that people:

who are "religious" are members of a social institution not substantially different from many other social institutions.

Mean =5.29
Std. Dev. =0.756
N =7

Mean =5.71
Std. Dev. =0.951
N =7

Mean =4.67
Std. Dev. =0.816
N =6

Stable:
Insufficient N to determine

Proportion of Non-experts (Panelist Ratings)

Proportion of Sociologists (Panelist Ratings)

Consequentiality of Shift (Panelist Ratings)
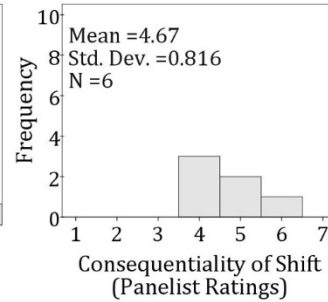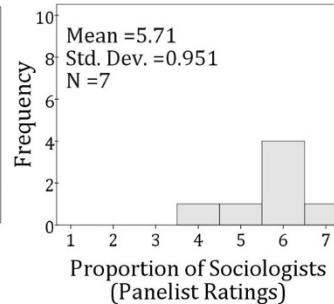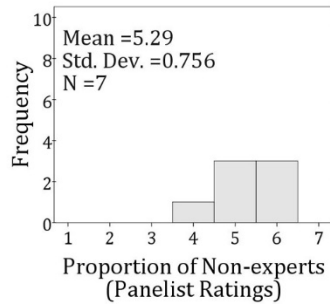
Clarifying Panelist Comments:
Panelist mentioned the phrasing was awkward, and should read "other social institutions" for both items.

Other Issues:
This was a proposed revision of an earlier version that was not parallel (the original compared people to institutions). However, it is likely that the panelist intended "other" rather than "another."

Figure 11.19: In-Progress Sociology Comparison 10

Many non-experts think that people:

who are homeless are generally drug addicts and people with serious psychiatric disorders.

But Sociologists think that people:

who are homeless are people whose circumstances disallow them access to stable, affordable housing.
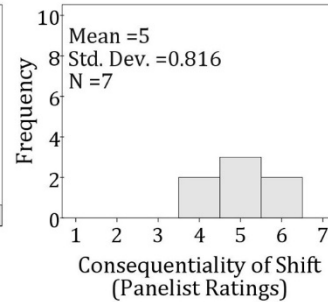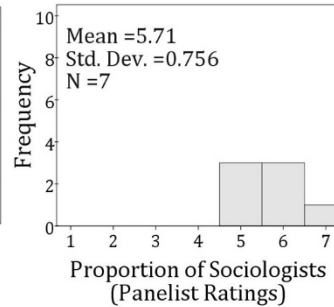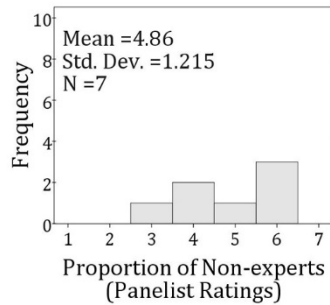
Mean =4.86
Std. Dev. =1.215
N =7

Frequency

Proportion of Non-experts
(Panelist Ratings)

Mean =5.71
Std. Dev. =0.756
N =7

Frequency

Proportion of Sociologists
(Panelist Ratings)

Mean =5
Std. Dev. =0.816
N =7

Frequency

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
None for this item.

Other Issues:
None for this item.

Figure 11.20: In-Progress Sociology Comparison 11

214

Many non-experts think that people:

are rational, acting in their own interest.

But Sociologists think that people:

sometimes act irrationally, on feelings rather than logic, but generally act to further a variety of competing interests.
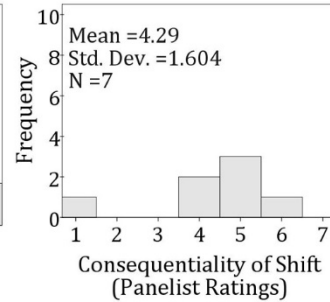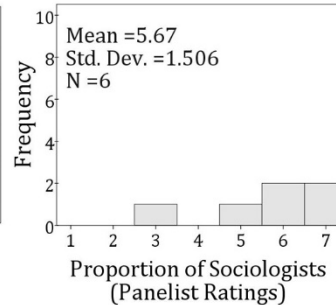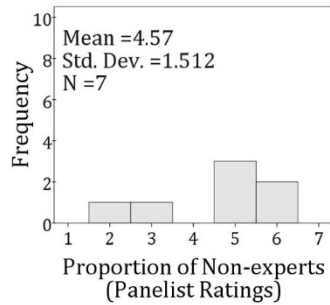
Mean =4.57
Std. Dev. =1.512
N =7

Frequency

Proportion of Non-experts
(Panelist Ratings)

Mean =5.67
Std. Dev. =1.506
N =6

Frequency

Proportion of Sociologists
(Panelist Ratings)

Mean =4.29
Std. Dev. =1.604
N =7

Frequency

Consequentiality of Shift
(Panelist Ratings)

Stable:
Insufficient N to determine

Clarifying Panelist Comments:
One panelist commented that the first statement was "a simplified principle from economics, maybe also philosophy, I'm not really sure how many lay people think this way, or even economists for that matter."

Other Issues:
None for this item.

Figure 11.21: In-Progress Sociology Comparison 12

CHAPTER 12

Agreements and disagreements across fields

In previous sections, we looked at suggested comparisons between experts and non-experts in their totality and all comparisons included both a non-expert perspective about people as well as a contrasting expert perspective. However, in order to look at agreements and disagreements across fields, in this section we are only looking at suggested ways that experts *within a field* understand people.

We do this for two reasons. First, this lets us compare how fields understand people, regardless if fields disagree about how non-experts understand people or whether the difference between experts and non-experts is considered consequential. That means that expert statements that were part of comparisons dropped from the study may nonetheless appear here, as long as these statements were well-rated by panelists as representative of researchers in their fields. Second, this greatly simplifies comparing statements for each field, as each statement now only has one component rather than two.

However, many expert statements are difficult to interpret by themselves, as they were initially written as contradicting a non-expert statement (i.e. non-experts think that people work in some particular way, but we do not). For statements such as these, where knowledge of the non-expert statement is necessary to understand the expert statement, we have added a brief summary of what the expert statement was intended to contrast with in order to provide context.

Given the difficulty of determining whether different statements are discussing the same concepts, we have included the actual statements in this section, so that readers may

judge for themselves if statements truly agree or disagree with one another. Some statements, if they contain two different concepts that each overlap with concepts from other fields, may appear twice in these comparisons.

Note that in this section, we are limited only to those concepts that panelists mentioned at some point during their panel (either as a comparison or as a comment elsewhere). If a field never mentioned a concept during the study, we cannot state how the field positions itself in regard to a particular concept. This may mean that a concept is not a focus of their field, or it may simply mean that no panelist happened to suggest or comment on that concept. As such, this section offers a glimpse into some conceptual agreements and disagreements across fields that appeared over the course of this project but should not be understood as a comprehensive summary of all such debates.

Agreements across fields

Several concepts did span across fields. That is, different fields independently mentioned that certain ways of looking at the human world were important to their fields.

Essentialism

Many of the statements about how social scientists think about the human world differently than non-experts centered around the concept of essentialism. Social scientists largely rejected the idea that people were driven to act in certain ways because of their identities, instead arguing that people are complex:

- "Anthropologists think that people: are complex and that human behavior is not easily explained by any one factor" ($M = 6.67$, $SD = 0.65$, $N = 12$).

- "Geographers think that people: 's behaviors are not only shaped by their identities, but also by space and place. In other words, peoples' environments play a significant role in shaping behaviors and identities" (in contrast to being shaped by their identities) ($M = 6.33$, $SD = 0.58$, $N = 3$).

- "Historians think that people: are best understood by reference to the multiple contexts within which they live and act" (in contrast to a single explanatory principle) ($M = 6.20$, $SD = 0.84$, $N = 5$).

- "Sociologists think that people: who share a racial, gender, or sexual category are much more heterogeneous (dissimilar) than they might at first appear, and the categories themselves change" (in contrast to people being fundamentally understandable by their racial, gender, or sexual categories). ($M = 6.00$, $SD = 0.50$, $N = 9$).

   Likewise, several fields rejected the idea that people could be easily divided into "good" and "evil":

- "Anthropologists think that people: display a range of acceptable (good) and unacceptable (bad) behaviors, which are defined differently by various cultural groups (and in different situational contexts)" (in contrast to people being good or evil based on universal morals) ($M = 6.33$, $SD = 0.49$, $N = 12$).

- "Historians think that people: cannot be easily summarized and are complex" (in contrast to being either "good" or "evil") ($M = 5.63$, $SD = 1.60$, $N = 8$).

- "Psychologists think that people: act in prosocial or antisocial ways depending on the situational pressures" (in contrast to being either inherently good or evil) ($M = 5.22$, $SD = 0.67$, $N = 9$).

Several fields rejected the idea of biological race and emphasized that people are biologically similar to one another:

- "Anthropologists think that people: cannot be classified into different races because socially-constructed racial categories have an arbitrary and variable relationship to biology, given the lack of substantial biological variation between populations and the continuous nature of what variation does exist" ($M = 6.75$, $SD = 0.45$, $N = 12$).

- "Geographers think that people: are genetically similar and therefore the concept of race is invalid" ($M = 5.00$, $SD = 1.73$, $N = 3$).

- "Historians think that people: are very uniform biologically, and that visible "racial" differences have no intrinsic significance" ($M = 5.25$, $SD = 2.05$, $N = 8$).

- "Sociologists think that people: who share a racial, gender, or sexual category are much more heterogeneous (dissimilar) than they might at first appear, and the categories themselves change" (in contrast to people being fundamentally understandable by their racial, gender, or sexual categories) ($M = 6$, $SD = 0.5$, $N = 9$).

The Anthropology, History and Sociology panels all emphasized that race is socially constructed. For example, the Historian panel rejected the idea that people in the past were universally less tolerant and argued that people "acquire and express such mistrust in specific historical situations and for reasons that vary with historical circumstances" ($M = 6.00$, $SD = 0.76$, $N = 8$).

Likewise, several fields rejected simple biological dichotomies of male and female, instead emphasizing variation:

- "Anthropologists think that people: have variation in both sex (phenotypes) and gender" (socially-defined roles) (M = 6.67, SD = 0.49, N = 12).

- "Psychologists think that people: have characteristics that vary along a continuum" (in contrast to being either male or female) (M = 5.33, SD = 0.5, N = 9).

- "Sociologists think that people: become "female" or "male" through a combination of their hormonal, chromosomal, social, and environmental experiences" (in contrast to being born definitively male or female) (M = 5.78, SD = 1.09, N = 9).

Two fields rejected the idea that cultures themselves have unchanging essences, instead describing culture as something that can be modified and reshaped, and that doing so was how people could stay resilient in the face of environmental or social upheaval:

- "Anthropologists think that people: 's cultural identity is resilient and can be reshaped yet persist over generations, especially in the face of adversity" (in contrast with being fragile and easily lost) (M = 6.00, SD = 0.63, N = 11).

- "Geographers think that people: can adapt to changes and modify their culture" (M = 4.67, SD = 0.58, N = 3).

Other forms of essentialism were rejected as well, but these did not cross fields. The Psychology panel rejected what might be called biological destiny, emphasizing that people "become who they are in part due to the circumstances they encounter" (M = 5.11, SD = 1.05, N = 9). Similarly, the Sociology panel also rejected the idea that personalities have immutable qualities, instead arguing that people "are dynamic in their personalities, experiencing changes in who they are based on life experiences and contact with other

groups of people" (see Figure 11.8). The Sociology panel also rejected the idea that disadvantaged groups were disadvantaged because of essential qualities.

The Political Science panel rejected the idea that people can be "neatly divided into racists or non-racists" (see Figure 9.13). Likewise, the Political Science understanding of identity seemed to reject essentialistic understandings of people (see Figure 9.6), and one panelist commented: "Many political scientists don't see identity as being terribly important or, when they are, they are not intrinsic to people but malleable and manipulatable be elites."

Rejection of group bias

Several of the comparisons suggested by panelists concerned non-experts tending to be biased towards their own group and tending to unfairly dismiss other groups and their behaviors. Two panels suggested people naturally support their in-groups:

- "Anthropologists think that people: are fundamentally driven to identify themselves as a member of one group and not others, which underlies our capacity for both conflict and cooperation, but which can also be reshaped through experience, culture, education, and public sentiment" ($M = 5.78$, $SD = 0.44$, $N = 9$).

- "Political Scientists think that people: are mostly partisans, even if they claim otherwise" (in contrast to being political independents) ($M = 5.20$, $SD = 0.84$, $N = 5$).

Two panels suggested that non-experts are uncharitable when explaining the behaviors and beliefs of people in other cultures:

- "Anthropologists think that people: in other cultures do things that make sense in the context of those cultures" (in contrast to doing crazy things) ($M = 6.73$, $SD = 0.47$, $N = 11$).
- "Historians think that people: who believe things we think of as superstitious have those beliefs as part of coherent ways of understanding the world" (in contrast to being ignorant or stupid) ($M = 5.80$, $SD = 0.45$, $N = 5$).

However, it is not entirely clear whether Historians generally think people are naturally biased against outgroups. One Historian supported this idea, commenting that "People are inclined to be biased agains [*sic*] people from other groups, but how the groups are constituted varies." However, some statements from the Historian panel (for example, see Figure 8.2) imply that people learn to be biased. This could imply that Historians think that people could, in theory, not be biased at all. However, this could equally well imply (in accordance with the previous quote) that Historians think that people are always biased in some manner, just not in any way consistent across time and space.

Two panels also suggested that non-experts often think that their own family structure is natural and correct and ignore cross-cultural variety in family structures:

- "Anthropologists think that people: live in family structures that vary cross-culturally" (in contrast to naturally living in one family structure) ($M = 6.70$, $SD = 0.48$, $N = 10$).
- "Sociologists think that people: have different forms of family, based on social-structural factors, and that there is no "ideal" form of family" (in contrast to a singular understanding of family) ($M = 5.71$, $SD = 0.95$, $N = 7$).

Two panels also rejected the idea that some groups are simply "better" than others:

- "Anthropologists think that people: are biologically similar, but culturally different with no culture considered more advanced than another" ($M = 6.39$, $SD = 0.60$, $N = 9$).

- "Historians think that people: in the "civilized" world came to dominate everyone else for very specific reasons" (in contrast to naturally dominating) ($M = 4.4$, $SD = 1.52$, $N = 5$).

However, an Anthropologist commented that while groups do differ in terms of technological achievement, they emphasized that "That does not mean the culture is somehow 'better' but each has adapted according to its environment and needs, and access to technologies." Likely related to this, the Anthropologist panel rejected the idea that any group of humans (or their languages) are "primitive" (see Figures 5.4 and 5.6). The Sociology panel also suggested that that people were more wary of strangers and immigrants than was warranted, which also potentially taps into in-group biases (see Figures 11.5 and 11.11).

It is unclear where Economics falls on this question. However, the comparison "Many non-experts think that people: in one nation only gain at the expense of people in other nations when free trade is introduced. But Economists think that people: in both countries could benefit from free trade" (Figure 6.1) is interesting because it implies that people see economic exchange as a competition between groups and may view economic exchange with in-groups as inherently different from economic exchange with out-groups. This is in line with Rubin (2003) who argues that people's intuitive economics presents

people from understanding economics well, and that therefore educating people about economics would likely lead to increased trading.

People are part of the natural world.

A handful of comparisons imply that social scientists think that people are more of a part of the natural world than non-experts tend to think. One aspect of this is that people are animals and descended from other animals:

- "Anthropologists think that people: are primates, and as such are products of that lineage" ($M = 6.40$, $SD = 1.26$, $N = 10$).

- "Psychologists think that people: are animals, though more complex and intellectually advanced, and possessing psychological mechanisms that make us uniquely human" ($M = 6.18$, $SD = 0.60$, $N = 11$).

While both of these statements were considered representative of experts in their respective fields, they were both dropped as comparisons during the course of the study. For the Anthropologist statement, the contrasting non-expert statement "Many non-experts think that people: are not animals, and are exceptional in that our understanding of how other species evolved and how they think and behave does not apply to us" was rated as not representative of non-experts. However, it was unclear why Anthropologists thought this was an unrepresentative non-expert view or what they thought was a more popular non-expert view.

For the Psychology panel, though they did rate this comparison as representative of Psychologists, there was some disagreement about this comparison. While rephrasing this statement, two Psychologists suggested rephrasings that emphasized that people had

superior abilities to animals ("more complex…with more flexible brains," "more intellectually advances [*sic*]"), two suggested rephrasing that emphasized that people were different without comparing superiority ("different psychological mechanisms," "particular set of skills and abilities.") and one took issue with the idea of superiority ("the definition of "sophisticated" depends on environmental demands").

While no fields argued against the position that humans are part of the natural world in this way, some panelists mentioned that their field has not yet taken a strong position on humanity's evolutionary history. One Sociologist commented: "It's not clear to me that Sociologists in general think about evolution and a possible genetic basis vs technical/structural factors. I don't feel like the field of Sociology offers specific principles on how to understand human evolution or the lack thereof." Likewise, one Economist commented: "I think evolutionary history is probably under[-]considered in economics but considered by some."

Two fields also suggested that people think they are more removed from the natural world than social scientists do:

- "Anthropologists think that people: are always engaged in a dynamic coupling with natural systems, despite considerable technological achievements" (in contrast with being separate from the natural world) ($M = 6.27$, $SD = 0.79$, $N = 11$).

- "Geographers think that people: are deeply connected to and reliant upon the natural environment through their resource use and consumption habits" (in contrast with being separate from natural environments) ($M = 6.33$, $SD = 0.58$, $N = 3$).

However, panelists in these panels did mention that people living in rural areas are less likely to think of themselves as separate from the natural world compared to urban dwellers. This accords with the idea that intuitive theories are partially experiential.

People underestimate the importance of context

The Geography and Sociology panels suggested that people underestimate their interconnectedness with others:

- "Geographers think that people: operate within complex and interconnected systems" (in contrast to acting as individuals) (M = 6.33, SD = 0.58, N = 3).

- "Sociologists think that people: are interdependent" (in contrast to self-reliant) (M = 6.33, SD = 0.50, N = 9).

Both the Psychology and Sociology panels likewise thought that people overestimate their agency, in particular their ability to determine whether or not they succeed in life:

- "Psychologists think that people: need lots of things (such as luck and circumstances) to go their way in order to succeed; willpower alone is insufficient" (in contrast to succeeding through willpower) (M = 5.70, SD = 0.67, N = 10).

- "Sociologists think that people: of different backgrounds face different obstacles to success, many outside of their control" (in contrast to succeeding through hard work) (see also figures 11.1 and 11.10) (M = 6.56, SD = 0.53, N = 9).

The Psychology panel also suggested other ways that people tend to overestimate agency, including that non-experts overestimate how consistent people are across situations. Interestingly, some psychologists went even further in questioning agency,

echoing eliminative materialism (discussed previously). One Psychologists suggested the comparison: "Many non-experts think that people: have free will. But Psychologists think that people: are influenced by the environment and other factors beyond their conscious control, so that free will is not actually occurring." This was considered weakly Psychologists (M = 4.44, SD = 1.33, N = 9), though one panelist noted that this was an academic understanding, and that they "suspect most psychologists believe in free will, and would only endorse a deterministic view if really asked to consider it thru their scholarly lens." Relatedly, Political Scientists emphasize that leaders are more restrained in their agency than many people think (see Figure 9.2).

People likely overemphasize agency because our "theory of mind" would tend to cause us to overemphasize intentions in our explanations of the world, as it requires little mental effort for us to think in those terms and thinking in those terms will feel correct to us.

People resist changing their minds

Three fields mentioned that people are resistant to changing their beliefs or opinions, even when doing so is warranted.

- "Geographers think that people: have opinions based on a wide range of factors which may or may not have to do with factual information (in contrast to changing their opinion based on facts)" (M = 5.00, SD = 1.00, N = 3).

- "Psychologists think that people: have many ways to defend their beliefs, therefore attitudes are highly stable and difficult to change." (M = 5.42, SD = 1.31, N = 12).

- "Sociologists think that people: are very unlikely to adopt new ideas even if they are really good ones unless there is mass pressure to do so and the new ideas do not challenge key identity markers" ($M = 5.29$, $SD = 0.76$, $N = 7$).

The Psychology panel further argued that people are self-serving in their beliefs and are motivated to seek out supporting information. This seems to be a way of thinking about people characteristic of Psychology, and which may have spread to other fields. One Anthropologist explained the human tendency to resist changing their minds in terms of "cognitive dissonance" (a term originally from Psychology):

> While humans have the capacity to use evidence and logic, opinions are often emotionally embedded and difficult to change, even when evidence is to the contrary. Humans have a high capacity for cognitive dissonance, which allows people to conform to social norms and agreed-upon values and beliefs, even when evidence or logic indicates these are inaccurate or unhelpful ways of approaching reality.

However, panelists also mentioned that people do sometimes change their minds. One Sociologist commented that they see people as being between rational and irrational. An Anthropologist also mentioned this depended on perspective, commenting that (capitalization as in original) "People of course use different logic and reasoning, which may seem rational to THEM but irrational to other people in another society or cultural context."

Opposites do not attract

Both the Psychology and Sociology panel rejected the idea that "opposites attract" and argued that people tend to prefer others like themselves:

- "Psychologists think that people: who are similar are more likely to be attracted to one another" ($M = 5.18$, $SD = 1.47$, $N = 11$).

- "Sociologists think that people: are attracted to people who share their interests, personality types, etc." ($M = 5.18$, $SD = 1.66$, $N = 11$).

Both statements were contrasted with the idea that opposites attract, but both comparisons were dropped as inconsequential. Panelists also mentioned that opposites can attract but are less likely to lead to stable long-term relationships. Panelists did not explain why non-experts thought that opposites attract, though they did note that non-experts also thought that "birds of a feather flock together."

Disagreements across fields

Some fields did suggest ways of understanding people that seemed to directly contradict ways suggested by other fields. In particular, the History and Economics panels stood apart from the other fields in how they understood people. For History, this centered around the universality of human nature. For Economics, this centered around how people made decisions.

Some of these debates seem to be fields talking past one another, as they interpret terms in different ways. However, there does seem to be some underlying differences in how people are conceptualized across fields. Consequently, we would argue that these areas where we should expect to see movement within the social sciences in the near

future as terminology is clarified and opposing models are tested. Resolutions of these debates may end with one side becoming dominant, or with both sides being rejected in favor of a new perspective that supersedes both.

Is there a human nature?

Panelists across fields disagreed about whether people could be said to have a shared human nature, though this was also often intertwined with whether people could be said to be more similar or more different to each other (and for History in particular, questions about what motivates people). The History panel strongly rejected the concept of human nature:

- "Historians think that people: see the world differently and want different things from it in different cultures and periods (i.e. human nature is not always the same)" ($M = 5.00$, $SD = 0.82$, $N = 4$).

However, several other fields suggested ways of understanding people that imply some form of shared human nature:

- "Anthropologists think that people: share many cultural features (such as language, religion, kinship, marriage, etc), although manifested in different ways across cultures" (in contrast to being completely different across cultures) ($M = 6.70$, $SD = 0.48$, $N = 10$).
- "Geographers think that people: are basically the same at the core and differ only in superficial ways, shaped by social, structural, and environmental conditions" ($M = 4.67$, $SD = 2.31$, $N = 3$).

- "Psychologists think that people: are more similar (than different) to each other, in that people will generally respond in the same way to stimuli or act in accordance to rules of their culture" (in contrast to being unique) ($M = 5.00$, $SD = 0.94$, $N = 10$).

    One Anthropologist elaborated on this understanding further:

    People are similar in broad attributes across time and location (such as having a family structure, having a spirituality, and having an economy that provides ways to exchange resources). People differ widely in the details of how they meet these similar needs and structures.

    Both Anthropology and Psychology made it clear that while they think people are similar to each other, they rejected the idea of nature alone as a sufficient explanation for human behavior:

- "Anthropologists think that people: behave in ways that are reflections of complex interactions between culture and biology" ($M = 6.33$, $SD = 0.71$, $N = 9$).
- "Psychologists think that people: have behaviors/traits/etc. that are caused by an interaction of genes and environment" ($M = 6.33$, $SD = 0.78$, $N = 12$).

    One Anthropologist elaborated on how nature and nurture interact to create behavior: "People and non-human animals all "learn" how to navigate their social and physical environments, but do not forget that they use learning "rules" and "biases" that are themselves (in part) genetically programmed."

    Interestingly, some Historians disagreed with the Historian view that human nature does not exist, instead espousing views similar to those above in which inherent similarities between humans play out differently in different contexts. Given this, it is

possible that Historians tend to understand "human nature" in terms of the debate over "nature verses nurture," which is roughly equivalent to biological essentialism verses cultural essentialism. Thus, it is possible that Historians are rejecting biological essentialism rather than the more modern understanding of human nature that emphasizes the indivisible contributions of nature and nurture and that nature both drives humans to create culture and be profoundly shaped by it. It is not currently clear whether Historians would also reject this understanding of human nature.

It is not clear how Economics as a field approaches this issue, however, one Economist described an understanding similar to that of the other fields: "I would argue folks are the same in that they seek what is best for them. However, the framing of that goal will depend upon local tastes which are influence by culture and other local conditions."

It is not clear where Political Science stands on this issue, as neither human nature nor whether people are similar came up in any suggested comparisons, nor did panelists comment on this issue elsewhere.

The Sociology panel seemed to not oppose the idea of human nature but tended to not find the question relevant to their field. One Sociologist, discussing the study as a whole, commented: "I also am suspicious of the degree to which sociologists actually talk about human nature as opposed to specific mechanisms of various locally situated social processes. That's more something for psychologists or anthropologists than sociologists." Another Sociologist commented: "I think most would say that talking about human similarities is a bit like oncologists talking bout [*sic*] a thing called cancer. Sure, fine, it's a

thing, but what's really interesting is how the differences work." This emphasis on differences is similar to a comment made in the History panel: "Historians generally don't explain people/humanity in a broad way. We deal in specifics." This implies that these two fields are generally more interested in differences compared to the rest of the social sciences.

Are people self-interested?

Several fields rejected the idea that people are primarily motivated by self-interest. However, the Economics panel, did consider self-interest to be the primary motivation of people:

- "Economists think that people: are primarily, but not entirely, motivated by self-interest" ($M = 6.29$, $SD = 0.76$, $N = 7$).

Other fields tended to disagree with this, downplaying the importance of self-interest (particularly economic self-interest):

- "Historians think that people: act on the basis of values and goals that are not limited to economic interest" (in contrast to acting on the basis of economic self-interest) (see also figure 8.11) ($M = 5.20$, $SD = 0.84$, $N = 5$).
- "Political Scientists think that people: vary greatly in terms of what their motivations are" (in contrast to being primarily interested in benefiting themselves) ($M = 4.67$, $SD = 0.52$, $N = 6$).
- "Psychologists think that people: have many goals that are more important, including feeling that they are respected and doing something worthwhile" (in contrast to money) ($M = 5$, $SD = 1.41$, $N = 12$).

233

- "Sociologists think that people: take action on the basis of psychological and sociological factors, including rational and nonrational aspects of decision making" (in contrast to personal self-interest) (M = 5.71, SD = 0.76, N = 7).

When debating suggested comparisons within their own field, several panelists mentioned that viewing people as completely self-interested was characteristic of Economics. This debate extended within the fields as well and panelists commented that people in their own fields trained in Economics tended to disagree with those in their fields trained in other fields, such as Psychology, in how to understand human motivation. However, as Economics does not seem to argue that people are entirely self-interested, this implies that the social sciences may differ in degree rather than kind, but that they are not currently aware of this. Economists also likely think about self-interest in broader terms than just finances. This is implied by one suggested comparison (dropped due to low ratings for consequentiality) in the Economics panel: "Many non-experts think that people: derive happiness from money. But Economists think that people: derive happiness from the freedom that money represents" (M = 4.75, SD = 1.82, N = 12).

While Economics does seem to emphasize self-interest more than the other fields, it is not currently possible to disentangle this debate further, given that researchers differ in how broadly they conceptualize self-interest (for example, whether values are also a form of self-interest). One Anthropologist commented: "People are motivated by a variety of things that act in a complex relationship with each other because humans are social animals. This means that self-interest overlaps considerably with group interest (and these may include several different competing or collaborating groups)." Likewise, a Sociologist

commented: "If following social norms keeps us from experiencing social condemnation, I guess you could argue that we are motivated to follow social norms out of self-interest in not experiencing social sanctions." It is unlikely that this debate will be resolved until self-interest can be specified more precisely.

Do people act in their own best interests?

The Economics panel differed from other fields in that they saw people as consistently acting in terms of their own best interests:

- Economists think that people: almost always act in their own best interests, based on the information and resources they have at the time of the decision ($M = 5.60$, $SD = 0.55$, $N = 5$).

In contrast, the History and Psychology panels did not see people as acting in their own best interest, because they do not see people as knowing what their best interest is:

- "Historians think that people: do not usually know what is in their interest and act at the behest of a wide array of cultural impulses" ($M = 5.50$, $SD = 0.58$, $N = 4$).

- "Psychologists think that people: are pretty bad at understanding themselves and knowing how best to get what they want" ($M = 5.08$, $SD = 1.31$, $N = 12$).

One Historian added that people: "are seldom motivated by rational self-interest and, in any case, typically cannot discern exactly what it would be in their interest to do."

For Psychology, part of this may be due to Clinical Psychologists seeing people when they are struggling. As one Psychologist commented: "I would argue these are main reasons people seek therapy." Similarly, a Psychologist commented that whether people act in ways that are likely to achieve their desired outcomes "Depends on one's level or degree of

psychopatholog[y]." However, another Psychologist commented that Social Psychologists (who do not see patients) would also think in this way, so this is likely a way of thinking about people broadly characteristic of Psychologists.

For Economists, this was a perspective with important implications. Economists rejected the idea that people should have their lives decided by experts, as they argue that people "are qualified and properly motivated to act in their own best interests" ($M = 5.43$, $SD = 0.79$, $N = 7$). Economists commented that this was because "Individuals have more-personal information than someone else, how would another person know more about others?" However, one panelist qualified the Economists view, commenting: "I think this depends greatly on the context. Few economists would say this is universally true. Many would say this is generally true."

Unsurprisingly, as Political Science was described by panelists as having groups influenced by either Psychology or Economics, it seemed split on this question. For example, one Political Scientist suggested a comparison that rejected the idea of people being instrumentally rational, while another Political Scientist commented that the opposite was true. Another Political Scientist implied that people try but do not always succeed in advancing their interests, commenting that "People act in ways that are most likely to further their interests but the nature of politics is that there are always some winners and some losers." However, Political Scientists did agree that people rarely understand the world well enough to make informed decisions about it (see figure 9.5), and it is hard to see how uninformed decisions would help people achieve their goals.

When asked whether people tended to act in ways that were likely to achieve their desired outcomes, a Sociologist commented "People may not know what their desired outcome is. Also, they may act in ways likely to achieve some desired outcomes and unlikely to achieve others. People have complex and competing interests." This is somewhat similar to the view espoused by the History and Psychologist panel.

However, an Anthropologist seemed to be closer to the view of the Economics panel, commenting:

> People are often seeking to achieve multiple goods that interact in complex and sometimes conflicting ways. People generally act in ways to maximize desired outcomes, but this may come at a cost to another desired outcome and seem, without all the contextual information, to be self-sabotaging. Even self-sabotaging or self-harming behavior usually serves the actor in some way that is understandable if thoroughly studied."

Do people make choices carefully?

The Economics panel seemed to consider human decision making more deliberative than other fields did. The following statement was rated as being representative of Economists, though only weakly and with several panelists questioning this statement:

- "Economists think that people: make every choice by weighing the costs and benefits of all available choices and choosing the best available option" (in contrast to making choices without much consideration) ($M = 4.57$, $SD = 1.62$, $N = 7$).

Other fields disagreed, emphasizing that people often made decisions in an unconscious and suboptimal manner:

- "Anthropologists think that people: make decisions based on partial and selective information, in which predictive ability is limited, based on emotion and cultural "rules of thumb" that they have learned" (M = 5.67, SD = 0.87, N = 9).

- "Geographers think that people: often respond in seemingly irrational ways to environmental threats and hazards due to imperfect knowledge, conflicting goals, or larger structural constraints" (in contrast to responding rationally) (M = 6.00, SD = 0.00, N = 3).

- "Political Scientists think people: make decisions (such as who to vote for) by relying on heuristics, intuitions, and predispositions" (M = 6.20, SD = 0.45, N = 5).

- "Psychologists think that people: rely on emotions, cognitive shortcuts, and stereotypes" (in contrast to being rational and logical) (M = 5.75, SD = 0.75, N = 12).

- "Sociologists think that people: are mostly creatures of habit and emotion" (in contrast to acting strategically) (M = 4.78, SD = 1.20, N = 9).

Two fields emphasized that they see people as primarily acting unconsciously and habitually, which would tend to preclude careful and strategic behavior:

- "Anthropologists think that people: behave unconsciously and habitually, on the basis of acquired cultural patterns of thought and behavior" (in contrast to being fully conscious of their attitudes and behaviors) (M = 5.18, SD = 1.40, N = 11).

- "Psychologists think that people: engage in behaviors and thoughts that are mostly unconscious" (M = 4.25, SD = 1.42, N = 12).

However, when asked whether people tended to make decisions consciously or unconsciously, the Anthropology, Geography, Psychology, and Sociology panels all clarified

that their fields believe that people do think in *both* conscious and unconscious ways. This was described as people being deliberative for some decisions and impulsive for others (though tending to make important decisions more deliberately), but this was also described as conscious and unconscious ways of thinking working in tandem. One Anthropologist commented:

> People generally make decisions in a two-part process: it begins with unconscious assumptions that guide elimination of options (which are never consciously considered) and then further refined through conscious and deliberative decision making of apparent options. Both parts of the decision making process are guided by both emotion and reason; these are not mutually exclusive cognitive tasks.

> However, another Anthropologist commented that they "would not chracterize [*sic*] human decision-making as 'impulsive,' because the factors that shape decisions can be studied and understood in terms of the broad patterns that characterize individual and social behaviors." In our view, this argument was reminiscent of how Economists discussed decision making. Given this, and given that Economists considered this item only slightly representative of Economists and (as discussed in the divisions within Economics section) that Economists themselves disagree about how deliberative people are when they make decisions, it is likely that this is not as much of a point of disagreement across fields as it may have been in the past.

CHAPTER 13

Agreements and disagreements within fields

Panelists were also asked to rate whether their field is best thought of as sharing a single way to understand people or whether their field tends to have multiple competing views. Panelist ratings for this question can be seen in Figure 13.1.

Panelists were also asked about the trajectory of their fields regarding their understanding of people. That is, whether their fields were in the process of converging on a single understanding of people or whether they were in the process of diverging instead. Panelist ratings for this question can be seen in Figure 13.2.

Below, we have summarized panelists' comments about the divisions within their fields. These comments are drawn from many places across the study, including comments on other questions and comments panelists made about such divisions during the actual process of rating and commenting on comparisons.

Given the complexity of the divisions within fields, we have divided the panelists' responses by field. Note that as these sections are drawn from panelist comments they reflect divisions that panelists felt were important but do not cover all divisions within a field. Furthermore, given that panelist comments were often ambiguous or lacking in detail, these summaries cannot be considered a full accounting of the divisions presented.

Anthropology

Though Anthropologists often use the four field approach (Archaeology, Cultural Anthropology, Physical/Biological Anthropology, and Linguistic Anthropology) to divide their field, the major division described by panelists was a philosophical one. Several

panelists described a profound division between "science-minded Anthropologists" and

"humanist-minded ones", arguing that "Anthropologists start from different paradigms in

creating knowledge that can be philosophically very different and in opposition." This

division was "why the Society for Anthropological Sciences formed and many biological

anthropologists stopped integrating with the AAA[12] -- not because of *field* but because of

theoretical paradigmatic *orientation* -- that is, whether anthropology is a science, a

humanities, or both at once." These two groups have different goals, as "Science-minded

anthropologists have often looked for general trends and theory building to explain human

behavior and evolution. Humanist-minded anthropologists have often focused on specific

traits of specific cultures, and try to understand the meaning of these things in their own

context." This panelist described the split as "fairly major" but said that "the split is healing

in recent years." However, another panelist divided the field into three approaches, and

argued that "There is an ongoing divergence between materialist, interpretivist, and

evolutionary approaches."

> One panelist rejected the idea of divisions within the field:
>
> I don't think there are fundamentally different understandings that are mutually
>
> exclusive. I think it's more like a spectrum of how much there is an emphasis on
>
> individual agency and meaning as opposed to evolutionary/biological drivers in
>
> human behavior. This overlaps significantly with a methodological spectrum of
>
> empiricism on the one hand and interpretivism on the other.

---

[12] American Anthropological Association

This potential division is related to, but not the same as, a divide described as being between Cultural (particularly Sociocultural) Anthropology and Biological (particularly Evolutionary) Anthropology. While some panelists appeared to consider the division between Cultural and Biological Anthropology to be equivalent to the division between Humanistic and Scientific Anthropology, others described the debate between Humanistic and Scientific Anthropology as occurring *within* these fields as well. For example, one panelist stated that "some Biological Anthropologists agree with Evolutionary Psychology" which puts them "fundamentally at odds with many Cognitive Anthropologists (who are generally in the cultural anthro field)" while other Biological Anthropologists are not "at odds with many cognitive anthropologists at all." Likewise, while "some of the heavily humanistic cultural anthropologists are at odds with some biological anthropologists…many of the cultural anthropologists who employ and understand scientific approaches do not feel that they are differentiated substantially from biological anthropologists." Another panelist mentioned that within Cultural Anthropology, there is a difference:

> between those who think that human thought and behavior can and should be
> approached scientifically (at least some of the time) and can and should be modeled
> for description and prediction (such as cultural modeling theory, many cognitive
> and linguistic anthropologists, many applied anthropologists) and those who think
> that human thought and behavior is best approached humanistically, as a dense web
> of meaning and story.

Nonetheless, several panelists did contrast Cultural and Biological Anthropology, arguing that these fields were different in interests, methods, and philosophical orientation. Another panelist commented (ellipses as in original):

The differences are pretty real and often a point of contention among faculty/departments. I would say that it largely has to do with topics and methods (e.g., cultural anthropology tends to work qualitatively, whereas biological anthropologists and archaeologists tend to work quantitatively). We ultimately want to understand the human condition, but how we do it varies considerably...and thus do our questions.

Another panelist argued that the difference was multifaceted, and touched on interests, methods, and how they explained human behavior:

[i]n large part the difference reflects interest in different questions: what makes us different versus what makes us the same? Cultural anthropologists are interested in cultural variation and attempting to understand that variation (its history, context, political-economic forces that shape it, etc). Biological anthropologists are more interested in how similar human processes are shaped by different cultural and environmental contexts, such as hemoglobin production in low vs high altitude environments or sickle cell trait in agricultural versus non-agricultural societies. The methods used to answer these questions are largely different. That being said, there are some core philosophical differences as well--for example, how much (not if) of a behavior is learned (cultural) versus innate (biological), etc[.]

243

This division between whether Anthropologists favor nature or nurture as an explanation for human behavior was also raised by other panelists. One panelist argued: "The main difference seems to be about how much emphasis is put onto culture as autonomous to biology." Another commented that "Some people see human behavior resulting from the confluence of nature and nurture, while others stand doggedly on one side or the other (though I would argue there are few on the nature-only side of things)." Panelists considered Biological Anthropology to be more on the nature side of the divide, with one panelist commenting that "biological anthropologists agree that there are some shared characteristics of humans (some of which are shared with other primates) that are robust across cultures. For example, human prosociality or pointing, which are markedly [d]ifferent from other species but appear in all studied human societies." Another panelist argued that Biological Anthropologists are more likely than Sociocultural Anthropologists to "believe in innate sex differences." Another panelist elaborated that the divide between Anthropologists who "think that sexual differences are largely *gender* differences" and Anthropologists who think "that sexual differences are partly gender differences…is a fundamental difference in thinking among anthropologists."

One panelist argued that these divides were real but partially driven by the structure of academia, commenting that Anthropologists (emphasis as in original):

> do parcel out along the nature/nurture, cultural/biological divide" and that while Anthropologists theoretically explain human behavior as a combination of nature and nurture "few anthropologists actually take a truly balanced perspective. This is in part a function of how academia has evolved... one has to 'focus' to succeed, etc.

244

That 'third possibility' is THERE but there isn't a journal for it and so nobody REALLY does it.

However, one panelist argued that "the contrast between biological and cultural anthropology is artificial," though it was not clear why they felt this way.

Relatedly, some panelists commented on a division between "those who view people as primarily shaped by evolution and biology (such as evolutionary psychologists, sociobiologists) and those who view people as primarily shaped by culture and history (almost everyone else)." According to panelists, Evolutionary Anthropologists "think that people's individual and cultural beliefs and behaviors have been shaped by natural and cultural selection, but are flexible and strategic, and can be understood in those terms." In contrast, Sociocultural Anthropologists "think that people must be understood in terms of the structures that predispose them to behave in certain ways, and the influences of history, globalization, and modernity on those structures, in order to understand cultural diversity today."

Aside from the aforementioned debates, several other points of contention between Anthropologists were mentioned. For example, that "Osteo/forensic folks within bio anth 'believe in race' while those informed by genetics do not (nor do sociocultural anthropologists)". Another debate was over whether people should be thought of as complex adaptive systems, though this debate mainly seemed to be panelists talking past one another over terms rather than a conceptual disagreement.

One panelist described a debate about rationality within Economic Anthropology, saying that: "formalists would see economizing rationality as a human universal;

substantivists would see economic action as culturally embedded."[13] Another panelist

expanded on this, commenting that "the debate over rationality (as [universal] vs[.]

culturally specific) dominated economic anthro ca. 1955-1980s [without] being decisively

resolved."

As in other panels, the concept of rationality proved difficult to disentangle for this

panel. One panelist commented that Cognitive Anthropologists think about rationality in

nuanced terms, aspects of which are similar to how Political Scientists think about

rationality (emphasis as in original):

> The problem is how we define "rational/irrational" and "informed/uninformed."
>
> Cognitive anthropologists largely agree that people primarily make decisions based
>
> on emotion and cultural models or "rules of thumb" -- not based on a great capacity
>
> for rationally sorting through data in a predictive fashion (i.e., as many economists
>
> used to think). However, this doesn't mean that people are irrational or uniformed.
>
> It means that they often use irrational means to rational ends, letting culture (which
>
> is often emotionally charged) guide complex decision-making, which saves time and
>
> also guides people toward predictable, understandable behavioral patterns within a
>
> group. In this way, people are BOTH rational and informed (but in socially logical
>
> ways, not purely logical, data-driven ways) AND they are irrational and uninformed
>
> (they're mostly making their decisions based on emotional responses to context,

---

[13] Formalism is related to Neoclassical Economics, while Substantivism is focused on

exchange as part of social relationships embedded in a larger cultural framework.

which is culturally guided, not guided accurately by logical assessments of data).

This is why policy (and the process to form, evaluate, etc. it) is so often different

from politics (and the way that policy comes to be viewed and engaged with by the

public).

Several panelists also mentioned that the divides discussed above left out

Archaeology but did not clarify how Archaeology fit into these debates. One panelist quoted

above mentioned that Archaeologists "tend to work quantitatively." However, another

panelist commented that "Archaeologists may draw from evolutionary theory,

sociocultural theory, or both to understand the lives of people."

One panelist summarized the field of Anthropology as follows:

If you step back and look at human behavior with a broad anthropological lens, you

must come to the conclusion that much of human behavior is similar--for example,

all have languages, all have religions, all have ethnomedicine, all have folk

taxonomies, etc. This universality of cultural systems is of more interest to biological

anthropologists, but just because it is not of interest to other anthropologists does

not make it not true of human behavior. With our lens, we must also admit that

much of human behavior does vary from culture to culture--we have different

religious beliefs, different languages, different disease etiologies and treatments, etc.

This is what cultural anthropologists are more interested in. How much is similar,

and how much is different, from culture to culture is not well known or agreed upon,

and has been and continues to be an area of much debate in anthropology.

Economics

Many of the comments that Economists made were not about how Economists differ from one another, but how Economists share a viewpoint that differs from other viewpoints. One panelist commented:

Economists are different because they "tend to look at costs, particularly opportunity costs, in a [way] that others do not. I would argue that opportunity costs are fairly unique. Non economists will look at direct costs but not the value of what is given up in comparison or instead.

Another panelist considered marginality to be a key unifying concept that differentiated Economists from non-economists, arguing that: "Non economists are likely to look at total costs and total benefits and decide accordingly. Economists will look at costs and benefits but on a unit by unit basis."

One panelist commented: "Regardless of whether people are 'rational,' most Economists agree that much of people's behavior can be modeled/predicted through a lens of rationality" (Elsewhere in the study, panelists described this as the concept of people being "Predictably Irrational"[14]).

However, panelists did mention differences between Economists as well. One panelist commented that "there is large agreement over the major theories in terms of behavior. In that respect, there is convergence. The exceptions in behavior, however, are probably generating some divergence in how we explain them."

---

[14] Panelist credited this concept to Dan Ariely. See Ariely (2008).

The largest divide mentioned by panelists was the divide between Neoclassical Economics and Behavioral Economics. According to a panelist, Behavioral Economics differs from Neoclassical Economics[15] because it:

> allows for people to make decisions that are seemingly contradictory. For example, the lowest amount you would be willing to sell something you own is often much larger than the highest amount you would be willing to pay to buy the exact same thing (classical theory suggests this should be the same number).

One panelist emphasized this division but added that "[t]here are some heterodox departments/economists, but those are a small minority of the profession." From what we can tell, two subfields suggested by panelists (Original Institutionalist Economics and Marxist Economics) appear to be forms of Heterodox Economics (because of their rejection of Neoclassical Economics), but no details as to the nature of this difference were provided by panelists. Lastly, one panelist mentioned a division between Micro and Macro Economics but did not provide details.

Other divisions between Economists appeared during this project. For example, panelists disagreed over the degree to which prices drive behavior, though they mentioned that prices are not the only factor. One panelist commented: "If you don't think prices influence behavior, then you are not an economist. Economists also recognize how habit

---

[15] Generally speaking, Neoclassical Economics is heavily mathematical and considers people to be rational actors. In contrast, Behavioral Economics overlaps with Psychology heavily (both in terms of methods and understanding of people).

formation, social influence and behavior biases can systematically influence economic decision-making." However, another panelist commented: "I think economists may well focus too much on just the prices. Price elasticities and subject to tastes and preferences which grow in part out of the habits, customs, and cultures of a society in which a market is found."

Panelists also seemed to disagree about the degree to which people can be trusted to make their own choices, as opposed to having experts make their choices for them. One panelist commented: "Individuals have more-personal information than someone else, how would another person know more about others?" This seems to be a general statement, which conflicts with the comments of another panelist: "I think this depends greatly on the context. Few economists would say this is universally true. Many would say this is generally true."

Panelists also disagreed about whether Economists think of people as making careful and deliberative choices or not. The suggested comparison "Many non-experts think that people: make choices without much if any consideration. But Economists think that people: make every choice by weighing the costs and benefits of all available choices and choosing the best available option" was dropped from the study due to low ratings for the non-expert statement, but also caused some debate over the expert statement. Two panelists seemed to support this statement, but in qualified ways. One commented "Economists claim this is done subjectively not explicitly" and another commented "Economists do think this way but I also think we realize that a complete calculus is not

performed each time." In contrast, one panelist commented "I disagree that this how economists think literally people behave."

Another division between Economists was that "[s]ome Economists think that people will mostly do what is in their best short run interest rather than their long run interest." However, one panelist cautioned that different Economists define best interest in different ways, and "Neoclassical economics mostly defines best interest in terms of material interests. Behavioral economics defines best interest more situationally and in ways that can change depending on the social and psychological context."

Geography

Geographers mentioned a bewildering array of subfields that have viewpoints that differ from those of other Geographers. Some of the fields mentioned by panelists (but not elaborated on) include Emotional Geography, Psychoanalytic Geography, Political Ecology, and Postmodern Geography. Panelists also mentioned Humanists, Post-Structuralists, and Feminists as having different theoretical orientations.

Several panelists mentioned that Landscape Geography was different from mainstream Human Geography as they emphasize "the built environment as a force affecting what we know and how we act than those who are not among this group." Marxist Geography were also described as different as they "see the role of the economic system as consistently more influential than other elements of society in terms of how to understand people."

Several panelists mentioned Critical Geography as different from Human Geography generally (some panelists mentioned Critical Human Geography and Critical Social

Geography). One panelist commented that Critical Geographers "tend to think of power as residing in the contest over meaning (via economics) more consistently and more deeply than the rest of human geographers." One panelist described themselves as a Critical Human Geographer, and described their position as such:

> I would label myself as a critical human geographer. This means that I believe that social and structural issues have profound impacts on human behavior. To explain poverty in Central America, for instance, I would look to the region's broader colonial and imperial history and geography. I also care deeply about social and environmental issues and believe that my research should contribute to social and environmental change. I don't believe that objectivity is truly possible as I believe there is inherent bias and subjectivity in all research. More and more, most human geographers would agree with me on these issues. However, there is a more traditional realm of the discipline that stands by efforts for objective and apolitical research.

This view (that objectivity is not possible and that research should improve the world) is likely the same critical perspective mentioned in other panels (such as Sociology).

Panelists disagreed about how significant these divisions were. Some panelists considered it a difference in kind, arguing that Geographers "live in methodological and theoretical silos" and "use different epistemologies, which lead them to ask different questions and get different answers." Another panelist echoed this, arguing that Geographers "may reach vastly different conclusions because they will ask different questions about human thought and behavior." One panelist argued:

Geography is a very divided discipline, as we have people who specialize in more

quantitative climate modeling, geospatial information systems, remote sensing, and

then those who specialize in more qualitative emotional geographies, children's

geographies and political ecology. While we may share similar foundational courses

from our undergraduate studies, our ways of understanding the world are very

different.

However, other panelists seemed to downplay these divisions. One commented that

Geographers did not "disagree in their understanding of people, they just [focus] in

different aspects of society. Human geographers may look at the characteristics of a human

group while regional geographers may be interested in comparing them according to the

regions they occupy." Another panelist commented that "Human geographers trace

connections between seemingly unrelated events around the world. As a discipline, we

study interconnections among people, places and environments. The difference is the

ability to see these interconnections and interpret why they matter." Another panelist

commented that former divisions may be subsiding (ellipses as in original):

I think many of the rigid dogmas that once characterized our field are melting in the

light of evidence that dogmatic positions aren't very useful... Today's thinkers are

more humbled by their inability to explain [the] chaotic nature of human systems.

History

Some panelists were unsure that it was possible to generalize about how Historians

think about people. One panelist commented: "It's very, very hard to generalize about

historians, given that Africanists, Europeanists, pre-modern scholars, ancient world

253

scholars, etc. have vastly different training and historiographies." Another argued that they were "not sure that historians ever think of themselves as understanding "people": we are always specifying which people." However, one panelist did generalize about how Historians feel about people, commenting that "Historians tend to have more empathy and understanding toward people who are different because their study exposes them to diversity across time and space."

Panelists disagreed about whether Historians themselves tend to generalize in their research. One panelist commented that "Historians are prone to many varieties of reductionism" while another commented that "It seems to me that what historians do in general is make things more complicated!"

Panelists did note several ways that Historians differ from other Historians in how they think about people. One panelist commented that Historians differ in "[w]hether they emphasize individual agency or systemic factors." Another panelist tied this to the division between Humanities-focused Historians and Science-focused historians, commenting that: "Historians who tend to be more humanities-oriented emphasize individual agency. Historians who have a more social science approach tend to emphasize systemic factors."

Historians also debated the degree to which knowledge of the past matters today. The suggested comparison "Non-experts think that people: in the past knew little about sex. But Historians think that people: have always been aware of a wide range of sexual behaviors" was eventually dropped from the study due to low ratings by panelists (panelists rated this as representative of Historians but not of non-experts nor consequential). However, in discussing their ratings, panelists strongly disagreed. One

254

panelist stated "[k]nowledge of the history of sexuality challenges the terms of current cultural debates around for instance, gay marriage and trans identity." In contrast, another panelist commented that they did not "see how one's view of people's knowledge of sex in the past affects perceptions of the human world today."

One issue that caused a lot of debate among panelists was whether History could be said to repeat itself, though part of this debate seemed be over how literally one should interpret "repeat." Some panelists rejected this because they believe that people "can learn from history and thus avoid the mistakes of the past" or that people "adapt, change, and do not repeat the same mistakes." Another commented that Historians "believe that in complicated situations people often behave unpredictably -- i.e., history does not repeat itself to any great degree." However, one panelist commented that they "think people do repeat the same mistakes." Likewise, another panelist commented that they "think people rarely repeat the same mistakes because everything around a decision changes in a new situation. But they make the same kinds of mistakes all the time."

Rationality was also a point of contention among Historians, however, rationality seemed to only be part of a larger debate among Historians about motivation. This debate over motivation was complex and multi-faceted and touched on many other debates across and within fields. Panelists agreed that Historians thought about motivation in a complex way. One panelist commented that "Historians give different weights to particular dimensions of motivation: most of us see motivation as complex, but we often think one thing is more important than others."

255

However, panelists disagreed about whether non-experts thought about motivation in simple or complex ways. One panelist commented: "In general, I'd say historians think about motivation in more complex ways than do non[-]historians." However, another panelist commented "I don't know of a single human being 'out there' who thinks that 'people are motivated by single causes...Both historians and non-historians think similarly about causation in human action."

Panelists also noted that "some historians use psychology to generalize about humans." While another panelist rejected an item because it suggested "psychological causation, and historians for the last half century have been most attentive to the cultural determinants of behavior." Another added that: "[m]ost historians would think that people act according to their circumstances, in ways that are shaped by issues like gender, class and race." It is unclear how these Historians interpret culture influencing behavior and why they consider this inherently in opposition to psychological explanations.

Another mentioned that the debate over motivation touched on issues of rationality and nature vs. nurture, commenting that:

Historians differ greatly in their understanding of what motivates people to take action in response to changing circumstances. Some assume they are dealing with rational calculators of self-interest, some see prisoners of inherited cultural traditions, some see stand-ins for a particular view of human nature (i.e. see people who are inherently greedy, violent, etc.)

This disagreement over nature vs. nurture appeared several times in the rounds. In particular, a comparison arguing that there is no such thing as "human nature" caused a lot

of debate among Historians. While this comparison was well-rated by panelists, several disagreed and argued that there is a consistent human nature. Likely related to this, one panelist proposed the comparison: "Many non-experts think that people: acted very differently in the past than they do now. But Historians think that people: acted similarly to [how] people act today." However, this comparison was dropped as it received some of the lowest ratings of any comparisons in any panels. Together, these imply that there is a profound divide among Historians on whether there is a human nature and that most Historians think that there is not.

A dropped item about whether people were naturally peaceful or violent caused a similar debate, with a panelist commenting that "the human propensity to violence is a subject that is debated by both historians and non-historians, and probably not as big a difference between them."

Panelists disagreed about how to define self-interest. Historians felt the statements that people "do not usually know what is in their interest and act at the behest of a wide array of cultural impulses" and the statement that people "act on the basis of values and goals that are not limited to economic interest" were strongly representative of Historians. However, one panelist questioned the contrasting of personal self-interest against "values (religious, political, economic)," asking whether these would not also be self-interest.

One panelist commented on a division between Cultural Historians and other (unspecified) Historians, commenting that "Some cultural historians rely heavily on postmodernist theories. Most historians do not, and find postmodernism a heavy-handed attempt to twist words around, putting modern theory before historical evidence."

One panelist commented that these debates are driven by theoretical frameworks that Historians use:

It's impossible to write about the past without some theoretical framework that guides your thinking. Whether it's explicit or not, there's a theory behind our work. It's as much a theoretical perspective to say people were always the same as to say that they are shaped by cultures. So clearly, different theoretical perspectives shape interpretations.

This does raise the question of where Historians get their theoretical frameworks and whether such frameworks are coming from inside the field of History, or whether such frameworks are derived from other fields and then applied to the study of History.

Historians also disagreed about where to position History as a field. Generally speaking, panelists rejected the idea that History should be considered a social science. There were no comments by panelists arguing that History was unambiguously a social science. However, one panelist took a middle position, commenting "I don't think historians trade in generalizations about "people" and "human nature" -- which is why we sit between the social sciences and humanities."

Other panelists argued that History is not a social science. One panelist commented "I am [a] historian, not a social scientist." Another panelist commented "I don't consider history as social science, so I don't think I can accurately comment on this. Historians generally don't explain people/humanity in a broad way. We deal in specifics."

Lastly, one panelist elaborated on how they saw the study of History, which they contrasted sharply with the social sciences:

History is not a social science and does not look to create or test generalizable truth claims about people, or to synthesize laws which have predictive value. History is the anti-social science -- an empirical field [refocusing] on what is specific and particular in any situation -- and so it is a softer, less generalizable, but more trustworthy kind of wisdom. The attempts to make general claims about human nature on which this study has [centered] don't really come into history, except perhaps when we are trying to get students to generate theses in undergraduate essays. Above that level it is very rare. Rather than the truth tests of social science, history relies on imaginative, convincing uses of evidence well expressed -- an aesthetic, humanistic standard. History is a strongly empirical and evidence-based field, but not a science. Science is not the only kind of empiricism.

Political Science

There were several debates within the Political Science panel, but these are difficult to untangle, as it is difficult to tell when they are debates over definitions, debates over how people work, or both at once.

The major division seems to be between Political Scientists who are similar to Economists and Political Scientists who are similar to Psychologists. One panelist commented "Some political scientists understand people as seeking to maximize their own interests, based on the incomplete information they have at the time, while others tend to understand people as being motivated primarily by unconscious responses to psychological cues." Other panelists echoed this division, describing "'rational choice' theory (economic self-interest dominates)" as being in competition with

"psychological/social/behavioral theories (in which interests beyond economics (tribalism, culture, class, etc... dominate)."

Unsurprisingly, this led to disagreements about whether people were rational. One panelist explained:

> I think political scientists are pretty split here - but, again, in part because people are taking about different things when they talk about rationality. Strictly speaking - in terms of rational choice theory - rationality only means that people try to choose the best thing for themselves given their option or information. Thus, none of the factors mentioned as deviations from rationality aren't, i.e., most of them speak to the information available to the individuals or their capacity to process that information.

Another panelist explained (though they called this a "slight exaggeration") that rational choice theory "leads one to think about problems in a different way rather than simply resorting to concluding that people are stupid b/c the[y] don't do the right thing." However, some of the suggested comparisons did not seem to hold standard human decision-making in high regard, though they often emphasized this was due to a lack of information on people's part. Panelists gave the impression of perhaps being overly familiar with having these debates within their field. When we asked panelists whether people tended to make decisions in a conscious and deliberative manner, one Political Scientist commented "I'm sure some of my colleagues will be difficult and quibble over those specific terms."

One panelist noted that non-experts have a different, somewhat self-serving definition of rationality, and that "many (or most) non-experts are happy to consider people non-rational, but it is based on the particular version of 'rationality' that basically translates into not seeing things the same way they do."

A related disagreement was whether people are "instrumentally rational" which means that they act in ways that will tend to achieve their goals. While an item about whether people are instrumentally rational was well-rated by panelists, the ratings for the Expert statement had an unusually wide distribution, and one panelist commented that the comparison should be reversed.

Panelists also disagreed about how motivated people were by self-interest. For example, the following statement caused considerable debate: "Many non-experts think people: make decisions that put their own needs and preferences first. But Political Scientists think people: take factors, such as empathy, into account when deciding what to do and are not limited to their own needs and preferences." This statement split Political Scientists and was dropped from the study due to not representing Political Scientists (and also being considered inconsequential).

One problem, as seen in other panels, was that panelists disagreed on how broadly to construe self-interest (ellipses as in original):

I think most [Political Scientists] probably think that people are primarily interested in the things that benefit them - what benefits them may consist of different things but...I guess I find it hard to think of someone having motivations that don't benefit them. The motivations may not be directly self-serving but if one's motivation is to

help others, isn't basically because the derive some pleasure from it (even if it is something like a sense of justice being served)?

Because panelists may interpret "self-interest" differently, it is hard to disentangle how they view people.

There were several other divisions mentioned within Political Science. One panelist commented that four subfields "(American, comparative, IR, theory) differ pretty substantially in how we view things. While there are some commonalities, there's a lot of differences." Others mentioned that "Americanist" and "American Political Development" were different, and panelists mentioned that an item on partisanship was a "US-centric view." Another panelist mentioned that, rather than subfields, they would "propose constructivism as an approach that often seems to rely on a very different perspective of how 'people are.'"

There were also many topics that Political Scientists disagreed on. For example, the importance of gerrymandering, whether people are manipulated by the media, whether campaigns affect voters, whether education leads to being more politically informed, and how important the state of the economy is for elections.

One panelist commented that Political Scientists disagreed about the importance of "identity." Another commented that:

There are some who believe that "culture" is not particularly explanatory (especially when comparing politics/governing in different parts of the country or world) and there are others who think that culture is very important to understanding differences (even if difficult to measure and therefore difficult to study).

Lastly, gender proved a point of contention among Political Scientists. One panelist phrased this in terms of "some folks stuck in older thinking in which gender does not seem to matter" while another phrased this as a difference between those who "believe gender is centrally important in explaining political outcomes, while other political scientists think that gender is only important in selective circumstances."

Psychology

Compared to other panels, the Psychology panel had few internal disagreements, and the disagreements that did occur tended to be technical rather than philosophical. For example, panelists disagreed about whether infants could remember events in their first year or whether people stopped gaining happiness from additional income after reaching a certain income level or whether they merely gained happiness more slowly after that level.

The only major division described by panelists in how they understand people was the division between research Psychologists and Psychologists who do counseling/therapy. One panelist commented that they "suspect counseling psychologists' training is quite different from psychologists who are focused on research, and thus they likely understand people in different ways." One panelist commented that there are "some traditions in clinical psychology, such as psychoanalytic theory, that will never converts[*sic*] with modern, data driven theories." In addition, Psychologists who do counseling/therapy can be further divided in how they understand people. One panelist contrasted Psychoanalysts,

Behaviorists[16], and Cognitive Therapists, and said that "Psychologists who subscribe strictly to one of these theoretical orientations very much differ in how they understand the cause-and-effect of emotional, behavioral, and cognitive outcomes."

In contrast, some areas of research Psychology appear to be converging. One panelist commented that "Developmental psychologists are converging towards seeing development as a dynamic, changing system, with each level interacting with the others: cells to society." This is likely a reference to thinking about people as Complex Adaptive Systems (an idea which was also mentioned by the Anthropology panel).

Another relatively technical division mentioned by panelists is that "Connectionists think that people learn and develop through detecting patterns in numerous, related experiences in an automatic sort of way." As Connectionists typically study neural networks as the basis for mental processes, this was likely intended to be contrasted with Computationalists, who think of mental processes as the logical manipulation of symbols. However, in recent years, researchers have argued these perspectives are not necessarily contradictory.

Many of the differences within Psychology were described as a difference in degree rather than kind. For example, one panelist commented "Although virtually all

---

[16] This panelist was likely not referring to Behaviorism (a type of research Psychology popular in the first half of the 20th century), but to Behavioral Therapy (a group of therapy methods derived from Behaviorism). Practitioners of both sometimes call themselves "Behaviorists."

psychologists would agree that genes and environment interact to product[*sic*]

development and behavior, they differ some on the degree to which they emphasize the

influence of each." Likewise, "The field disagrees on the extent to which identifying the

neural pattern of activation correlated with a behavior actually "explains" the behavior.

Neuroscientists are more likely to see the biological basis of behavior as a better (more

scientific) explanation of behavior."

Lastly, one panelist mentioned that some Psychologists "focus on the causes of

behavior being internal - others say it is external" and another panelist commented that

"Psychologists differ in they focus on people as individual versus a person-in-context."

However, it was unclear whether these were intended as differences of degree or of kind

for Psychologists.

Sociology

While other fields had relatively clear and agreed upon divisions within them, the

picture for Sociology is more complicated. While some panelists described divisions within

Sociology, others preferred to think of divisions within Sociology as ones of degree rather

than kind. One panelist commented that not only is Sociology diverse as a field, "Even the

groups can be pretty diverse."

Panelists did draw a clear distinction between Sociologists and non-sociologists,

saying that "non-sociologists often see people as autonomous actors while sociologists

stress the constrains [*sic*] imposed by various social institutions, structures and patterns of

bias and discrimination." Another panelist commented that "Most non-sociologists operate

off of ideology in developing their understanding of people. Sociologists use theory and

social facts to understand people." One panelist also commented that Sociologists distinguish between prejudice and racism (as a combination of prejudice and power that leads to systematic advantages) but that most non-experts do not.

In contrast to describing clear division between Sociologists and non-Sociologists, several (but not all) panelists described differences within Sociology as a matter of degree rather than kind. One panelist commented:

I'm not sure about profound differences in how Sociologists understand people, but there are differences in the degree to which we think people are able to freely make choices compared to how much social structure shapes limits our actions. We also differ in the degree to which we think biology matters compared to the social environment - the nature vs nurture debate. Again, these are differences of degrees and not profound differences, but they do present conflicts in the field.

Another panelist likewise rejected viewing differences within the field as dichotomous, commenting that while there are factions and differences in methods among Sociologists, the relationship between these groups are complicated:

As there are a plethora of social theories used to explain social behavior, sociologists develop into different factions, depending on what theories they embrace to explain and understand people. In the macro theory camps, we have structural functionalists and Marxists, for example, while in micro theory, we have feminism, and its many variants, symbolic interactionism, social psychology, rational choice theory, and a host of others. These groupings are further complicated by research methodology, with some approaching an understanding of humanity from a

quantitative approach, others from a qualitative approach, and still others

embracing both (mixed methods approach). Depending on one's theoretical

orientation and questions of interest, the form of understanding of people may be

quite distinct from that of a fellow researcher who embraces a different

methodology (and is asking very different kinds of questions about social behavior).

To frame this as a binary distinction is problematic, as these differences discussed

above produce a multi-faceted set of nuanced distinctions between sociologists in

their understanding of people.

Other panelists elaborated on some of these distinctions. One panelist commented

that "Micro Sociologists (or 'interactionist') tend to understand how people create their

own reality in social interaction and through symbols. Macro Sociologist tend to

understand how people are impacted by social structures. These views aren't necessary

incompatible or oppositional, merely they have a difference in focus." Likely referring to

the same divide, another panelist commented, without clarifying, that there was a divide

between "functionalists vs. symbolic interactionists." In addition to micro and macro,

another panelist commented that some Sociologists embrace theories in the "mid range."

Other panelists did seem to imply fairly profound divisions within the field. One

panelist commented that Sociologists do not all share a single epistemology, and that

"Much of our methodological logic is based on positivism, which has includes[*sic*] an

assumption about scientific progress. Many Sociologists reject positivism, but not all."

Another panelist commented that some Sociologists "view humans [as] passive recipients

of social forces whereas others see humans as active agents in shaping their environment."

And another panelist commented that they thought that "critical theory/Marxist/political economy views are somewhat distinct," though they did not explain how.

Many panelists mentioned other divisions within the field without clarifying whether they saw these as differences of degree or of kind. For example, one panelist commented that Sociologists "differ in explaining behavior along the axes of order (individualistic to collectivistic) and action (rational to nonrational)" and another mentioned differences in whether Sociologists "are willing to organize for social change." Another mentioned, but did not describe, a divide between "sociologists of medicine vs. sociologists in medicine." Panelists also disagreed about the degree to which culture causes sexual assault relative to other factors. Another panelist commented that Sociologists differ on "whether racism can be talked about generally, or whether it cannot because the concept of race has changed over time."

One panelist summarized Sociology as follows:

Sociology is a heterodox discipline. We don't have only one of understanding people. We do agree that there are forces beyond individuals (social structure and norms) that guide, but do not completely dictate our actions. Beyond that, we tend to think of people as complex and rarely offer singular explanations that we expect to apply to all people.

Additional Questions

Are the social sciences converging or diverging?

Panelists were asked whether the fields of the social sciences are currently converging or diverging across fields on how they understand and explain people. Panelist ratings for this question can be seen in Figure 13.3.

One Geographer commented that they "think Economics, and perhaps political science have cultures that keep them from converging in the way that geography, sociology and anthropology have over the years. Economics in particular seems politicized in an almost religious fashion." Several historians rejected this question entirely as not being applicable to their field. One Political Scientist was unsure that they "know enough about the other fields to say. Lastly, one Psychologist commented that there was "[p]ossibly some convergence in awareness of implicit bias and the difficulty of changing many people's minds."

How well do non-experts naively understand people?

Panelists were asked whether people who have not formally studied a social science have an understanding of the human world that tends to be accurate. This gives something of a baseline for how much it is possible to change from studying a field. The more sophisticated non-experts are at understanding people without a formal education, the less such an education can change how they understand people. Panelist ratings for this question can be seen in Figure 13.4.

Several panelists commented that studying a social science is not the only way to understand people better. A Geographer commented "I think a clearer understanding of the

269

world can be gained with an open mind, travel and/or a humanities education as well."

Similarly, a Historian commented that they were "not sure I would weigh formal study

more than the evidence of experience."

One Sociologist commented: "This question assumes a tautological fallacy. I have to

know what the world is like in order to decide if what I know is accurate." Another

Sociologist attributed the non-expert view to socialization, commenting "the simplest ideas

that blame another vulnerable group seems to be the guiding force for much of modern,

American understanding of the human world."

How did studying their field change their views about people?

Panelists were asked how much their understanding of people changed as a result of

studying their field. Panelist ratings for this question can be seen in Figure 13.5.

An Anthropologist noted that they were raised by a Sociologist, so "my field didn't

change my views as much as it might for most." This implies that Anthropology and

Sociology have at least a somewhat overlapping understanding of people. However, in

contrast to this, one Sociologist argued against the idea that studying their field changed

how they understood people. Instead, they commented "It was always common sense to

me. Sociology provided me with a language and tools to explore things I'd felt intuitively

and allowed me to test those ideas." If the fields of the Social Sciences are codified common

sense (rather than having a profoundly different understanding of people), this would

argue against the idea that the fields have a scientific understanding of people that differs

from people's intuitive theories, or that the social sciences are converging on a scientific

understanding of people.

An Economist commented that they were "Not sure my understanding has changed, but this study has made me think about how we frame or phrase things." A Geographer focused on their own research, rather than the field as a whole, that changed how they understood people as their: "understanding of people was influenced by my experiences doing qualitative interviews with people as it gave me a chance to ask deep and meaningful questions about their lives" though they did not clarify how this changed their understanding.

A Political Scientist commented "I used to think people were mostly rational." A Geographer concurred with this, commenting (ellipses as in original):

I've studied people's economic decision making for years. I've grown increasingly convinced that many people are irrational actors in a system they don't understand, but that academics often suppose that people are ration[al] actors and that the systems in which they act are also rational...and fair.

Another Political Scientist, commenting on a comparison, noted: "One of the key insights social science gives us is that people are terrible judges of themselves."

A Historian commented "As someone whose research is distant in both time and space from contemporary California, I'm very aware of the ways categories - of identity, self-interest, etc. -- change over time. So I believe more in the possibility of change."

A Political Scientist commented "What really changed for me was the understanding that context matters a great deal. Humans may be the same all over but the circumstances they are in shapes their views and behavior.' Similarly, a Psychologist commented that they "have a much greater understanding of the complexity of the interacting causes of

behavior." Lastly, one Psychologist commented that social psychology prompted a crisis of faith that "profoundly changed my understanding of the world."

Did studying their field make them more optimistic about people?

Panelists were asked whether studying their field made them more optimistic or pessimistic about humankind. Panelist ratings for this question can be seen in Figure 13.6.

An Economist criticized the vagueness of the phrase "about humankind."

An Anthropologist commented that "anthropology contributes greatly to people's understanding of human similarities and differences, especially in education, and that the generational shift to greater awareness and acceptance of difference is cause for optimism."

A Geographer commented that they were "Equal proportions optimistic and pessimistic - which one dominates depends on the specific question at hand. More optimistic within the sphere of one's influence, less optimistic about larger structural things. This may have a relationship with larger political contexts." Another Geographer seemed more pessimistic, commenting that "Humans are short sighted and rarely learn from our past mistakes. I'm very concerned about the environmental catastrophe at our doorstep."

One Historian rated that they were a little more optimistic and commented "Ironically – I think because I see the possibility of change."

Lastly, a Psychologist commented that they were "more aware of the difficulty of changing people's minds and of confirmatory biases in people's thinking."

Did panelists like participating?

The final questions asked panelists whether they liked participating in this project and whether they had any additional comments about the study. Panelists in different fields differed in how much they liked participating (See Figure 13.7). While panelists tended to like participating, it should be noted that the panelists who answered this question were those who chose to take the final round and were likely the panelists who were the most positive about the project.

Some panelists were positive about the project. One commented that the project was a "Very interesting exercise. Sometimes a bit time consuming, but overall enjoyed the process and opportunity to reflect on my field."

The most common complaint about participating in this project was that it was repetitive and time-consuming. Panelists also mentioned that the questions were difficult to answer. One reason given was that the questions "take a lot of mental energy." Another panelist commented that it was "challenging to think about how non-experts think about other people as opposed to how they think about themselves. E.g., most people think they are rational while they are quite willing to think that other people are not."

Another complaint was the dichotomous nature of the study. One panelist commented that "The nonexpert and expert statements were too strongly stated; most nonexperts and most experts would agree somewhat with both statements."

Another panelist questioned whether their field really fit into the framework used in this study, commenting "I think this is a very hard project to undertake given that some fields (political science) don't view the role of people in the same way. We're more about

systems & process so the phrasing of your initial questions was difficult to put into our language."

Another reason given for why questions were difficult was the lack of specificity or precision in the questions. One panelist commented: "A lot of this study seemed (to me) to be very general. It was hard to give a quantitative ranking for most of the questions. I think the use of specific examples (or debates within the field) would have made the questions easier to answer." The lack of specificity was a common complaint throughout the study. However, since the purpose of the project was to distill general understandings about people, this lack of specificity was somewhat unavoidable. In addition, some of the comparisons that panelists felt lacked specificity were their suggestions from colleagues in their field, and we could not have made such comparisons more specific had we wanted to, given our lack of disciplinary knowledge.

Lastly, panelists also mentioned that they were not confident in their answers, or that they often wanted to add caveats to their answers. Others mentioned that they felt "unsure speaking for an entire and vast discipline."
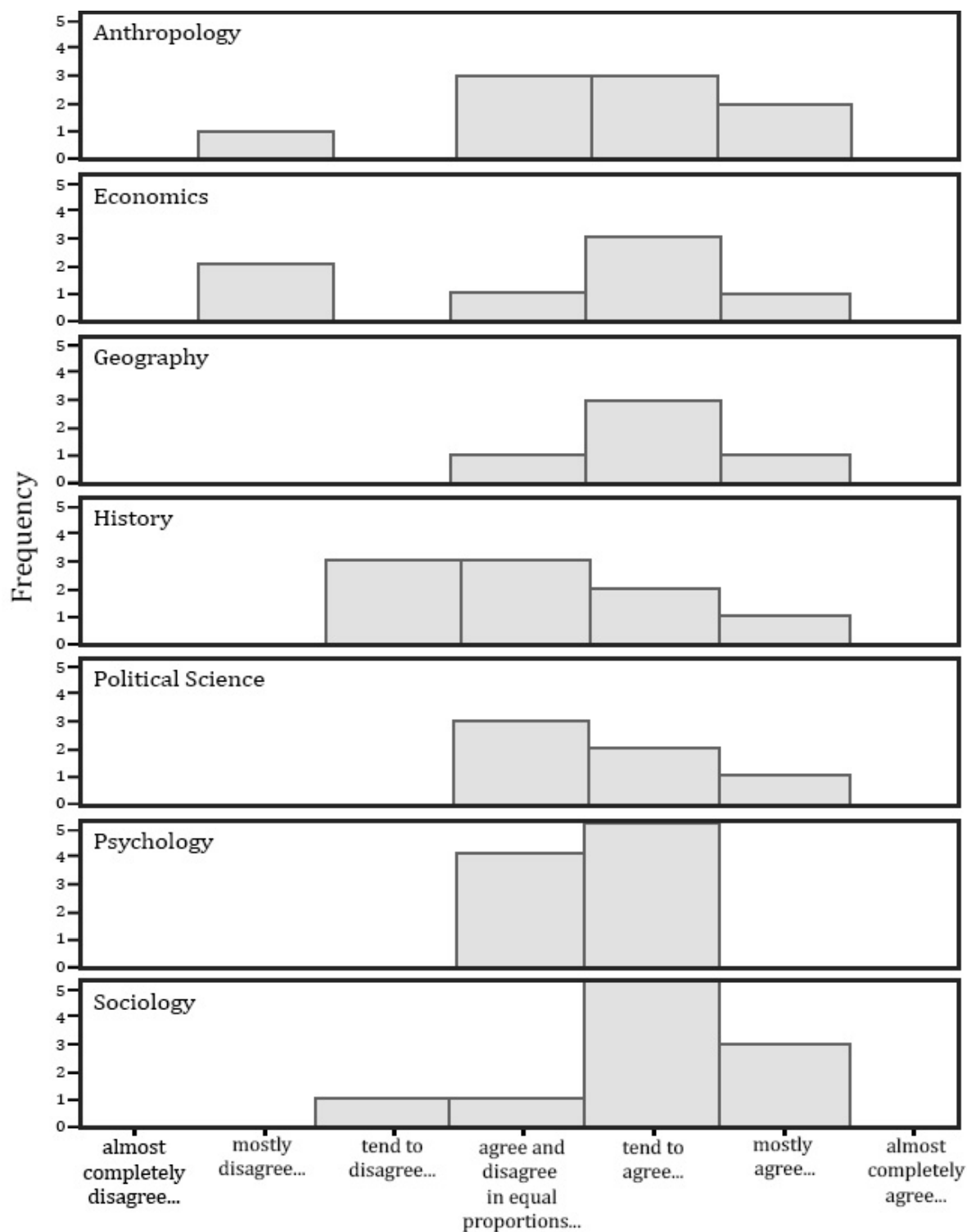
Note: All answers choices concluded with "on how best to understand people" except for "tend to agree" and "mostly agree," which erroneously omitted "best."

Figure 13.1: Panelist ratings of whether their field tends to agree.
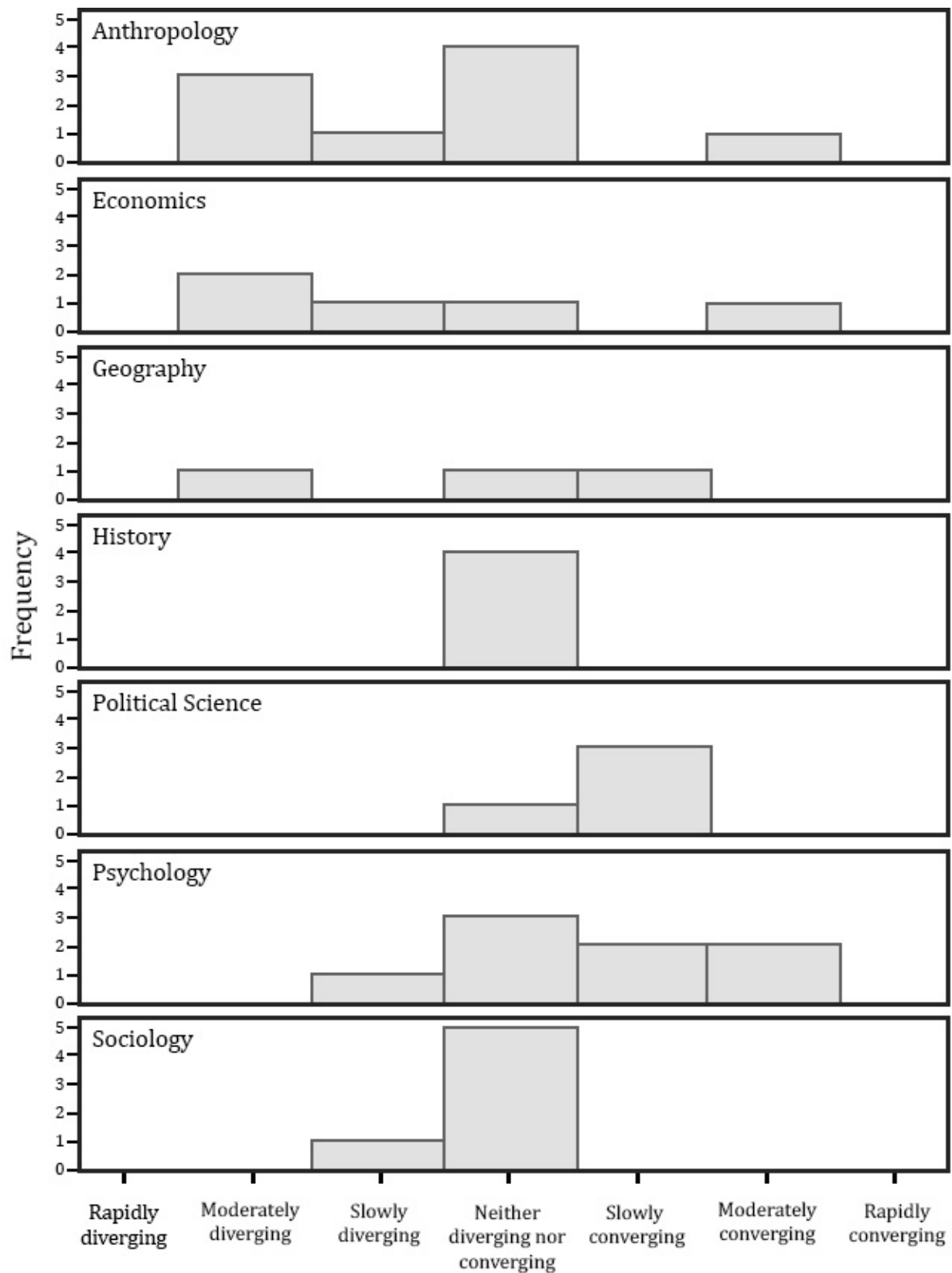
Figure 13.2: Panelist ratings of whether their field is converging.
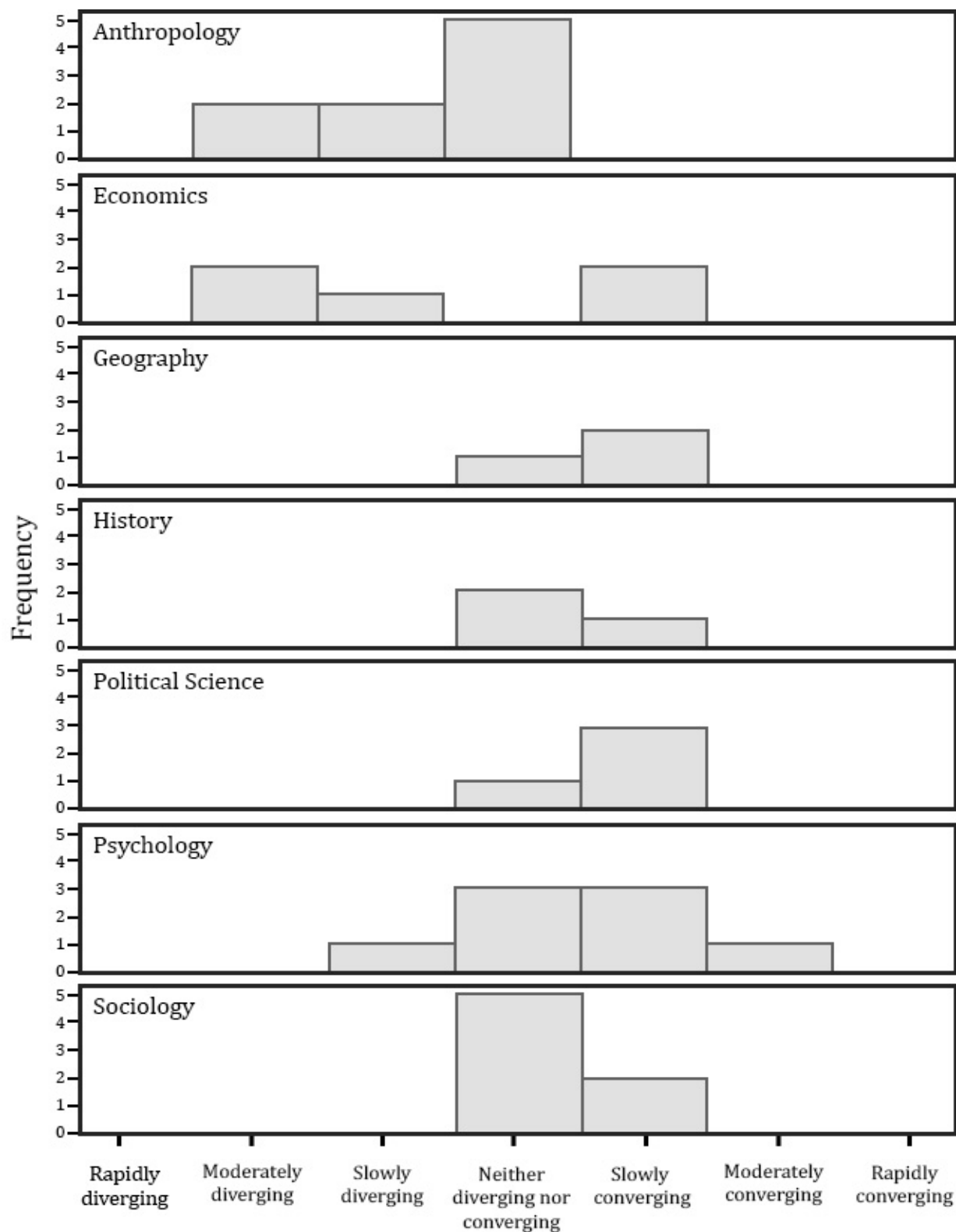
Figure 13.3: Panelist ratings of whether the social sciences are converging.
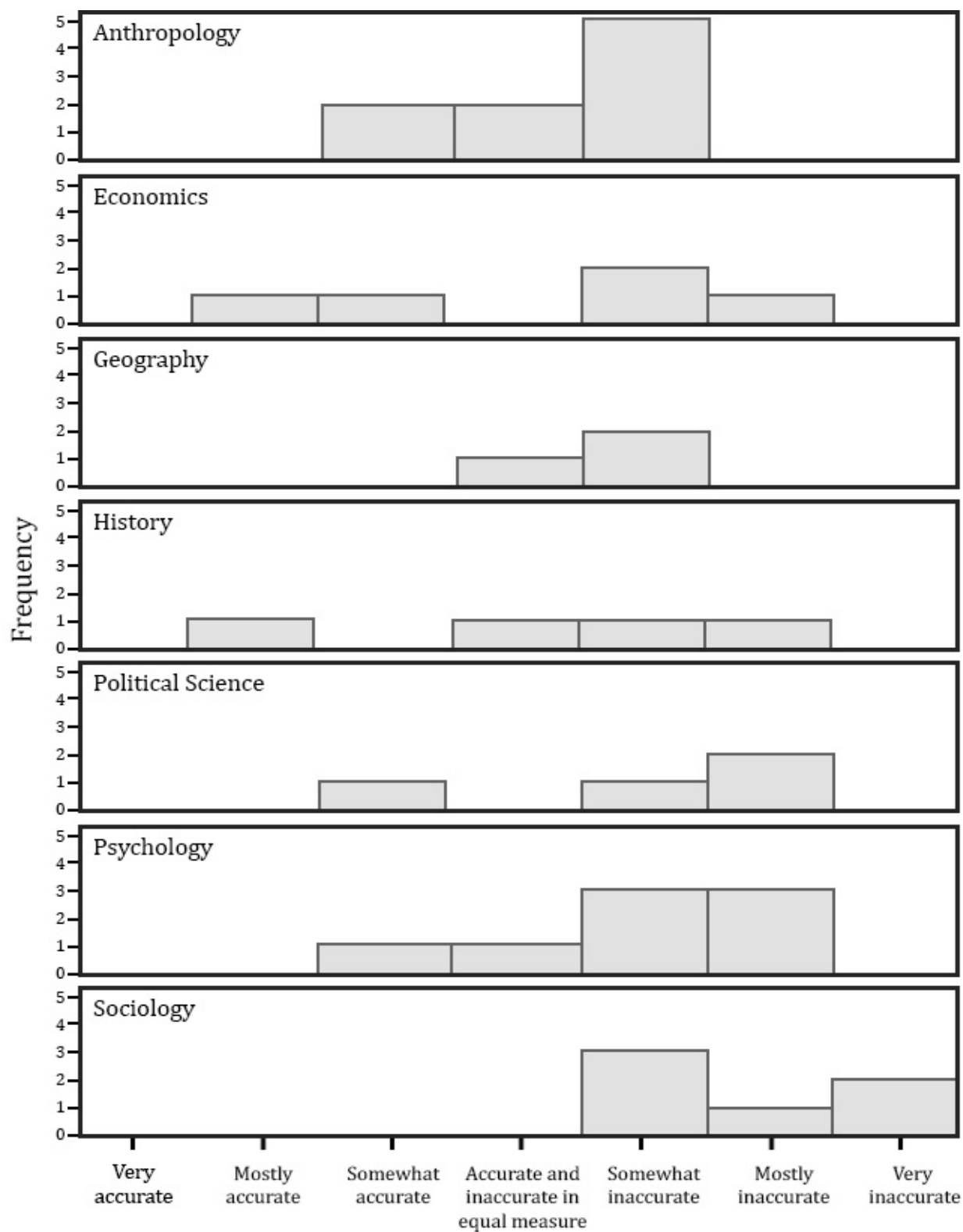
Figure 13.4: Panelist ratings of accuracy of non-social scientists about people.
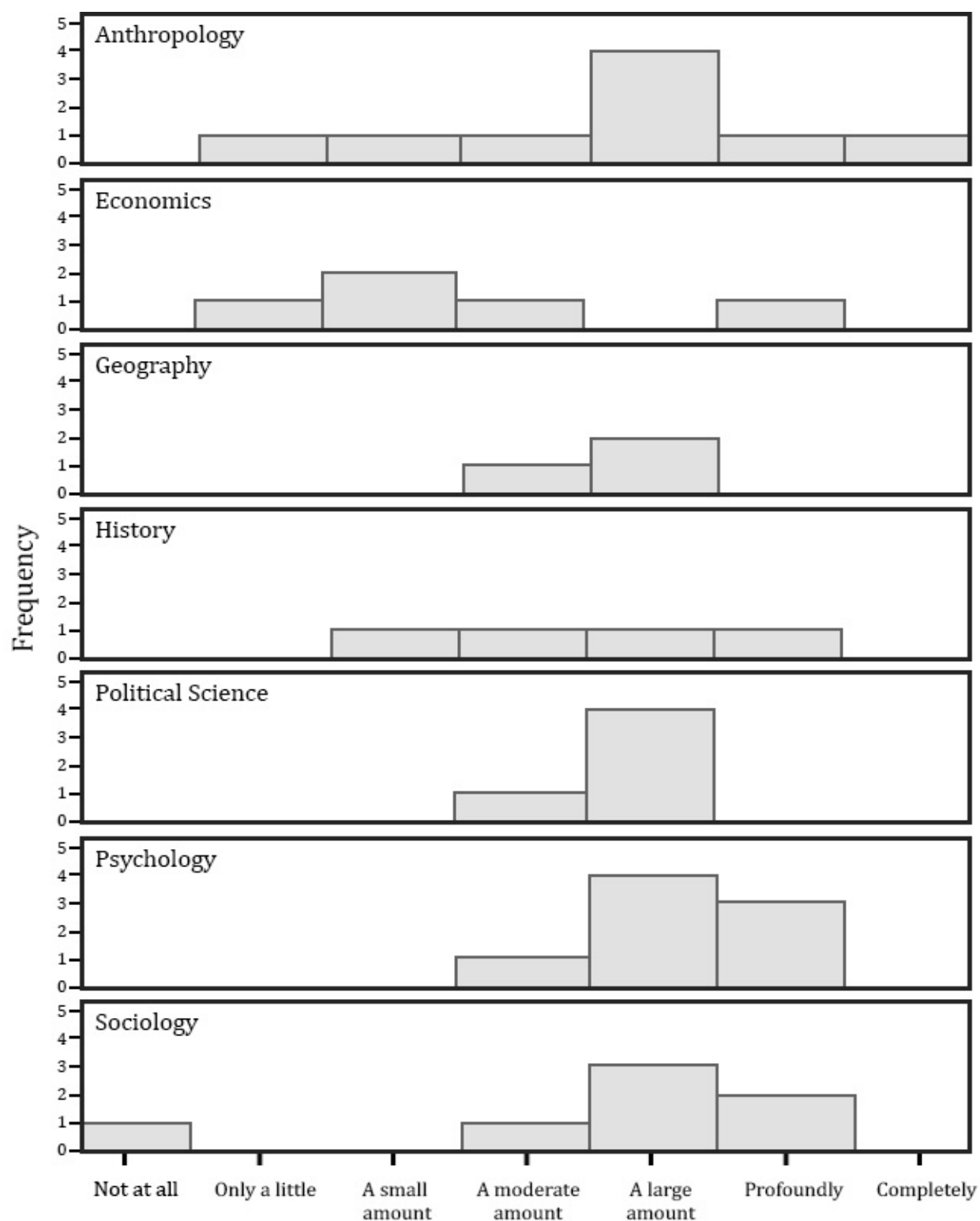
Figure 13.5: Panelist ratings of how much their field changed their understanding of people.
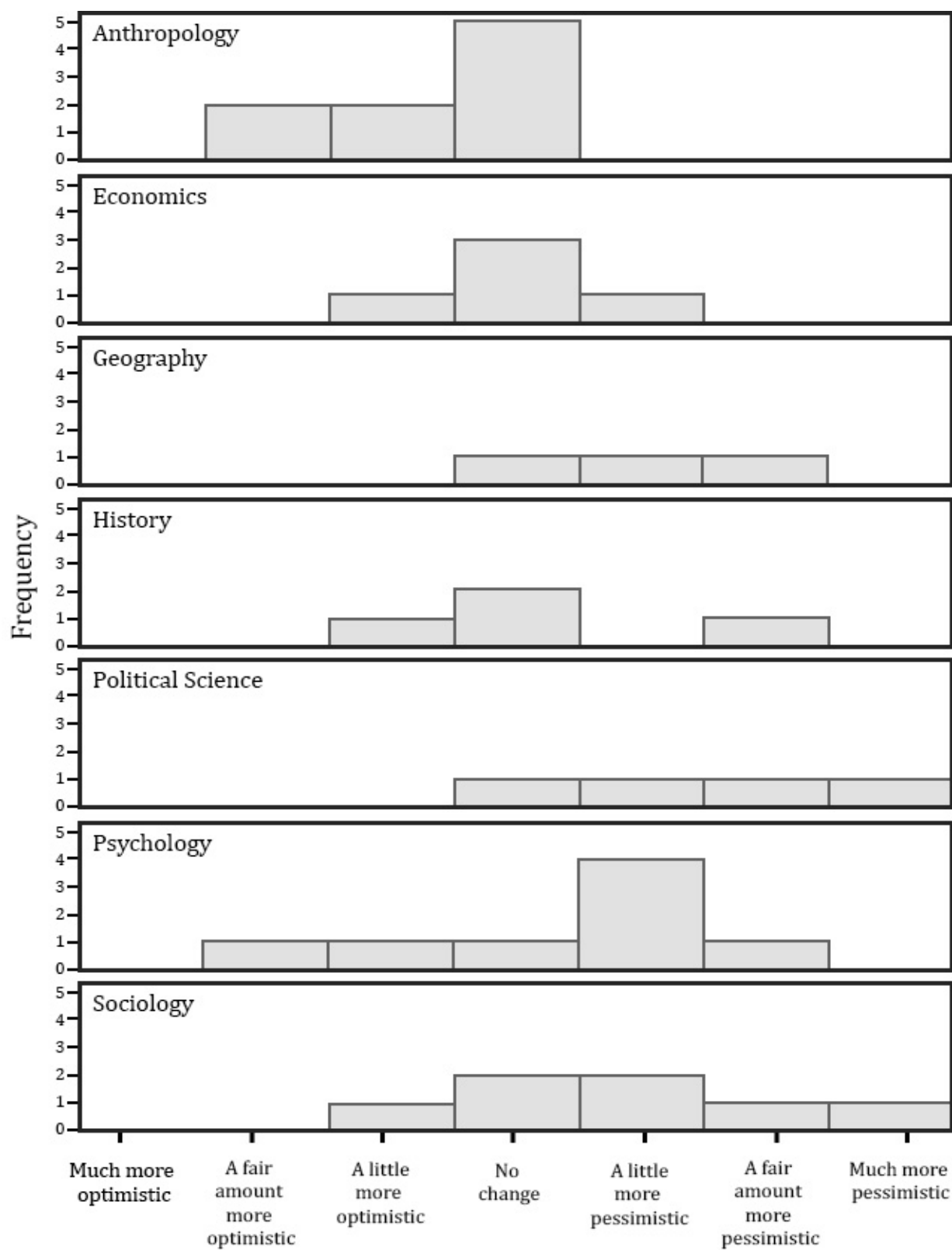
Figure 13.6: Panelist ratings for whether their field made them optimistic about humankind.
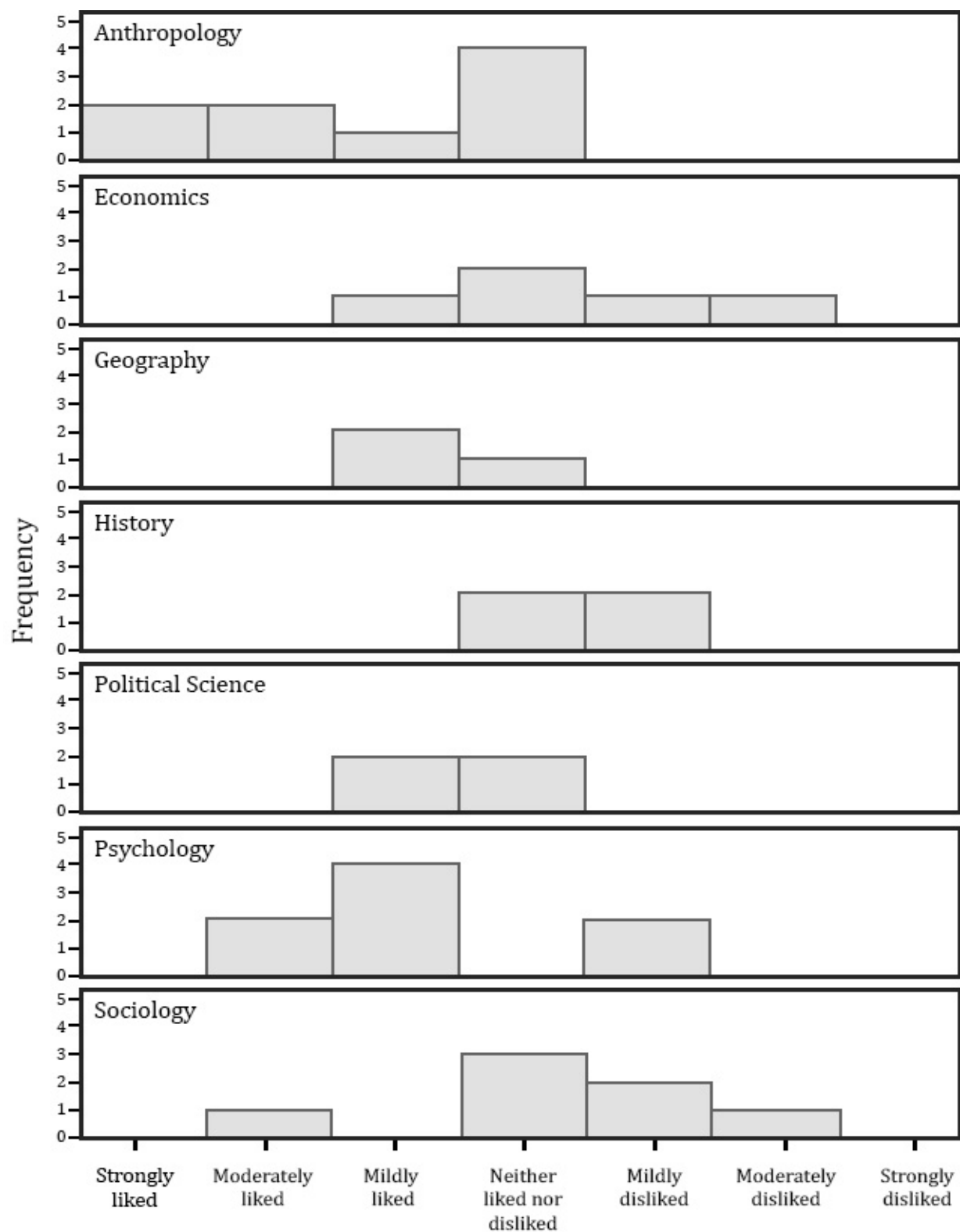
Note: Anthropologists saw the third choice as "Mostly liked."

Figure 13.7: Panelist ratings of whether they liked participating.

CHAPTER 14

Discussion

Overall, this project achieved some of our aims, but not others. We believe that the results demonstrate that social scientists have an understanding of the human world that differs substantially from the understanding of those who have never studied a social science, that this understanding does sometimes span the social sciences (or at least much of it), and this difference can be fruitfully described despite the complex nature and terminology of the social sciences.

While the view that people have intuitive theories about the world is not yet widespread across the social sciences, this project does imply that researchers in the social sciences do notice that students differ from experts in their fields in consistent and predictable ways. Given this, it is likely that researchers would be more receptive of this view if it was framed as an explanation of why students will struggle to understand their field and why the concepts that students struggle with stay the same year-to-year, even as the students themselves change.

Though the social scientists in this project suggested many ways that studying their fields changes how one understands the human world, this does not prove that these changes are actually *caused* by studying a social science. Panelists may have interpreted the question more loosely as how experts differ from non-experts, rather than thinking of it in terms of a progression as we intended. If so, it could be that these differences are caused by people who already hold a specific perspective sorting into those fields, rather than the act of studying the field causing the change. However, panelist comments give some limited

support to the idea that studying a field does cause one to change how they understand the human world, as several panelists credited their studies with profoundly changing their understanding. That being said, several panelists commented that their field did not really change how they understood people. It is not currently clear why some panelists felt their studies changed their perspective and others did not, and determining whether studying the social sciences only changes some people's understanding of the human world, whether such changes are dependent on which social science is studied, or whether people do change their understanding but tend to forget this, will likely require a longitudinal study of people as they are studying a social science rather than the retrospective measures employed in this study.

The Delphi Method

Though the Delphi method was difficult to employ and did not produce results as streamlined as we would have liked, we think that using it in this project (rather than simply asking experts what they thought) was the correct decision.

For one, had we simply asked experts to suggest comparisons and accepted all such suggestions, this would have led to accepting numerous comparisons that others in that field did not actually agree with or feel were representative of their field. Just as importantly, by allowing panelists to improve and refine comparisons, the Delphi method spurred the creation of some comparisons that were extremely popular with panelists by the end of the study, but which no panelist had originally suggested. Likewise, several comparisons became much clearer than their original formulation due to rounds of revisions.

Additionally, while asking for three ratings for each comparison (representativeness of the non-expert statement, representativeness of the expert statement, and consequentiality) did increase the complexity of the study considerably, it also proved useful. Simply directing panelists to only suggest comparisons that were representative and consequential did not work, and without these ratings, unrepresentative or inconsequential comparisons would not have been dropped.

That being said, our instantiation of the Delphi method could have been improved. It is likely that our threshold for ratings (a minimum of four out of seven) was too strict. This is particularly true for the non-expert statement – many of the comparisons that were dropped from the study were dropped only because they fell below an average rating of four out of seven on this statement. However, many of these dropped items fell below the threshold only slightly, and many if not most were above 3.5. In such cases, it probably would be best to explore whether this low rating was due to other prominent non-expert perspectives competing for non-expert allegiance, whether it was due to the statement being sub-optimally worded, or whether this was due to experts not agreeing that a large number of non-experts held that particular view.

Though our creation of comparisons with three components was unusual, we believe it was the correct decision. For one, these comparisons were well-rated by panelists. Secondly, it was not only our judgment that experts both agreed on the expert perspective and disagreed on the non-expert one, some panelists stated this as well. One Geographer commented: "More than anything, it seems that these geographers disagree on what non-experts think, but seem to agree (more or less) on what human geographers

284

think." Likewise, an Anthropologist commented: "I'm torn on this one. I think a lot of laypeople will think this way, but others will think completely the opposite: humans are mostly biology and a little culture." Given this, we believe it was justified to create comparisons with two non-expert perspectives, as comparisons with only one non-expert perspective would only partially represent the differences between experts and non-experts on these issues. However, even two perspectives does not fully capture how non-experts think, as there are likely many different competing perspectives of varying popularity. Going forward, even more of these perspectives about people should be limned; though the endpoint for non-experts learning the field is the same (holding the expert perspective), the trajectory to getting there is different depending on which conception a person begins with.

Are these shifts conceptual changes?

Can it be said that the shifts described by panelists (that is, a shift from the non-expert perspective to the expert perspective in their fields) are a form of conceptual change? It is difficult to answer this for two reasons. First, compared to the physical sciences, it is harder to say what the correct "scientific" perspective is given that the fields of the social sciences are messier and do disagree in places. Second, conceptual change researchers do not agree on what it means to undergo conceptual change, as seen by the existence of multiple competing models of conceptual change. For example, the Framework Theory Approach (Vosniadou, Vamvakoussi, & Skopeliti, 2008) differs from the "knowledge-in-pieces" approach (diSessa, 1988, 2008) in how they conceptualize mental models and thus, what changes when students undergo conceptual change. Likewise, it is

often difficult to distinguish between conceptual change and the more general process of learning. For example, Thagard (2008) lists nine different degrees of conceptual change, such as adding new concepts or adding new instances of a concept. However, many researchers would likely consider adding a new instance of a concept to be simply learning and not conceptual change at all.

As mentioned previously, since conceptual change touches on many fundamental and still-as-yet-understood aspects of cognition, a full accounting of concepts and conceptual change is not currently possible. Nonetheless, while this makes it impossible to create a definitive test to determine whether a change in understanding about the world qualifies as a conceptual change, we believe it is possible to demonstrate that the shifts mentioned by panelists are likely conceptual changes by analogy. Even though many researchers in the social sciences are unaware of the field of conceptual change or its findings, they have nonetheless noticed many odd features of the process through which people shift their understanding of the human world, and these odd features echo ways that conceptual change researchers have noted in people undergoing conceptual change in the physical sciences.

Though conceptual change researchers differ in how they conceptualize the process of conceptual change, most agree that students have mental models of the world and that the process of conceptual change involves these models becoming more similar to the models that scientists use. That is, conceptual change involves more than just differences over facts, and panelists noted that non-experts disagreed with experts due to these deeper conceptual disagreements. One Anthropologist commented: "There is a deeper issue rather

than just facts (although it has to do with access to or proper understanding of facts). The problem is that non-experts do not fully understand evolutionary principles and processes."

We can also see that there are some concepts that people struggle to understand, which affects their mental models of the physical and social world alike. For example, people struggle to understand extremely long stretches of time. People struggle with "deep" or geological time (Trend, 1998; Trend, 2000), which acts as a critical barrier for understanding geoscience. Similarly, people struggle to think about the long timescale during which people have existed, which, while brief compared to geologic time, is still so large as to be difficult to conceptualize. One Anthropologist commented: "Most people cannot think in terms of the long span of time that humans have been on the planet, but Anthropologists think in terms of the evolution [of] humans over the long-term."

Additionally, many conceptual change researchers argue that the process of conceptual change involves both learning new conceptions and unlearning intuitive ones. However, these older conceptions never fully disappear. Panelists noted that even though their field rejected non-expert ways of viewing people in favor of the scientific conception, learning the scientific conception did not completely overwrite the previous understanding. For example, on a comparison about the importance of peer influence, one panelist commented: "My husband, for example, is a developmental psychologist and knows that peer influence [is greater than] parental influence. However, at least once a month he says something at home to suggest otherwise." In a comparison in which Psychologists rejected free will, one panelist commented: "I suspect most psychologists

believe in free will, and would only endorse a deterministic view if really asked to consider it thru their scholarly lens." One Historian argued that this causes their field to not think about race in as sophisticated a manner as they otherwise might: "Historians continue to use 'race' as a necessary category of understanding. They know the biology, but have not yet embraced the implications of the idea that there is only one human race. The historical record does not reflect the biological insight."

Just as non-experts have intuitive theories that are non-deterministic but predictable (such as models of the shape of the Earth), we see similar patterns in people's theories about the human world. Several panelists noted that non-experts did not all hold the same perspective and that there were several popular and mutually exclusive non-expert perspectives (see the comparisons with three components for examples of these). Which particular model a person will hold is unpredictable, but context and life experience will tend to cause some mental models to be more attractive to individuals than others. In line with this, panelists mentioned that some non-expert perspectives (such as the idea that people are separate from the natural world) would be more popular in urban areas than in rural ones.

In addition, panelists mentioned that ways of organizing the world spontaneously suggest themselves to people rather than needing to be learned, just as for many mental models in the physical sciences. For example, one Anthropologist commented "in my experience, few non-experts have ever thought about the language spoken by people living in small groups (although I agree that many non-experts who think about it, do make assumptions that the language is 'primitive.')"

People also suffer from an "illusion of explanatory depth" (Rozenblit & Keil, 2002; Keil, 2003), which leads them to grossly overestimate the degree to which their models of the world explain complex phenomena. According to Keil (2003):

> The illusion of understanding has been most extensively documented for our understandings of devices and then, secondarily, for knowledge of some biological organs and some non-living natural phenomena, such as the tides. It is likely to also hold for other complex causal systems, such as those governing human behaviour. (p. 369)

This does appear to be the case, as the Political Science panel suggested that "people overestimate their own understanding of the government/political system" (see Figure 9.1). Likewise, several suggested comparisons in the Psychology panel imply that people overestimate the degree to which they understand the reasons for people's behaviors.

The intuitive theories people hold about the social world, much like those they hold for the physical world, also echo the scientific theories of the past. For example, one Psychologist commented: "non[-]experts likely think people do things unconsciously as well, but may think about it more like a [F]reudian, motivated unconscious rather than the less intentional unconscious most experts think about."

Furthermore, many of the non-expert perspectives are based on ways that people are known to intuitively reason about the world. These ways of thinking drive their mental models of the physical world and seem to drive their models of the social world as well. For example, many suggested comparisons are about non-experts thinking essentialistically and overweighting identity as a predictor or explanator of behavior. Similarly, non-experts,

as they do for the physical world, also seem to be drawn to thinking in terms of entities rather than processes when thinking about the social world. As one Anthropologist commented: "Non-experts often find a very basic form of functionalism rather intuitively comprehensible, but struggle with culture or society as emergent phenomena." Likewise, Historians also mentioned that non-experts sometimes subscribe to a "great man" theory of History, which emphasizes the agency of powerful individuals rather than processes and context. Finally, people tend to reason teleologically, and we see evidence of this form of thinking in these comparisons as well. For example, one proposed comparison in the Psychology panel (dropped due to a consequentiality rating of 3.9) was: "Many non-experts think that people: have things happen to them for reasons. But Psychologists think that people: strive, sometimes arbitrarily, to place structure on random events."

From these, we can see some evidence that the intuitive theories people have about the social world are, in many ways, similar to the intuitive theories people have about the physical world. Consequently, even though this process is messier and harder to specify than it is for the physical world, it seems fair to characterize the process of shifting from a popular non-expert understanding of the social world to the corresponding expert understanding as a conceptual change.

Limitations

There were numerous limitations in this project. One major limitation is that there were often too few panelists participating. This was particularly acute for the Geography panel, but other panels suffered from low participation as well, especially as the project progressed. By the last round, the number of participating panelists had seriously dropped

across all panels. Coupled with panelists changing or forgetting their anonymous codes, this meant for several panels that we could no longer perform statistical tests on the results. Because of this, several comparisons were left in a state where they may or may not represent their field, but we are unable, within this study, to determine whether or not this is the case.

This loss of panelists may have been important in another way as well. It is possible that panelists left the study not in a random fashion (such as due to an increased workload), but for reasons related to the panel itself (such as dissatisfaction with the direction of the panel). This kind of attrition is problematic because it can mean that even if a panel began as a reasonable cross-section of the field, it may not have ended that way. Due to the anonymity of the panel, it is not possible ascertain why many panelists chose not to continue participating. However, the fact that most of the panelists who dropped did so before the first round implies it was not due to dissatisfaction with the panels themselves or their direction.

Another limitation was that we based our understanding of how non-experts think about the world entirely on how the experts said they did. That is, we could not see whether non-experts actually endorsed the non-expert views. However, this is not as large of a limitation as it may seem. Conceptual change researchers generally only query experts to understand common misconceptions and intuitive theories. This is done because non-experts are generally unaware of their models in detail. Indeed, many non-experts simply assume that their understanding of the world is simply a direct and unfiltered portrayal of reality, not a model, and often only reconsider this when their expectations are thwarted by

surprising events (and sometimes not even then). Secondly, our panel was based, by design, on professors as they are not merely experts in their field but are also expected to teach this understanding to others, and in so doing learn the ways that students tend to misunderstand important concepts. The value of teachers as experts on intuitive theories about the world has not gone unnoticed by researchers. For example, Jarrett, Takacs, and Ferry (2011) performed a Delphi to understand what concepts were required to understand climate change and included both members of climate research groups and high school teachers with experience teaching the topic. Indeed, the original studies underlying the field of conceptual change were performed, in some instances, by high school science teachers (though academically trained), precisely because they noted the discrepancy between how scientists model the world and how students do.

Another limitation was that not all panelists seemed engaged with the task, which likely reduces the depth and accuracy of suggested comparisons. While many panelists clearly engaged deeply with the project and gave thoughtful comments and clarifications on the comparisons, not all panelists did so. In another sign of a lack of engagement, several panelists gave responses that implied that they had not read the directions carefully. For example, they asked questions that had already been explained earlier in the study or responded to suggested comparisons and comments as though they had come directly from the researchers rather than from their fellow experts. This lack of engagement was likely due, in part, to the nature of the Delphi Method, as panelists commented that the study was repetitive. However, given that this repetition is essential to the Delphi Method, it is difficult to see how to prevent this.

Another limitation was the suboptimal phrasing of the questions. Partially this was due to the use of an online survey platform, which impacted how questions and answers were phrased, almost always in a manner that made them worse from a psychometric viewpoint. While we did pilot questions with colleagues, the need to convey information compactly combined with the limited control we had over displaying questions, meant we were often unhappy with the phrasings of questions.

However, many other problems with phrasings were not due to the format of the study, but rather, were due to suboptimal choices or mistakes on our part. For example, the use of the word "people" was somewhat ambiguous. This was done in order not to force panelists to think only about one particular scale (as a field may not focus on that scale) but came at the cost of occasionally making statements ambiguous as to which scale they referred to. Instead, it would have probably been better to give panelists a list of possible scales (individuals, dyads, groups, etc.) and let them select the correct one. This also would have had the benefit of eliminating comparisons that referenced different scales in their two components, and such comparisons were suggested several times (though they were typically rephrased by panelists in subsequent rounds). Likewise, as mentioned earlier, "non-expert" was probably not the optimal choice of phrase, and merely explaining to panelists our intended meaning was not sufficient to cause them to understand the term in the same way.

One important issue, mentioned by panelists repeatedly but never suggested as a comparison, is that people have profoundly different theories for how they work and how other people work (i.e. they are rational but others are not, they are good judges of

informational quality but others are not, they are not swayed by advertising but others are, etc.) whereas social scientists would probably consider people more similar. It is likely that this omission was caused by it being difficult to fit such comparisons into the format of the study.

Another limitation was caused by randomizing the order in which panelists saw the comparisons for their field. While randomizing the order is important to ensure that panelist's responses were not influenced by where the comparison fell in the study, randomizing the order also made it difficult to interpret several comments by panelists. Often, panelists would contrast comparisons with "earlier" comparisons, but we had no way of knowing the order in which they saw comparisons, and thus, which comparison they were referring to.

The largest limitation of this project, in our opinion, was that the project was mainly performed by a single researcher. This was problematic for two main reasons: the logistical requirements of the project, and the nature of the project. The logistical requirements of the project exceeded our estimates, and we failed to catch all mistakes before sending surveys to panelists.

The Delphi method required far more judgment calls than we anticipated. We chose the Delphi method originally because we felt it would reduce the degree to which our views (and our ignorance of other fields) would influence the end results of the study (compared to doing a literature review, a single survey, or interviews). While this likely was the case, the project nonetheless required an enormous number of judgment calls, some extremely subtle (for example, whether two suggested comparisons were actually different iterations

of the same underlying idea). Numerous times throughout this project, we wished that we had a room of philosophers we could have tossed such problems into (though our experience with philosophers makes us question whether this would have led to fewer conceptual debates or more).

In the end, while our perspective likely shaped the final output, we believe that the results do largely represent the viewpoints of the panelists for three reasons. Several of the suggested comparisons surprised us, many of the successful suggested comparisons are at odds with our own understanding of people, and several comparisons we would argue are critical for understanding people were dropped over the course of the study.

Conclusion

While this project demonstrates that the social sciences do agree in some ways on how to understand people, given the numerous debates within and across fields, we would have to agree with Wilson (1998) that the fields of the social sciences are far from reaching consilience.

However, the seeds of what this eventual unity might look like can already be glimpsed. For one, several long-standing debates within and across fields are (according to panelists) starting to ebb (for example, the debate between nature and nurture, which is in the process of being obviated in several fields). Secondly, we believe that new perspectives can shed light on the debates that remain, such as whether people are self-interested and whether they are rational. As it stands, such debates are nearly impossible to resolve given the lack of specificity, and virtually any action that might be considered irrational or selfless from one viewpoint can be recontextualized so that it is no longer so. New

perspectives may suggest better ways of framing such problems such that they become tractable.

It is likely that these new perspectives will have emergence as a key concept. According to Wilson (1998): "The greatest challenge today, not just in cell biology and ecology but in all of science, is the accurate and complete description of complex systems" (p. 85). Indeed, several of the new perspectives discussed here already embody this concept, such as the dynamic systems approach to conceptual change (Brown & Hammer, 2008), the Causal Bayesian Net interpretation of the theory theory (Gopnik & Wellman, 2012), and the Cultural Attraction Model (Claidière, Scott-Phillips, & Sperber, 2014). The concept of emergence allows these perspectives to understand trajectories (for learning, for culture, etc.) which have a mixture of regularities and inherent randomness, and while these theories should enhance our ability to predict which intuitive theories will become popular and which will not, the inherent randomness means they will never be able to do so perfectly.

From panelist comments, we can see that some panelists already embrace a complex systems perspective, but we can also see that some panelists do not do so. However, it seemed that panelists who did not agree with the complex systems perspective were unfamiliar with this perspective, rather than being familiar with the perspective and then rejecting it.

The idea of complex adaptive systems came up in the Anthropology panel[17], as one panelist suggested the comparison: "Non-experts think that people: work like mechanical

---

[17] This idea was also briefly mentioned in the Psychology panel (see Chapter 13).

systems. But Anthropologists think that people: work like complex adaptive systems."

Other panelists agreed with this, and suggested rephrasings that extended and clarified this comparison. For example, one panelist suggested the following comparison:

> Many non-experts think that human societies can be broken down into a series of component parts and studied as one studies a machine. But Anthropologists think that human societies are interconnected systems that express behavior at the aggregate level that is not expressed or expected at the individual level. That is, they are complex adaptive systems.

While this is a fairly strong and straightforward application of complex systems approaches to understanding people, not all Anthropologists agreed, seemingly because they were not familiar with this perspective. One panelist commented that "the original item does not seem to make much sense." Another commented that "The idea is that culture works like a complex adaptive system, but this idea is outdated in anthropology." Another Anthropologist strongly criticized this comment, arguing:

> The person who said that a complex adaptive systems approach in Anthropology is "outdated" clearly has no idea what a complex adaptive system is. It is a relatively recent body of theory that is only just now being applied in Anthropology. Likely s/he is thinking of the general "systems theory" approaches in the 1970's, which is a completely different body of thought.

From this, we can see that the complex systems perspective, while considered useful by some Anthropologists, is still far from being a dominant perspective. Given that people struggle with complex systems (Chi, 2005), it is not surprising that this perspective has met

with confusion and a lack of acceptance. Nonetheless, we would argue that the future scientific understanding of why people tend to think about the world in the ways they do will focus on emergence and on perspectives that embrace this concept, given the inherent complexity of people across scales.

Given the disagreements within and across fields, it is clear that the social sciences are far from having a complete understanding of people. For the foreseeable future, the social sciences will likely be a mixture of sophisticated version of intuitive theories (especially an overemphasis on intuitive entities such as personality), true scientific theories about people, and idiosyncratic theories, much as the physical sciences were for many centuries.

However, we should also appreciate how far the social sciences have come. Many of the insights the social sciences now offer for understanding people may have been literally unimaginable to scientists of the past, or imaginable only in the most metaphorical sense, much as general relatively would have been to early physicists. I believe that social scientists, despite the messiness that characterize these fields and the work still to be done, can be justifiably proud of what they have already accomplished.

In the end, whether we understand ourselves well or whether we fail to do so is more than just an academic question. Many of the problems we now face as a species (such as climate change) are not just (or even primarily) scientific or engineering problems, but problems of organizing humans on a never-before-attempted scale. As such, gaining an accurate understanding of ourselves, and teaching this understanding to the next generation, grows ever more important.

REFERENCES

Amsel, E., Baird, T., & Ashley, A. (2011). Misconceptions and conceptual change in undergraduate students' understanding of psychology as a science. *Psychology Learning & Teaching*, *10*(1), 3-10.

Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. New York: HarperCollins.

Ball, W., & Tronick, E. (1971). Infant responses to impending collision: Optical and real. *Science*, *171*(3973), 818-820.

Bardecki, M. J. (1984). Wetland conservation policies in Southern Ontario: A delphi approach. *Geographical Monographs, 16*.

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Journal of Developmental and Learning Disorders*, *5*(1), 47-78.

Bensley, D. A., & Lilienfeld, S. O. (2017). Psychological misconceptions: Recent scientific advances and unresolved issues. *Current Directions in Psychological Science*, *26*(4), 377-382.

Bernstein, A. C., & Cowan, P. A. (1975). Children's concepts of how people get babies. *Child Development*, *46*, 77-91.

Bloom, P., & Weisberg, D. S. (2007). Childhood origins of adult resistance to science. *Science*, *316*(5827), 996-997.

Boden, M. A. (1979) *Piaget.* Brighton, Sussex: The Harvester Press.

Brady, S. R. (2015). Utilizing and adapting the Delphi method for use in qualitative research. *International Journal of Qualitative Methods*, *14*, 1-6.

Brown, D. E., & Hammer, D. (2008). Conceptual change in physics. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 127-154). New York: Routledge.

Bryman, A. (1984). The debate about quantitative and qualitative research: a question of method or epistemology?. *British Journal of Sociology*, *35*, 75-92.

Buckland, L. A., & Chinn, C. A. (2010, June). Model-evidence link diagrams: A scaffold for model-based reasoning. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Proceedings of the 9th International Conference of the Learning Sciences: Volume 2* (pp. 449-450). Chicago, IL: International Society of the Learning Science.

Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*(5), 187-192.

Carney, P. A., Reisch, L. M., Piepkorn, M. W., Barnhill, R. L., Elder, D. E., Knezevich, S., Geller, B. M., Longton, G., & Elmore, J. G. (2016). Achieving consensus for the histopathologic diagnosis of melanocytic lesions: use of the modified Delphi method. *Journal of Cutaneous Pathology*, *43*(10), 830-837.

Carifio, J., & Biron, R. (1978). Collecting sensitive data anonymously: The CDRGP technique. *Journal of Alcohol and Drug Education, 23*(2), 47-66.

Carifio, J., & Biron, R. (1982). Collecting sensitive data anonymously: Further findings on the CDRGP technique. *Journal of Alcohol and Drug Education, 27*, 38-70.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carey, S. (1986). Cognitive science and science education. *American Psychologist*, *41*(10), 1123-1130.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*(1), 59-68.

Carey, S., & Spelke, E. (1996). Science and core knowledge. *Philosophy of Science*, *63*, 515-533.

Chaffin, W. W., & Talley, W. K. (1980). Individual stability in Delphi studies. *Technological Forecasting and Social Change*, *16*(1), 67-73.

Chi, M. T. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences, 14(2)*, 161-199.

Chi, M. T., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, *4*(1), 27-43.

Childress, V. W., & Rhodes, C. (2008). Engineering student outcomes for grades 9-12. *The Technology Teacher*, *67*(7), 5.

Chinn, C. E. (1998). A critique of social constructivist explanations of knowledge change. In B. Guzzetti & C. Hynd (Eds.), *Perspectives on conceptual change: Multiple ways to understand knowing and learning in a complex world* (pp. 77–115). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Christian, D. (2017). What is big history?. *Journal of Big History*, *1*(1), 4-19.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, *78*(2), 67-90.

Clayton, M. J. (1997). Delphi: a technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology*, *17*(4), 373-386.

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How Darwinian is cultural evolution?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1642). 20130368.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, *50*(1), 66-71.

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition*, *50*(1), 41-77.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*(3), 458-467.

Dailey, A. L., & Holmberg, J. C. (1990). Delphi – A catalytic strategy for motivating curriculum revision by faculty. *Community College Journal of Research and Practice,* *14*, 129-136.

Dajani, J. S., Sincoff, M. Z., & Talley, W. K. (1979). Stability and agreement criteria for the termination of Delphi studies. *Technological Forecasting and Social Change*, *13*(1), 83-90.

Davidson, P. L. (2013). The Delphi technique in doctoral research: Considerations and rationale. *Review of Higher Education and Self-Learning*, *6*(22), 53-65.

Davies, P., & Mangan, J. (2007). Threshold concepts and the integration of understanding in economics. *Studies in Higher Education*, *32*(6), 711-726.

de Ribaupierre, A. (2015). Piaget's theory of child development. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed.), *18* (120-124). Amsterdam, NL: Elsevier.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P.B. Pufall (Eds.), *Constructivism in the computer age* (pp. 49-70). Hillsdale, NJ: Erlbaum.

diSessa, A. A. (2008). A bird's-eye view of the" pieces" vs" coherence" controversy (from the "pieces" side of the fence). In S. Vosniadou (Ed.), *International handbook of research on conceptual chang*e (pp. 35-60). New York: Routledge.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist*, *33*(2-3), 109-128.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7(10)*, 454-459.

Filyushkina, A., Strange, N., Löf, M., Ezebilo, E. E., & Boman, M. (2018). Applying the Delphi method to assess impacts of forest management on biodiversity and habitat preservation. *Forest Ecology and Management*, *409*, 179-189.

Flavell, J. H. (1996). Piaget's legacy. *Psychological Science*, *7*(4), 200-203.

Gelman, S. A. (2000). The role of essentialism in children's concepts. *Advances in Child Development and Behavior, 27*, 55–98.

Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology (Vol. 3: Cognitive development).* New York: Wiley.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165-193.

Gerstenberg, T., & Tenenbaum, J. B. (2017). Intuitive theories. In M. Waldmann (Ed.), *Oxford handbook of causal reasoning* (pp. 515-548). New York: Oxford University Press.

Gruber, H. E., & Vonèche, J. J. (Eds.). (1977). *The essential Piaget*. New York: Basic Books.

Gnatzy, T., Warth, J., von der Gracht, H., & Darkow, I. L. (2011). Validating an innovative real-time Delphi approach – A methodological comparison between real-time and conventional Delphi studies. *Technological Forecasting and Social Change*, *78*(9), 1681-1694.

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science*, *20*(4), 480-487.

Goldman, K., Gross, P., Heeren, C., Herman, G., Kaczmarczyk, L., Loui, M. C., & Zilles, C. (2008). Identifying important and difficult concepts in introductory computing courses using a delphi process. *ACM SIGCSE Bulletin*, *40*(1), 256-260.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257-293). New York: Cambridge University Press.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The scientist in the crib: Minds, brains, and how children learn*. New York: William Morrow & Company, Inc.

Gordon, T. J. (2009). The Delphi method. In J. C. Glenn & T. J. Gordon (Eds.), *Futures research*

    methodology: Version 3.0 (chapter 4). [Electronic Book]. The Millennium Project.

    Available from http://www.millennium-project.org/publications-2/

Gordon, T., & Pease, A. (2006). RT Delphi: an efficient, "round-less" almost real time Delphi

    method. *Technological Forecasting and Social Change*, *73*(4), 321-333.

Gordon, T. J., & Helmer, O. (1964). *Report on a long-range forecasting study*. Santa Monica,

    CA: RAND Corporation.

Graefe, A., & Armstrong, J. S. (2011). Comparing face-to-face meetings, nominal groups,

    Delphi and prediction markets on an estimation task. *International Journal of*

    Forecasting, *27*(1), 183-195.

Green, R. A. (2014). The Delphi technique in educational research. *Sage Open*, *4*(2), 1-8.

Halldén, O. (1997). Conceptual change and the learning of history. *International Journal of*

    Educational Research, *27*(3), 201-210.

Helmer, O. (1967). *Analysis of the future: The Delphi method*. Santa Monica, CA: Rand

    Corporation.

Helmer, O., & Rescher, N. (1959). On the epistemology of the inexact sciences. *Management*

    Science, *6*(1), 25-52.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American*

    Journal of Psychology, *57*(2), 243-259.

Hill, K. Q., & Fowles, J. (1975). The methodological worth of the Delphi forecasting

    technique. *Technological Forecasting and Social Change*, *7*(2), 179-192.

Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds.* Cambridge, MA: MIT Press.

Holden, M. C., & Wedman, J. F. (1993). Future issues of computer-mediated communication: The results of a Delphi study. *Educational Technology Research and Development, 41*(4), 5-24.

Hughes, S., Lyddy, F., & Lambe, S. (2013). Misconceptions about psychological science: A review. *Psychology Learning & Teaching*, *12*(1), 20-31.

Hynd, C. (1998). Conceptual change in a high school physics class. In B. Guzzetti & C. Hynd (Eds.), *Perspectives on conceptual change: Multiple ways to understand knowing and learning in a complex world* (pp. 27–36). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Jarrett, L., Ferry, B., & Takacs, G. (2012). Development and validation of a concept inventory for introductory-level climate change science. *International Journal of Innovation in Science and Mathematics Education, 20*, 25-41.

Jarrett, L. E., Takacs, G., & Ferry, B. (2011, August). What scientific concepts are required to understand climate change?. In M. Sharma, A. Yeung, T. Jenkins, E. Johnson, G. Rayner, & J. West (Eds.) *Proceedings of The Australian Conference on Science and Mathematics Education 2011.* Paper presented at the 17th Annual Uniserve Science Conference, Sydney: Universe Science.

Jorm, A. F. (2015). Using the Delphi expert consensus method in mental health research. *Australian & New Zealand Journal of Psychiatry*, *49*(10), 887-897.

Kamps, F. S., Julian, J. B., Battaglia, P., Landau, B., Kanwisher, N., & Dilks, D. D. (2017).

Dissociating intuitive physics from intuitive psychology: Evidence from Williams syndrome. *Cognition*, *168*, 146-153.

Kaplan, A., Skogstad, A. L., & Girshick, M. A. (1950). The prediction of social and technological events. *Public Opinion Quarterly*, *14*(1), 93-110.

Kaufman, D. R., Keselman, A., & Patel, V. L. (2008). Changing conceptions in medicine and health. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 295-327). New York: Routledge.

Kearney, K. A., Hopkins, R. H., Mauss, A. L., & Weisheit, R. A. (1984). Self-generated identification codes for anonymous collection of longitudinal questionnaire data. *Public Opinion Quarterly*, *48*(1B), 370-378.

Keil, F. C. (2003). Folkscience: Coarse interpretations of a complex reality. *Trends in Cognitive Sciences*, *7*(8), 368-373.

Kelemen, D. (1999). The scope of teleological thinking in preschool children. *Cognition*, *70*(3), 241-272.

Keller, A. (2001). Future development of electronic journals: a Delphi survey. *The Electronic Library, 19*(6), 383-396.

Kelly, G. J., & Green, J. (1998). The social nature of knowing: Toward a sociocultural perspective on conceptual change and knowledge construction. In B. Guzzetti & C. Hynd (Eds.), *Perspectives on conceptual change: Multiple ways to understand knowing and learning in a complex world* (pp. 145-181). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (pp. 263-314). Bingley, UK: Emerald Group Publishing Ltd.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: University of Chicago Press.

Land, R., Cousin, G., Meyer, J. H. F. & Davies, P. (2005) Threshold concepts and troublesome knowledge (3): implications for course design and evaluation, in: C. Rust (Ed.) *Improving student learning: diversity and inclusivity* (pp. 53–64). Oxford: Oxford Centre for Staff and Learning Development.

Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change*, *73*(5), 467-482.

Laudan, L. (1971). William Whewell on the consilience of inductions. *The Monist*, *55*, 368-391.

Leinhardt, G., & Ravi, A. K. (2008). Changing historical conceptions of history. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 328-341). New York: Routledge.

Leiser, D. & Krill, Z. (2016) How Laypeople Understand Economics. In R. Ranyard (Ed.), *Economic Psychology: The Science of Economic Mental Life and Behaviour* (pp. 139 – 154). Hoboken: Wiley-Blackwell.

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119-148). Cambridge: Cambridge University Press.

Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design evidence of a preference for teleological explanations in patients with Alzheimer's Disease. *Psychological Science*, *18*(11), 999-1006.

Looney, S. W., & Jones, P. W. (2003). A method for comparing two normal means using combined samples of correlated and uncorrelated data. *Statistics in Medicine*, *22*(9), 1601-1610.

Manizade, A. G., & Mason, M. M. (2011). Using Delphi methodology to design assessments of teachers' pedagogical content knowledge. *Educational Studies in Mathematics*, *76*(2), 183-207.

Martino, J. P. (1993). *Technological forecasting for decision making* (3rd edition). New York: McGraw-Hill, Inc.

Mazens, K., & Lautrey, J. (2003). Conceptual change in physics: Children's naive representations of sound. *Cognitive Development*, *18*(2), 159-176.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. *Science*, *210*(4474), 1139-1141.

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299-323). Hillsdale, NJ: Erlbaum.

McDermott, L. C. (1984). Research on conceptual understanding in mechanics. *Physics Today*, *37*, 24-32.

McElhinney, E. (2010). Factors which influence nurse practitioners ability to carry out physical examination skills in the clinical area after a degree level module–an electronic Delphi study. *Journal of Clinical Nursing, 19*, 3177-3187.

McInerney, J. O., O'Connell, M. J., & Pisani, D. (2014). The hybrid nature of the Eukaryota and a consilient view of life on Earth. *Nature Reviews Microbiology*, *12*(6), 449.

Meyer, J. H. F. & Land, R. (2003) Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising within the disciplines, in: C. Rust (Ed.) *Improving student learning: theory and practice—10 years on* (pp. 412-442). Oxford: Oxford Centre for Staff and Learning Development.

Mills, R., Tomas, L., & Lewthwaite, B. (2016). Learning in Earth and space science: A review of conceptual change instructional approaches. *International Journal of Science Education*, *38*(5), 767-790.

Minstrell, J. (1982). Explaining the" at test" condition of an object. *Physics Teacher*, *20*(1), 10-14.

Miton, H., & Mercier, H. (2015). Cognitive obstacles to pro-vaccination beliefs. *Trends in Cognitive Sciences*, *19*(11), 633-636.

Miton, H., Claidière, N., & Mercier, H. (2015). Universal cognitive mechanisms explain the cultural success of bloodletting. *Evolution and Human Behavior*, *36*(4), 303-312.

Mozuni, M., & Jonas, W. (2018). An introduction to the morphological delphi method for design: A tool for future-oriented design research. *She Ji: The Journal of Design, Economics, and Innovation*, *3*(4), 303-318.

Murphy, P. K., & Alexander, P. A. (2008). The role of knowledge, beliefs, and interest in the conceptual change process: A synthesis and meta-analysis of the research. In S. Vosniadou (Ed.), *International handbook of research on conceptual change*, (pp. 583-616). New York: Routledge.

Nancy, L., & Ogborn, J. (1994). Students as expert system developers: A means of eliciting and understanding commonsense reasoning. *Journal of Research on Computing in Education, 26*(4), 497-513.

Nixon, H. K. (1925). Popular answers to some psychological questions. *The American Journal of Psychology*, *36*(3), 418-423.

Nussbaum, E. M., Sinatra, G. M., & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education*, *30*(15), 1977-1999.

Ohlsson, S. (2009). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist*, *44*(1), 20-40.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management, 42*(1), 15-29.

Oreskes, N., (2007) The scientific consensus on climate change: How do we know we're not wrong? In J. F. C. DiMento & P. Doughman (Eds.) *Climate change: What it means for us, our children, and our grandchildren* (pp. 65-99). Cambridge, MA: MIT Press.

Pascual-Leone, J., & Smith, J. (1969). The encoding and decoding of symbols by children: A new experimental paradigm and a neo-Piagetian model. *Journal of Experimental Child Psychology*, *8*(2), 328-355.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L.

    Berkowitz (Ed.), *Advances in Experimental Social Psychology, Vol. 19* (pp. 123-205).

    New York: Academic Press.

Piaget, J. (1952). *The child's concept of number.* (C. Gattegno & F. M. Hodgson, Trans.). New

    York: Routledge. (Original work published 1941)

Piattelli-Palmarini, M. (1994). Ever since language and learning: Afterthoughts on the

    Piaget-Chomsky debate. *Cognition*, *50*(1), 315-346.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role

    of motivational beliefs and classroom contextual factors in the process of conceptual

    change. *Review of Educational Research, 63(2)*, 167-199.

Plantinga, C. (2011). Folk psychology for film critics and scholars. *Projections*, *5*(2), 26-50.

Pluta, W. J., Buckland, L. A., Chinn, C. A., Duncan, R. G., & Duschl, R. A. (2008, June). Learning

    to evaluate scientific models. In *Proceedings of the 8th international conference on*

    International conference for the learning sciences-Volume 3 (pp. 411-412).

    International Society of the Learning Sciences. Utrecht, The Netherlands.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a

    scientific conception: Toward a theory of conceptual change. *Science Education,*

    66(2), 211-227.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?.

    Behavioral and Brain Sciences, *1*, 515-526.

Rauch, W. (1979). The decision delphi. *Technological Forecasting and Social Change*, *15*(3),

    159-169.

Rhodes, M. (2013). How two intuitive theories shape the development of social categorization. *Child Development Perspectives*, *7*(1), 12-16.

Rhodes, M., & Wellman, H. (2017). Moral learning as intuitive theory revision. *Cognition*, *167*, 191-200.

Rieger, W. G. (1986). Directions in Delphi developments: Dissertations and their quality. *Technological Forecasting and Social Change*, *29*(2), 195-204.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting*, *15*(4), 353-375.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*(5), 521-562.

Sackman, H. (1974). *Delphi assessment: Expert opinion, forecasting and group procedure.* Santa Monica, CA: RAND Corporation.

Sadler, P. M. (Producer) & Schneps, M. H. (Director). (1989). *A private universe.* [Video]. *USA: Annenberg Media.*

Samawi, H. M., & Vogel, R. (2014). Notes on two sample tests for partially correlated (paired) data. *Journal of Applied Statistics*, *41*(1), 109-117.

Schnell, R., Bachteler, T., & Reiher, J. (2010). Improving the use of self-generated identification codes. *Evaluation Review*, *34*(5), 391-418.

Séré, M. G. (1982). A study of some frameworks used by pupils aged 11 to 13 years in the interpretation of air pressure. *European Journal of Science Education*, *4*(3), 299-309.

Shepardson, D. P., Choi, S., Niyogi, D., & Charusombat, U. (2011). Seventh grade students' mental models of the greenhouse effect. *Environmental Education Research*, *17*(1), 1-17.

Shtulman, A. (2009). Rethinking the role of resubsumption in conceptual change. *Educational Psychologist*, *44*(1), 41-47.

Shtulman, A., & Calabi, P. (2012). Cognitive constraints on the understanding and acceptance of evolution. In K. S. Rosengren, S. K. Brem, E. M. Evans, & Sinatra, G. M. (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 47-65). Cambridge, UK: Oxford University Press.

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, *8*(1), 118-137.

Shtulman, A., & McCallum, K. (2014). Cognitive reflection predicts science understanding. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2937–2942). Quebec City: Cognitive Science Society.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, *32*(6), 1049-1062.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*(2), 209-215.

Sinatra, G. M., Kardash, C. M., Taasoobshirazi, G., & Lombardi, D. (2012). Promoting attitude change and expressed willingness to take action toward climate change in college students. *Instructional Science, 40(1)*, 1-17.

Sinatra, G. M., & Mason, L. (2013). Beyond knowledge: Learner characteristics influencing conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change,* (pp. 377-394). New York: Routledge.

Sinha, I. P., Smyth, R. L., & Williamson, P. R. (2011). Using the Delphi technique to determine which outcomes to measure in clinical trials: Recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*, *8*(1), 1-5.

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education*, *6*, 1-21.

Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition*, *50*(1), 431-445.

Spelke, E. S., Katz, G., Purcell, S. E., Ehrlich, S. M., & Breinlinger, K. (1994). Early knowledge of object motion: Continuity and inertia. *Cognition*, *51*(2), 131-176.

Stich, S. P. & Nichols, N. (2003) Folk Psychology. In S. P. Stich & T. A. Warfield (Eds.), *The Blackwell Guide to Philosophy of Mind (pp.* 235-255). Malden, MA: Blackwell Publishing Ltd.

Strand, J., Carson, R. T., Navrud, S., Ortiz-Bobea, A., & Vincent, J. R. (2017). Using the Delphi method to value protection of the Amazon rainforest. *Ecological Economics*, *131*, 475-484.

Streveler, R. A., Litzinger, T. A., Miller, R. L., & Steif, P. S. (2008). Learning conceptual knowledge in the engineering sciences: Overview and future research directions. *Journal of Engineering Education*, *97*(3), 279-294.

Streveler, R. A., Olds, B. M., Miller, R. L., & Nelson, M. A. (2003, June). Using a Delphi study to identify the most difficult concepts for students to master in thermal and transport

science. In *Proceedings of the 2003 Annual Conference of the American Society for Engineering Education*. Nashville, TN: American Society for Engineering Education.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschi & R. J. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice* (pp. 147 – *176)*. Albany, NY: SUNY Press.

Subbotsky, E. (1997). Explanations of unusual events: Phenomenalistic causal judgments in children and adults. *British Journal of Developmental Psychology*, *15*(1), 13-36.

Thagard, P. (2008). Conceptual change in the history of science: Life, mind, and disease. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 374-387). New York: Routledge.

Townsend, L., Hofer, A. R., Hanick, S. L., & Brunetti, K. (2016). Identifying threshold concepts for information literacy: A Delphi study. *Communications in Information Literacy*, *10*(1), 23.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275-1289.

Trend, R. (1998). An investigation into understanding of geological time among 10-and 11-year-old children. *International Journal of Science Education*, *20*(8), 973-988.

Trend, R. D. (2001). Deep time framework: A preliminary study of UK primary teachers' conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching, 38*(2), 191-221.

Turoff, M. (1970). The design of a policy Delphi. *Technological Forecasting and Social change*, *2*(2), 149-171.

Ungar, S. (2000). Knowledge, ignorance and the popular culture: Climate change versus the ozone hole. *Public Understanding of Science*, *9*(3), 297-312.

Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education*, *1(2)*, 205-221.

von der Gracht, H. A. (2012). Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological Forecasting and Social Change*, *79*(8), 1525-1536.

Vosniadou, S. (2007). Conceptual change and education. *Human Development*, *50*(1), 47-54.

Vosniadou, S. (2008). Conceptual change research: An introduction. In S. Vosniadou (Ed.), *International handbook of research on conceptual chang*e (pp. 127-154). New York: Routledge.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535-585.

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual chang*e (pp. 3-34). New York: Routledge.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*(1), 337-375.

Whewell, W. (1840). *The philosophy of the inductive sciences, founded upon their history.* London: J. W. Parker.

White, R., & Gunstone, R. (2008). The conceptual change approach and the teaching of

> science. In S. Vosniadou (Ed.), *International handbook of research on conceptual*
>
> *chang*e (pp. 619-628). New York: Routledge.

Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York: Alfred A. Knoff, Inc.

Woudenberg, F. (1991). An evaluation of Delphi. *Technological Forecasting and Social*

> *change*, *40*(2), 131-150.

Zawacki-Richter, O. (2009). Research areas in distance education: A Delphi study. *The*

> *International Review of Research in Open and Distributed Learning*, *10*(3).

Ziglio, E. (1996). The Delphi method and its contribution to decision-making. In M. Adler &

> E. Ziglio (Eds.), *Gazing into the oracle: The Delphi method and its application to social*
>
> *policy and public health* (pp. 3-33). London: Jessica Kingsley.

APPENDIX A

Directions for Suggesting Comparisons (Round 1)

How do experts in your field, compared to non-experts, understand people?

Social scientists often spend decades studying people, and through this process develop a sophisticated understanding of people and why they think and act in the ways that they do. In this project, we seek to understand what these sophisticated understandings look like. That is, what do experts in your field learn about people in the course of becoming an expert?

For this project, you are going to be asked to contrast **commonly-held misconceptions** about people with your field's **sophisticated conceptions** about people. It may help to imagine that you have been asked to give a public lecture on a common misconception regarding people. What would your lecture be about?

For this project, these **commonly-held misconceptions** about people should be:
-Held by many (but not necessarily all) non-experts.
-Important to laypeople – A way of thinking that laypeople rely on to explain, predict, and make sense of how people think and act.

In contrast, these **sophisticated conceptions** about how people work should be:
-Broadly agreed upon by most (but not necessarily all) experts in your field.
-Central to your field – A way of thinking that experts in your field rely on to explain, predict, and make sense of how people think and act.

Directions:
On the next page, you will answer a few questions of the following form:

Example of a Comparison:

Many non-experts think that people:
But experts in my field think that people:

Please answer these in **plain language**, avoiding reference to specialist terms or specific theories. As much as you can, describe the misconceptions and correct conceptions in such a way that people lacking expertise in your field could understand them.

Please press the next button to begin.

APPENDIX B

Example of Rating a Comparison (Round 2)

The following has been proposed as a **shift** caused by becoming an expert in History:

Many non-experts think that people <u>are either "good" or "evil."</u>
But Historians think that people <u>cannot be easily summarized and are complex.</u>

As best as you can **estimate**:

	Almost none	A small minority	A large minority	Half	A small majority	A large majority	Nearly all
What proportion of **non-experts** think in the **first** way?	○	○	○	○	○	○	○
What proportion of **Historians** think in the **second** way?	○	○	○	○	○	○	○

Assuming the shift above is real, how much would this **change** how one understands and explains the human world?

Not at all	Slightly	A little	Moderately	A lot	Greatly	Completely
○	○	○	○	○	○	○

(optional) If you would like to assist **the researchers** by clarifying or rephrasing this item, please do so here:

Many non-experts think that people:

But Historians think that people:

(optional) If you would like to explain your ratings **to your colleagues**, please do so here. (be sure to note which rating you are explaining)

APPENDIX C

Example of Rating a Comparison (Round 3/4/5)

Many **non-experts** think that people <u>generally know what's good for them in terms of health, relationships, happiness, etc.</u>

But **Psychologists** think that people <u>are pretty bad at understanding themselves and knowing how best to get what they want.</u>
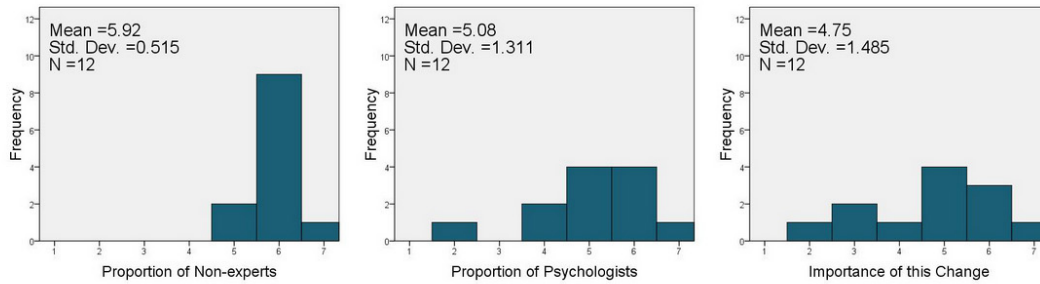
── Your Ratings for this Round ──

As best as you can estimate:

	Almost none	A small minority	A large minority	Half	A small majority	A large majority	Nearly all
What proportion of **non-experts** think in the **first** way?	○	○	○	○	○	○	○
What proportion of **Psychologists** think in the **second** way?	○	○	○	○	○	○	○

Assuming the shift above is real, how much would this **change** how one understands and explains the human world?

Not at all	Slightly	A little	Moderately	A lot	Greatly	Completely
○	○	○	○	○	○	○

── Panelist Ratings from Previous Rounds ──

Mean =5.92
Std. Dev. =0.515
N =12
Proportion of Non-experts

Mean =5.08
Std. Dev. =1.311
N =12
Proportion of Psychologists

Mean =4.75
Std. Dev. =1.485
N =12
Importance of this Change

── Panelist Comments from Previous Rounds ──   Add Comments for Future Rounds (Optional) ──

Panelist comments about the **non-expert statement**:
None

Your comments about the **non-expert statement**:

Panelists comments about the **Psychologists statement**:
-I suspect most social psychologists would agree with the second statement; I'm not sure whether psychologists in other areas would be aware of the relevant findings.

Your comments about the **Psychologists statement**:

Panelist Comments about the **importance of this change**:
None

Your comments about the **importance of this change**: