University of Oxford

DPhil in Public Policy

Culture and Capability:

Essays in behavioural economics

Author:

James Walsh

Supervisors:

Professor Sir Paul Collier

Professor Stefan Dercon

Submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

Blavatnik School of Government

February 24, 2022



UNIVERSITY OF OXFORD

Abstract

Blavatnik School of Government

Doctor of Philosophy

Culture and Capability:
Essays in behavioural economics

by James Walsh

How do mental models guide behaviour? Although a great deal of attention has been paid to mental models in psychology, relatively little is understood about how they influence incentivised decision-making or their relevance for policy. In this dissertation, I explore this question over the course of four papers. In the first paper, I synthesize the literature on mental models and aim to situate it in standard decision theory. In paper two, I look at how people use mental models of social groups to determine strategies of action under incentivized conditions. In paper three, I look at how unemployed young people's mental model of their agency within the world influences their decisions to search for work. In paper four, I look at how mental models of technology influence vaccine hesitancy.

Contents

Abstract iii

0.1 Paper 1 summary: Mental models . . . 3
0.2 Paper 2 summary: Social groups . . . 3
0.3 Paper 3 summary: The self . . . 4
0.4 Paper 4 summary: Technology . . . 4

1 Mental Models 5

1.1 Introduction . . . 5
1.2 An introduction to mental models . . . 7
1.3 Mental models in action . . . 13
1.4 Where mental models come from . . . 19
1.5 Mental models and capability development . . . 22
1.6 Policymakers' mental models . . . 26
1.7 Conclusion . . . 29

2 Social groups 31

2.1 Introduction . . . 31
2.2 Stereotyping and Discrimination . . . 33
2.3 Contribution . . . 34
2.4 Overview of methods . . . 36
2.5 Summary of experiments . . . 37
2.6 Hypotheses . . . 38
2.7 Specification . . . 38
2.8 Results . . . 40
2.9 Discussion . . . 46
2.10 Conclusion . . . 48
2.11 Appendix . . . 49

3 The self 115

3.1 Introduction . . . 115
3.2 Conceptual framework . . . 118
3.3 Intervention and experiment . . . 119
3.4 Data . . . 124
3.5 Empirical Strategy . . . 127
3.6 Results . . . 128
3.7 Conclusion . . . 135
3.8 Appendix . . . 137

4 Technology **157**
4.1 Introduction . . . 157
4.2 Intervention and experiment . . . 161
4.3 Data and Outcomes of interest . . . 168
4.4 Empirical approach . . . 169
4.5 Results . . . 171
4.6 Discussion . . . 175
4.7 Conclusion . . . 177
4.8 Appendix . . . 178

Bibliography **195**

In memory of Fergal

Acknowledgements

Pursuing this DPhil at Oxford has been an extraordinary privilege, which was made possible by the support and sacrifice of many people.

First, I am extremely grateful to my supervisors, Paul Collier and Stefan Dercon, for their support and encouragement throughout the DPhil. Whenever I asked, Paul and Stefan were on hand to help; whether it meant reviewing papers over weekends or even making a trip to Ireland. No question was too big or too small. They helped me to structure papers, fix regressions, apply for funding, and think through research ideas. When my interest in development economics began, *The Bottom Billion* was one of the first books I read. While working at the World Bank, Stefan played a key role (both through his academic and policy work) in introducing psychology into the operations of the Bank. I never imagined I would have the honour of working under the guidance of such incredible and generous scholars.

Two people whose ideas had a profound influence on this thesis were Karla Hoff and Susan Fiske. Karla, my boss while at the World Bank, introduced the idea of mental models while she was Co-Director of the *World Development report 2015*. During my time at the Bank, Karla was an outstanding boss and mentor to me, and I had the opportunity to co-author two papers with Karla during the DPhil – in the process learning so much about how to write and how to think about social science. I had admired Susan's work on stereotyping for years before we met. When she visited Oxford, I was delighted to have the opportunity to persuade her to collaborate. It was not until we began to work together, however, that I realised that her expertise extended far beyond stereotyping and prejudice. In fact, due to the astonishing breadth and depth of her ideas, Susan stands out as one of the great social psychologists. Susan also provided invaluable mentorship through the DPhil.

There are many people at Oxford who I would like to thank – most of all Rose Page, the Administrator of CSAE, who shepherded me through a number of projects. I am also deeply grateful to the staff at Nuffield College, who made a home for me over 4 years. I am especially appreciative to the staff at the Buttery, the Lodge, and in Housekeeping. Eleni and Steve were wonderful sources of advice and pastoral care.

I am very grateful to the team at the Blavatnik School of Government and to be part of this incredible department and community. In particular, Anandi Mani, Kate Orkin, and Clare Leaver provided exceptional support through the seminars and Transfer and Confirmation examinations. Julien Labonne and Pepper Culpepper were wonderful supports as co-ordinators of the DPhil programme.

I would also like to thank friends and co-authors from Oxford and elsewhere, Vafa Ghazani, Ashley Pople, Jasmine Bhatia, Ruari Maguire, Naomi Vaida, Carolina Meijia-Mantilla, Sonja Vogt, Thokozani Chikuni, and Salman Asim.

I would like to thank my family: my mother Colette, my father Liam, and my brother Rob. During the lockdown, we ended up staying in Ireland for 4 months due to travel restrictions. Getting to spend so much time with family was one of the few silver linings of the pandemic.

Most importantly, I am sincerely thankful to my wife Laura, who moved to the United Kingdom to be with me while I pursued the DPhil. Laura uprooted a happy life in Washington DC to move to Oxford. As only Laura can do, she found ways to turn the move into an adventure. For the immense support she has given me and all that she has had to put up with, I am extremely grateful.

Introduction

This thesis is about the scope for applying the construct of mental models in economic and policy analysis. It includes 4 papers, which are briefly summarised below. The first paper is a conceptual piece that situates 'mental models' within behavioural economics and discusses its policy implications. The second paper reports findings from a series of experiments conducted in the United States in India intended to test the implications of mental models of social groups (i.e., stereotypes) on distinct forms of discriminatory behaviour. The third paper reports findings from a field experiment in Ethiopia that tested the effect of workshops intended to shift unemployed young people's mental models of themselves and their job search. The fourth paper reports finding from an experiment conducted in the United Kingdom that shared information about the COVID-19 vaccine with people who are vaccine hesitant.

0.1 Paper 1 summary: Mental models

The first paper situates the "mental models" construct within the economics and policy literature. The paper has several goals. First, it aims to define what mental models are, explain where the mental model construct comes from, and summarise how mental models are stored in memory. Second, it discusses how mental models are used to process information and make decisions under uncertainty. Third, it goes on to discuss the information sources through which people acquire mental models. Finally, it addresses policy implications through two lenses. It looks at how people's capabilities can be expanded through interventions targeting their mental models. It then considers the role of policymakers' mental models in designing good social programmes.

0.2 Paper 2 summary: Social groups

The second chapter describes 6 experiments conducted in the United States and India that were designed to investigate whether the dimensions of the Stereotype Content Model

(warmth and competence) elicit distinct behaviours under incentivized conditions. The paper is co-authored with Susan Fiske and Naomi Vaida.[1] It is currently being prepared for submission to the *Proceedings of the National Academy of Science Nexus*. The paper shows that warmth and competence guide distinct discriminatory behaviour.

0.3 Paper 3 summary: The self

The third chapter reports results from a field experiment conducted in Addis Ababa, Ethiopia designed to expose young people to ways of seeing the world that restore their sense of self-integrity (i.e., self-affirmation exercises, goal-setting exercises). The paper is co-authored with Carolina Meijia-Mantilla.[2] The paper is being prepared for an economics journal such as the Journal of Development Economics or the Journal of Economic Behavior and Organization or a general science journal such as Nature Human Behavior or PLOS one.

0.4 Paper 4 summary: Technology

The fourth chapter reports results from a online experiment conducted in the UK that examined ways to communicate information about the risks and benefits of the COVID-19 vaccine. It compares information conveyed as objective knowledge with information embedded in story. In addition, it compares two strategies for building trust emphasizing good intentions (i.e., warmth) and expertise (i.e., competence). To test this these approaches, I conduct an online experiment with 1,042 vaccine-hesitant British participants. I oversample Black participants because they are five-times more likely to be hesitant to take the vaccine. I find that people who read the story are significantly more likely to favourably update mental models of vaccines and increase their trust in the government. I also find that story design matters. White participants only update in the condition where the doctor is conveyed as competent. On the other hand, Black participants only update in the condition where the doctor is conveyed as warm.

[1] I am the first-author and conducted all data collection and analysis. I took the lead on the writing with edits from Susan and Naomi.

[2] I led the implementation of the study in Ethiopia. We drafted the pre-anlaysis plan together and conducted the data analysis together. I took the lead on the writing with additions made by Carolina.

Chapter 1

Mental Models

1.1 Introduction

Policy analysis is generally conducted using economic theory, where human behaviour is modelled based on three variables: preferences, constraints, and beliefs. The model provides behavioural predictions that governments can use to design policy. Consider each variable – preferences, constraints, and beliefs – in turn. Preferences are the subjective comparative evaluations that people make about their desired state of the world. Governments can shift people's (induced) preferences by changing the incentive structure of the choice environment, for example by taxing or subsidising products. Constraints are factors that determine people's ability to satisfy their preferences, such as their budget. Governments can give people resources to help them overcome constraints, for example through welfare payments. Beliefs are people's subjective expectations about the likelihood of different states of the world. Governments can influence people's beliefs by sharing information, for example through public service announcements.

Despite the power of models of human behaviour based on preferences, constraints, and beliefs, their predictions are limited by the omission of a fourth key variable: mental models. Mental models are the cognitive structures people use to process information and make predictions. They include categories, causal narratives, and worldviews (see Table 1.1). Mental models are a central mechanism through which culture influences behaviour (DiMaggio, 1997; Hoff and Stiglitz, 2016). Governments can influence mental models through education, public communication, and choice architecture. The mental model construct is best captured formally by a sparsity-based model of bounded rationality, in which people construct simplified models of the world (Gabaix, 2014; Gabaix, 2018).

Rational expectations theory assumes that people hold mental models that structure data in ways that are consistent with the underlying structure of the world (Manski, 2004; Andre et al., 2019).[1] However, this is by no means guaranteed. For one thing, the underlying structure of the world is often unclear even to scientific and policy experts. Moreover, even when people formulate models of the world, they often disagree intensely about what the appropriate model should be. Indeed, scientists expend a great deal of effort coordinating on shared models of reality (Kuhn, 1970).

Mental models thus are key to explaining important social challenges at different levels of analysis, from scientific discovery and economic performance to polarization and social exclusion. In recognition of this analytical gap, economists have increasingly drawn on mental structures to explain social phenomena.[2] However, while economists have produced exciting theoretical and empirical work on the broad construct, thus far there has been no substantive effort to review what is known about mental models in psychology or other social scientific fields and how they should fit into economic and policy analysis. The contribution of this paper is to fill this gap by synthesizing the extant psychological research on mental models and to lay out mechanisms through which they affect policy-relevant outcomes. This will help economists to draw on a complete and coherent body of theory and research on mental models.

The paper is structured as follows. In section 1.2, I define what mental models are, where the construct comes from, and summarise important characteristics of how they are stored in memory. In section 1.3, I discuss how mental models are used to process information and make decisions under uncertainty. In section 1.4, I review the literature on how people acquire mental models. In section 1.5, I discuss how mental models affect people's capabilities in social environments. In section 1.6, I consider the role of mental models in the policymaking process. I conclude the paper in section 1.7 The remainder of section 1.1 summarises the paper.

Mental models are cognitive structures that organise information in memory. People use them to process information and make decisions when faced with uncertainty. The Cambridge psychologist, Frederic Bartlett, introduced the idea to experimental psychology in the 1930s. Following the cognitive revolution, there was a spurt of theoretical research on the representation of information in the mind in the 1970s and 1980s, followed by a body of

[1] People may hold different beliefs, but the group as a whole converges on a representation of the underlying model.

[2] For example, Mukand and Rodrik (2018) and Rodrik (2014) discuss the role of ideas as a variable political economy. Bordalo et al. (2016) and Bonomi, Gennaioli, and Tabellini (2021) explore the effect of identities and stereotypes on belief distortions. Tabellini, 2010 looks at the role of generalized trust and locus of control on economic performance.

empirical work aimed at showing how they affect judgments and decisions. The storage and retrieval of mental models in memory are influenced by two processes. Top down processes guide the mind to instantiate mental models as a coherent whole. Bottom up processes guide retrieval through associative networks. Mental models operate implicitly or explicitly. Explicit mental models can be brought to conscious attention and manipulated. Implicit mental models are beyond the reach of conscious attention. They include habits and skills and automatic associations.

Mental influence decision-making in two ways. First, they establish how people process information. They determine how people construct meaning from situations. They affect what information people attend to and what they filter out. Second, they establish how people make decisions by mapping the decision space. In some cases, mental models can enable calculated estimates of the likelihood of different events. In highly complex decision-environments, mental models offer rules of thumb (i.e., heuristics) to facilitate decision-making.

Mental models are developed using information from three sources. People are born with models of the world. In other words, they genetically inherit the information required for mental models. Second, people formulate mental models through personal experience – trial and error. Third, people acquire mental models cultural processes. Cultural acquisition of mental models occurs either through direct communication with other people (i.e., social learning) or through mental models symbolically encoded in cultural artefacts. While it is not known which source constitutes the dominant share of a person's stock mental models, it is generally recognized that cultural mental models are the most relevant for explaining social and economic outcomes.

By enabling people to use resources and navigate complex environments, mental models constitute a fundamental capability. Adaptive mental models – i.e., mental models that effectively help people solve their goals – facilitate access better jobs and educational opportunities. This has important distributional implications. It thus matters how mental models are formed. An exciting body of research has been produced over the past decade demonstrating that mental models (a) can be a constraint to human functioning and (b) can be changed through active intervention. We are at the early stages of understanding the practical and normative challenges associated with intervening on mental models at scale.

1.2 An introduction to mental models

1.2.1 Defining Mental Models

Mental models are cognitive structures that organise information in memory.[3] People use mental models to process information and make decisions under uncertainty. Mental models are instantiated in the mind as incomplete representations such as sketches, diagrams, or blueprints (Johnson-Laird, 2010). For example, individuals might hold a mental model of a doctor, informed by data collected over time and experience, and make choices about how to interact with their physician accordingly. Psychologists use the term "mental model" to describe the operation of mental structures in higher-order cognitive processes such as reasoning (Johnson-Laird, 1980; Gentner and Stevens, 1983). Psychologists use the term "schema" to describe the use of mental structures in lower-order cognition, such as perception, memory, and comprehension (Rumelhart, 1980; Brewer and Nakamura, 1984). Economists have generally avoided the term schema, instead using the terms "model" (Schwartzstein and Sunderam, 2021; Gabaix, 2014) or "mental model" (Hoff and Stiglitz, 2016; Denzau and North, 1994). For the purposes of this paper, I use term "mental model" and "schema" synonymously and will generally use the term mental model.

Three important types of mental models are categories, causal narratives, and worldviews. Table 1.1 defines each of these and provides examples. Already, rich literatures in economics deal with each kind of mental model. Economists have studied social categories at least since Gary Becker's classic work on the economics of discrimination (Becker, 1957). There is an active empirical (Bertrand and Duflo, 2017; Hoff and Pandey, 2014; Hoff and Pandey, 2006; Hoff and Walsh, 2018) and theoretical (Bordalo et al., 2016; Akerlof and Kranton, 2000) research agenda on identity and stereotypes. An emerging literature focuses on causal narratives and models in economics (Michalopoulos and Xue, 2021; Shiller, 2017; Eliaz and Spiegler, 2020). Rubinstein (2012) has pointed to the close relationship between narratives and models in economic theory. Finally, literature has focused on the role of generalised trust, moral value systems, and locus of control on political and economic outcomes (Tabellini, 2010; Enke, 2020; Caliendo, Cobb-Clark, and Uhlendorff, 2015).

Mental models are related to beliefs, but are distinct. In economics, beliefs are subjective expectations. They represent epistemic commitments in reference to specific empirical states of the world (see Manski, 2004, section 5). They are generally falsifiable. For example, "the percent chance that you will cast a vote for President" or the "percent chance that you will

[3] In this paper, I use the terms cognitive, mental, psychological, "in memory," and "in the mind" synonymously. They all describe informational activity in the brain.

TABLE 1.1: Types of mental models

Type	Definition	Examples
Categories	Stereotyped classifications of animate and inanimate objects and scenes.	Examples include "an immigrant," "money," and "the Brexit referendum."
Causal narratives	Models representing causal relationships in a system. Stereotyped scripts of scenarios.	Examples include scientific and folk theories, and scripts such as "applying for a job."
Worldviews	Default representation of the universe and of peoples' place within it.	Examples include religions, ideologies, moral value systems, as well convictions about locus of control, the malleability of intelligence, and generalised trust.

lose your job during the next 12 months."[4] Mental models, on the other hand, are generalized, abstract and rarely possible to falsify unconditionally. To illustrate the distinction and relationship between beliefs and mental models, consider the following example. A farmer might believe "her crop yield will increase if she uses fertilizer." Her belief may be derived from a mental model that "fertilizers increase crop yields." The farmer's belief is an empirical expectation while her mental model is a causal model or 'functional form' about the relationship between fertilizer use and crop yields.

Mental models are also related to, but distinct from, heuristics and biases. Heuristics are mental shortcuts for making judgments and decisions (Tversky and Kahneman, 1974; Gilovich, Griffin, and Kahneman, 2002). Heuristics describe how people operationalise mental models in decision-making. For example, the representativeness heuristic describes how people make decisions by comparing their situation to the most representative mental model. The availability heuristic describes how people make decisions based upon how easy it is to instantiate the mental model. Mental models that lead to systematic errors in judgment are biased. However, many of the most important applications of mental models pertains to complex decision environments where systematic errors are difficult to pin down. Under such circumstances, psychologists and decision theorists have focused on whether people's mental models are adaptive or ecologically rational. This describes whether mental models provide pragmatic or "good enough" solutions to the decision problem at hand (Fiske, 1992; Simon, 1955; Gigerenzer, 2000).

[4]Even when economists seek to understand how people think about causal relationships such as the returns to schooling, they generally ask people to imagine themselves in hypothetical scenarios and predict their earnings in such conditions. It is not an incidental point that beliefs refer to empirical states of the world. Beliefs were first operationalised in economics by the Cambridge economist, Frank Ramsey. Ramsey (1926) proposed to elicit beliefs by offering people bets on different outcomes. For bets and their corresponding beliefs to be credible, the states of the world must either happen or not happen.

1.2.2 History of Mental Models

Although Plato emphasized the importance of ideas, the proposal that cognitive structures filter and organise information in the mind was first discussed in detail by the German philosopher Immanuel Kant (Mandler, 2011; Gardner, 1987). Kant was interested in finding a middle-way between continental rationalism and British empiricism. Rationalists such as Descartes believed that thoughts were developed based on reason alone and were formed independent of sensory experience. Empiricists such as John Locke and George Berkeley disagreed. They argued that sensory experience was the ultimate source of our beliefs. In *The Critique of Pure Reason*, Kant (1781) proposed that schemas (or mental models) were the mediating cognitive processes that mapped sensory experiences to beliefs.

Frederic Bartlett (1932) introduced the concept of mental models to modern experimental psychology. Bartlett was interested in understanding how memory informed the interpretation and recollection of ideas. To this end he defined schemas (or mental models) as "an active organization of past reactions, or of past experiences, which must always be supposed to be operating in any well-adapted organic response" (pg. 201). In a set of studies on the reconstructive nature of memory, Bartlett had English children read the Native American folk tale, "War of the Ghosts." Bartlett found that the children replaced and modified parts of the story that did not fit with their mental models – for example, misremembering "canoes" as "boats."

Other psychologists have played an important role in developing the mental model construct. Jean Piaget theorised child development as the process through which a child fits new experiences into old mental models and modifies existing mental models to fit their new experiences (Piaget also used the term schema). Jerome Bruner emphasized that perceivers draw on structured past information to produce meaning (Bruner, 1957). Kenneth Craik (1943) proposed that the mind constructs *models* of reality: "organisms hold 'small-scale models' of external reality and its own possible actions within its head" (pg. 61). Charles Sanders Peirce and Ludwig Wittgenstein are regarded as philosophical precursors to the application of mental models in reasoning (Johnson-Laird, 2004).

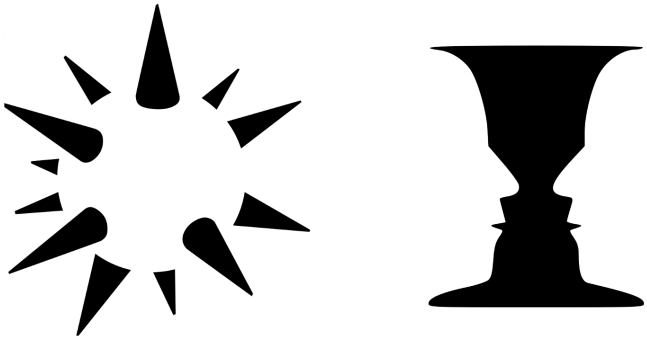
During the 1970s and 1980s, a number of influential theoretical papers were written on the representation of knowledge in memory (Rumelhart and Ortony, 1977; Rumelhart, 1980; Brewer and Nakamura, 1984; Minsky, 1975; Mandler, 1984). Although the terminology that cognitive scientists used varied, their ideas formed the basis of a contemporary understanding of mental models. These ideas were soon incorporated into social psychology, and form the basis of the field of social cognition (Fiske and Linville, 1980; Fiske and Taylor, 2021). The

cognitive approach afforded by schemas (or mental models) laid the ground work for advances in understanding into how we see others (Fiske and Neuberg, 1990), how we ourselves (Markus, 1977) and our ability to change (Dweck and Leggett, 1988), and how culture shapes cognition (Markus and Kitayama, 2010; Kitayama and Park, 2010).

1.2.3 How Mental Models Organise Information

Mental models organise data through a combination of two processes: one "top-down" and the other "bottom-up." The top-down process involves fitting and filtering new information or data based on existing mental models. This feature of cognitive structures was first theorised in detail by the Gestalt psychologists. Categories, one type of mental model, are generally thought to be stored as prototypes or exemplars (Lakoff, 1987). Prototypes are summary representations or "best" instances of a type (Rosch, 1999) – for example, a Ferrari for a car. Exemplars are examples that easily come to mind – such as the car one drives every day (Medin and Schaffer, 1978). Mental models are also theorised to contain information about how to use them, what to expect from them, and what to do if expectations are violated (Rumelhart and Ortony, 1977).

FIGURE 1.1: Top down processes of cognition



The figure on the left contains a series of variants triangle-like shapes. Due to the way they are organised, people automatically perceive them as a whole, coming from three-dimensional sphere. In the figure on the right, people tend to see either the vase or the faces.

One area in which top down pressure on mental models has been shown to matter is in hiring decisions. For example, in one field experiment, Bertrand and Mullainathan (2004) sent CVs to employers in the United States. Some CVs had White-sounding names and some CVs had Black-sounding names. On average, the CVs with White and Black sounding names were statistically identical. Despite this, they found that job candidates with White-sounding names received 50 percent more call-backs for interviews than candidates with Black-sounding names, despite the fact that on average their CVs were identical. This,

and other work, suggest that employers' mental model of Black candidates works to filter out conflictual individuating information (Pager, Western, and Bonokowski, 2006). The top-down process of fitting information to existing mental models demonstrates their molar nature. "[P]eople simplify reality by storing knowledge at a molar, inclusive level rather than squirreling away, one-by-one, all the original individual experiences in their raw forms, which would be pure data-driven processing" (pg. 118) (Fiske and Taylor, 2021).

However, mental models are also created and changed by a "bottom-up" process of constructing associative networks of concepts (Collins and Loftus, 1975; Collins and Quillian, 1969; Ortony, 1978; Smith and Queller, 2001). Concepts, such as "red" and "truck" become associated together from experience to create a mental model (i.e. fire truck). When a part of the network is activated through perception (i.e. seeing a red truck) or thinking, excitation is passed to other interlinked concepts. Situational influences exert a powerful influence on which interlinked concepts (mental models) are activated (Mendoza-Denton et al., 2001).

The links between concepts in a mental model strengthen with co-activation (see Strauss and Quinn, 1997, chapter 4). Stronger links create more excitation and thus trigger related ideas to come to mind more easily. The more links connected to a particular concept, the more likely the concept is to come to mind and the more complex the conceptual categorization. Therefore, adding links can influence mental models, proving relevant for policy interventions. For example, Er-rafiy and Brauer (2013) found that exposing French secondary school students to posters with diverse depictions of young Arabs led the students to update their mental models of Arabs to be more diverse and more favourable.

1.2.4 Mental Models and dual-process cognition

Mental models may be implicit or explicit. Explicit mental models can be consciously accessed and articulated from memory, such as a fire truck. Implicit mental models are non-conscious and therefore cannot be deliberately recalled. Implicit mental models facilitate the development of refined motor skills, actions, and perceptual abilities that are difficult to consciously regulate. For example, people struggle to access their identity consciously (Wilson and Dunn, 2004). Implicit mental models are also generated gradually and incrementally - that is, by practise, experience, or many trials.

One area where considerable research has been conducted is on implicit attitudes toward social groups (Banaji and Greenwald, 2016). Implicit mental models were thought to be largely unchangeable (Bargh and Chartrand, 1999). However, research over the past decade shows targeted interventions can elicit short-term changes in implicit mental models (Lai

et al., 2016). Moreover, Charlesworth and Banaji (2019) has shown, based on 4.4 million tests of implicit attitudes over 14 years in the United States, that implicit attitudes toward sexual orientation, race, and skin-tone attitudes have become more neutral. Meanwhile, implicit attitudes toward age and disability have remained constant and attitudes toward body-weight have become less neutral.

1.3 Mental models in action

In this section I discuss the role that mental models play in comprehension, perception, memory, and decision-making. Table 1.2 provides a definition for each cognitive process and sketches the implications for information processing. Throughout the section, I show how strategies for managing information affect policy-relevant outcomes.

TABLE 1.2: How mental models affect perception, memory, comprehension, and decision-making

Function	Definition	Implications
Perception	The process or result of becoming aware of objects, relationships, and events by means of the senses, which includes such activities as recognizing, observing, and discriminating.	People use mental models to determine what information to attend to, what information to filter out, and what information to add to make sense of situations.
Memory	The ability to retain information or a representation of past experience, based on the mental processes of learning or encoding, retention across some interval of time, and retrieval or reactivation of the memory.	People use mental models to store information in memory. Mental models that are more easily stored and retrieved are more likely to influence behaviour. Cues influence how mental models are retrieved.
Comprehension	The act or capability of understanding something, especially the meaning of a communication	People use mental models to construct meaning out of situations. The mental models they draw determine their decisions and affect their ability to achieve their goals.
Decision-making	The cognitive process of choosing between two or more alternatives, ranging from the relatively clear cut (e.g., ordering a meal at a restaurant) to the complex (e.g., selecting a mate).	Mental models enable people to construct and deliberate between alternative choices when faced with uncertainty.

Source: APA Dictionary of Psychology. These activities enable organisms to organize and interpret the stimuli received into meaningful knowledge and to act in a coordinated manner.

1.3.1 Comprehension

The main role of mental models is to enable people to establish meaning by organising information coherently. To illustrate this, consider the following text:

> The procedure is actually quite simple. First you arrange things into different groups depending on their makeup. Of course, one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities that is the next step, otherwise you are pretty well set. It is important not to overdo any particular endeavor. That is, it is better to do too few things at once than too many. In the short run this may not seem important, but complications from doing too many can easily arise. A mistake can be expensive as well. The manipulation of the appropriate mechanisms should be self-explanatory, and we need not dwell on it here. At first the whole procedure will seem complicated. Soon, however, it will become just another facet of life. It is difficult to foresee any end to the necessity for this task in the immediate future, but then one never can tell. (Reported in Bransford and Johnson (1972)).

Without an organizing mental model, people find it very difficult to establish the meaning of this paragraph. The problem is not that there is incomplete information, but that it does not hang together coherently. Simply cueing a mental model for "washing clothes" serves this function.[5]

The process of meaning construction is generally left out of decision-analysis. It is assumed that people receive information preassigned with meaning. In economic theory, agents have at least a stochastic understanding of the relationship between actions and outcomes. In experimental games, participants generally must demonstrate that they understand the game before they can play.

The simplification helps us understand rational action in simple decision-contexts, but it tells us little about behaviour when people don't understand their decision environment and thus are unable to undertake goal-oriented decisions. To illustrate the importance of meaning construction in a policy-relevant scenario, consider the case of voting in Brazil. Until 1998, about a third of voters were unable to complete their ballots in Brazilian elections. Brazil's elections used paper ballots. Because many people could not read and write, they were unable to understand their options in the ballot box. Their votes were thrown out.

In 1998, Brazil began transitioning to electronic voting. This meant that people did not need to read and write to be able to vote. The voter just needed to enter the candidate's ID number and the candidate's photograph came up. Voters could push the green button to confirm their selection or the orange button to cancel their choice. The change in technology de facto enfranchised 11 percent of the electorate. With more votes from lower-income people

[5] When study participants had the previous statement available to them, they were twice as likely to be able to comprehend and remember the material (Bransford and Johnson, 1972).

counted, more candidates from pro-poor parties were elected to state legislatures. In the 8 years that followed, states increased the budget for health expenditures by more than a third. The proportion of lower-educated pregnant women receiving regular prenatal care doubled and babies' health outcomes improved (Fujiwara, 2015).

1.3.2 Perception

Mental models also serve as guideposts for selecting what to attend to and what to filter out (Johnston and Dark, 1986). The process of selecting information to be stored in working memory is highly competitive (Desimone and Duncan, 1995) and the goals embedded in mental models provide a retrieval plan for allocating attention (Neisser and Becklen, 1975). This is demonstrated in famous experiment known as the "invisible gorilla" (Simons and Chabris, 1999).[6] Participants are shown 6 players on a hardwood floor with a red curtain behind them. Three players are wearing white t-shirts and three are wearing black t-shirts. Participants are instructed to "count how many times the players wearing white pass the ball." The players to proceed to walk around the screen for about 20 seconds and pass the ball to others wearing the same colour t-shirt. Most people are able to count the number of passes made by the team in white. Strikingly, very few notice the fact that a gorilla walks through the players and beats their chest during the game.

It is sometimes argued that when people have skin in the game, they notice all relevant features. Mistakes arising from poor selective attention might "wash out" in equilibrium (Fehr and Tyran, 2005). However a large body of experimental evidence shows that people persistently employ mental models that filter out relevant information (Hanna, Mullainathan, and Schwartzstein, 2014; Lacetera, Pope, and Sydnor, 2012; Beaman, Magruder, and Robinson, 2014). Theoretical work suggests that market forces may exacerbate – not alleviate – the problem (Akerlof and Shiller, 2015).

Mental models provide default values to support sense making. Default values 'fill in the gaps' to bring information together in a coherent whole. This is especially evident in areas of sensation and perception, where the brain goes beyond the information given to construct coherent representations of the real world (see Pinker, 1997, chapter 1). One thumbprint of this is that default values last in memory. This is demonstrated in studies of inaccurate recall. In experiments based on the D-R-M "dream" paradigm, participants are asked to recall a list of associated words unified by a term that is not on the list. Experimenters examine how commonly participants list the non-presented term (Roediger and McDermott, 1995;

[6] Video available at: $https://www.youtube.com/watch?v=IGQmdoK_ZfY$

Stadler, Roediger, and McDermott, 1999). In the original experiment, the words table, sit, legs, seat, soft, desk, arm, sofa, wood, cushion, rest and stool are linked to chair (Deese, 1959). Although the rate naturally varies based on the list, false recall as high as 65 percent has been documented (Stadler, Roediger, and McDermott, 1999).

The default values that people use to 'fill in the gaps' can lead people to derive meaning from communication that is at odds with what is intended. A study of day care centres in Israel that began fining parents who were late to collect their children provides a case in point. The initiative was intended to discourage late pickups by imposing a monetary cost on parents for arriving late. But it had the opposite effect: the intervention instead increased the number of late parents. One explanation for this is that parents did not interpret the fine as an independent cost. Rather, they made sense of the charge in a holistic manner – when the day cares introduced a monetary cost, it instantiated a mental model of an economic transaction, transforming the situation from one governed by a moral conduct (be punctual) to one governed by market conduct (parents could "purchase" additional time) (Gneezy and Rustichini, 2000).

1.3.3 Memory

Mental models are central to the action of encoding and retrieving information in memory (Gilboa and Marlatte, 2017; Bartlett, 1932), but the process is not straight-forward. On the one hand, research often finds that people are better at remembering information that conflicts with their instantiated mental model. In a set of experiments, study participants are shown a room – for example, a graduate student's office – containing typical items (e.g., a stapler) and atypical items (e.g., a horseshoe). They are more likely to remember items that conflict with their mental model (Pezdek et al., 1989; Lampinen, Copeland, and Neuschatz, 2001). An explanation for this is that people encode information consistent with mental models automatically, while inconsistent information must be explicitly stored (Graesser, Gordon, and Sawyer, 1979). On the other hand, mental models provide essential leverage for the retrieval of information from long term memory. One signature of this is that people hold false memories of information consistent with mental models. A large body of theoretical and experimental literature show that accessible mental models that come to mind influence judgments and decisions (Tversky and Kahneman, 1973; Gennaioli and Shleifer, 2010)

1.3.4 Decision-making under uncertainty

In models of decision-making under uncertainty, people aim to maximise their expected utility based on their beliefs about the likelihood of possible outcomes and the utility that they would get from each outcome. The role of mental models is generally not made explicit in standard decision-analysis. However, even in simple games mental models play a crucial role in decision-making by enabling people to understand the decision-context.

Consider first objective probability.[7] Objective probability describes cases where probability distributions are driven by random physical processes, such as the role of a die or lottery. It is implicitly assumed that people simply know that each number has an equal probability of landing face-up on fair die. But this is not the case. People do not form such expectations from nowhere. They derive them from a mental model of how that physical system works.

The importance of mental models becomes more obvious as the complexity of the physical system increases. In such cases, people may draw on sub-optimal models even when they have extensive experience with the decision-problem at hand. An example of this can be seen in a study of seaweed farming practices in Indonesia. In field experiment, Hanna, Mullainathan, and Schwartzstein (2014) showed that experienced farmers in Indonesia made consistently sub-optimal decisions about farming inputs because they did not notice important factors that affected their yields. Even when farmers had access to the experimental data, the farmers didn't update their mental models. Instead, the farmers only changed their behaviour when they were presented with summaries that highlighted previously unattended-to relationships in the data. This updated their mental models.

In most cases of uncertainty, it is not practical or feasible to draw on a systematically derived model of the decision-context. One example is the expected earnings from different career choices. As Manski (2004) points out, economics research has often left these questions unresolved. It makes little sense to think that individuals know these probability distributions. In such circumstances, people may derive their expectations from a mental model based on

[7] Decision-theory based on this kind of uncertainty traces its roots to the 18th century when the Swiss mathematician Daniel Bernoulli formulated a solution to the St. Petersberg paradox. Bernoulli's ideas were operationalized in decision-analysis and game theory by von Neumann and Morgenstern (1944, 1947). The paradox concerns an open-ended game of chance at a betting house. In the first round, the stake is 2 pounds and a coin is tossed. If the coin falls on heads, the amount is doubled. If the coin falls on tails, the game ends and the gambler receives the stake. Thus, if this first toss lands on tails, they win 2 pounds. If it lands on heads, the next round is for 4 pounds. If the second rounds on heads, they go on to 8 dollars in the next round. And so on. How much should a risk-neutral gambler be prepared to bet on this? Assuming the game can go on forever, the expected value in this kind of game is infinite. Bernoulli solved the paradox by delineating between the expected value of the bet (the outcome) and the expected utility of the bet (how much benefit the gambler would get from each potential prize). Bernoulli proposed that people derived diminishing marginal utility from money, which would bound the amount of money they would bet.

subjective probabilities or expectations. Depending on the degree of complexity, subjective probabilities can be quantified as "degrees of belief" in the likelihood of events occurring.[8]

In cases of radical uncertainty, the decision environment may be so complex that no quantifiable probability distribution can be derived. This is called "Knightian" uncertainty (King and Kay, 2020; Knight, 1921). An emerging literature on decision-making under radical uncertainty suggests that people make decisions using narrative scripts (Tuckett and Nikolic, 2017; Johnson and Tuckett, 2018).

A rich theoretical literature on interdependent decision-making supports the idea that narratives and scripts guide behaviour in strategic contexts (See chapter 2 Bicchieri, 2005; see also Ostrom, 2009). An important fingerprint of the relevance of scripts is that they can be activated based on cultural cues – shifting behaviours even when incentive structures remain constant. In a series of landmark studies with economic games in 15 small-scale societies, Henrich et al. (2005) found that societies varied significantly in terms of how they played canonical economic games, such as the ultimatum game, the dictator game, and the public goods game. In their discussion of the results, the authors considered whether these difference reflected deep-seeded preferences – i.e. "generalized behavioural dispositions (equity, altruism, etc.)" or culture cues that elicit mental models of the decision problem. They describe one of their studies in detail, suggesting that cultural cues mattered:

> "In these situations, subjects in some places were first identifying the kind of situation they were in, seeking analogues in their daily life, and then acting appropriately. ... [One of the groups] "the Orma," made a connection between the public goods game and their local practice, the harambee. The Orma believe that wealthier households should make larger contributions to the harambee than poorer households. Multivariate regressions involving wealth, age, education, and income indicate that wealth is the only significant predictor of public goods game contributions among Orma individuals. The more wealth a person has the more they contribute to the common pool, just as in the real harambee."

The Orma did not perceive a similar connection between the harambee and the ultimatum game and wealth was not a significant predictor of offers in the game. While this is just an anecdote, systematic evidence highlights the importance of frames on game behaviour (Dufwenberg, Gächter, and Hennig-Schmidt, 2011). To show how these games work, consider the following example. Psychologists invited participants to play the prisoner's dilemma.

[8] Subjective expectations also traces its roots to early modern thought and the first writings on probability but came to prominence through the work of Frank Ramsey (1926) and Bruno Finetti (1937) in the late 1920s and 1930s. Ramsey and de Finetti conceived of people betting on the truth of propositions and used this to define the idea of choice-based subjective probability. Just as people "reveal" their preferences in Samuelson's model of welfare analysis, people reveal their level of confidence in the truth of a proposition through their betting behaviour. Savage (1954) built on this early work in The Foundations of Statistics, introducing an analytical framework consisting of sets of states of the world, consequences, and acts that map the set of states to the set of consequences. Acts correspond to strategies of action, consequences describe what happens to the person, and states are possible resolutions of uncertainty. Savage envisioned this as a model of how people actually make decision but how people should make decisions.

One group of participants were told they were playing "the Wall Street Game." The other group were told they were playing "the Community Game." Although the structure for all participants were the same, the framing of the event dramatically influenced how people behaved. Roughly seventy percent of participants who were told that they were playing the Wall Street Game chose to betray the other prisoner. The proportion was the inverse for those told they were playing "the Community Game." Among this group, roughly seventy percent of participants chose to cooperate (Liberman, Samuels, and Ross, 2004).

1.4 Where mental models come from

The information that constitutes mental models comes from three sources: genes, culture, and personal experience. Information transmitted through genes is generally studied by biologists, as well as evolutionary and some developmental psychologists. Information acquired through culture is generally studied by anthropologists, sociologists, and psychologists. Information acquired through personal experience is generally studied by economists and decision-researchers.

1.4.1 Genetic inheritance of information

People genetically inherit information embedded in DNA through their biological parents. This provides the physical foundation for human cognitive capacities. The transmission of these capabilities emerged and adapted over long run time scales via natural selection (Barkow, Cosmides, and Tooby, 1992; Barrett, 2014). Evolutionary psychologists propose that this includes specialized cognitive capacities – for example, for the acquisition of language (Chomsky, 1980; Pinker, 1993) and various other socio-cognitive capabilities such as theory of mind (Dunbar, 1998).

There is good reason to believe that biological processes not only produce the physical machinery to process information, but also some of the information itself. The best evidence for this comes from empirical work by development psychologists, which suggests that humans are born with a repertoire of 'primitive' or 'core' knowledge structures (Gopnik and Meltzoff, 1997; Spelke and Kinzler, 2007). According to this research, infants are born preprogrammed with a basic conceptual repertoire (e.g., object, agency, numbers, cause) that enables them to bootstrap other concepts (Carey, 2004; Carey, 2009).

Core knowledge structures are implicit. They are building blocks for higher-level cognitive processes – reasoning, problem solving, planning, and language use, which we have virtually

no direct introspective access to (Nisbett and Wilson, 1977). Other evidence shows that humans also genetically inherit more concrete informational structures. This is demonstrated, for example, by our attentional bias toward faces, which develops before birth. In a study in the UK, researchers used advanced ultrasound technology to project patterned stimuli through pregnant women's abdomens toward foetuses. In one condition, they project images configured like a face (∵). In another condition, they project the inverse configuration (∴). Only in the condition with the face like configurations did the foetuses turn their heads (Reid et al., 2017).

1.4.2 Cultural acquisition of information

People also acquire information through social and cultural learning (Bandura, 1977; Heyes, 1994; Tomasello, 1999).[9] Scholars of cultural evolution propose that information transmitted and acquired during social learning forms the basis of culture (Mesoudi, 2011, pg. 2-3). This is broadly consistent with models of culture in cognitive sociology (DiMaggio, 1997; Swidler, 1986). It contrasts with the conception of culture as a homogeneous macro-level trait held by groups (i.e., the idea, popular among classical sociologists such as Weber and Parsons, that there is a cohesive, stable, and distinct Irish, Jewish, or Japanese cultures, for example).

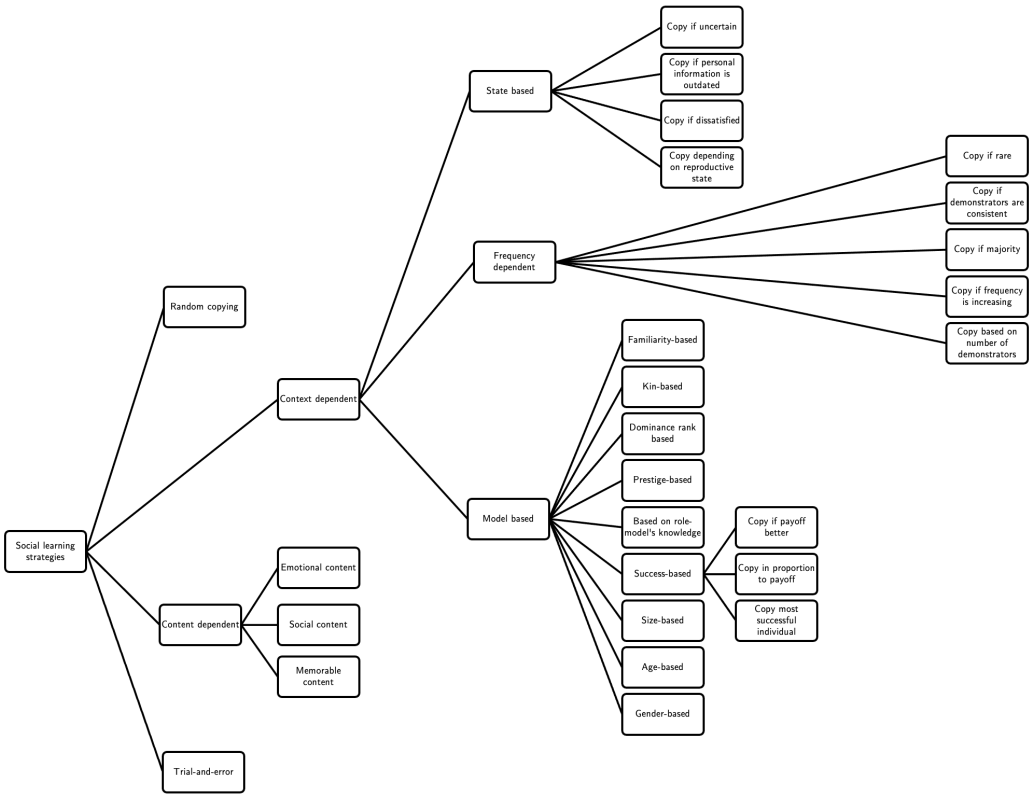
Research on social learning suggests that much of human beings' distinctive cognitive abilities over other species appears to be derived not from genetic inheritance of superior mental faculties (i.e., executive functions), but through vastly better capacity for social learning (Henrich, 2017). Studies of young children, for example, show that human children are no better than Chimpanzees or Orangutan at special, quantitative, or causal reasoning, but significantly better at tasks involving social cognition (mimicry, communication, theory of mind) (Herrmann et al., 2007). In studies with college students, chimpanzees score higher on some measures of executive function than humans (Inoue and Matsuzawa, 2007).

One implication of humans' knack for social learning is that we tend over-imitate others, replicating even causally irrelevant behaviours (McGuigan, Makinson, and Whiten, 2011; Whiten et al., 2009). Our strong tendency to rely on others also means that that mass behaviour can rapidly shift in response to relatively small shocks that produce information cascades (Bikhchandani, Hirshleifer, and Welch, 1992).

[9] A small definitional note is required here. The fields biology, psychology, and anthropology adopt a general understanding of social learning as "learning assisted by other agents" (Heyes, 2018). This definition encompasses all economic behaviour, even in Robinson Crusoe economies with one producer and one consumer. It is too broad to distinguish forms of learning in social science. Scholars in these fields generally define information acquired through sophisticated social cognition such imitation, education, and language as cultural learning (Henrich, 2017; Heyes, 2018; Tomasello, Kruger, and Ratner, 1993).

Humans are selective in their learning strategies. People generally do not replicate behaviour at random. Rather, they select who to replicate based on factors such as their psychological state, the content of the information, and the qualities of the role model (Rendell et al., 2011). Figure 1.2 schematizes some of the social learning strategies that have received theoretical and empirical support to date. More work is needed to understand the policy relevance of these processes by testing them in lab and field contexts.

FIGURE 1.2: Social learning Strategies



Source: Adapted from Rendell (2011, Figure 1)

1.4.3 Information through personal experience

People acquire the information through individual learning (Erev and Barron, 2005; Hertwig et al., 2004; Roth and Erev, 1995). This learning comes through experience gained from trial and error – learning by doing. A major paradigm for studying the means through which individuals gain information through experience is "reinforcement learning" (Erev and Roth, 1998). Reinforcement learning models how people map situations to actions in order maximise pay-offs. The agent is not instructed which activities to perform, but must determine which behaviours provide the greatest reward through trial and error. In some instances, actions may have an effect not just on the immediate reward but also on the

future circumstance and, thus, on subsequent payoffs. Trial-and-error search and delayed reward are the two primary features of reinforcement learning.

1.5 Mental models and capability development

One way of engaging with the policy-relevance of mental models is through a capability approach (Sen, 1999; Nussbaum, 2011). The capability approach maintains that the "success term" for public policy (e.g., justice, social welfare) should be based on the *de facto* freedoms that people have to achieve functionings have reason to pursue. Functionings are states of being and attainment such as getting married, being educated, and being healthy. Capabilities capture a person's ability to convert resources (e.g., money, entitlement to social services, information) into functionings. Human agency is characterized and sustained by mental capacities (Bandura, 2001). By enabling people to use resources and navigate complex environments, mental models constitute a fundamental capability; "thinking is for doing" (Fiske, 1992).

One implication of this is that mental models guide a vast range social processes and outcomes. Many of the original and most developed ideas about how mental models effect social outcomes come from largely theoretical work in cognitive sociology (Zerubavel, 2009; Cerulo, 2002) and anthropology (d'Andrade, 1995). A notable exception to this is work by Margaret Frye, who integrates survey data with qualitative work to understand how narratives and aspirations affect educational outcomes in Malawi (Frye, 2017; Frye, 2012). However, the availability of causal and quantitative empirical evidence on mental models is mixed and still emerging (see Hoff and Walsh (2018) for a review).

1.5.1 Cultural capital

Mental models enable people to navigate resource-rich social environments such as educational institutions and workplaces. The French sociologist, Pierre Bourdieu, is credited with first proposing and analysing the implications of this. He referred to the advantages gained by mental models as "cultural capital" (Bourdieu, 1986).[10] Similarly, Ann Swidler (1986) proposed that culture provides people with "strategies of action" that enable them to solve the problems in their day to day life. These ideas lay the foundation for applying mental

[10] For Bourdieu, cultural capital described more than mental models. It included the whole symbolic repertoire that a person acquires from their social position and can put to use to get ahead. Cultural capital refers to personal qualities such as talents, tastes, posture, clothes, mannerisms, material items, credentials, and so on. Bourdieu emphasized physical embodiment of cultural capital, which he called *habitus*. These are deeply ingrained habits, skills, preferences and dispositions that, equivalent to implicit mental models. Bourdieu emphasized that habitus was implicit – using metaphors such as having a "feel for the game." Bourdieu also emphasized that cultural capital could be objectified – for example, by wearing particular clothes or driving a particular kind of car.

models in policy analysis. At the heart of their conceptual approach is the implication that mental models are not just a way of explaining basic cognitive processes in the lab. Nor is culture a monolithic homogenizing entity. Rather, mental models and culture are an active and key resource – similar to money – that people can put to work in achieving their goals. Swidler called this "culture in action."

One implication of cultural cognition is that mental models often reflect and instantiate social inequalities (Hoff and Walsh, 2021). Appadurai (2004) writes of the cultural advantages enjoyed by more privileged groups:

> The better off ... have a more complex experience of the relation between a wide range of ends and means, because they have a bigger stock of available experiences of the relationship of aspirations and outcomes, because they are in a better position to explore and harvest diverse experiences of exploration and trial, because of their many opportunities to link material goods and immediate opportunities to more general and generic possibilities and options. (pg. 68)

These ideas have been formalised by (Dalton, Ghosal, and Mani, 2016; Garance and Debraj, 2017).

A key policy goal that has emerged over the past decade is to understand how "adaptive" mental models can be provided to vulnerable groups. Policymakers are in the early stages of understanding the strengths and limitations of these approaches. As will be discussed in the next section, an key concern with interventions focused on persuading people to adopt particular mental models is whether those mental models are adaptive. We are only at the early stages of developing frameworks for assessing this. Several field experiments have sought to equip people with the mental tools for upward mobility.

One approach is to shift people's causal models through brief informational interventions. In one experiment, Paunesku et al. (2015) designed a short online intervention composed of two 45 minute videos; one communicating the growth-mindset while the other communicating a sense-of-purpose intervention. They then delivered the modules to 1,594 students in 13 geographically diverse high schools in the United States. About a third of the sample were at risk of dropping out of high school. Both interventions raised these students' grade point averages in core academic courses. It also increased the rate at of satisfactory completion in core courses among these students by 6.4 percentage points.

Another approach is to role model the key behaviours, equipping people with new scripts. A field experiment conducted by Tanguy et al. (2014) examined the impact of showing inspiring videos to poor villagers' aimed at increasing their aspirations. The team randomly selected 64 villages for the intervention. The location was remote, requiring 4x4 vehicles, and only 10

percent reported watching TV once a week or more. The intervention increased aspirations by 12 percent. Six months later, these effects (though smaller) were still significantly higher than among the control group (3 percent of a standard deviation). The intervention also caused changes to future-oriented behaviour. School enrollment of children was 17 per cent higher than the control group at the end of the intervention. Savings were also higher.

A third potential way to help people to develop adaptive mental models is to help them recognise situations that require deliberative thinking. In a series of large randomized controlled trials (RCT), Heller et al. (2017) tested a number of interventions aimed at reducing automaticity among low-income youths from disadvantaged areas of Chicago in the United States. The activities focused on giving people new mental models for dealing with conflict scenarios. In one intervention, arrest rates for students fell by 44 percent and 36 percent for violent and other (non-violent, non-drug, non-property) crimes respectively. The participants also improved on measures of school engagement by 0.14 standard deviations in the first year and 0.19 in the year after, which the authors forecast could translate into increases in graduate rates of about 7-22 percent. In another intervention, arrests fell by 31. In a third evaluation of a program at a Juvenile Temporary Detention Center, youths who had been treated were 16 percentage points less likely to be re-admitted to the center.

Finally, a fourth option for influencing individuals' mental models is mentoring and coaching. Bettinger and Baker (2014) evaluated a mentoring and coaching service for non-traditional U.S. college students. Coaches worked with students to help them clarify their aspirations, link their day-to-day activities to long-term plans, and build time management and self-advocacy skills. Coached students were 14 percent more likely to persist in school after 24 months and four percentage points more likely to complete their degree within four years of receiving the treatment.

1.5.2 Status and identity

Another way in which mental models affect social outcomes is through their effect on status and identity. The most obvious way in which mental models of identity affect social outcomes is through stereotyping. People use stereotypes to assess others' intent and capabilities. As discussed later in the dissertation, this informs how people make decisions about the other group. A growing empirical literature is generating evidence on ways to tackle prejudice. The evidence from this literature suggests that increasing interactions, perspective taking, choice architecture, and awareness raising can reduce prejudice (for a review see Hoff and Walsh, 2018 pgs 5-12). Another pathway recently identified is through positive celebrity

role models. Just as role models can encourage people to formulate new expectations and goals, role models also have the capacity to improve attitudes toward stigmatized groups. A recent study examined the effect of the Egyptian footballer, Mo Salah, joining Liverpool football club on attitudes toward Muslims. They find that hate crimes in the Liverpool area fell by about 16 percent and the rate of anti-Muslim tweets fell by about a half (Alrababa'h et al., 2021).

Another important implication of status is that it shapes how people see themselves. When people are socially excluded, the stigma can "get into people'sheads" and degrade their self-concept (Goffman, 1963).Paul DiMaggio (2012) writes:

> Social scientists ordinarily define social class and related concepts in terms of such tangible measures as occupation, income, years of schooling, and home ownership. Humans experience inequality, however, in social interactions, which produce the orderings that our variables index. We learn where we stand by observing how others treat us: deferentially or dismissively; with interest in or indifference to our ideas or desires; with apprehension, reserve, or openness to intimate exchange (pg. 15).

Evidence suggests that mental model based interventions to support stigmatized groups can have significant influence on their capabilities. In a recent study in India, Ghosal et al. (2020) show that a series of 8 weekly sessions focused on restoring the integrity of sex workers' mental models of themselves led the sex workers to increase their savings and engage in more preventative health behaviour. Although the intervention was purely psychological, the effects lasted for over a year. In another study, Blattman, Jamison, and Sheridan (2017) looked at the effect of providing cognitive behavioural therapy alongside cash supports to criminally engaged men in Liberia. The cash by itself and the therapy by itself reduced crime in the short-run, but the effects dissipated after. However, when combined, the cash and therapy had large effects on crime reduction. The authors suggest that the effects were mediated in part by restoring self image.

1.5.3 An agenda

In summary, field experiments show that interventions can improve people's capabilities by influencing their mental models. The evidence presented here represents the beginning of an exciting agenda. Much more must be learned before mental models can be systematically incorporated into policy design. In particular, more research is needed on implementation to understand what strategies have greatest impact, are scalable, and are most cost effective. Notably, many design features of interventions are selected based on theoretical synthesis and intuition. While this is pragmatic, testing key assumptions during the implementation

process may improve design choices. Diverse methods – from Bayesian (Caria et al., 2021) learning to human centred design (Yeager et al., 2016) – offer promise in this area.

1.6 Policymakers' mental models

A key determinant of the effectiveness of government programmes is the quality of the mental models that policymakers can bring to bear in policy design. Sub-fields within sociology, psychology, anthropology, and especially economics are concerned with how governments form and employ models of policymaking. Critical engagement with policymakers' mental models can be organised into two broad categories. One body of work is concerned with whether policymakers have a good enough model of the problem they are trying to solve. This argument assumes that policymakers have good intent in that their goal is to improve social welfare. Policies fail because policymakers don't understand the problem they are trying to solve. A second body of work is concerned with whether policymakers' mental models are embedded with either ill intent or lack of concern for ordinary people. Policies fail, according to this argument, because policymakers don't value or respect socially marginalised groups. Standard political economy models hold similar assumptions to the second approach but propose "interests" rather than "mental models" as the mediating process (e.g., Acemoglu and Robinson, 2012).

1.6.1 Bottom up epistemic constraints

A rich literature focuses on the epistemic constraints to designing good policy. One concern is that policymakers lack effective "bottom-up" processes for refining their mental models. In other words, policymakers don't understand the challenges ordinary people face because they lack information. This view has been put forward by ideologically diverse thinkers such as Frederick Hayek and James Scott. Hayek emphasized the relative quality of individuals' personal mental models. According to Hayek (1945), policymakers simply lack the information to formulate high quality mental models of people's individual circumstances. On the basis of this, Hayek famously argued for limited government.

In *Seeing Like a State*, Scott (1998) makes a related but distinct argument. Scott draws on case studies of efforts at Soviet collectivism and villagization in Tanzania to propose that centrally planned designs fail to understand the importance of local mental models, what he calls mētis. Policy failures occur not (necessarily) because policymakers have ill-intent, but because government policy is formulated in political and economic centres such

as Washington D.C., Beijing, New Delhi, and London, where worldviews, categories, and causal models differ from the communities the policies are designed for.

Scott provides an evocative example to illustrate his point. When the first European settlers came to North America, they struggled to successfully cultivate the land. In particular they had a challenge growing corn. To overcome the challenge, they employed the latest scientific methods and published specific formulas in journals such as the *Farmer's Almanac* advising the precise date to plant. But the results failed to generalize across New England. When they sought help from the Native Americans living there, they were simply advised "to plant corn when the oak leaves were the size of a squirrel's ear." The advice produced successful yields across the region. The reason for this was that the causal narrative helped avoid frost. Scott writes:

> "Embedded in this advice, however folkloric its ring today, is a finely observed knowledge of the succession of natural events in the New England spring. For Native Americans it was this orderly succession of, say, the skunk cabbage appearing, the willows beginning to leaf, the red-wing blackbird returning, and the first hatch of the mayfly that provided a readily observable calendar of spring" (pg. 311).

1.6.2 Top-down epistemic constraints

A second challenge is concerned with "top-down" processes in mental model formation. The decision environment involved in policy design is socially complex. It entails high levels of uncertainty. To make predictions under such conditions, policymakers are trained to approach problems using stylized models. Both the advantage and disadvantage of these models is their simplification. A major critique of economic models in recent decades has been that the models available to policymakers have historically been under- or misspecified (Rodrik, 2010) failing to take importnat contextual features into consideration in policy design. Ostrom and Cox (2010) describes this as the 'panacea problem.' Under- or misspecified models of the problem may manifest themselves in everyday biases and mistakes in policymakers' decision-making. A recent set of experiments by (Banuri, Dercon, and Gauri, 2019) suggest that this is the case. In a series of experiments conducted with staff from the World Bank and DFID, they found that policy professionals are vulnerable to decision-making traps, such as confirmation bias based on ideological predisposition.

1.6.3 The intent of policymakers' mental models

A related concern is that policymakers' mental models may be embedded with power structures that influence the design of policy. The french philosopher, Michel Foucault, was

an early and influential proponent of this view. In a series of lectures at the Collège de France, Foucault proposed a theory of 'governmentality' arguing that policymaking is informed by reciprocal relationship between mental models ('mentalité') and power (Foucault, 2014). These ideas, based on his work on madness and sexuality, argued that people naturalise mental models (i.e., absolute, objective, universal and true) that actually developed through historically and culturally contingent processes (Foucault, 1978; Foucault, 2003). In particular, he emphasised that societies' power structures are central to the development of mental models (Foucault, 2019).

1.6.4 Improving policy models

Concerns about the quality of policymakers' mental models have motivated major reforms in global policymaking, especially at institutions such as the World Bank. Hakeyian concerns can be seen in much of the World Bank's lending strategies through the late 20th century, where financial support was made conditional on cutting government programming. They are also an important basis for contemporary proposals by economists advocating that supports to vulnerable groups be delivered in cash rather than in kind (Lowrey, 2018).

Scott's concerns were highly salient during the late 1990s and early 2000's when the World Bank began investing in community driven development programmes. These ideas were largely focused on localising decision-making along the lines advocated by Scott. Scott's concerns can also be seen in contemporary research advocating for the allocation of discretion in decision-making to local front-line workers (Honig, 2020; Honig, 2018).

Concern with "one-size-fits-all" approaches have motivated an influential agenda on incorporating diagnostics into policy design (Rodrik, 2010) and on learning through experimentation (Andrews, Pritchett, and Woolcock, 2013). The World Bank has adopted these concerns by introducing systematic country diagnostics that now inform country strategies.

Finally, Foucaultian concerns that policy makers' mental models perpetuate inequalities by reflecting and instantiating power structures have motivated arguments for quotas (Walsh, 2013). Evidence from India on gender quotas suggests that quotas do affect policy outcomes, and probably through through cognitive processes as Foucault suggested. For example, when women are elected to local government, police are more likely take reports of violence against women seriously (Iyer et al., 2012).

1.7 Conclusion

The paper has aimed to present the applications of the mental models construct to economic and policy analysis. It made the following contributions. It proposed a definition of mental models based on a synthesis of research in cognitive psychology. It identified three important mental models – categories, causal narratives, and worldviews – and described key ways in which mental models work. It then went on to summarise the information source for mental models, noting that information comes from three sources: biology, individual learning, and social learning. Finally, it discussed applications to policy – considering the the scope for policy to influence people's mental models and the scope for improving policymakers' mental models.

Like other organizing frames in economics and psychology (e.g., dual-process cognition, rational action), mental models are an instrumental construct (Jr, 2007). Their utility in behavioural science is not evaluated against their ability to describe the actual physiology of the brain, though this is an active research agenda in neuroscience.[11] Rather, they aim to describe a tractable model for generating theoretical insights and making reliable empirical predictions. In their application to behavioural economics, the goal of introducing the construct is to better understand how human cognition and behaviour influence social outcomes and structures, and to develop frameworks for policy to foster human capability and flourishing (Camerer and Loewenstein, 2004).

[11] See, for example, (Gilboa and Marlatte, 2017; Kesteren et al., 2012). For an accessible history of neuroscientists' work in this area see Eric Kandel (2007), chapter 22.

Chapter 2

Social groups

Economists have produced a rich literature documenting the prevalence of discrimination in consequential contexts. While this research shows that people use mental model (i.e., social categories) to inform how they engage with people – for example, when making hiring decisions – it sheds little light on how categorical content informs behaviour. To address this, we ran six experiments to operationalise the two stereotype dimensions—warmth and competence—using consequential behaviours toward varied societal groups. Across over 18,000 decisions, American and Indian participants (N=1,281) rated their respective society's distinctive stereotype content (perceived warmth and competence) for distinct target groups. Reported societal stereotypes robustly appear in behaviour toward individual group members even when actors have skin in the game. Online participants played economic games with pre-recorded actual members of societal groups representing the four quadrants of warmth-by-competence space. For Americans, stereotypically: A middle-class person and a parent both seem warm and competent; an alcoholic and a homeless person both seem neither warm nor competent; a learning-disabled person and a Parkinson's patient both seem warm but incompetent; an Ivy League graduate and an Asian person both seem competent but not warm. Such stereotyped group members served as participants' game partners. Three studies used the trust game – a canonical experiment in economics – as a behavioural operationalization of warmth. As predicted, participants shared more with players from groups they believed society perceives to be warm. Three more studies used a novel puzzle game as a behavioural operationalization of competence. Participants bet that players from groups society perceives as competent would solve more puzzles. With nontrivial incentives on the line, participants used the content of their mental models of social groups to guide their own distinctive behaviour toward individual group members.[1]

2.1 Introduction

Since Hobbes, modern social science has preoccupied itself with how people navigate encounters with strangers (Hobbes, 1651). With often little information at their disposal, people set out to manage vulnerable situations with unfamiliar others by pragmatically anticipating what those others will do (Abele et al., 2020). How do people solve these social challenges? Social scientists have converged on the central role of social structure and institutions in making others' actions predictable and intelligible. Like Hobbes, scholars of macro level behaviour (economists, sociologists, and political scientists) emphasize that societies set stable

[1] With Naomi Vaida (Princeton) and Susan Fiske (Princeton). The paper is being revised for submission to PNAS Vistas. The original layout is presented for PNAS. I have made modifications to adjust the paper to the format of an economics paper.

expectations by creating formal and informal structures (institutions such as laws, contracts, and social norms) that establish incentives to behave cooperatively (Ostrom, 1990; North, 1990; Coleman, 1994). But this account is incomplete. No legal diktat can be perfectly enforced (Becker, 1968), no contract can specify every contingency (Hart, 2017), and few norms can reliably regulate private behaviour (Bicchieri, 2005). Instead, people must draw on their repertoire of mental models to make pragmatic judgments about the character of others.

Thus, social psychology plays a central role in organizing the daily reality of collective life (Fiske and Taylor, 2020). People use social structures, such as interdependence and hierarchy, to inform their expectations about others. One fingerprint of this is that people's attention focuses on those who control valued resources (Erber and Fiske, 1984). In other words, people pay attention when they must: in the vocabulary of economics, when facing principal agent relations; in psychologese, under outcome-dependent situations where their goals are contingent on someone else.

What characteristics do people seek to evaluate in such circumstances? Social cognition researchers emphasize people's inference of the other's intent. When constructing mental models, people focus on two fundamental qualities: warmth and competence (Fiske et al., 2002; Fiske, 2018). Perceptions of warmth guide expectations as to whether the other intends to play nicely. These assessments gauge whether the target will be trustworthy, honest, sincere, sociable, and friendly. Perceptions of competence guide beliefs about whether the other has the capacity to deliver. They instruct whether the other will be intelligent, skilled, efficient, determined, and confident.

People infer intent from social structure. Through societal arrangements and incentives, formal and informal institutions cue mental models embedded with information about the warmth and competence of social groups (Hoff and Walsh, 2021; Durante et al., 2013; Fiske, 2012). People use socio-economic status to infer competence; meritocracy rules. Interdependent relations indicate whether the target appears to be a competitor or cooperator (i.e., signal warmth). In cross-national studies, rich people stereotypically are competent but cold (especially in more unequal countries), whereas poor people stereotypically are untrustworthy, incompetent, and less human (Durante et al., 2013; Durante and Fiske, 2017; Durante, Tablante, and Fiske, 2017). Older people, who no longer control resources, stereotypically are incompetent but harmless, well-intentioned but irrelevant (Cuddy, Norton, and Fiske, 2005). Skilled helping professions, such as medicine and teaching, are regarded as both warm and competent (Figure 2 in (Fiske and Dupree, 2014)).

2.2 Stereotyping and Discrimination

People make pragmatic determinations about the warmth and competence of individuals using whatever scaffolding is available (Fiske, 1992). When they lack information, capacity, or motivation to individuate, people form impressions of targets by drawing on mental models that categorize them (Neuberg and Fiske, 1987; Bruner, 1957). Categories are mental models that enable people to filter out information deemed irrelevant and render discriminably different targets equivalent by grouping them into classes. Grouped categories are then treated "in terms of their class membership rather than their uniqueness" (Bruner, Goodnow, and Austin, 1956). The other's apparent group membership is a key mental model: a social category with information attached in the form of a stereotype specifying warmth of intent and competence to enact. Thus, the mental model tells the perceiver whether and how to interact with the stranger. Categories cue a stereotype of the target's most immediately recognizable social group. Categories such as gender, age, and race are visually observable and immediately available, but other categories matter too – for example, professional occupation, nationality, station in life. Although research emphasizes negative prejudices, stereotyping need not be harsh and unflattering. A target's group identity (e.g., nurse, parent) may lead people to view them in a more positive light. Many stereotypes are mixed, high on warmth or competence, but low on the other. Ever since Katz and Braly first measured racial and ethnic stereotypes in 1933 (Katz and Braly, 1933)—and going through Gordon Allport's 1954 account of social categories as the "nouns that cut slices" (Allport, 1954)—social psychology has understood the cognitive processes and content of stereotyping. Discrimination, not so much.

On first impression, one might expect economics and social psychology to offer incommensurate accounts of discrimination. In the standard neoclassical economic model, people are rational actors who use all information at their disposal to pursue autonomous preferences. Categories – the tool of cognitive misers – have no place in such a world. Nor do social motives. Yet, economic theory has operationalized social categories to provide two distinct explanations for discrimination. The first account, offered in the 1950s by Gary Becker, is that some people have a "taste for discrimination" (Becker, 1957). In other words, those people are prepared to incur cost to avoid interacting with a member of a particular group. In a perfect market, prejudice is costly. For a bigoted White employer, this means hiring a White person over a more productive Black person. Thus, the prejudiced must forego expected income to avoid hiring a member of a group they do not like. This account accords

with the psychology of individual differences in overt bias (Allport, 1954). The second account is that people are "statistical" discriminators (Phelps, 1972; Arrow, 1973). Without individuating information, people look to the target's social group to infer characteristics. If a target's group appears less productive on average, the person assumes that the target is also less productive. This account accords with the cognitive miser view of stereotypes as categories standing in for knowledge about individuals (Fiske and Taylor, 2020).

Stereotypes shortchange perceivers, as well as targets, when people fail to see others as individuals, but the process is not simple. Individuating and category-based processing lie on a continuum (Fiske and Neuberg, 1990). Over the last twenty years, economists have generated a rich body of evidence from the labor market showing that stereotypic information affects behaviour even when compelling individuating information is available to the perceiver(Bertrand and Duflo, 2017). Audit (or resume) studies can precisely investigate discrimination by triggering a single categorical cue. One such field experiment sent statistically identical CVs to employers, but randomly varied whether the applicants had Black-sounding names or White-sounding names. A Black sounding name was about one third less likely to get a callback (Bertrand and Mullainathan, 2004). Another study traced hiring practices of symphony orchestras during the period when "blind" auditioning became popular. About 30 percent of the increase in women in symphony orchestras could be attributed to the advent of the blind audition (Goldin and Rouse, 2000). These studies find that mental models have a statistically and economically large effect on decision-making, even in the presence of rich information, but they tell us little about how the content of those categories guided the decision. Without understanding how the structure of mental models affect behaviour, it is difficult to make meaningful predictions about discrimination.

2.3 Contribution

The research traditions on discrimination in both psychology and economics leave a gap at their intersection. On the one hand, economic research shows that social categories affect decision-making even when the stakes are high, but elucidate little on *how* stereotype *content* affects judgment. As Bertrand and Duflo recently pointed out, economists lack a theoretical explanation for how distinct features associated with the social category trigger distinct forms of discrimination (Bertrand and Duflo, 2017). For example, homeless people are dehumanized, whereas the elderly are infantilized. In this vein, the stereotype content model's two central dimensions of stereotype content, warmth and competence, can explain distinct forms of discrimination (Fiske et al., 2002). As noted, some groups are liked but

disrespected (older people), and some are respected but disliked (rich people). Of course, some groups seem high on both (doctors, teachers) or low on both (homeless, alcoholics). Likable warmth and respected competence anticipates individuals' behavioural tendencies toward different groups (Cuddy, Fiske, and Glick, 2007).

On the other hand, while psychological research provides rich insight into the nature of prejudice, the studies' dependent variables are often economically inconsequential. Though this is beginning to change (Jenkins et al., 2018), a notable self-critique lamented that psychology had become the "science of self-reports and finger movements" (Baumeister, Vohs, and Funder, 2007). We aim to fill this gap.

This is important because social psychology's long history showing that self-reports may not predict intended or actual behaviour (Fiske and Taylor, 2020) (Chapter 15). That distinct stereotypes will correspond to distinct behaviours is not a foregone conclusion. In fact, it may even be an improbable, risky prediction, for several reasons.

First, people may not know what they think or what they will do (Nisbett and Wilson, 1977). Second, self-reported predispositions—for example, attitudes—do not predict behaviour unless they are strong, accessible, important, and relevant (Petty and Krosnick, 2014). Stereotypes may not be so cognitively robust. Third, self-reports frequently measure both attitudes and behavioural tendencies, but both are potentially biased by social desirability. This is especially true in a domain such as prejudice, where people want to show their best face. Given that self-reports are more controllable than behaviour, self-reports might be neutral or positive, where behaviour would be avoidant or hostile.

Finally, the field of psychology learned that attitudes predict behaviour only if they match levels of specificity (Fiske and Taylor, 2020)(Chapter 15). For example, an attitude toward Black people in general does not predict behaviour toward a particular Black person. A general attitude predicts general behaviour across several encounters. Equally, a specific attitude toward a particular outgroup member predicts behaviour toward that individual. Thus, for stereotypes and behaviour to align, the present paradigm demands much: a generic judgment about a group must operationalize as a specific behaviour toward a unique individual. This seems like a long shot. Worse yet, in the case of the SCM, the canonical stereotyping questions ask about how society judges groups, and here the criterion is how the individual participant behaves toward one individual group member. Prior evidence does not link individual actors' warmth/competence group stereotypes and their actual behaviour toward a member of that group.

Apart from audit studies and laboratory interactions, observing discriminatory behaviour requires a finer lens. On surveys, people may intentionally or unintentionally misreport how they would behave, when they face no costs (Bertrand and Mullainathan, 2001). Incentives to be accurate could make people think harder; even instructions to be accurate have this effect (Neuberg and Fiske, 1987). Thus far, the SCM has not been tested when people have skin in the game.

Predicting that distinct warmth-by-competence combinations (stereotypes) will operationalize as distinct forms of category-based behaviour (that is, discrimination) is not an inevitable outcome. Valenced attitudes have dominated respective accounts of bias, using a single dimension, to good-enough effect. The current question asks whether using more dimensions will more precisely predict distinct forms of category-based behaviour.

2.4 Overview of methods

This section summarises the overall set up of the study. Each experiment is described in detail along with summary statistics in subsection 2.11.5 of the appendix. The study consists of four experiments in the U.S. (N=717) and two in India (N=564). The experiments were conducted in three rounds. Each round consisted of a trust game and a puzzle game. All three rounds test the same concepts, but vary the context to explore the robustness of the constructs. They are not intended as precise replications nor are they designed to study the effects of the variations.

In each study, participants undertook two rounds of actions. In one round, they assessed the stereotype content of different social groups. Specifically, participants rated how warm and competent the groups were seen by society. The groups for each experiment are listed in subsection 2.11.3. The wording of the questions are presented in subsection 2.11.2. This is the standard approach in the literature on stereotype content (Fiske et al., 2002). These ratings were incentivised. In studies 1a and 1b, we used previous ratings to reward participants. From then onward, we used the averages from the study to reward participants.

In another round, participants played one of the two economic games with members from those social groups. The games were real and rewards were on the line: participants played with actual members of the social groups (responses were pre-recorded), knew they were doing so, and received payments on the basis of how they interacted with those players.

- In studies 1a, 2a, and 3a, the game participants played was the trust game. In the trust game, participants were given an endowment and told that they can share as

much of it as they like with the other player. The amount that they share would be tripled. The other player could return as many tokens as they like.[2]

- In studies 1b, 2b, and 3b, the game participants played was a puzzle game. In the puzzle game, participants were shown a number of puzzles and asked to guess how many of the puzzles different the other player was able to solve.

We note that we ran separate experiments for each behavioural outcome. The reason for this is that we were concerned that participants' behaviour in one economic game could influence their behaviour in another economic game. The case for both games would have been that it would have allowed us to take care of unobserved within-person factors driving the results. However, our fixed effects model accounts for these within-person factors. The regression tables can be found in the appendix, in section subsection 2.11.1.

All participants were recruited online and played the games on an instrument designed on Qualtrics. Any participants who took less than 5 minutes were removed from the sample.

2.5 Summary of experiments

Study 1a: American participants were recruited from the Cloud Research platform and 249 completed the study; 9 participants were dropped because they spent less than 5 minutes on the study. Participants answered questions about the stereotype content of 20 social groups and played the trust game with a member from one of those groups (the order was randomized). Participants were incentivized with tokens for a $500 lottery.

Study 1b: American participants were recruited from the Prolific platform and 196 completed the study; 1 participant was dropped from analysis for spending less than 5 minutes on the study. Participants answered questions about the stereotype content of 20 social groups and played the trust game with a member from one of those groups (the order was randomized). Participants were incentivized with tokens for a $500 lottery.

Study 2a: American participants were recruited from the Prolific platform, and 151 completed the study. Participants answered questions about the stereotype content of 12 social groups and played the trust game with a member from one of those groups (the order was randomized). Participants earned between $4-20 in the game (median amount $11).

[2]We captured this by asking people to indicate how much they would return if they received a given amount. For practical reasons, the unit of value that pre-recorded players made choices with was tokens for a lottery. This was the same store of value as in the first game. In games 2 and 3, participants played with a different form of value (USD and reward points).

Study 2b: American participants were recruited from the Prolific platform, and 131 completed the study. Participants answered questions about the stereotype content of 12 social groups and played the trust game with a member from one of those groups (the order was randomized). Participants earned between $2-19 in the game (median amount $9).

Study 3a: Indian participants were recruited from the IPSOS platform, and 351 completed the study. Participants answered questions about the stereotype content of 12 social groups and played the trust game with a member from one of those groups (the order was randomized). To give a sense of the magnitude of the incentives, participants generally receive 100 points for a short survey. The median participant who completed the study earned 1,450 points.

Study 3b: Indian participants were recruited from the IPSOS platform, and 213 completed the study. Participants answered questions about the stereotype content of 12 social groups and played the puzzle game with a member from one of those groups (the order was randomized). The median participant who completed the study earned roughly 1,000 points.

2.6 Hypotheses

Our first hypothesis held that a target's perceived warmth would predict incentivized behaviour in the trust game better than perceived competence perceptions. These perceptions comprised participants' beliefs about how society judged the target groups' warmth and competence. In other words, we expected participants to give more to members of groups that, in their view, society perceived as warm but to be relatively indifferent to whether players are from groups they think society perceives to be competent or incompetent.

Our second hypothesis was that participants' competence perceptions would predict incentivized expectations of the players' ability better than warmth perceptions. By perceptions, we mean participants' beliefs about society's judgments of groups' competence and warmth. In other words, we expected participants to guess that members of groups they perceive as competent solved more puzzles than members of groups they perceived to be incompetent, regardless of perceived high or low warmth.

2.7 Specification

In order to test the effect of warmth and competence on incentivised behaviour, we run three specifications. Equation 2.1 is the main specification. It includes participant fixed

effects. Equation 2.2 and Equation 2.3 are added to test for robustness. We have two outcomes: investments in the trust game and bets on performance in puzzle bets. The main specification is as follows:

$$Y_{sp} = \beta_0 + \beta_1 \cdot W_{sp} + \beta_2 \cdot C_{sp} + \gamma_p + \mu_{sp} \tag{2.1}$$

Where Y_{sp} is the outcome of interest with s indexing the stereotyped group and p indexing the participant group. β_1 is the coefficient of warmth, capturing the effect of warmth on the outcome. W_{sp} is the participant p's warmth rating of player s. β_2 is the coefficient of competence, capturing the effect of competence on the outcome. C_{sp} is the participant p's competence rating of player s. We include participant fixed effects γ_p. μ_{sp} is a mean zero error term.

Hypothesis 1 is tested in the trust games (1a, 2a, 3a). The parameter of interest for hypothesis 1 is $beta_1$ - $beta_2$. Hypothesis 2 is tested in the trust games (1b, 2b, 3b). The parameter of interest for hypothesis 2 is $beta_2$ - $beta_1$.

In a second specification, we add stereotype fixed effects. This controls for factors related to the other player's identity other than warmth and competence that may affect participants' behaviour. This is denoted by η_s in the equation below:

$$Y_{sp} = \beta_0 + \beta_1 \cdot W_{sp} + \beta_2 \cdot C_{sp} + \gamma_p + \eta_s + \mu_{sp} \tag{2.2}$$

Finally, in the third specification we investigate whether behaviour in the game might be affected by answering the questions (or vice versa). To do this, we develop new independent variables for warmth and competence. These variables are the average scores from participants who took the survey (Module B) first. Thus, the scores are based on responses from participants who are unaware of the game. The average ratings of warmth and competence are used to predict the behaviour of participants who played the game (Module A) first. The limitation of this is that it does not connect individual cognition and behaviour. the advantage of this is that it removes the concern that behaviour in the game and responses in the questionnaire effect one another. The specification is as follows:

$$Y_{sp} = \beta_0 + \beta_1 \cdot W_{s \cdot group} + \beta_2 \cdot C_{s \cdot group} + \gamma_p + \mu_p \tag{2.3}$$

In this specification, $W_{s \cdot group}$ is the average warmth rating of participants who took the

survey first. $C_{s \cdot group}$ is the average competence rating of participants who took the survey first.

2.8 Results

2.8.1 Trust game

The trust game is one of the most commonly studied experimental games in economics. In experiment 1a, participants played 20 games with different players. In each game, participants were given 5 tokens for a lottery for $500 and were told that they could share some or all of the tokens with another player. If they decided to share nothing, they could keep all 5 tokens. However, if they chose to share with the other player, the amount the participant shared was tripled. For example, if the participant gave 1 token, the other player received 3 tokens. The other player would then be given an opportunity to give tokens back. Participants received only one piece of information about the other player: the social group to which they belonged. (i.e., Hispanic person). In studies 2a and 3a, participants played with US dollars and points for an online marketplace, respectively, with the same incentive structure. The amount returned by the SCM group member was pre-recorded. Prior to the study, we recruited members of the SCM groups to play all potential plays (e.g., "The participant has 5 tokens and gives you 4 tokens. This is tripled: you receive 12 tokens. How many tokens will you give back?").

In Study 1a, 240 participants played 20 different players from different social groups (resulting in 4,800 decisions). They gave on average 2.1 of their 5 tokens (SD = 1.585), and the modal amount given was 1 token in about 28 percent of the games. Only 5 percent of the participants gave the same amount in all 20 games – indicating that the identity of the other participant influenced the vast majority of the focal participants' trust behaviour. On average, the teacher received the most tokens (2.9), while the alcoholic received the least (1.1). Table 2.3 documents the results. A one-point increase in warmth (on a 4-point scale) is associated with a 0.45 ($p < 0.001$) increase in the number tokens shared, while a one-point increase in competence is associated with 0.05 increase in the number of tokens shared ($p = 0.169$) (Figure 2.1). The effect of competence becomes significant once stereotype fixed effects are added but the post-estimation t-test shows that the difference between warmth and competence is statistically significant ($p = 0.002$).

Study 2a replicated the basic structure of the experiment with 151 participants playing with 12 players to test the robustness of the results (1,812 decisions). The study retained the

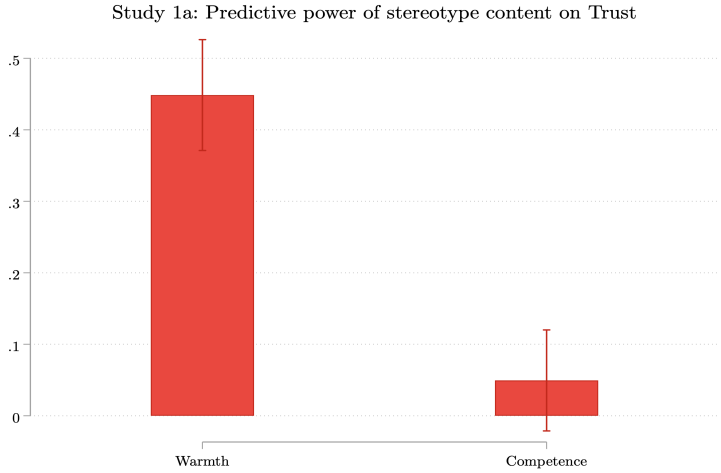


FIGURE 2.1: Results: Study 1a: Trust game

incentive structure of the game (i.e., the amount shared would be tripled, etc.), but changed some features. Notably, participants received a cash endowment of $5, had a two-week gap between the SCM ratings and the game, and were rewarded for one randomly selected game (methods are discussed in detail in SI). Participants in Study 2a gave on average about $2.80 (SD = 1.74). Table 2.4 documents the results. A one-point increase in warmth (on a 4-point scale) is associated with a 0.37 ($p < 0.001$) increase in the amount of shared, while a one-point increase in competence is associated with 0.11 decrease in the amount of money shared shared ($p = 0.005$) (Figure 2.2). The effect of competence washes out once stereotype fixed effects are added. The post-estimation t-test shows that the difference between warmth and competence is statistically significant ($p < 0.001$).

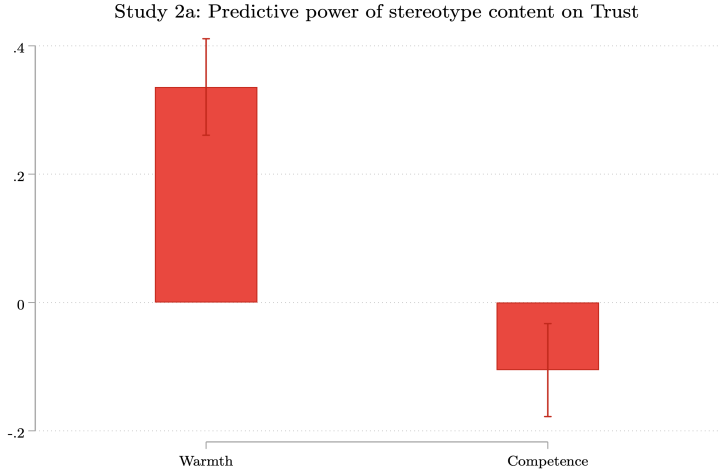


FIGURE 2.2: Results: Study 2a: Trust game

In Study 3a, 351 participants from India played the game with players from 12 social groups selected from each Indian SCM quadrant (15 and FiskeLab.com). It replicated the design

from 2a, except that participants played with points for an online platform instead of cash. Participants gave 1.95 points on average (SD = 1.56) and about 7 percent of the sample gave the same number of points to all players they played with.Table 2.5 documents the results. A one-point increase in warmth (on a 4-point scale) is associated with a 0.28 ($p < 0.001$) increase in the number points shared while a one-point increase in competence is associated with 0.12 decrease in the number of points shared ($p < 0.001$) (Figure 2.3). The effect of competence washes out once stereotype fixed effects are added. The post-estimation t-test shows that the difference between warmth and competence is statistically significant ($p < 0.001$).

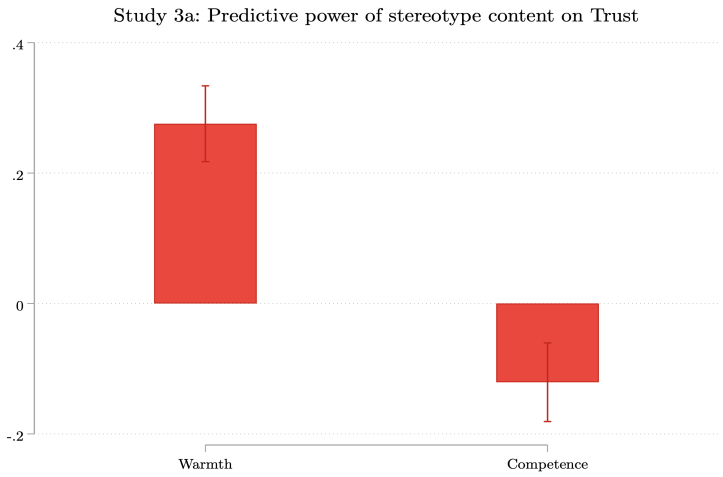


FIGURE 2.3: Results: Study 3a: Trust game

Table 2.1 shows the results from the combined studies. To summarize studies 1-3: In each study, there is a statistically significant difference between the Warmth and Competence coefficients and this is robust across specifications. The Warmth coefficient is robust to specification 1 and specification 2 while the Competence is not robust in any of the studies. In study 1, the coefficient is not significant under specification 1, but become significant in specification 2 (which adds stereotype fixed effects). On studies 2 and 3, the opposite is the case. Moreover, the direction of the point-estimate is unstable. In study 1, the point-estimate is positive while in studies 2 and 3 the point-estimate is negative.

Table 2.1: Full Sample (Studies 1a + 2a + 3a)

		Individual perceptions	Group Averages
	(1)	(2)	(3)
	Trust	Trust	Trust
Warmth	0.377***	0.231***	
	(0.000)	(0.000)	
Competence	-0.0266	0.0777***	
	(0.210)	(0.000)	
Group Warmth			0.391***
			(0.000)
Group Competence			-0.268***
			(0.000)
Constant	1.189***	1.306***	1.534***
	(0.000)	(0.000)	(0.000)
P: Warmth = Competence	0.000	0.000	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	10824.000	10824.000	5540.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.8.2 Puzzle game

Participants expected members of groups they believe society perceives as competent to do better at puzzles. Warmth, while statistically significant, was a far smaller predictor. The figures below show coefficient for fixed effects regression of warmth and competence on trust game behaviour. All confidence intervals are at the 95 percent level. Errors are clustered at the participant level.

In Study 1b, 195 participants played 20 different players from different social groups (3,900 decisions). Participants guessed on average that players solved 2.7 (SD = 1.3) puzzles. The modal guess was 3 puzzles, occurring in 30 percent of the games. No participant guessed that each player solved the same number of puzzles. On average, participants guessed that the Ivy League graduate solved the most puzzles (3.9), while the person with a learning disability received the lowest estimate (1.34). Table 2.7 documents the results. A one-point increase in competence (on a 4-point scale) is associated with a 0.57 ($p < 0.001$) increase in the expectations of number of puzzles solved while a one-point increase in warmth is associated with 0.11 increase in the number of tokens shared ($p < 0.001$) (Figure 2.4).The effect of warmth persists when stereotype fixed effects are added. The post-estimation t-test shows that the difference between warmth and competence is statistically significant ($p < 0.001$).

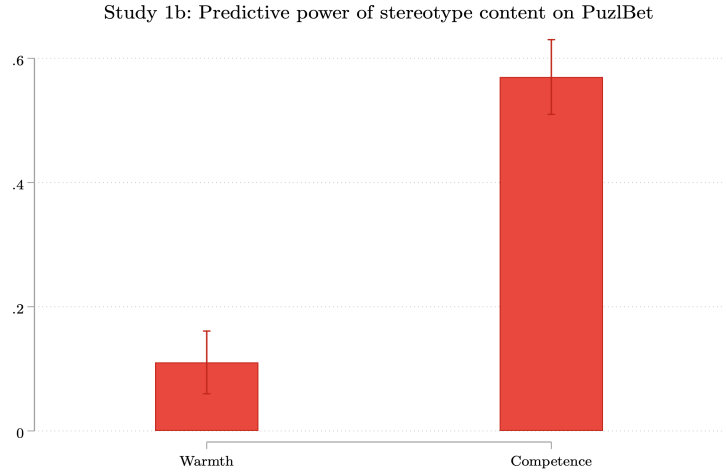


FIGURE 2.4: Results: Study 1b: Puzzle game

To test the robustness of the results, we replicated the study with similar conditions but with a number of changes. In Study 2b, 131 participants played the puzzle game with 12 players (1,572 decisions). This time, participants were given $1.25 for each correct guess, and we broke up the decision and social perception ratings by two weeks to mitigate order effects. Participants guessed on average that people solved 2.8 puzzles (SD = 1.2), and

nobody guessed that everybody solved the same number of puzzles. Table 2.8 documents the results. A one-point increase in competence (on a 4-point scale) is associated with a 0.60 ($p < 0.001$) increase in the expectations of number of puzzles solved while a one-point increase in warmth is associated with 0.07 increase in the number of tokens shared ($p = 0.021$) (Figure 2.5).The effect of warmth persists when stereotype fixed effects are added. The post-estimation t-test shows that the difference between warmth and competence is statistically significant ($p < 0.001$), though these results are not robust to the second specification.

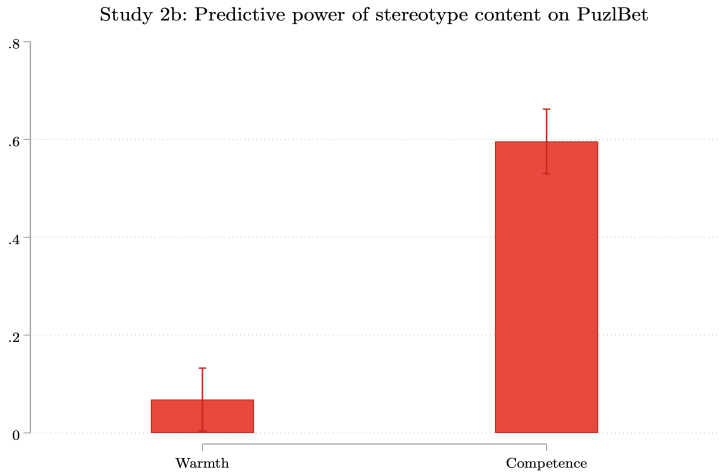


FIGURE 2.5: Results: Study 2b: Puzzle game

Study 3b replicated the procedure with participants from India, who played the game with players from 12 social groups from each Indian SCM quadrant; in 2,556 decisions, 213 players played with points for an online platform instead of cash. Participants guessed on average that people solved 3.1 puzzles (SD = 1.1), and only 6 guessed that everybody solved the same number of puzzles. Table 2.9 documents the results. A one-point increase in warmth (on a 4-point scale) is associated with a 0.29 ($p < 0.001$) increase in the number tokens shared while a one-point increase in competence is associated with 0.10 increase in the number of tokens shared ($p = 0.021$) (Figure 2.6). The effect of warmth washes out when stereotype fixed effects are added. The post-estimation t-test shows that the difference between warmth and competence is statistically significant ($p < 0.001$), though these results are not robust to the second specification.

Table 2.2 shows the results from the combined studies. To summarize studies 1-3: In each study, there is a statistically significant difference between the Competence and Warmth coefficients. This result is robust across specifications in study 1, but not studies 2 and 3. The Competence coefficient is robust to specification 1 and specification 2 in all studies. The Warmth Coefficient is robust only in study 2. On balance, this provides strong evidence that

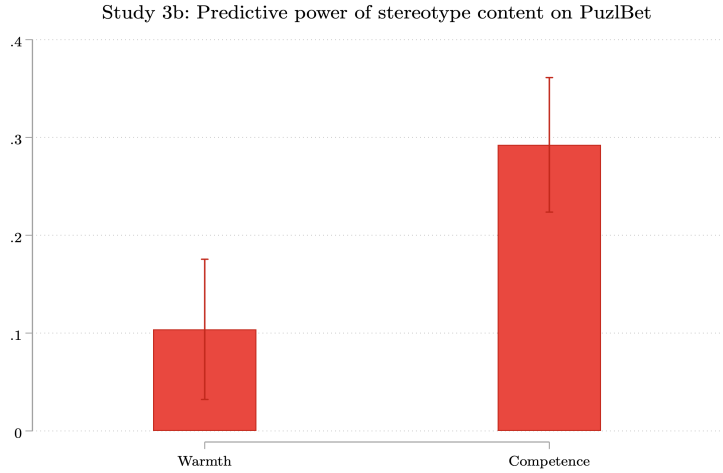


FIGURE 2.6: Results: Study 3b: Puzzle game

Competence is distinctly effects people's perceptions of targets performance in the puzzles. However, the results are not as robust as for the trust game.

TABLE 2.2: Full Sample (Studies 1b + 2b + 3b)

	Individual perceptions		Group Averages
	(1)	(2)	(3)
	Puzzle_Bet	Puzzle_Bet	Puzzle_Bet
Warmth	0.114***	0.0811***	
	(0.000)	(0.000)	
Competence	0.506***	0.190***	
	(0.000)	(0.000)	
SC_Grp_Warmth			0.0720***
			(0.001)
SC_Grp_Comp			0.476***
			(0.000)
Constant	1.187***	2.125***	0.659***
	(0.000)	(0.000)	(0.000)
P: Warmth = Competence	0.000	0.001	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	8016.000	8016.000	4080.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.9 Discussion

With evidence from six studies including 1,281 participants and over 18,000 decisions in two distinct cultural settings, the United States and India, these findings shed light on a number of features of the relationship between mental models and economic behaviour. The perceived warmth of social groups robustly predicts the degree to which people will share with members of that group in the trust game. Conversely, the perceived competence of

social groups predicts how people will bet on players' performance in challenging puzzles. In all studies, our primary specification supported our hypotheses that distinct stereotype content predicts distinct behaviour. In studies 2b and 3b these results were not robust to a secondary specification (though the difference in point estimates was relatively large). These results provide support for the core proposition of the stereotype content model – that warmth and competence differentially guide discrimination – even under incentivised conditions.

Second, as predicted, shared representations antecede economically relevant behaviour. In the original publication on the SCM, Fiske et al., 2002 consider the importance of participants' responses in surveys. They ask "[w]ere participants reporting the culture's, their group's, or their own personal stereotypes and prejudices?" By examining people's behaviour when value is on the line, our results indicate that people operationalise signals from shared mental models in making decisions of consequence.

2.9.1 Policy implications

The results have two main policy implications. First, they contribute to understanding the mechanisms guiding distinct forms of discrimination. As discussed at the outset, evidence from self-reports indicate distinct forms of prejudice for different groups – for example, people with disabilities (pity and avoidance) differ from prejudice against high-performing ethnic minorities (envy and sabotage) (Cuddy, Fiske, and Glick, 2007). This evidence bolsters the credibility of warmth and competence as fundamental and distinct mechanisms guiding social cognition. It suggests that interventions focused on tackling discrimination should first diagnose the kind of prejudice.

These results also shed light on the dynamics of trust in society. Trust plays a well-established role in nations' social and economic prosperity (Glaeser et al., 2000; Putnam, 1993; Fukuyama, 1995), serving – as Kenneth Arrow famously noted – to lubricate the social system (Arrow, 1974). Although we use the "trust game" to capture only intent, trust is conventionally understood to constitute an expectation both about the motive and competence of other actors to act to achieve particular results that affect the trustor (Gambetta, 2000; Castelfranchi and Falcone, 2010; Mayer, Davis, and Schoorman, 1995). The findings highlight that much of the important action and variation occurs not only between societies but within them. People take care to discern whom to trust, and predictably vary in what they expect different people to do. For policymakers aiming to foster and sustain high-trust

societies, understanding these dynamics – their causes and manifestations – is of central importance. This is especially important in contexts of conflict and instability (Collier, 2011; Bank, 2011).

2.9.2 Future directions

Several areas warrant future research. First, the model should be tested outside the laboratory in a more ecologically valid setting – for example, via audit studies. Second, while our findings show that the reported warmth and competence predict incentivized behaviour, future research needs to causally identify these relationships. Third, the model instruments could be tested and refined for robust use globally.

2.10 Conclusion

Stereotyping has profound implications for its targets. Such people are essentialized by their group membership in their day-to-day interactions with strangers. For stigmatized groups, stereotyping can rob them of control and a fair hearing (Hoff and Walsh 2018). When they seek jobs and education, their payoff may be lower than others (Quillian et al., 2017). When they seek jobs or housing, audit studies show, racial minorities receive fewer callbacks and offers (Massey and Lundy, 2001). Ultimately, experiencing bias is associated with lower health and well-being (Keyes, Hatzenbuehler, and Hasin, 2011). In sum, stereotypes play a central role in the mechanics of social exclusion (Hoff and Walsh, 2018; Massey, 2007).

These data offer an understandable, efficient, portable method for assessing distinct forms of discrimination directed to distinct societal groups. Rather than a single dimension of positive-negative valence (as in attitudes, feeling thermometers, or utility), respect for competence and liking for warmth respectively anticipate betting on someone's performance and trusting someone to share assets. These systematic patterns all result from individuals applying society's stereotypic mental models to individuals.

2.11 Appendix

2.11.1 Results

Trust Game

Table 2.3: Study 1a

		Individual perceptions	Group Averages
	(1)	(2)	(3)
	Trust	Trust	Trust
Warmth	0.449***	0.335***	
	(0.000)	(0.000)	
Competence	0.0494	0.159***	
	(0.169)	(0.000)	
Group Warmth			1.240***
			(0.000)
Group Competence			-0.925***
			(0.000)
Constant	0.755***	0.760***	1.056***
	(0.000)	(0.000)	(0.000)
P: Warmth = Competence	0.000	0.002	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	4800.000	4800.000	2600.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: Study 2a

	Individual perceptions		Group Averages
	(1)	(2)	(3)
	Trust	Trust	Trust
Warmth	0.336***	0.193***	
	(0.000)	(0.000)	
Competence	-0.105***	-0.0399	
	(0.005)	(0.410)	
Group Warmth			0.268***
			(0.000)
Group Competence			-0.112***
			(0.004)
Constant	2.219***	2.416***	1.970***
	(0.000)	(0.000)	(0.000)
P: Warmth = Competence	0.000	0.002	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	1812.000	1812.000	924.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5: Study 3a

	Individual perceptions		Group Averages
	(1)	(2)	(3)
	Trust	Trust	Trust
Warmth	0.276***	0.136***	
	(0.000)	(0.000)	
Competence	-0.121***	0.000775	
	(0.000)	(0.978)	
Group Warmth			0.459***
			(0.000)
Group Competence			-0.450***
			(0.000)
Constant	1.479***	1.544***	1.844***
	(0.000)	(0.000)	(0.000)
P: Warmth = Competence	0.000	0.005	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	4212.000	4212.000	2016.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.6: American Sample (Studies 1a + 2a)

		Individual perceptions	Group Averages
	(1)	(2)	(3)
	Trust	Trust	Trust
Warmth	0.424***	0.290***	
	(0.000)	(0.000)	
Competence	0.0121	0.134***	
	(0.658)	(0.000)	
Group Warmth			0.368***
			(0.000)
Group Competence			-0.197***
			(0.000)
Constant	1.137***	1.163***	1.555***
	(0.000)	(0.000)	(0.000)
P: Warmth = Competence	0.000	0.000	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	6612.000	6612.000	3524.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Puzzle Game

Table 2.7: Study 1b

		Individual perceptions	Group Averages
	(1)	(2)	(3)
	Puzzle_Bet	Puzzle_Bet	Puzzle_Bet
Warmth	0.110***	0.0994***	
	(0.000)	(0.001)	
Competence	0.570***	0.295***	
	(0.000)	(0.000)	
SC_Grp_Warmth			-0.00000706
			(1.000)
SC_Grp_Comp			0.640***
			(0.000)
Constant	0.888***	1.667***	1.048***
	(0.000)	(0.000)	(0.000)
P: Warmth = Competence	0.000	0.000	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	3900.000	3900.000	2040.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: Study 2b

	Individual perceptions		Group Averages
	(1)	(2)	(3)
	Puzzle_Bet	Puzzle_Bet	Puzzle_Bet
Warmth	0.0683**	0.0870**	
	(0.037)	(0.030)	
Competence	0.596***	0.178***	
	(0.000)	(0.001)	
SC_Grp_Warmth			0.00201
			(0.946)
SC_Grp_Comp			0.410***
			(0.000)
Constant	1.105***	2.103***	0.606***
	(0.000)	(0.000)	(0.000)
P: Warmth = Competence	0.000	0.229	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	1560.000	1560.000	732.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: Study 3b

	Individual perceptions		Group Averages
	(1)	(2)	(3)
	Puzzle_Bet	Puzzle_Bet	Puzzle_Bet
Warmth	0.104***	0.0296	
	(0.005)	(0.442)	
Competence	0.292***	0.0816**	
	(0.000)	(0.012)	
SC_Grp_Warmth			0.124***
			(0.001)
SC_Grp_Comp			0.442***
			(0.000)
Constant	1.989***	2.784***	-0.0374
	(0.000)	(0.000)	(0.870)
P: Warmth = Competence	0.001	0.363	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	2556.000	2556.000	1308.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.10: American Sample (Studies 1b + 2b)

	Individual perceptions		Group Averages
	(1)	(2)	(3)
	Puzzle_Bet	Puzzle_Bet	Puzzle_Bet
Warmth	0.0954***	0.101***	
	(0.000)	(0.000)	
Competence	0.581***	0.262***	
	(0.000)	(0.000)	
SC_Grp_Warmth			0.0438
			(0.107)
SC_Grp_Comp			0.494***
			(0.000)
Constant	0.949***	1.782***	0.970***
	(0.000)	(0.000)	(0.000)
P: Warmth = Competence	0.000	0.000	0.000
Participant FE	Yes	Yes	Yes
Stereotype FE	No	Yes	No
Observations	5460.000	5460.000	2772.000

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.11.2 Survey module

Following prior research on the stereotype content model[3], participants were asked to rate how society perceives the trustworthiness, warmth, competence and assertiveness of target groups. Participants rated groups social groups on a 4-point scale (0 not at all; 4 extremely). In Studies 1a and 1b, we only asked participants one question on the scale (trustworthy and competent). In the four other studies, we asked the four items listed above.

> *To what extent do most (Americans/ Indians) view an Target Group as (trustworthy, warm, competent, assertive)?*

In addition to this, participants were asked to rating each group according to the cooperative/competitive (on a 4-point scale); ideology (on a 5-point spectrum), as well as the MacArthur Scale of Subjective Social Status (10-point scale).

> *To what extent do most Americans view a middle class person as cooperative or competitive with the rest of society (for example, exploiting limited resources or not)?*

In the United States, we operationalized ideology in terms of party affiliation. Following consultation with Indian social scientists, we operationalized ideology as a spectrum from conservative to liberal in India.

> *In politics today, do most Americans view a parent as a Republican, a Democrat or an Independent? (US Sample)*

> *In politics today, do most Indians view an educated person as a Conservative, a Liberal or an Independent? (Indian Sample)*

[3] S. T. Fiske, A. J. C. Cuddy, P. Glick, J. Xu, A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. J. Pers. Soc. Psychol. 82, 878–902 (2002).

2.11.3 Full list of groups

Studies 1a & 1b

1. A banker
2. A doctor
3. A homeless person
4. A Jewish person
5. A nurse
6. A parent
7. A person in a wheelchair
8. A retired person
9. A Turkish person
10. African
11. An adult with learning disabilities
12. An alcoholic
13. An Asian person
14. An Ivy league graduate
15. A financial analyst
16. A hispanic person
17. An Italian person
18. A middle class person
19. A person with Parkinson's
20. A teacher

Studies 2a & 2b

1. A Jewish person
2. A person in a wheelchair
3. A homeless person
4. An Asian person
5. A retired person
6. A Turkish person
7. An Ivy league graduate
8. An adult with learning disabilities
9. An alcoholic
10. A parent
11. A person with Parkinson's

12. A middle class person

Studies 3a & 3b

1. Brahmin
2. Middle class person
3. Educated person
4. Hindu person
5. Housewife
6. Uneducated person
7. Rich person
8. Old person
9. Disabled person
10. Man
11. Person from a scheduled caste
12. Unemployed person

2.11.4 Validation

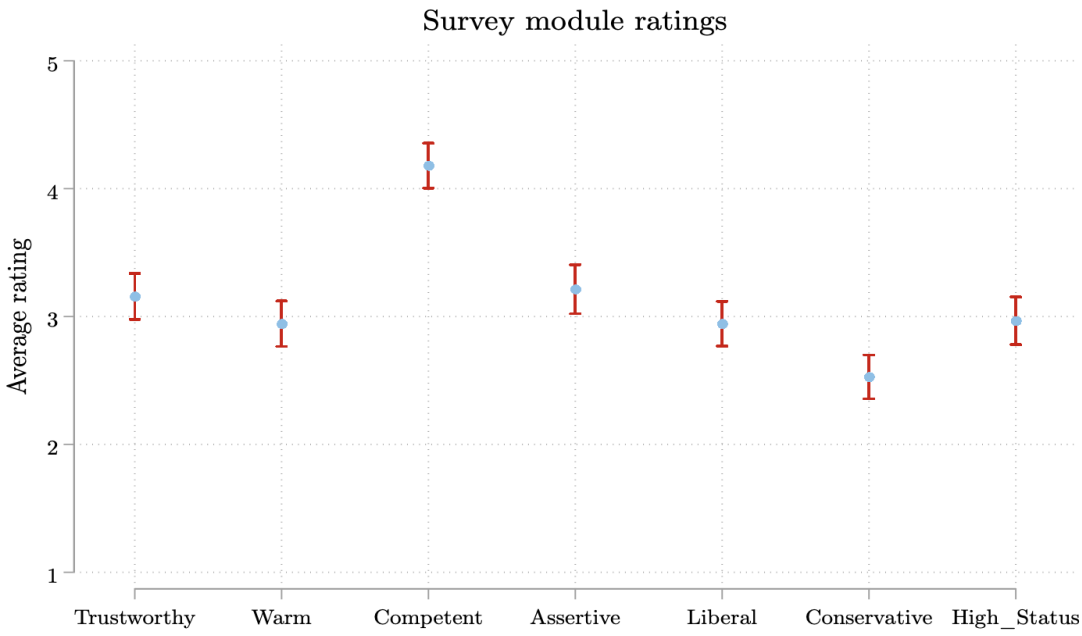
US Sample

89 participants from the US were recruited on the Prolific platform to validate the behavioral measure of competence. Participants were shown each of the puzzles and asked to write a word that describes a person who would be able to solve the them. Participants who tried to solve the puzzles were removed from the sample.

TABLE 2.11: Word frequency

Description	Frequency
Smart	49
Clever	47
Creative	21
Logical	19
Mathematical	17
Intelligent	14
Average	10
Perceptive	10

Note: Participants listed 89 unique words. 22 participants were dropped for attempting to solve the puzzles rather than answer the questions.



Participants were asked to rate a person who solved all puzzles on a five point scale according to the tr

Indian Sample

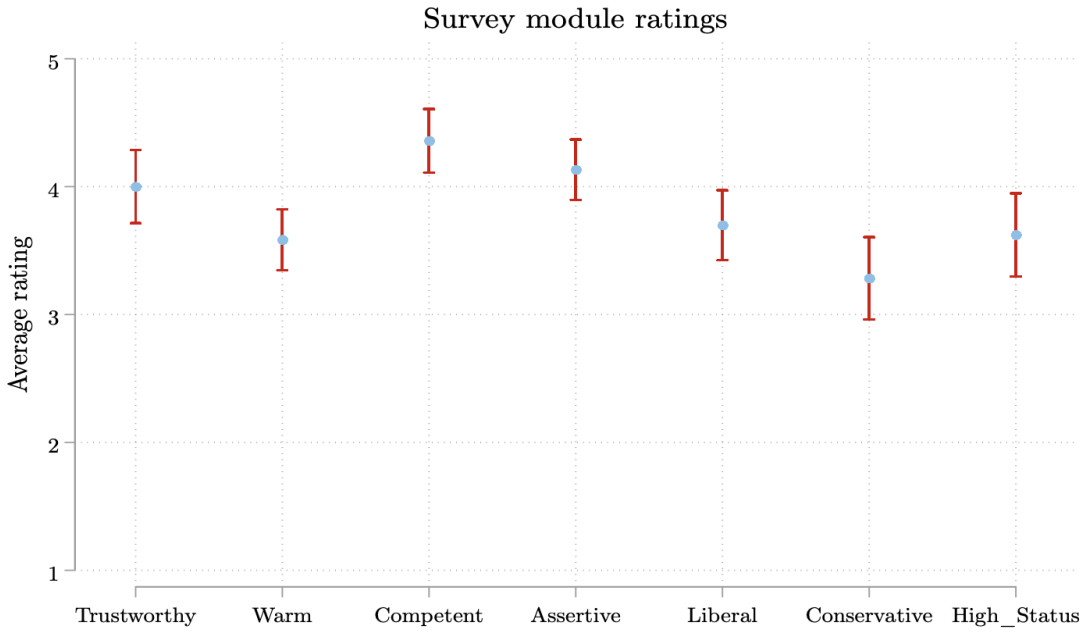
117 participants from India were recruited on the Ipsos platform to validate the behavioral measure of competence. Participants were shown each of the puzzles and asked to write a

word that describes a person who would be able to solve the them. Participants who tried to solve the puzzles were removed from the sample.

Table 2.12: Word frequency

Description	Frequency
Intelligent	40
Genius	29
Smart	17
Mathematician	11
Logical	10

Note: Participants listed 110 unique words. 64 participants were dropped for attempting to solve the puzzles rather than answer the questions.



Participants were asked to rate a person who solved all puzzles on a five point scale according to the tr

2.11.5 Detailed description of studies

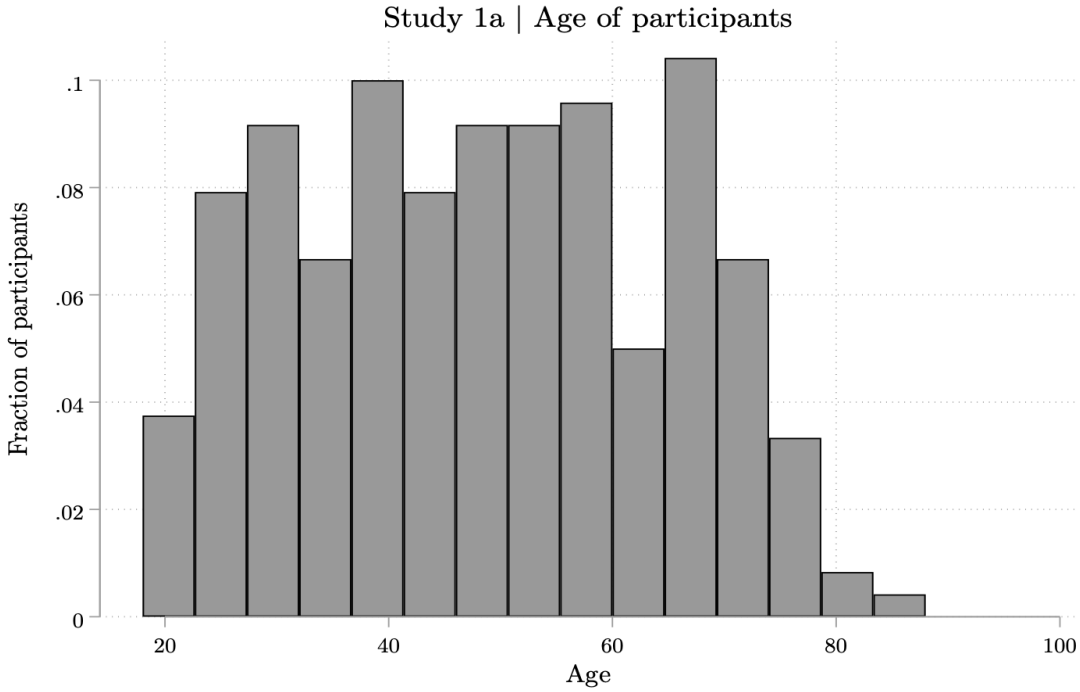
Study 1a

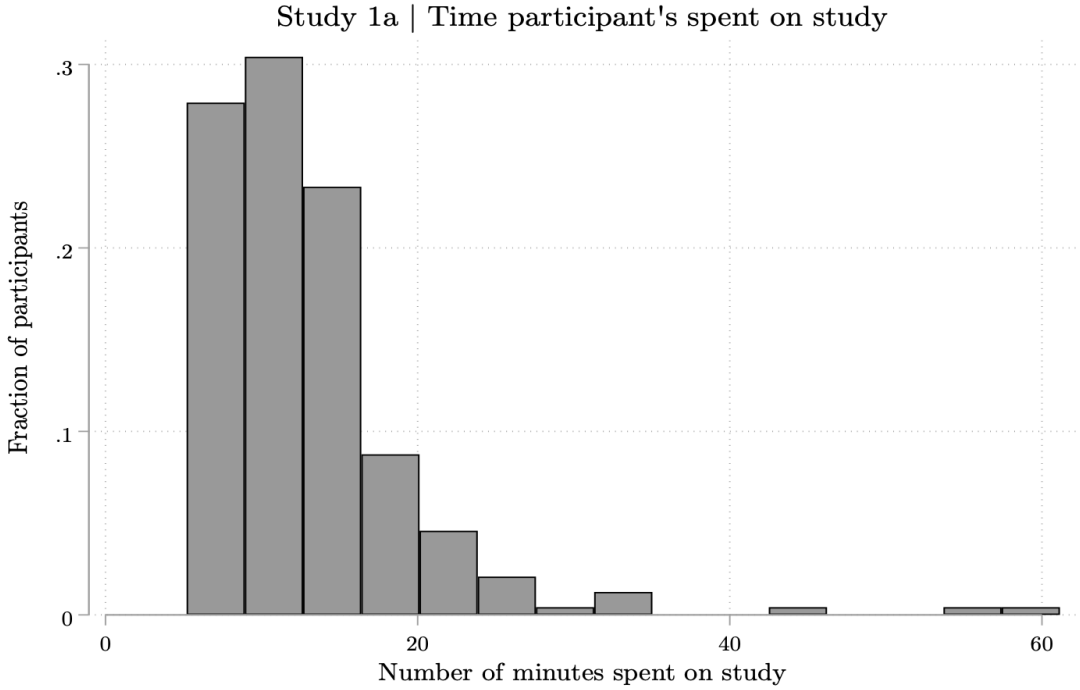
We conducted the study with an online sample of Americans (N=249), using the Turk Prime platform. 9 participants were dropped because they spent less than 5 minutes on the study. Participants played the game using a survey instrument created on Qualtrics. It took place in October 2019. The survey had two modules, Module A and Module B. Participants were randomly allocated to take Module A or Module B first. In both modules, participants were informed that they would have an opportunity to earn tokens for a lottery for $500. In Module A, participants were invited to play a trust game with 20 different individuals. In the trust game, participants were given an endowment of 5 tokens and told that they can share as many of them as they like with the other player. The amount that they share would be tripled. The other player can decide to give back as many tokens as they like. In each case, they were given stereotype-associated group about the person. In Module B, participants were asked to rate the 20 stereotypic categories on the dimensions of warmth and competence. If their guesses aligned with the findings of previous studies, they were rewarded with a token. Participants were also asked to provide demographic information (gender, age, political affiliation).

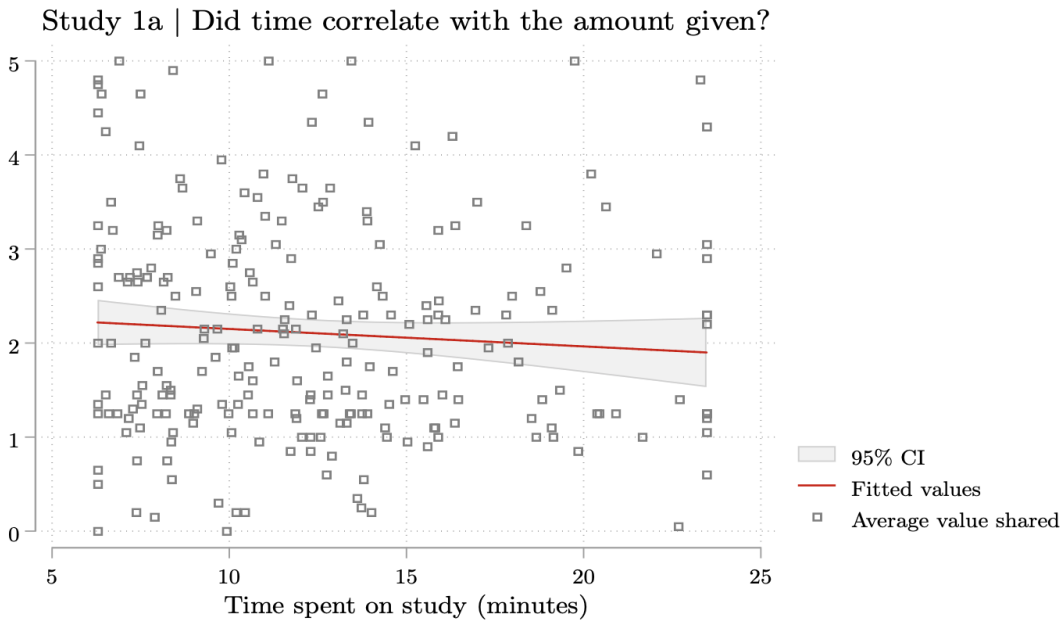
Table 2.13: Study 1a: Summary statistics and balance check on order

	(1) Survey first		(2) Game first		(3) Total		T-test P-value
Variable	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)
Age	110	48.445 (1.575)	130	47.915 (1.409)	240	48.158 (1.048)	0.802
Male proportion of sample	110	0.436 (0.048)	130	0.423 (0.043)	240	0.429 (0.032)	0.837
Time spent on study (minutes)	110	12.741 (0.606)	130	13.230 (0.646)	240	13.006 (0.446)	0.586
Average value shared	110	2.323 (0.115)	130	1.918 (0.095)	240	2.104 (0.075)	0.007***
Played warmth rating	110	2.841 (0.047)	130	2.582 (0.047)	240	2.701 (0.034)	0.000***
Played competence rating	110	2.877 (0.049)	130	2.673 (0.047)	240	2.767 (0.035)	0.003***

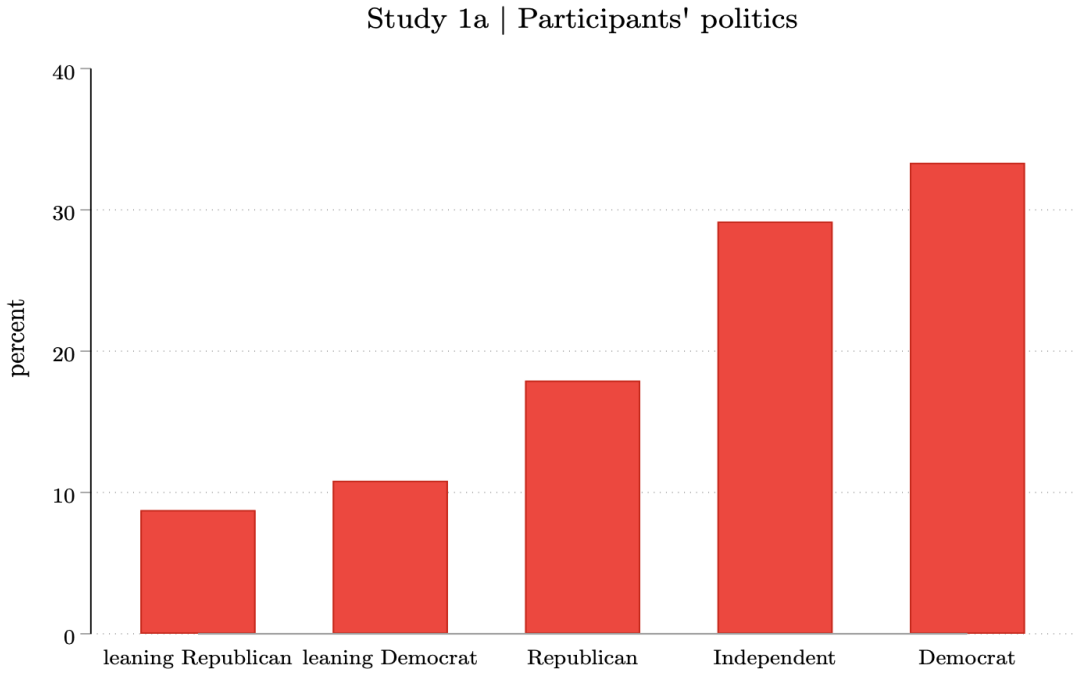
Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

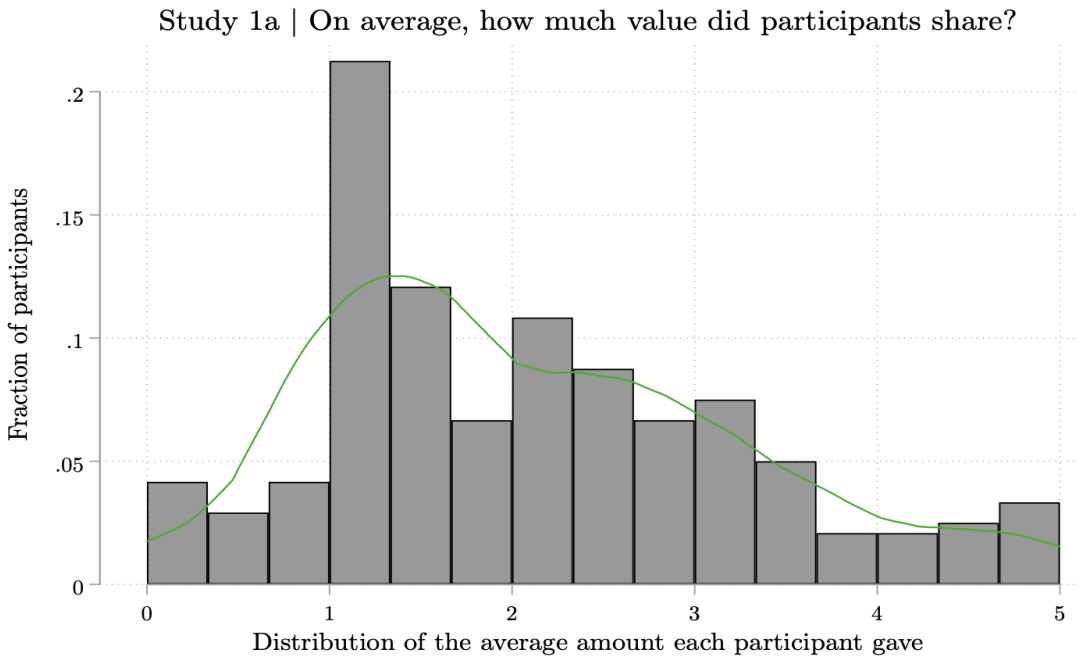

Study 1a | Age of participants
Fraction of participants
.1
.08
.06
.04
.02
0
20
40
60
80
100
Age

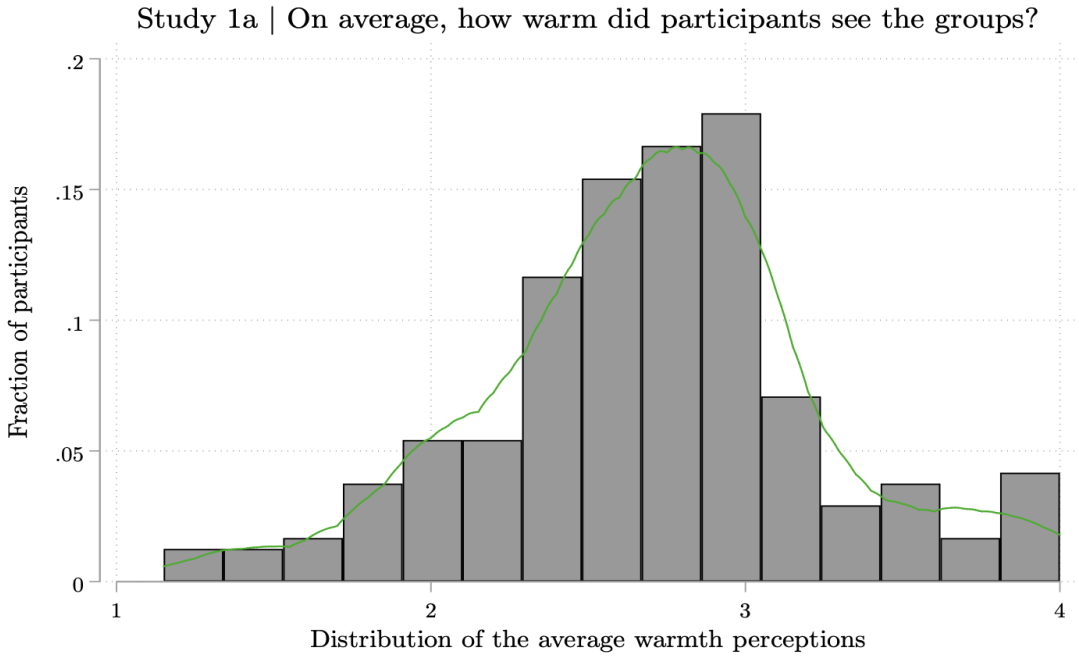

Study 1a | Time participant's spent on study
Fraction of participants
.3
.2
.1
0
0
20
40
60
Number of minutes spent on study

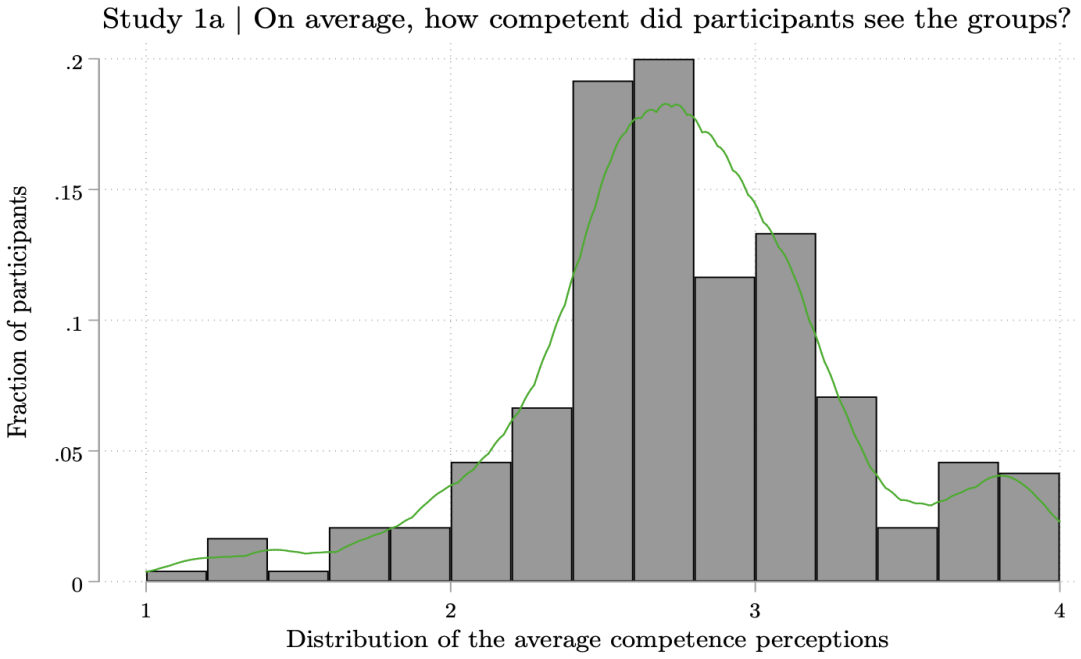


R-squared = 0.0059
Coefficient = -0.0185
Winzorized at 95 percent.




Study 1a | On average, how much value did participants share?
Fraction of participants
.2
.15
.1
.05
0
0
1
2
3
4
5
Distribution of the average amount each participant gave


Study 1a | On average, how warm did participants see the groups?
Fraction of participants
.2
.15
.1
.05
0
1
2
3
4
Distribution of the average warmth perceptions


Study 1a | On average, how competent did participants see the groups?
Fraction of participants
.2
.15
.1
.05
0
1
2
3
4
Distribution of the average competence perceptions

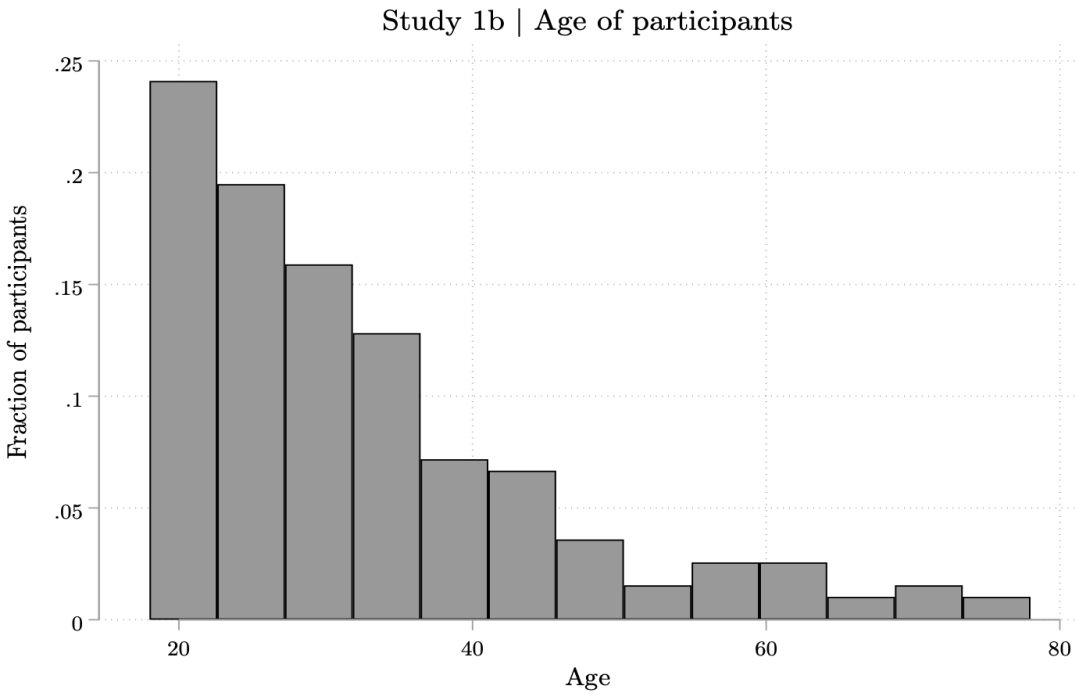
Study 1b

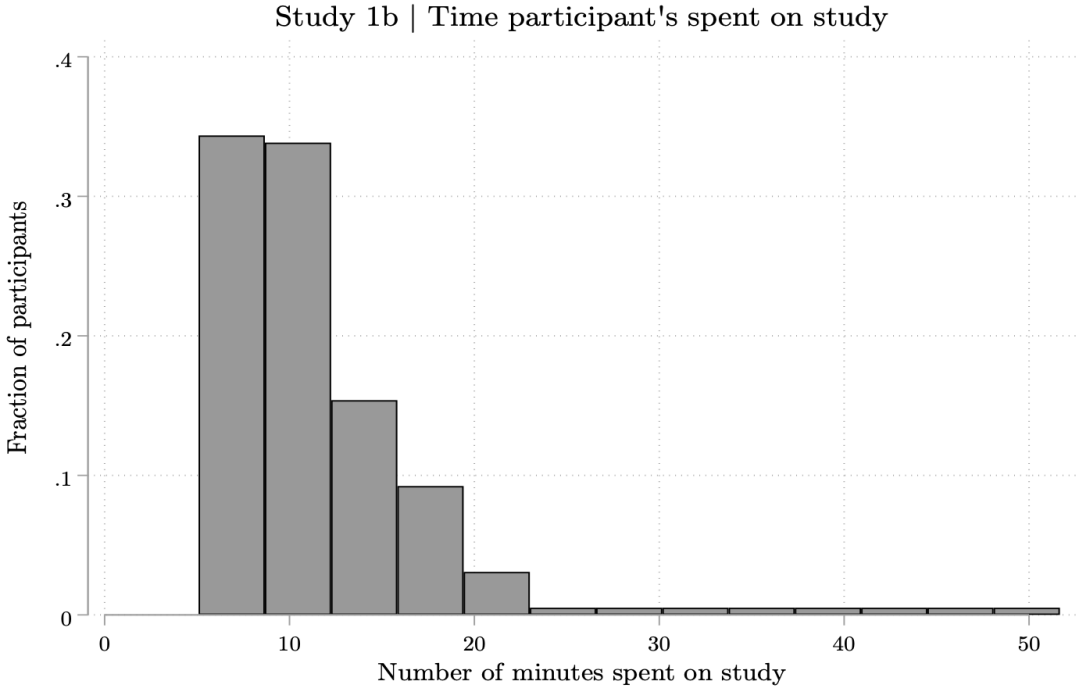
We conducted the study with an online sample of Americans,(N=196), using the Prolific platform. Participants played the game using a survey instrument created on Qualtrics. The study took place in September 2020. 1 participant was dropped from analysis because they spent less than 5 minutes on the study. The survey had two modules, Module A and Module B. We randomized the order of the modules so that some people took Module A first, while others took Module B first. In both modules, participants were informed that they would have an opportunity to earn tokens for a lottery for $500. In Module A, participants were shown a number of puzzles. They were then invited to guess how many of the puzzles different individuals were able to solve. In each case, they were given stereotypic information about the person (e.g., Asian, Doctor, African etc.). If they guessed correctly, they were be rewarded with a token. In Module B, participants were asked to rate the 20 stereotypic categories on the dimensions of warmth and competence. If their guesses aligned with the findings of previous studies, they were rewarded with a token. Participants were also asked to provide demographic information (gender, age, political affiliation).

Table 2.14: Study 1b: Summary statistics and balance check on order

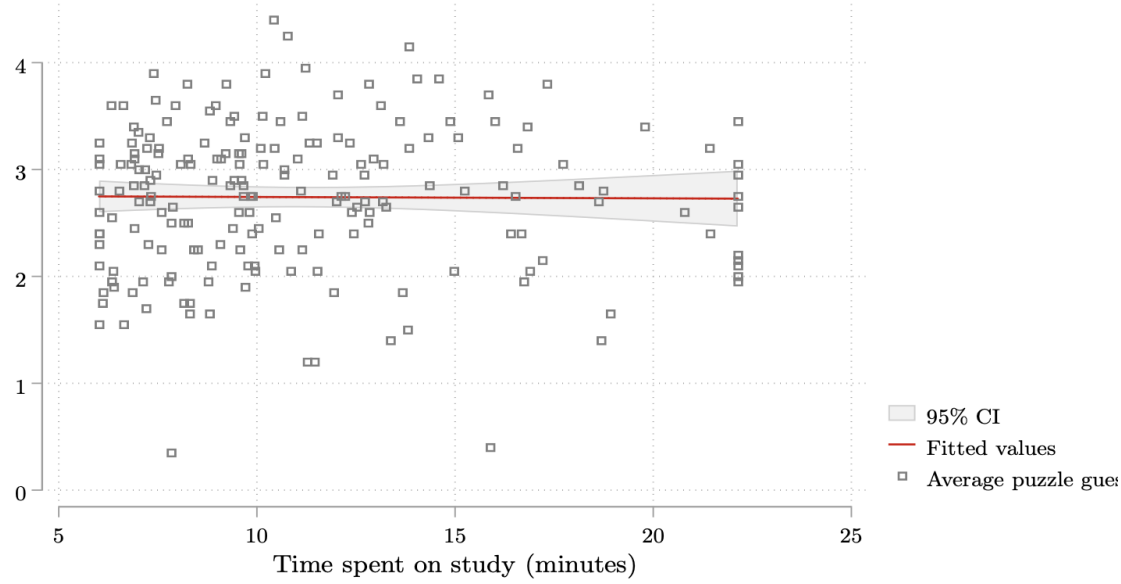
	(1) Survey first		(2) Game first		(3) Total		T-test P-value
Variable	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)
Age	93	32.849 (1.337)	102	32.520 (1.332)	195	32.677 (0.942)	0.862
Male proportion of sample	93	0.473 (0.052)	102	0.471 (0.050)	195	0.472 (0.036)	0.972
Time spent on study (minutes)	93	11.673 (0.766)	102	11.810 (0.604)	195	11.745 (0.482)	0.888
Average puzzle guess	93	2.691 (0.069)	102	2.788 (0.069)	195	2.742 (0.049)	0.327
Played warmth rating	93	2.691 (0.033)	102	2.727 (0.038)	195	2.710 (0.025)	0.472
Played competence rating	93	2.718 (0.031)	102	2.739 (0.036)	195	2.729 (0.024)	0.669

Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.


Study 1b | Age of participants
Fraction of participants
.25
.2
.15
.1
.05
0
20
40
60
80
Age

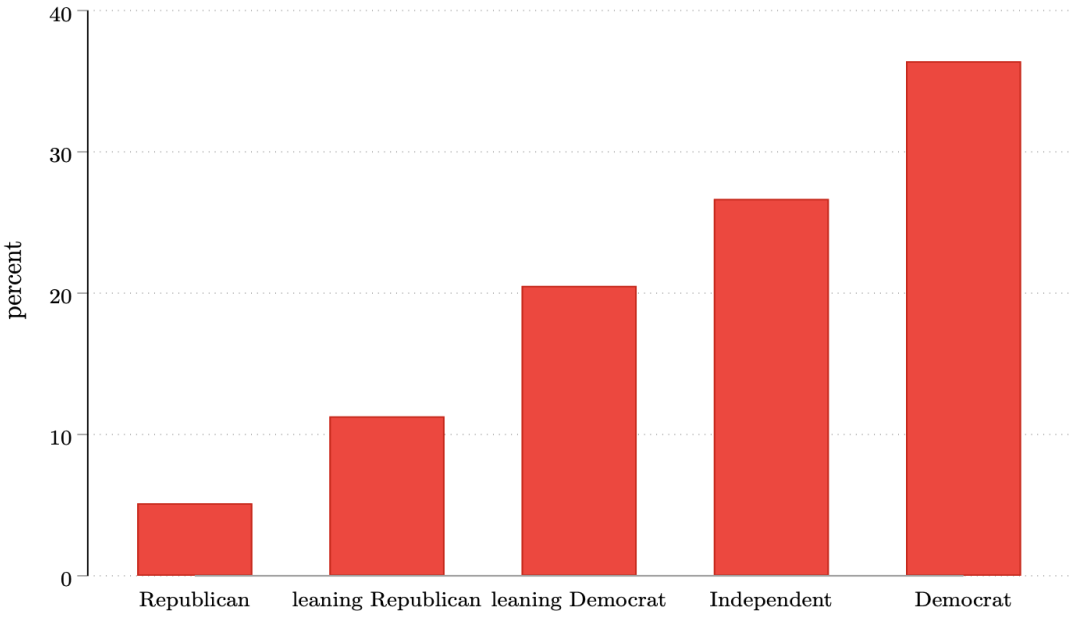

Study 1b | Time participant's spent on study
Fraction of participants
.4
.3
.2
.1
0
0
10
20
30
40
50
Number of minutes spent on study

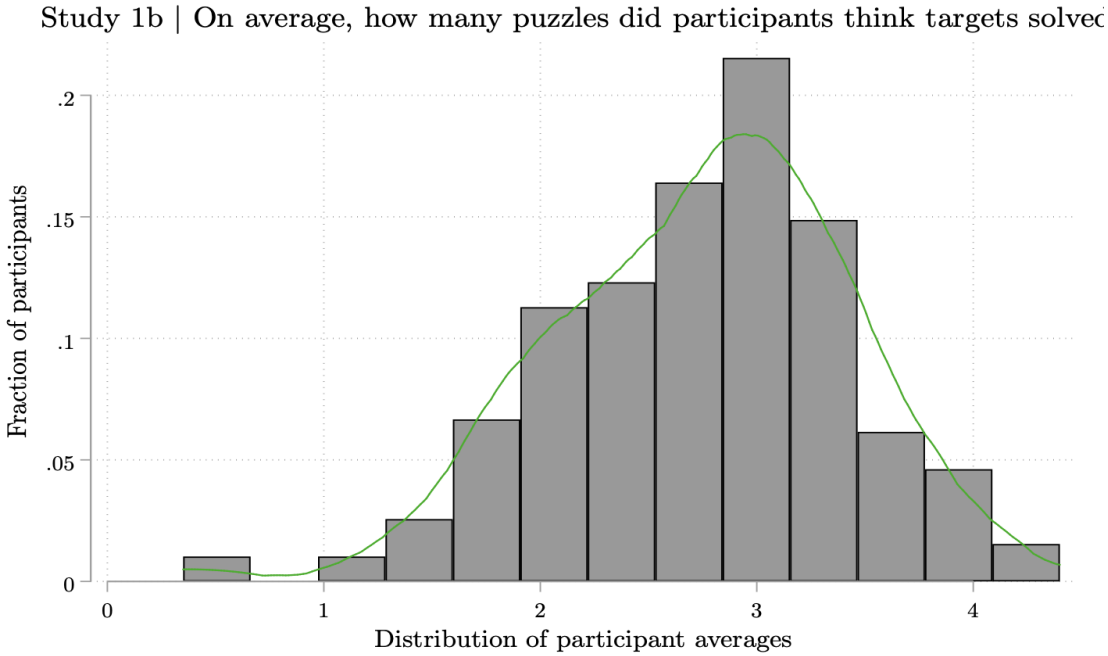
1b | Did time correlate with the average number of solved puzzles guessed?

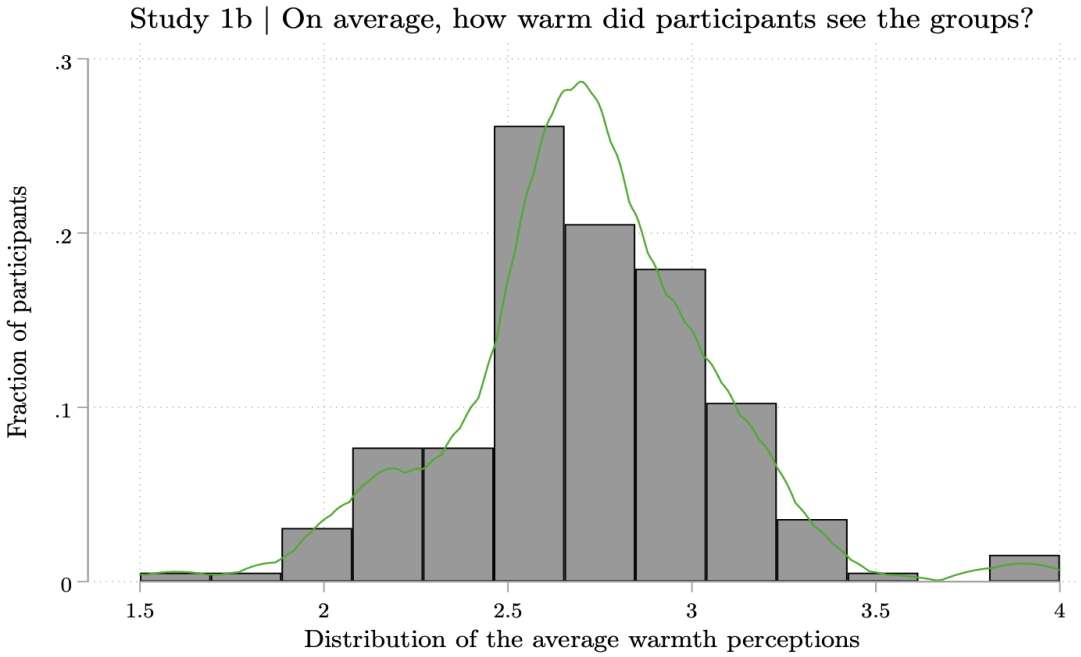


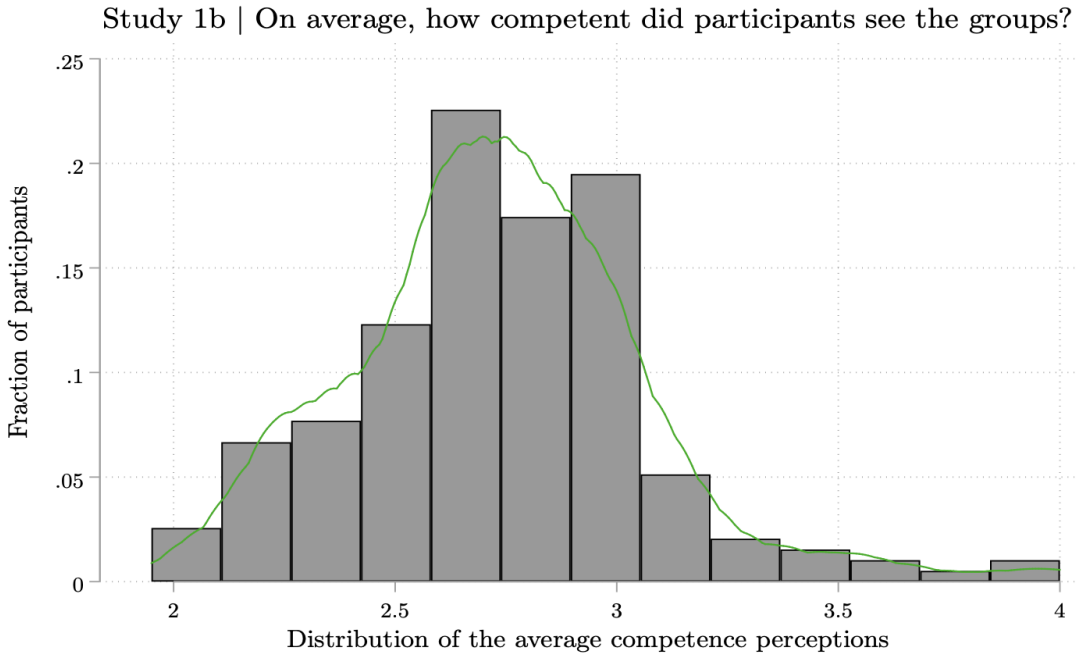
R-squared = 0.0001
Coefficient = -0.0013
Winzorized at 95 percent.

Study 1b | Participants' politics




Study 1b | On average, how many puzzles did participants think targets solvec
Fraction of participants
.2
.15
.1
.05
0
0
1
2
3
4
Distribution of participant averages


Study 1b | On average, how warm did participants see the groups?
Fraction of participants
.3
.2
.1
0
1.5
2
2.5
3
3.5
4
Distribution of the average warmth perceptions


Study 1b | On average, how competent did participants see the groups?
Fraction of participants
.25
.2
.15
.1
.05
0
2
2.5
3
3.5
4
Distribution of the average competence perceptions

Study 2a

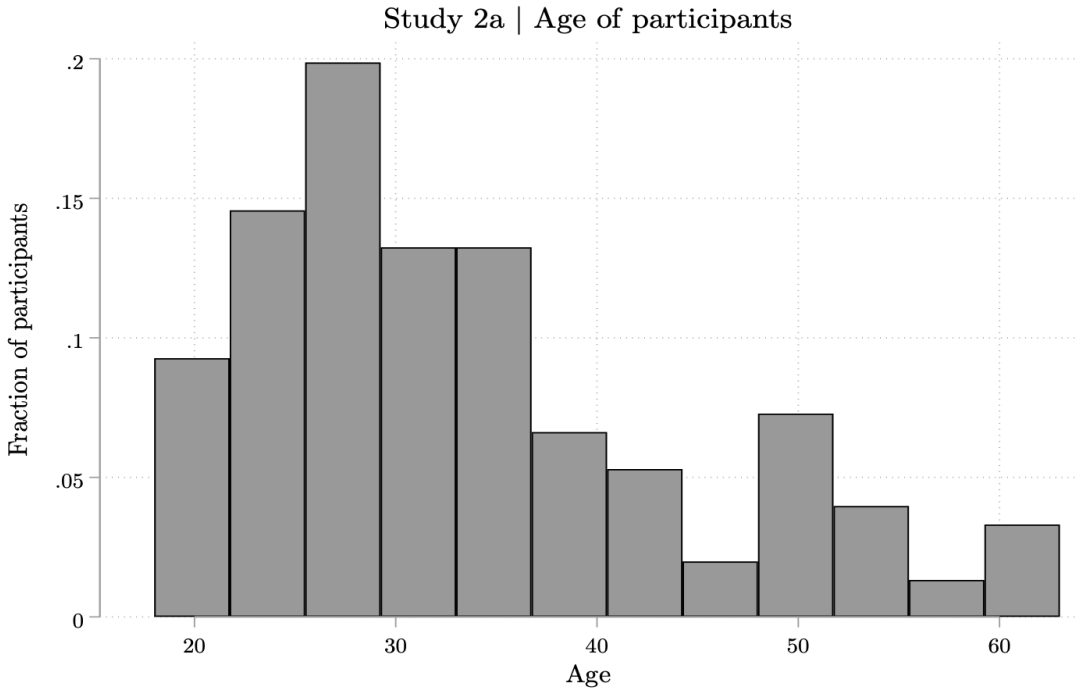
We conducted the study with an online sample of Americans (N=151), using the Prolific platform. Participants played the game using an online instrument that was designed on Qualtrics. The study took place in April 2020. The survey had two modules, Module A and Module B. In this study, the modules were completed 2 weeks apart. We randomized the order of the modules so that some people took Module A first, while others took Module B first. In both modules, participants were informed that they would have an opportunity to earn money in the study. In Module A, participants were invited to play a trust game with 12 different individuals (who had pre-recorded their responses). In the trust game, participants were given an endowment of $5 and told that they can share as much of them as they like with the other player. The amount that they share would be tripled. The other player can decide to give back as many dollars as they like. In each case, they were given stereotypic information about the person (e.g., Asian, Doctor, African etc.). They were informed that the earnings would be based on one randomly selected trust game. In Module B, participants were asked to rate the 12 stereotypic groups on the dimensions of warmth, competence, political affiliation, and status. One stereotypic category was selected at random to determine rewards. Participants' guesses were compared against the empirical average of the study sample's guesses in relation to the stereotypic category. Participants were rewarded $1 for answers on a four-point scale and $5 for answers that were on 5-point or more scale. Participants were also asked to provide demographic information (gender, age, political affiliation, self-rated status).

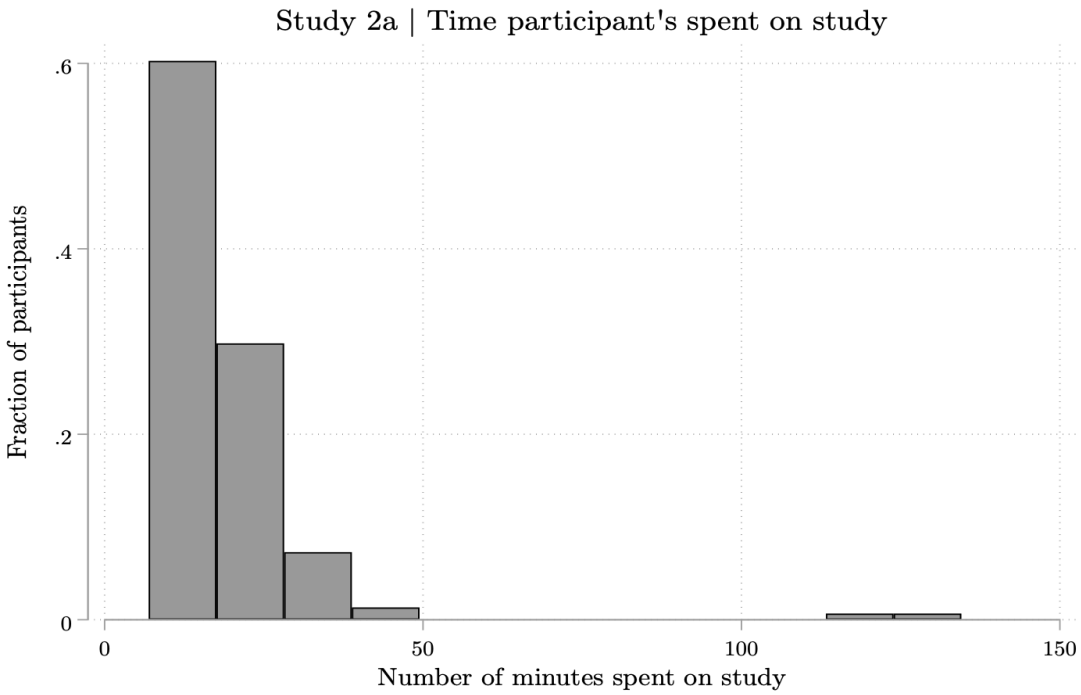
Summary of differences from Study 1a: (i) modules are 2 weeks apart; (ii) cash payment rather than tokens for lottery; (iii) smaller number of players with more questions on stereotype content and demographics.

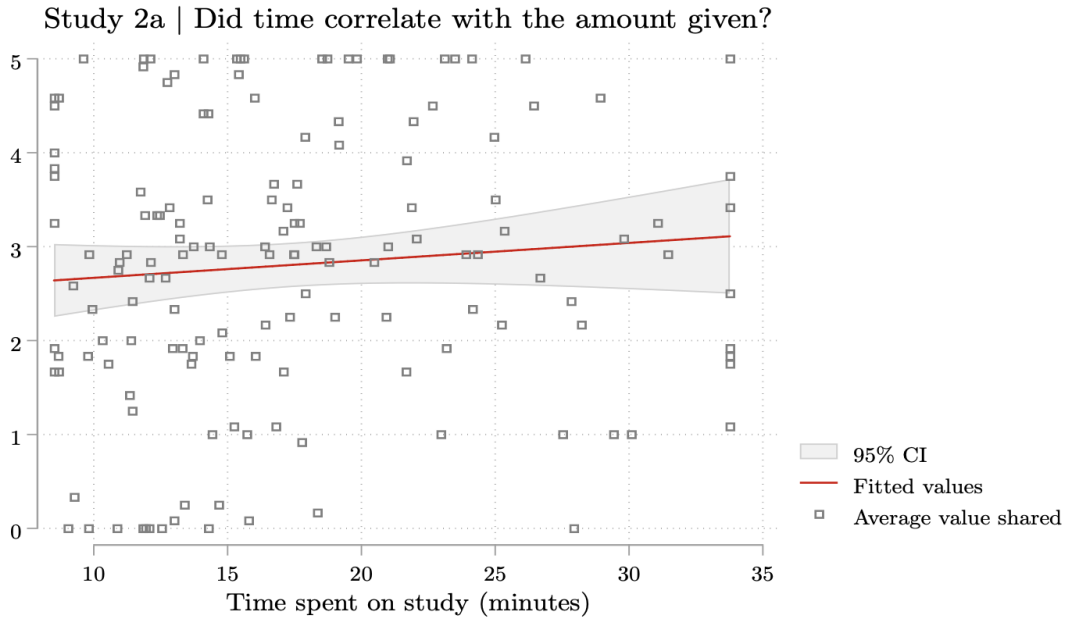
Table 2.15: Study 2a: Summary statistics and balance check on order

Variable	(1) Survey first N	(1) Survey first Mean/SE	(2) Game first N	(2) Game first Mean/SE	(3) Total N	(3) Total Mean/SE	T-test P-value (1)-(2)
Age	74	33.824	77	33.390	151	33.603	0.808
		(1.274)		(1.253)		(0.891)	
Male proportion of sample	74	0.392	77	0.364	151	0.377	0.722
		(0.057)		(0.055)		(0.040)	
Time spent on study (minutes)	74	19.270	77	18.397	151	18.825	0.713
		(1.775)		(1.572)		(1.180)	
Average value shared	74	2.842	77	2.772	151	2.806	0.769
		(0.168)		(0.171)		(0.120)	
Played warmth rating	74	5.142	77	5.026	151	5.083	0.314
		(0.081)		(0.082)		(0.057)	
Played competence rating	74	5.123	77	4.998	151	5.059	0.234
		(0.076)		(0.072)		(0.052)	

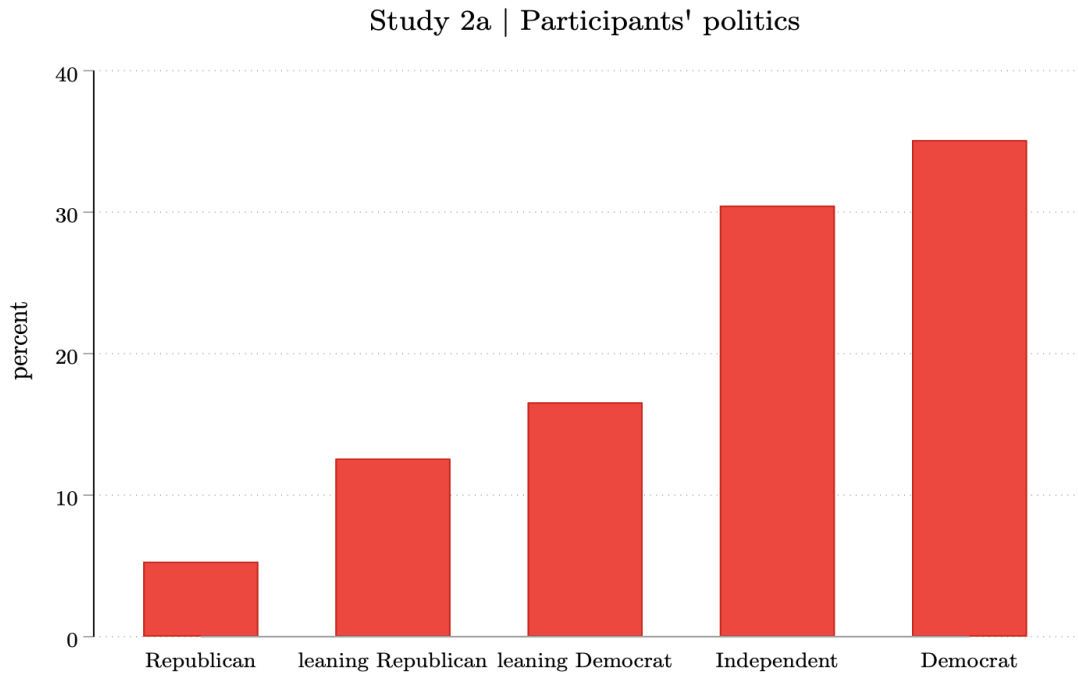
Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

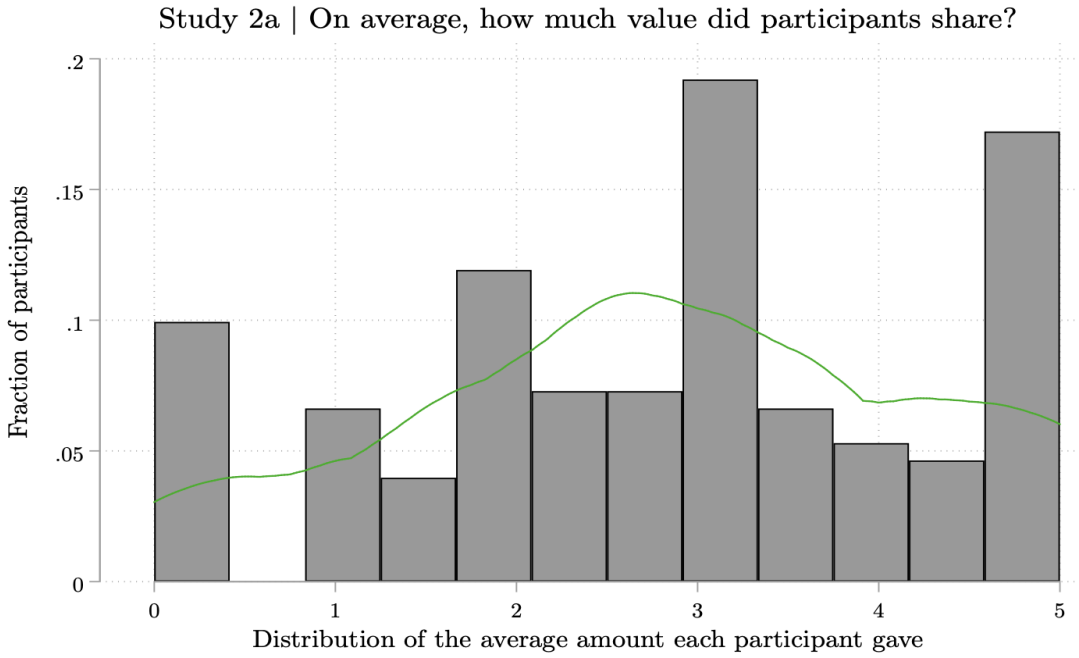


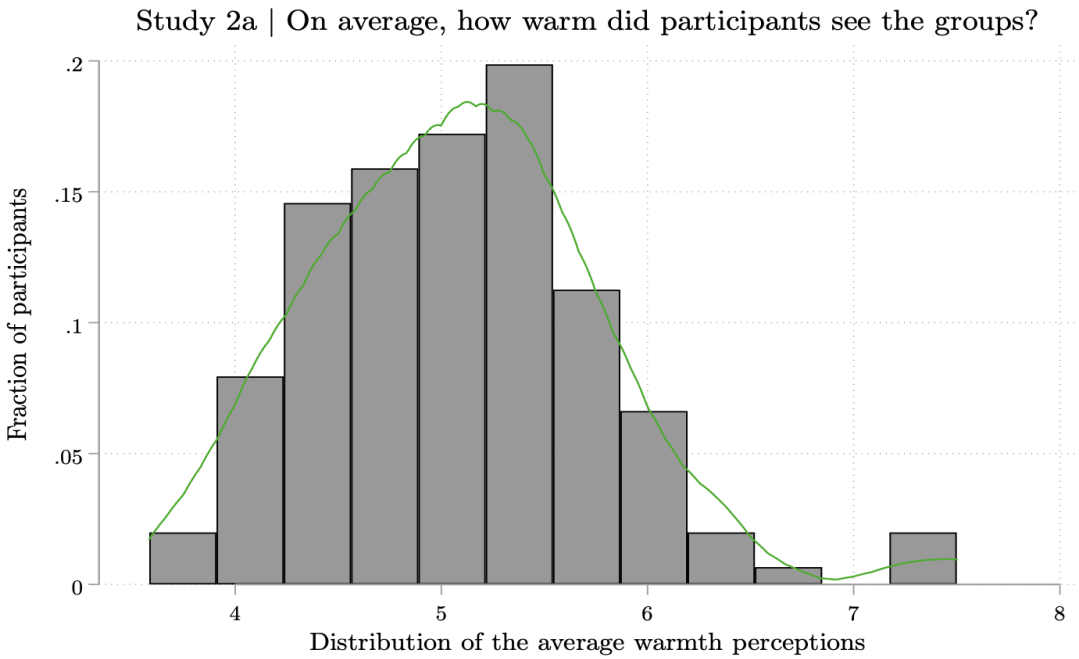


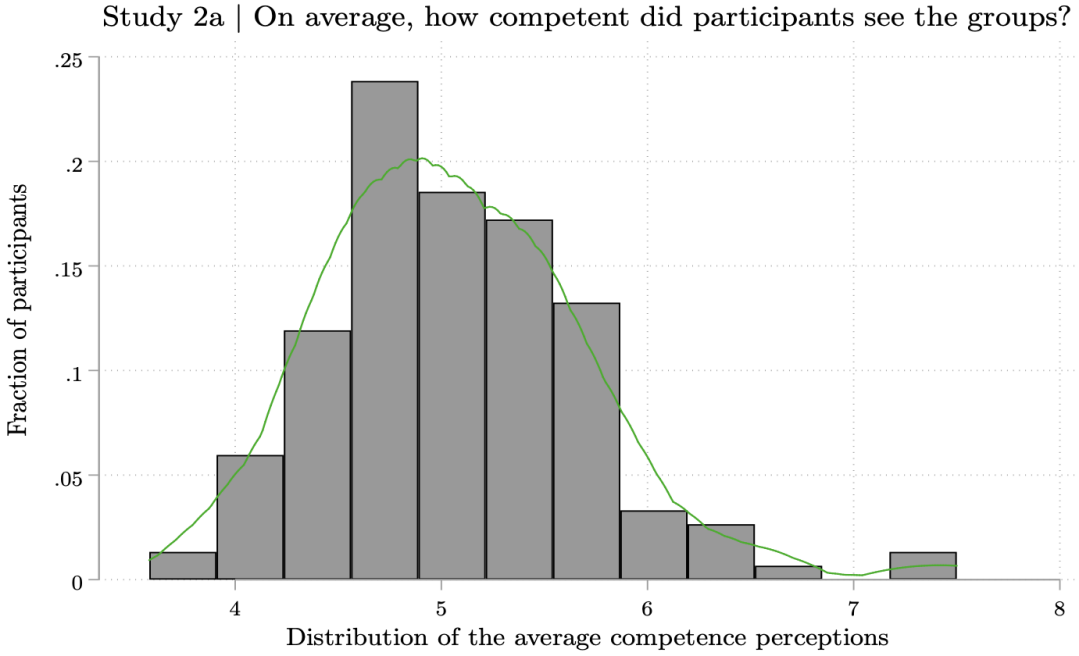


R-squared = 0.0076
Coefficient = 0.0186
Winzorized at 95 percent.


Study 2a | Participants' politics
percent
40
30
20
10
0
Republican
leaning Republican
leaning Democrat
Independent
Democrat


Study 2a | On average, how much value did participants share?
Fraction of participants
.2
.15
.1
.05
0
0
1
2
3
4
5
Distribution of the average amount each participant gave


Study 2a | On average, how warm did participants see the groups?
Fraction of participants
.2
.15
.1
.05
0
4
5
6
7
8
Distribution of the average warmth perceptions


Study 2a | On average, how competent did participants see the groups?
Fraction of participants
.25
.2
.15
.1
.05
0
4
5
6
7
8
Distribution of the average competence perceptions

Study 2b

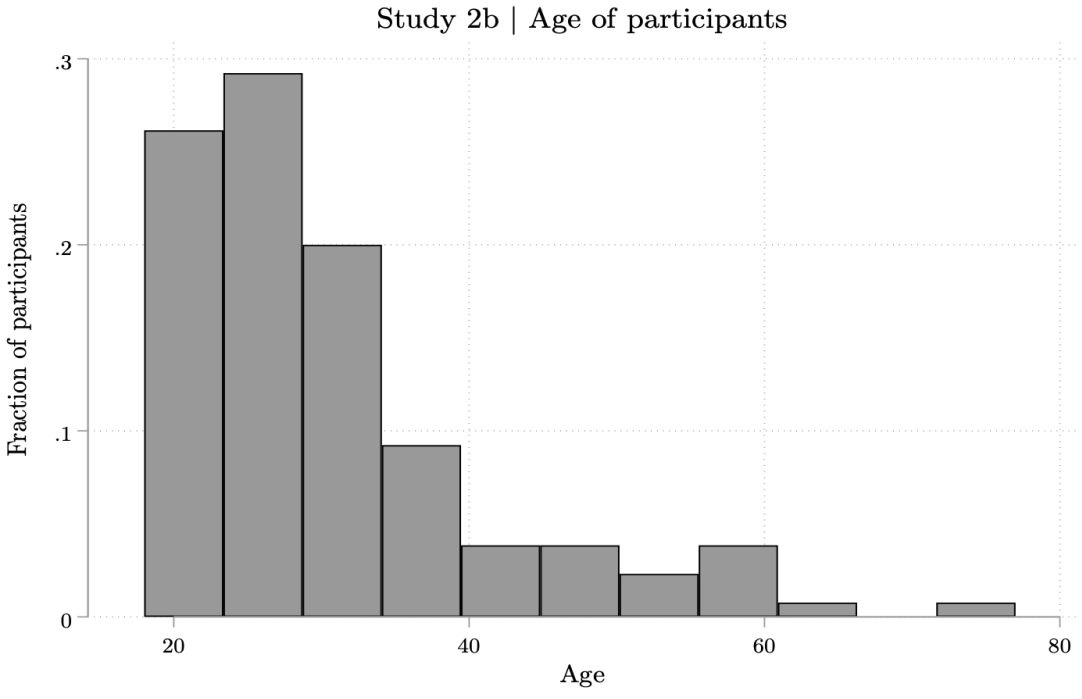
We conducted the study with an online sample of Americans (N=130), using the Prolific platform. Participants played the game using an online instrument that was designed on Qualtrics. The study took place in July 2020. The survey had two modules, Module A and Module B. The modules were completed 2 weeks apart. We randomized the order of the modules so that some people took Module A first, while others took Module B first. In both modules, participants were informed that they would have an opportunity to earn money in the study. In Module A, participants were shown a number of puzzles. They were then invited to guess how many of the puzzles different individuals were able to solve. In each case, they were given stereotypic information about the person (e.g., Asian, Doctor, African etc.). If they guessed correctly, they were be rewarded with a $1.25. In Module B, participants were asked to rate the 12 stereotypic groups on the dimensions of warmth, competence, political affiliation, and status. One stereotypic category was selected at random to determine rewards. Participants' guesses were compared against the empirical average of the study sample's guesses in relation to the stereotypic category. Participants were rewarded $1 for answers on a four-point scale and $5 for answers that were on 5-point or more scale. Participants were also asked to provide demographic information (gender, age, political affiliation, self-rated status).

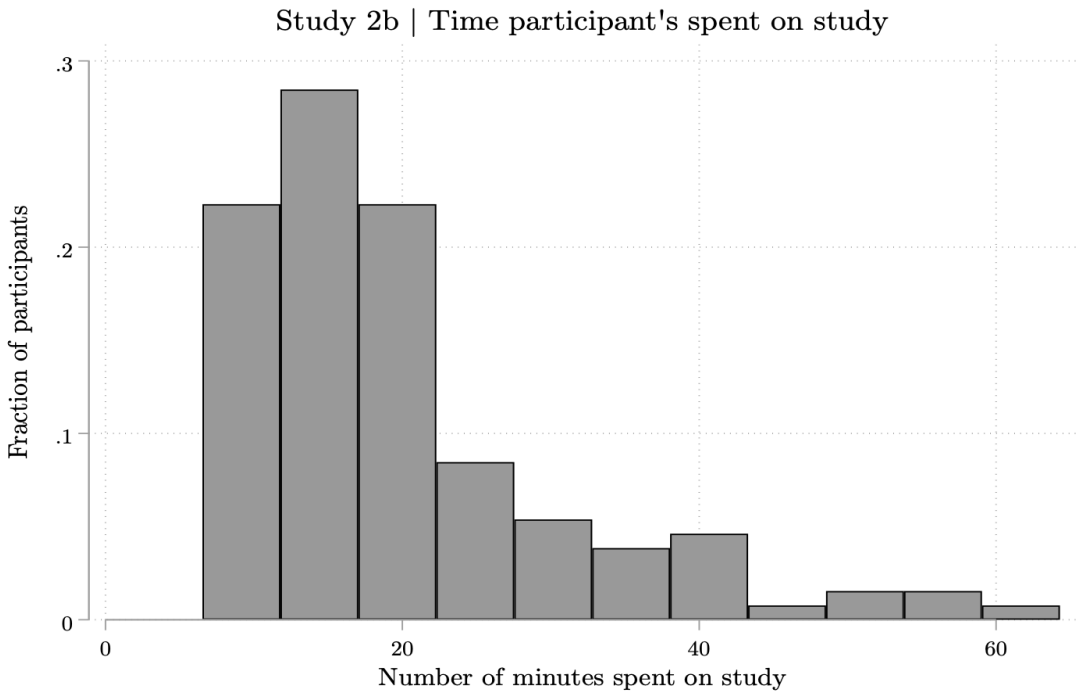
Summary of differences from Study 1b: (i) modules are 2 weeks apart; (ii) cash payment rather than tokens for lottery; (iii) smaller number of players with more questions on stereotype content and demographics.

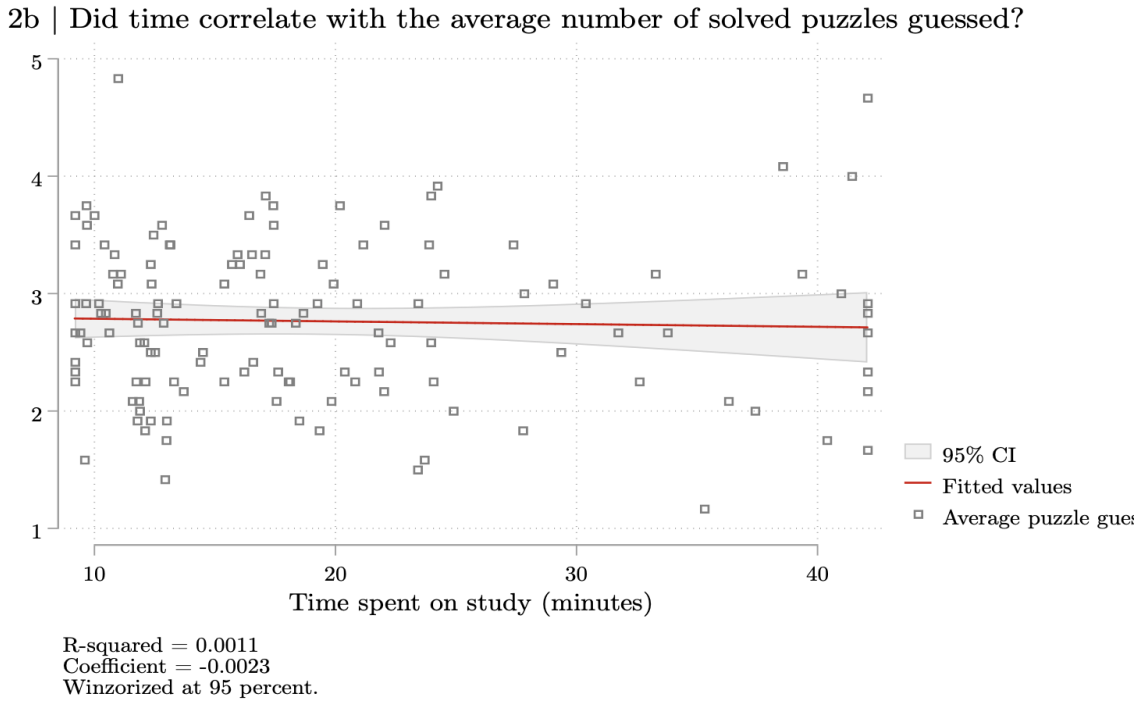
TABLE 2.16: Study 2b: Summary statistics and balance check on order

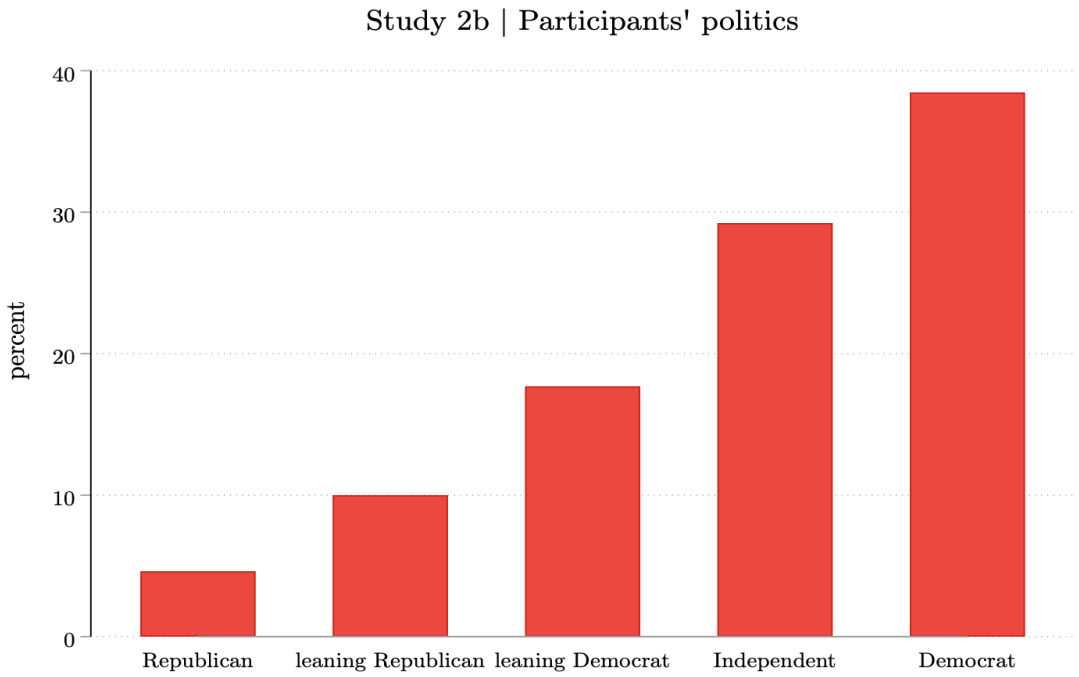
	(1) Survey first		(2) Game first		(3) Total		T-test P-value
Variable	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)
Age	69	30.623 (1.343)	61	30.459 (1.407)	130	30.546 (0.968)	0.933
Male proportion of sample	69	0.464 (0.060)	61	0.443 (0.064)	130	0.454 (0.044)	0.811
Time spent on study (minutes)	69	20.226 (1.293)	61	19.395 (1.541)	130	19.836 (0.994)	0.678
Average puzzle guess	69	2.862 (0.085)	61	2.654 (0.079)	130	2.765 (0.059)	0.077*
Played warmth rating	69	5.054 (0.073)	61	5.007 (0.080)	130	5.032 (0.054)	0.662
Played competence rating	69	4.971 (0.061)	61	5.015 (0.078)	130	4.992 (0.049)	0.653

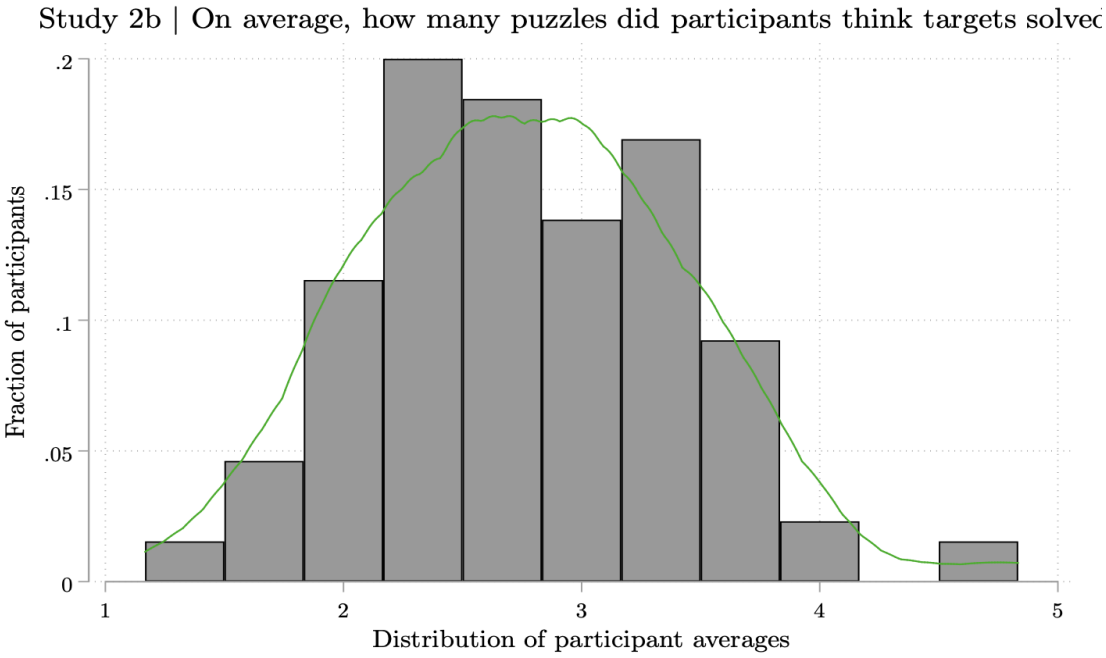
Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

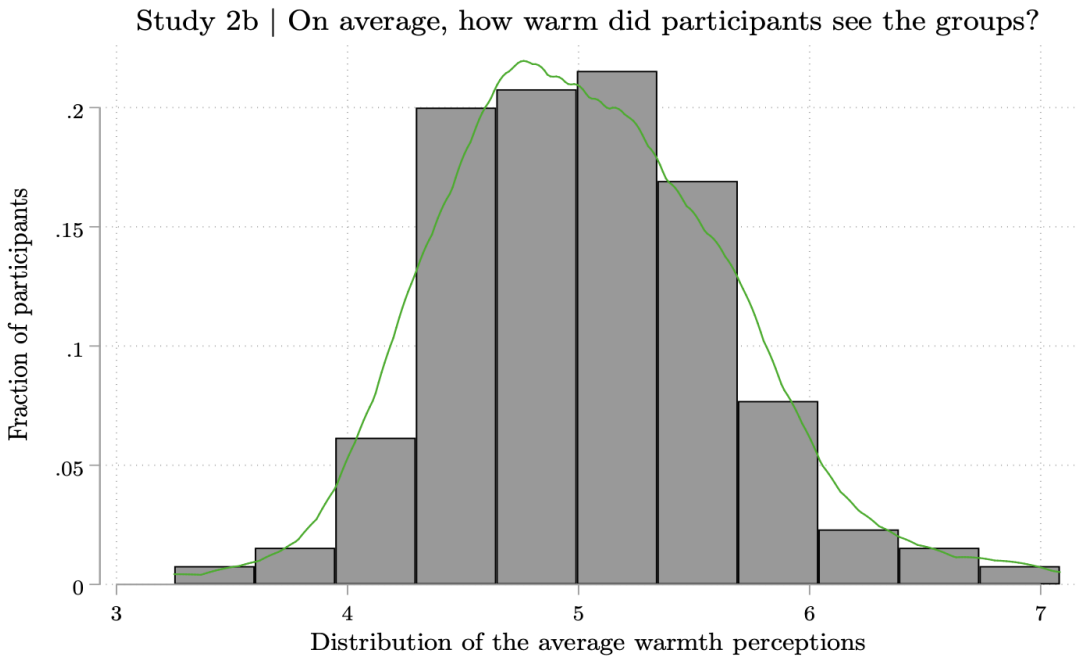


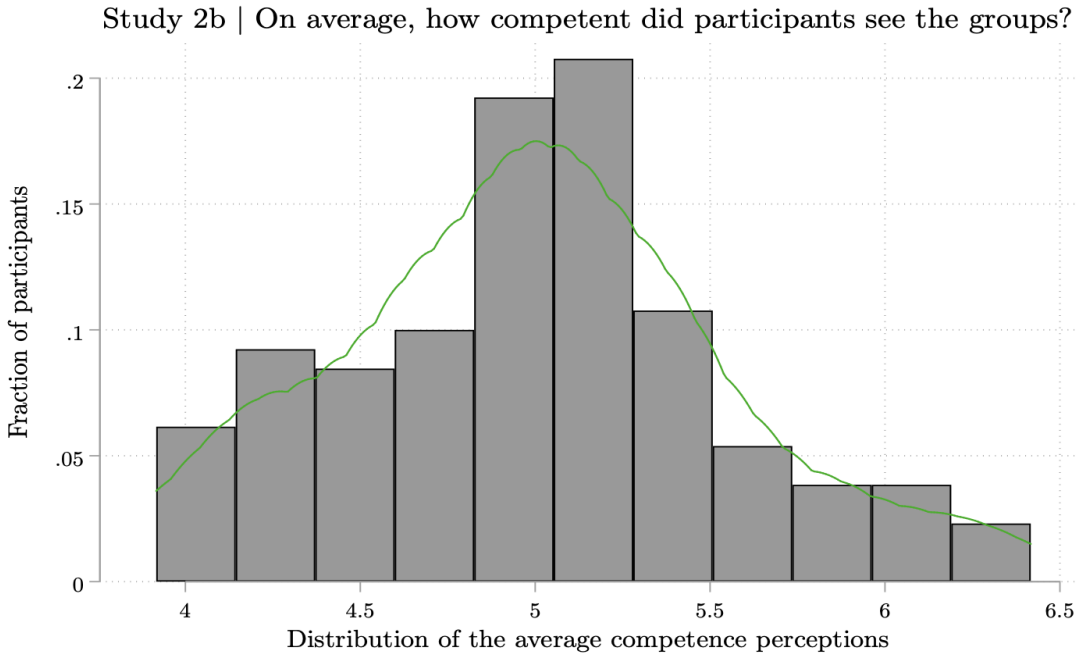

Study 2b | Time participant's spent on study
Fraction of participants
.3
.2
.1
0
0
20
40
60
Number of minutes spent on study


2b | Did time correlate with the average number of solved puzzles guessed?
5
4
3
2
1
10
20
30
40
Time spent on study (minutes)
95% CI
Fitted values
Average puzzle gues
R-squared = 0.0011
Coefficient = -0.0023
Winzorized at 95 percent.


Study 2b | Participants' politics
percent
40
30
20
10
0
Republican
leaning Republican
leaning Democrat
Independent
Democrat


Study 2b | On average, how many puzzles did participants think targets solved
Fraction of participants
.2
.15
.1
.05
0
1
2
3
4
5
Distribution of participant averages


Study 2b | On average, how warm did participants see the groups?
Fraction of participants
0
.05
.1
.15
.2
3
4
5
6
7
Distribution of the average warmth perceptions


Study 2b | On average, how competent did participants see the groups?
Fraction of participants
0
.05
.1
.15
.2
4
4.5
5
5.5
6
6.5
Distribution of the average competence perceptions

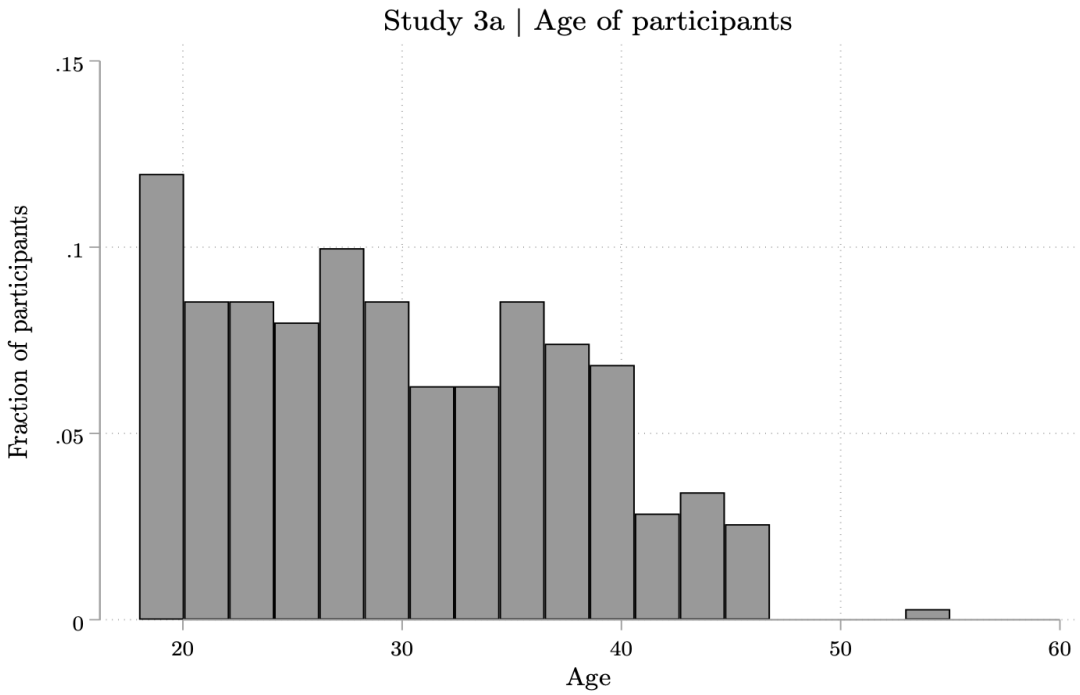
Study 3a

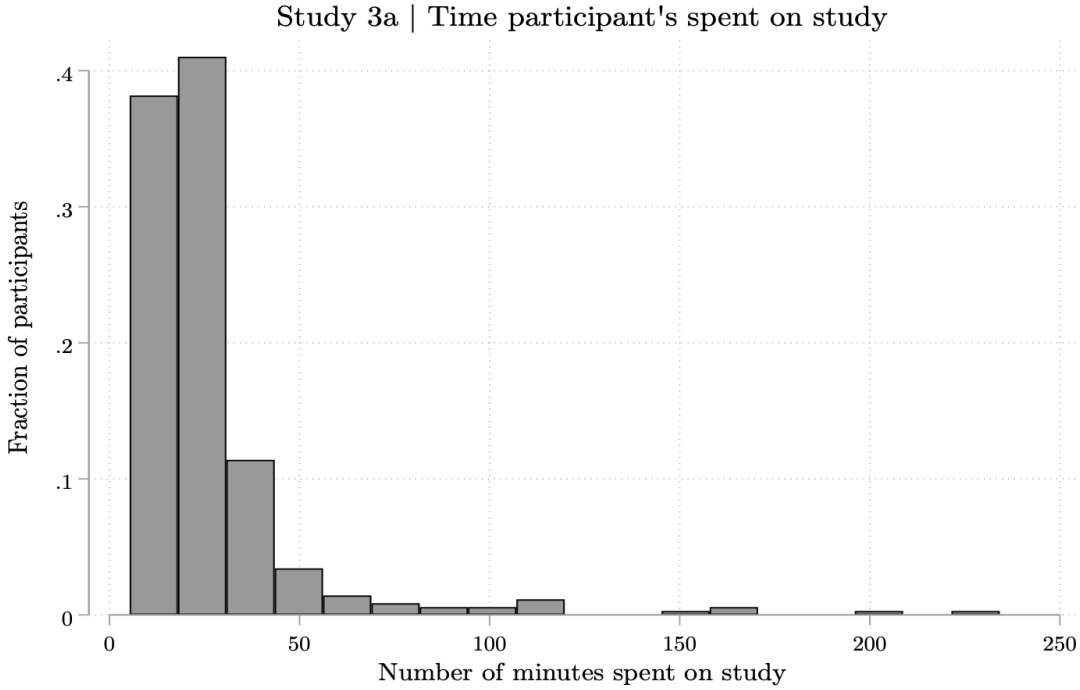
We conducted the study with an online sample of Indians (N=213), using a panel from IPSOS. Participants played the game using an online instrument that was designed on Qualtrics. The study took place in December 2020. The survey had two modules, Module A and Module B. The modules were completed 2 weeks apart. We randomized the order of the modules so that some people took Module A first, while others will take Module B first. In both modules, participants were informed that they have an opportunity to earn points for an online panel in the study. In Module A, participants were invited to play a trust game with 12 different individuals. In the trust game, participants will be given an endowment of 750 points for an online and told that they can share as much of them as they like with the other player. The amount that they share will be tripled. The other player can decide to give back as many dollars as they like. In each case, they will be given stereotypic information about the person (e.g., retired, unemployed, a Brahmin etc.). They were then informed that the earnings would be based on one randomly selected game. In Module B, participants were asked to rate the 12 stereotypic groups on the dimensions of warmth, competence, political affiliation, and status. If their guesses aligned with the empirical average of the study sample, they were rewarded with points.

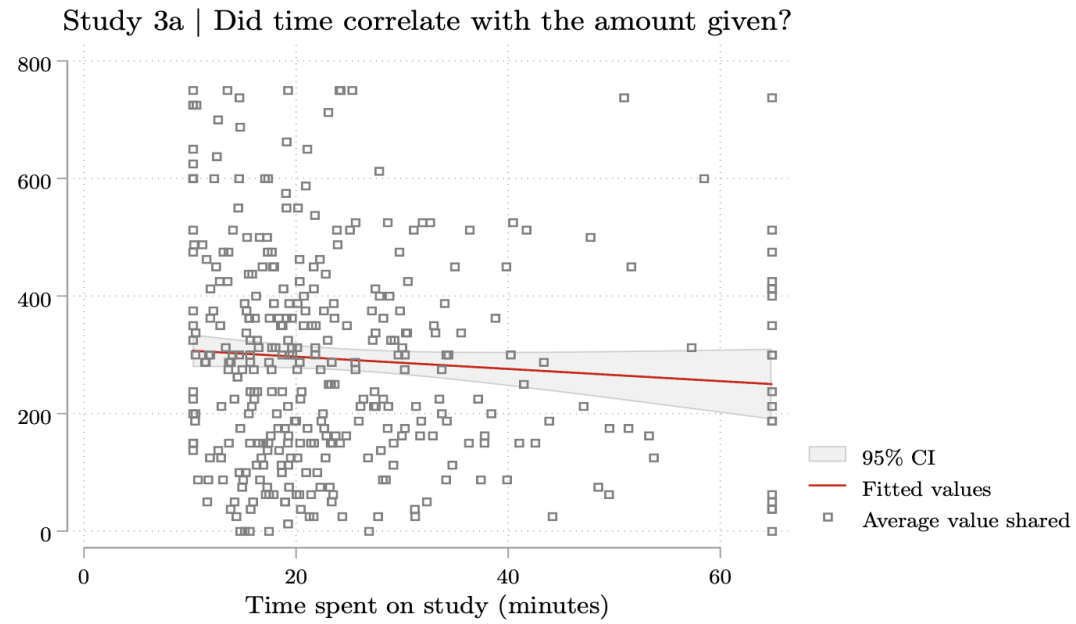
Table 2.17: Study 3a: Summary statistics and balance check on order

	(1) Survey first		(2) Game first		(3) Total		T-test P-value
Variable	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)
Age	183	30.251 (0.571)	168	29.935 (0.576)	351	30.100 (0.405)	0.697
Male proportion of sample	183	0.563 (0.037)	168	0.512 (0.039)	351	0.538 (0.027)	0.340
Time spent on study (minutes)	183	27.248 (1.953)	168	26.683 (1.845)	351	26.978 (1.346)	0.834
Average value shared	183	287.295 (12.935)	168	297.321 (14.690)	351	292.094 (9.732)	0.608
Played warmth rating	183	5.888 (0.071)	168	5.892 (0.066)	351	5.890 (0.049)	0.960
Played competence rating	183	5.700 (0.074)	168	5.675 (0.065)	351	5.688 (0.049)	0.802

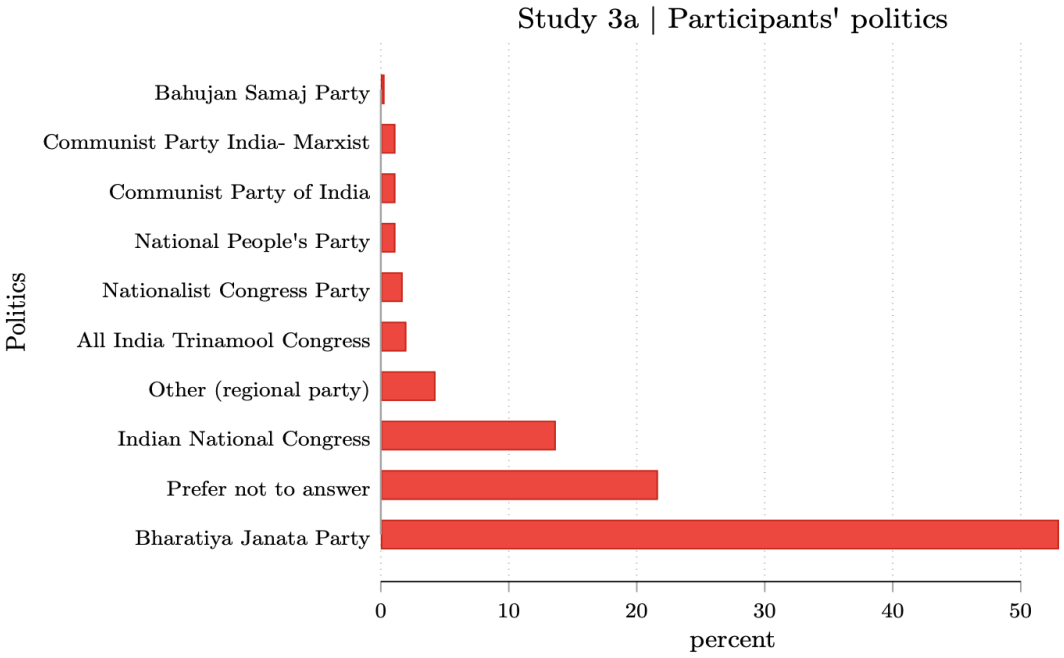
Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

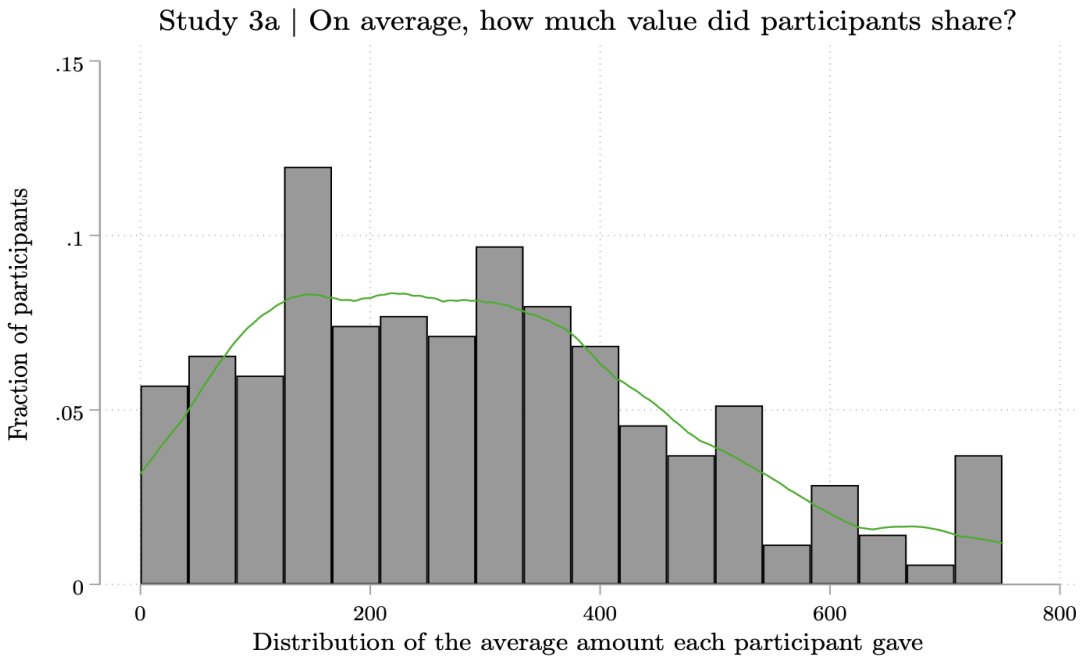

Study 3a | Age of participants
Fraction of participants
.15
.1
.05
0
20
30
40
50
60
Age

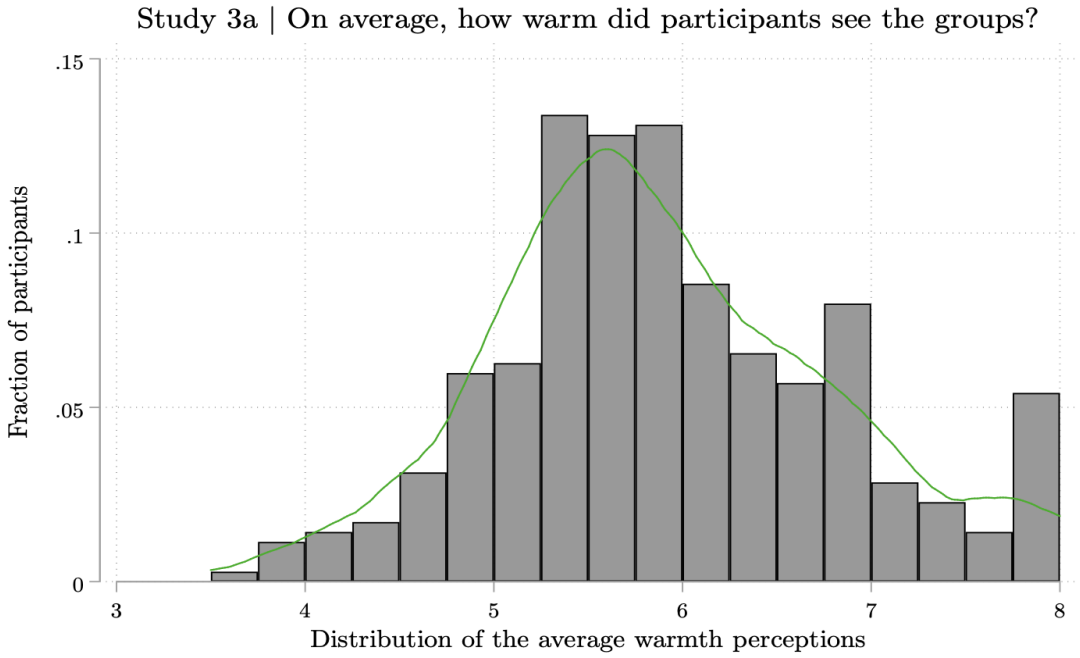

Study 3a | Time participant's spent on study
Fraction of participants
.4
.3
.2
.1
0
0
50
100
150
200
250
Number of minutes spent on study

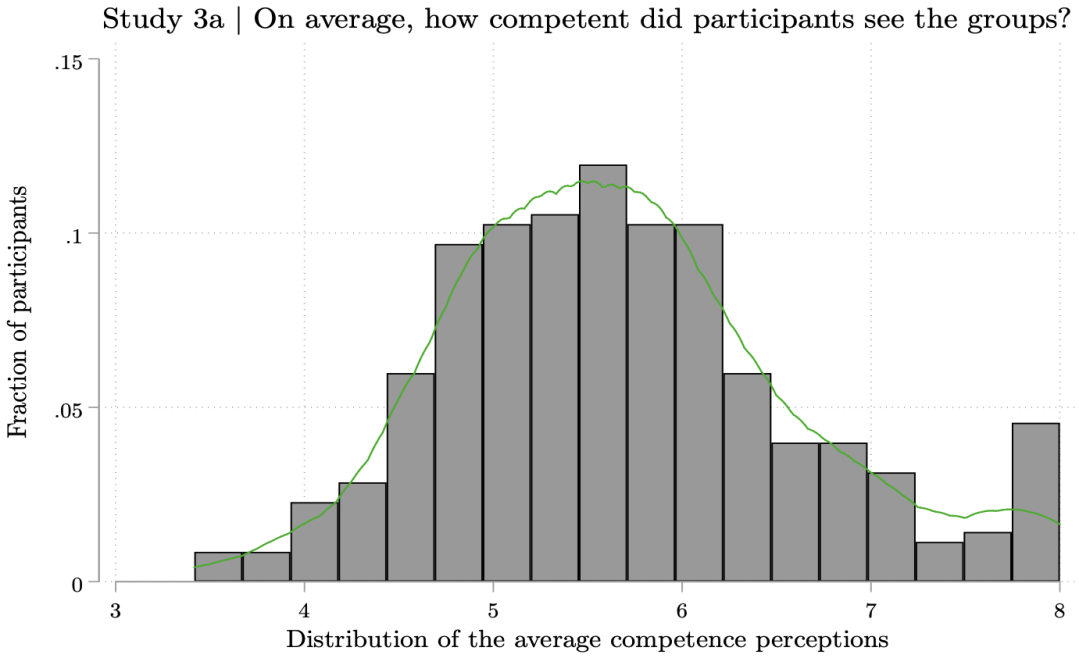


R-squared = 0.0059
Coefficient = -1.0388
Winzorized at 95 percent.




Study 3a | On average, how much value did participants share?
Fraction of participants
.15
.1
.05
0
0
200
400
600
800
Distribution of the average amount each participant gave


Study 3a | On average, how warm did participants see the groups?
Fraction of participants
.15
.1
.05
0
3
4
5
6
7
8
Distribution of the average warmth perceptions


Study 3a | On average, how competent did participants see the groups?
Fraction of participants
.15
.1
.05
0
3
4
5
6
7
8
Distribution of the average competence perceptions

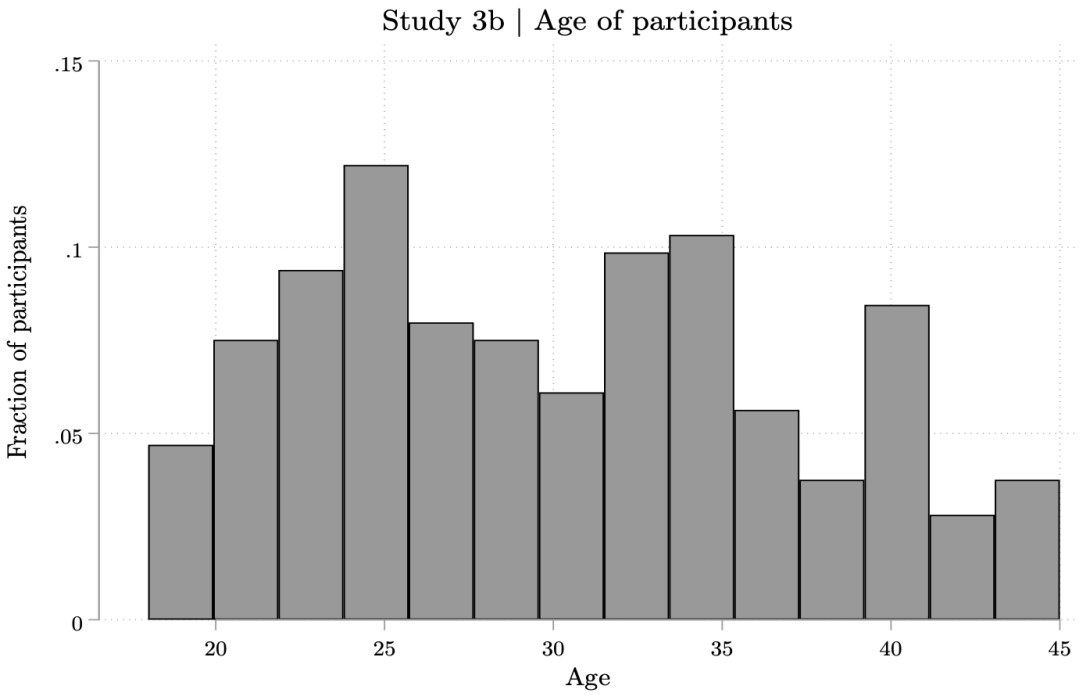
Study 3b

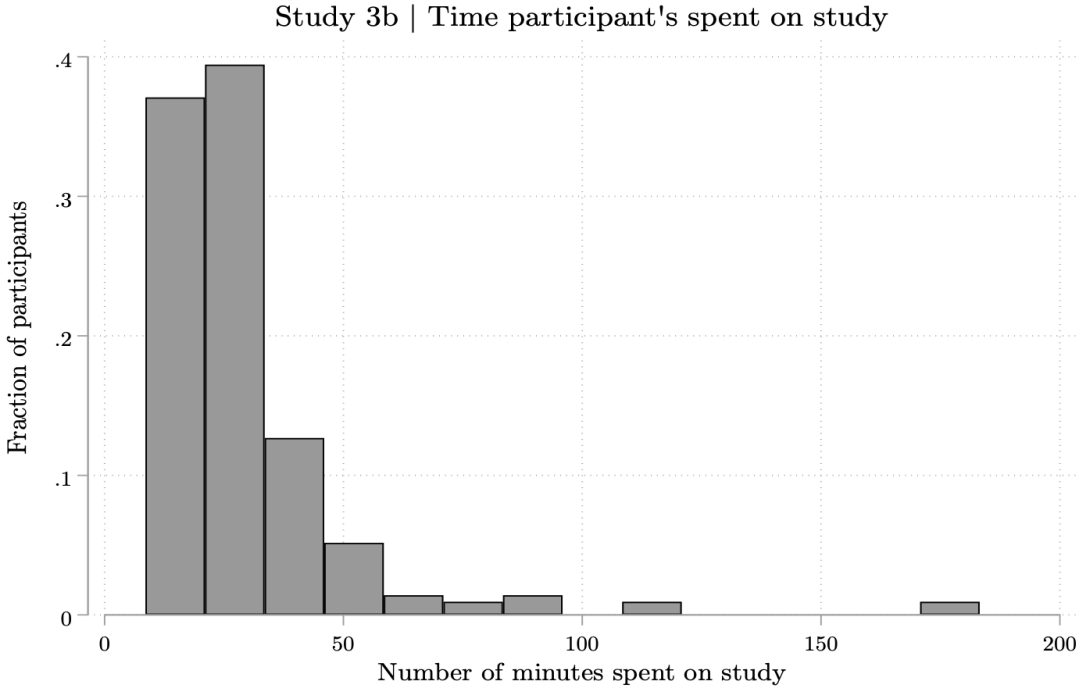
We conducted the study with an online sample of Indians (N=213), using a panel from IPSOS. Participants played the game using an online instrument that was designed on Qualtrics. The study took place in March-April 2021. The survey had two modules, Module A and Module B. The modules were completed 2 weeks apart. We randomized the order of the modules so that some people took Module A first, while others will take Module B first. In both modules, participants were informed that they have an opportunity to earn points for an online panel in the study. In Module A, participants were invited to play the puzzlebet game with 12 different individuals. In this game, participants were shown a number of puzzles. They were then invited to guess how many of the puzzles different individuals were able to solve. In each case, they will be given stereotypic information about the person (e.g., retired, unemployed, a Brahmin etc.). They were then informed that the earnings would be based on one randomly selected game. In Module B, participants were asked to rate the 12 stereotypic groups on the dimensions of warmth, competence, ideology, and status. If their guesses aligned with the empirical average of the study sample, they were rewarded with points.

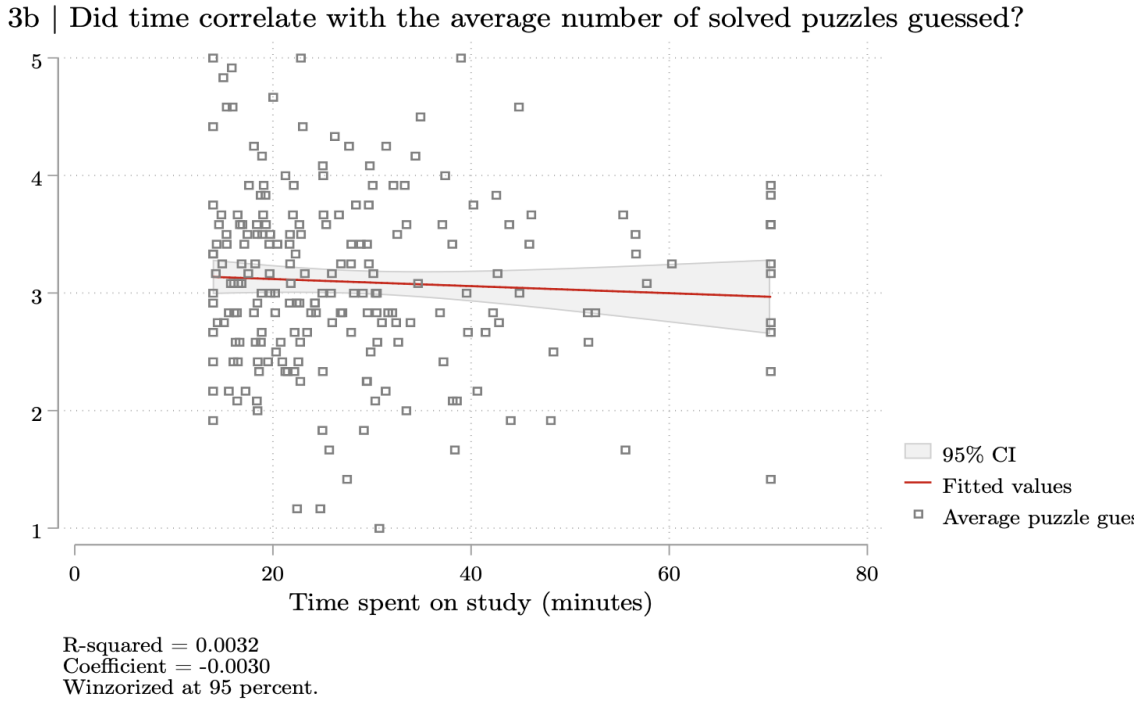
Table 2.18: Study 3b: Summary statistics and balance check on order

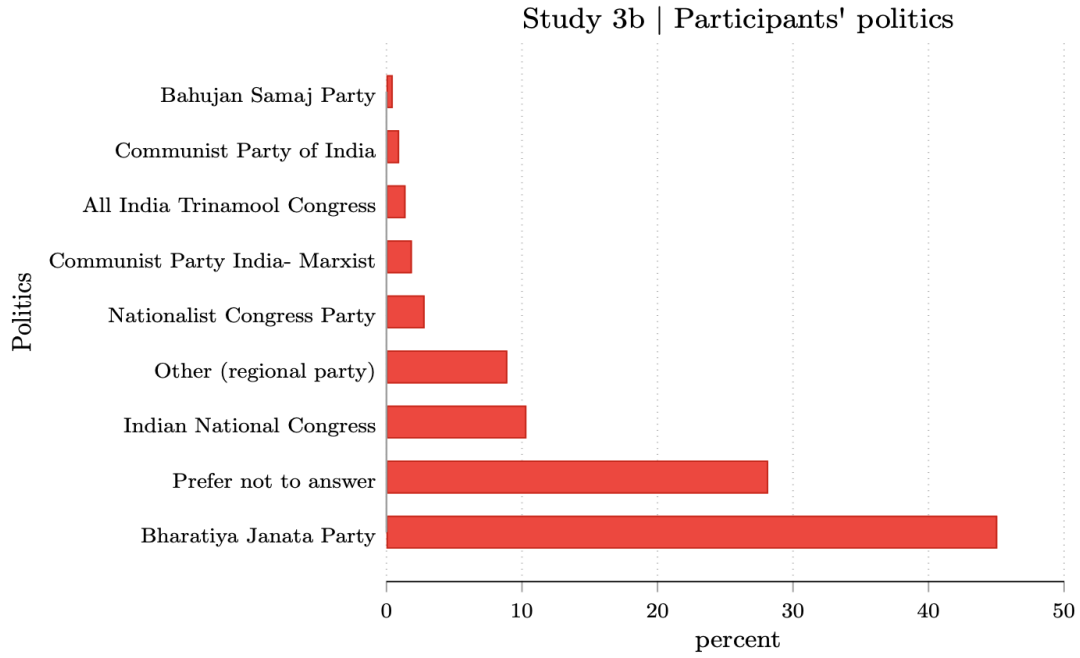
	(1) Survey first		(2) Game first		(3) Total		T-test P-value
Variable	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)
Age	104	29.894 (0.736)	109	30.330 (0.662)	213	30.117 (0.493)	0.659
Male proportion of sample	104	0.510 (0.049)	109	0.413 (0.047)	213	0.460 (0.034)	0.158
Time spent on study (minutes)	104	27.149 (1.442)	109	33.030 (2.608)	213	30.158 (1.519)	0.053*
Average puzzle guess	104	3.067 (0.067)	109	3.119 (0.078)	213	3.094 (0.052)	0.611
Played warmth rating	104	5.726 (0.076)	109	5.749 (0.070)	213	5.738 (0.051)	0.821
Played competence rating	104	5.534 (0.082)	109	5.497 (0.070)	213	5.515 (0.054)	0.727

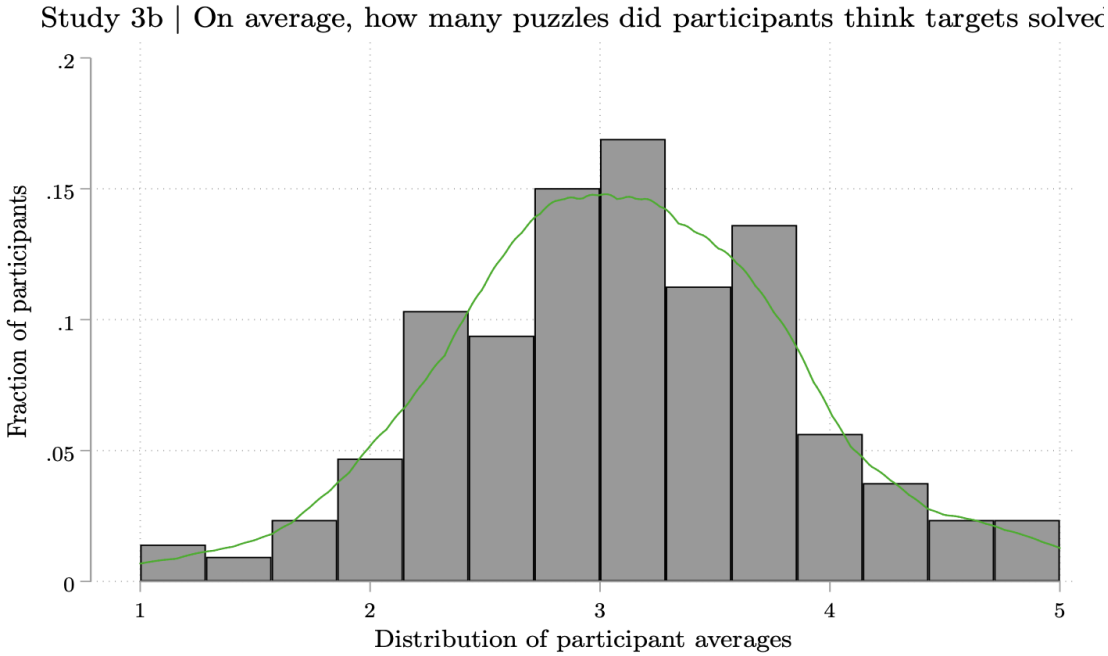
Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

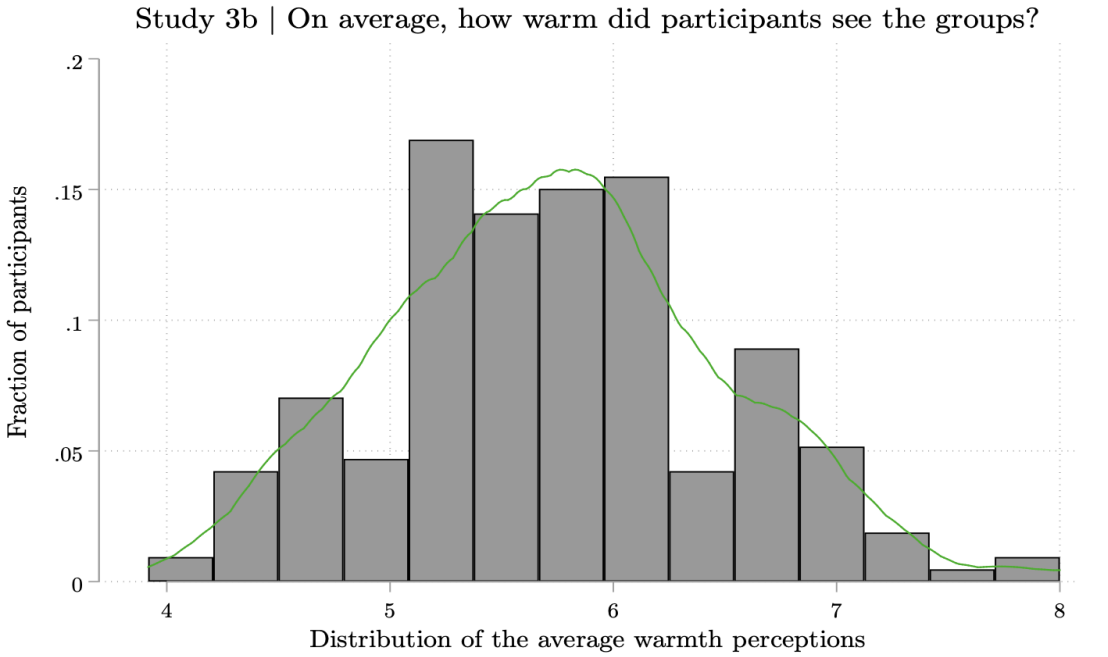

Study 3b | Age of participants
Fraction of participants
.15
.1
.05
0
20
25
30
35
40
45
Age

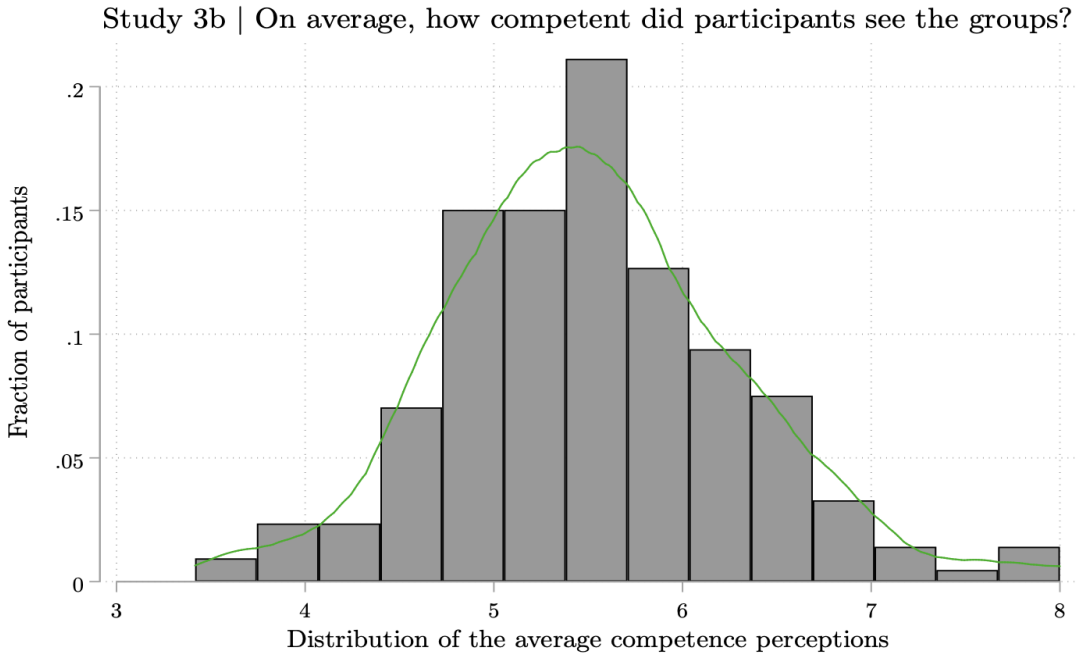

Study 3b | Time participant's spent on study
Fraction of participants
.4
.3
.2
.1
0
0
50
100
150
200
Number of minutes spent on study


3b | Did time correlate with the average number of solved puzzles guessed?
5
4
3
2
1
0
20
40
60
80
Time spent on study (minutes)
95% CI
Fitted values
Average puzzle gues
R-squared = 0.0032
Coefficient = -0.0030
Winzorized at 95 percent.


Study 3b | Participants' politics
Politics
Bahujan Samaj Party
Communist Party of India
All India Trinamool Congress
Communist Party India- Marxist
Nationalist Congress Party
Other (regional party)
Indian National Congress
Prefer not to answer
Bharatiya Janata Party
0
10
20
30
40
50
percent

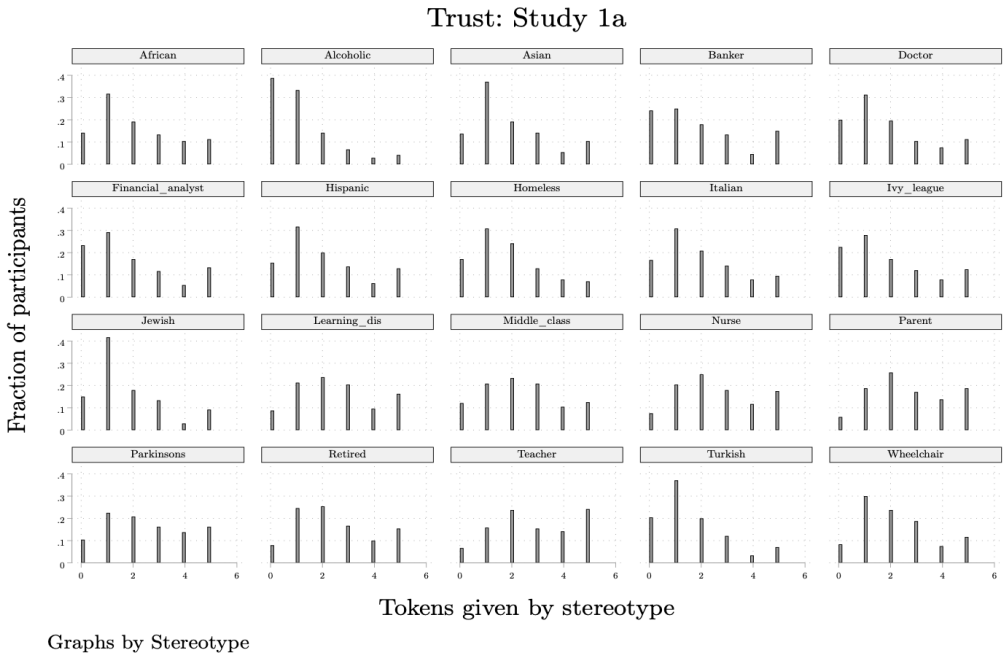

Study 3b | On average, how many puzzles did participants think targets solved
Fraction of participants
.2
.15
.1
.05
0
1
2
3
4
5
Distribution of participant averages

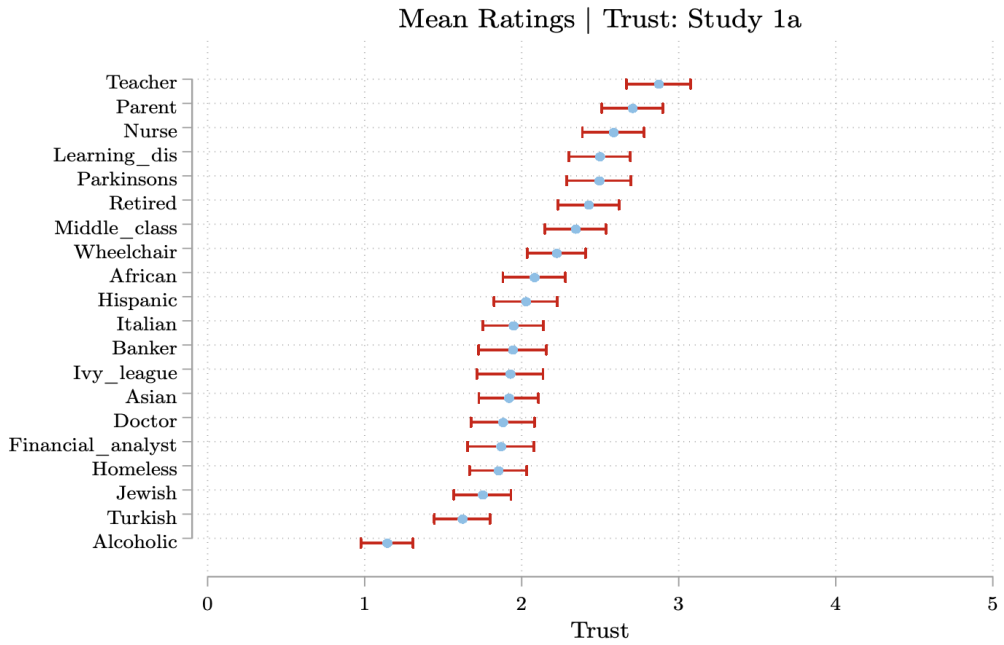

Study 3b | On average, how warm did participants see the groups?
Fraction of participants
.2
.15
.1
.05
0
4
5
6
7
8
Distribution of the average warmth perceptions

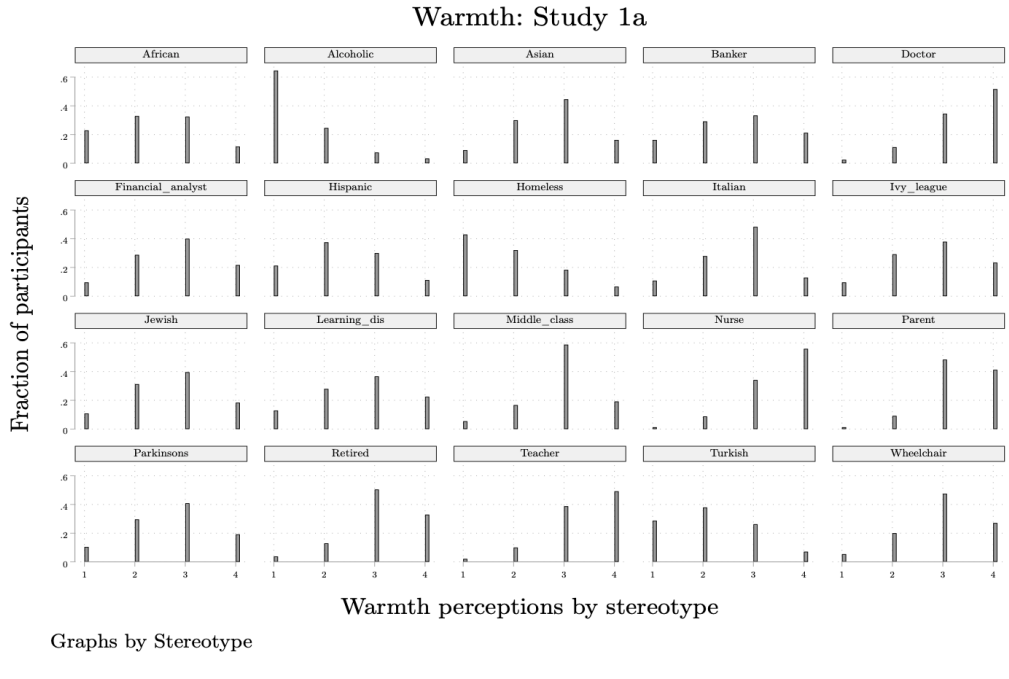

Study 3b | On average, how competent did participants see the groups?
Fraction of participants
.2
.15
.1
.05
0
3
4
5
6
7
8
Distribution of the average competence perceptions

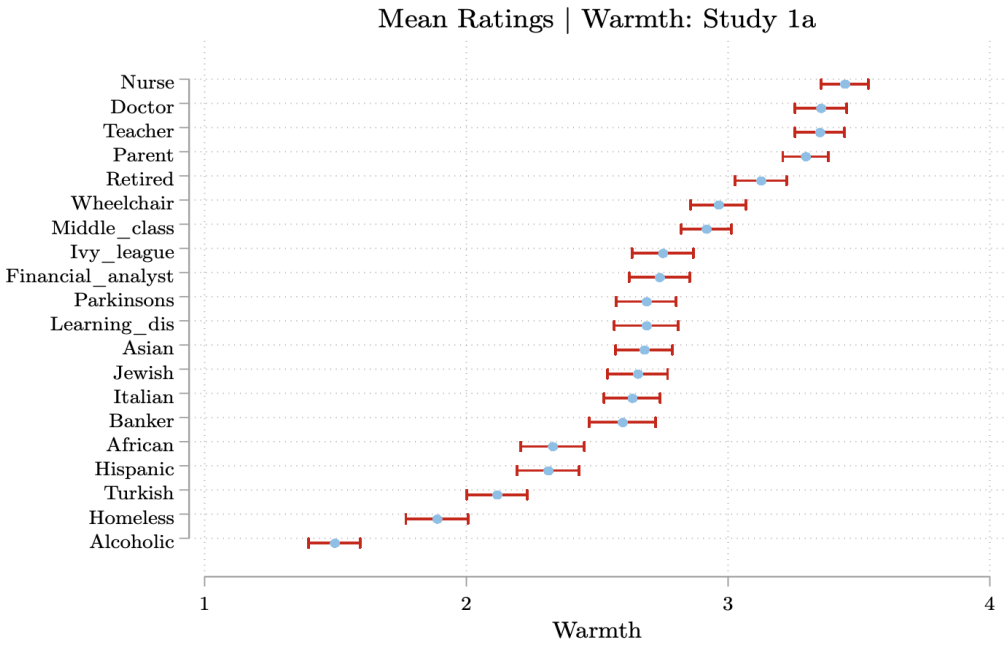
2.11.6 Descriptive statistics

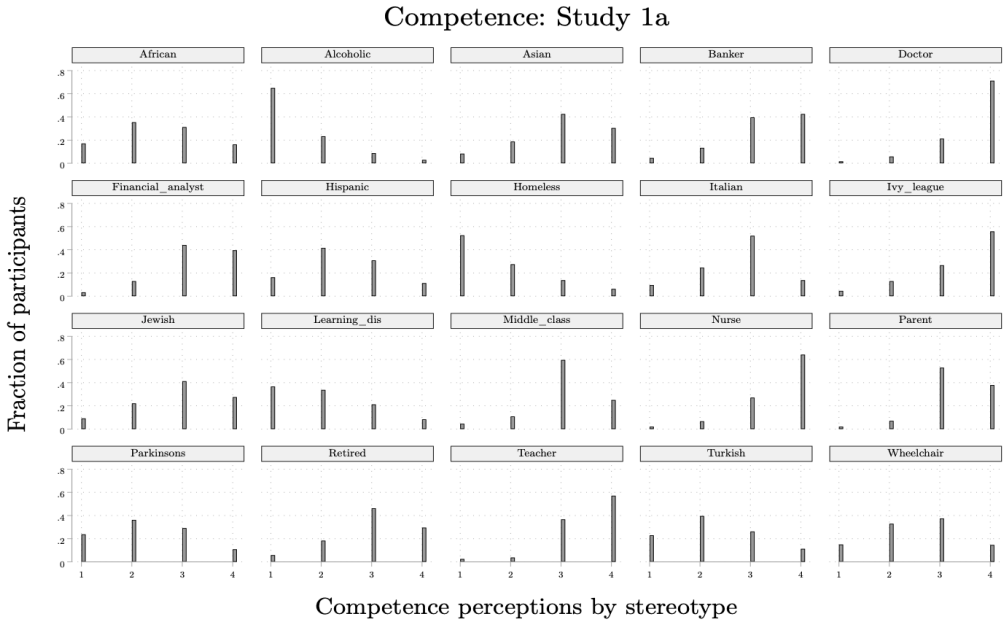
Study 1a

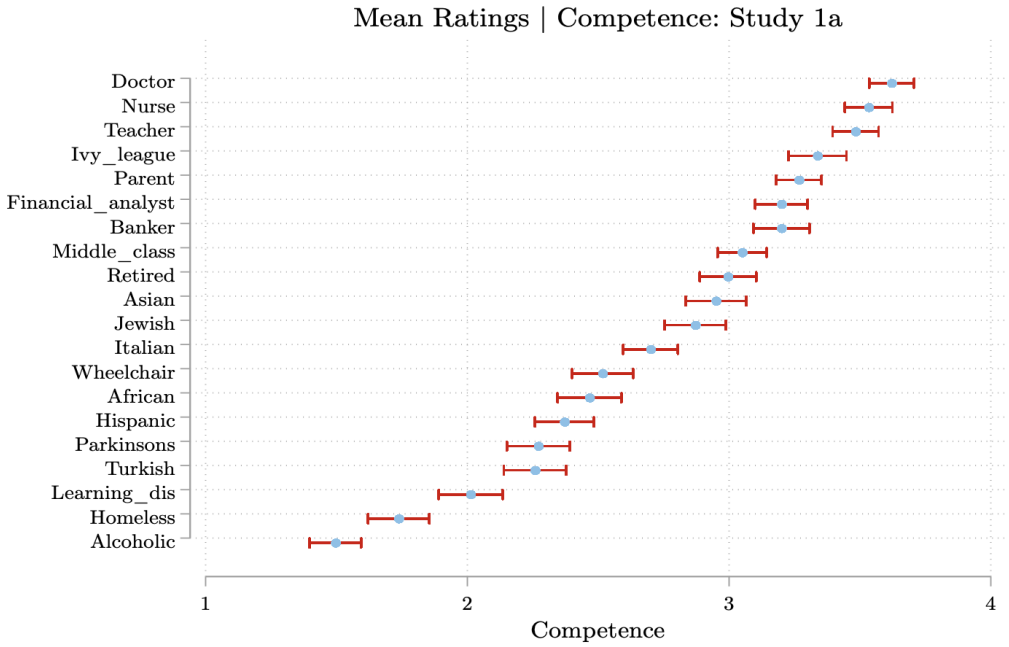



Mean Ratings | Trust: Study 1a
Teacher
Parent
Nurse
Learning_dis
Parkinsons
Retired
Middle_class
Wheelchair
African
Hispanic
Italian
Banker
Ivy_league
Asian
Doctor
Financial_analyst
Homeless
Jewish
Turkish
Alcoholic
0
1
2
3
4
5
Trust

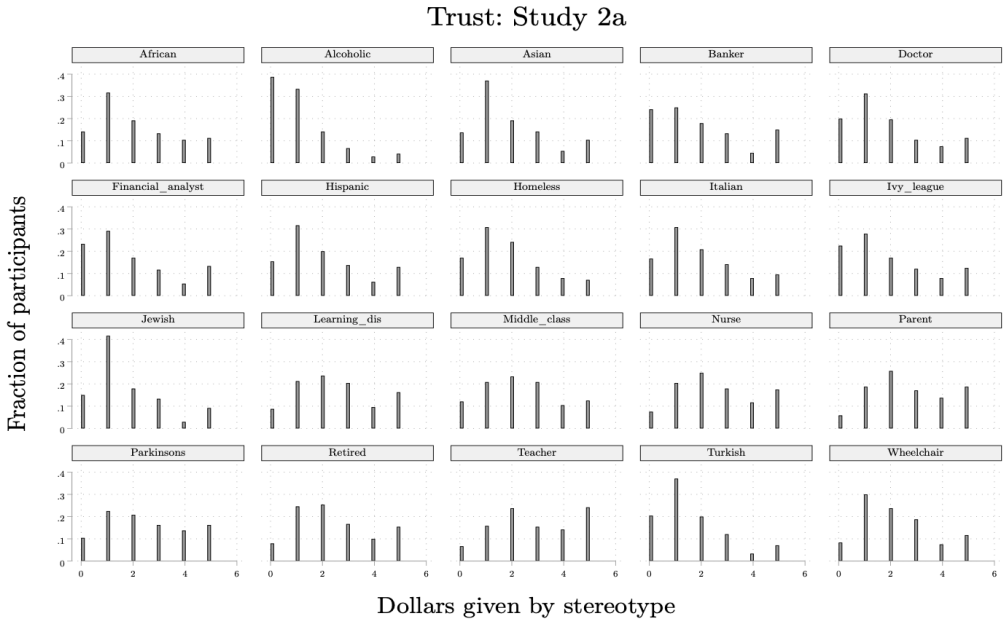

Warmth: Study 1a
African
Alcoholic
Asian
Banker
Doctor
Financial_analyst
Hispanic
Homeless
Italian
Ivy_league
Jewish
Learning_dis
Middle_class
Nurse
Parent
Parkinsons
Retired
Teacher
Turkish
Wheelchair
Fraction of participants
Warmth perceptions by stereotype
Graphs by Stereotype

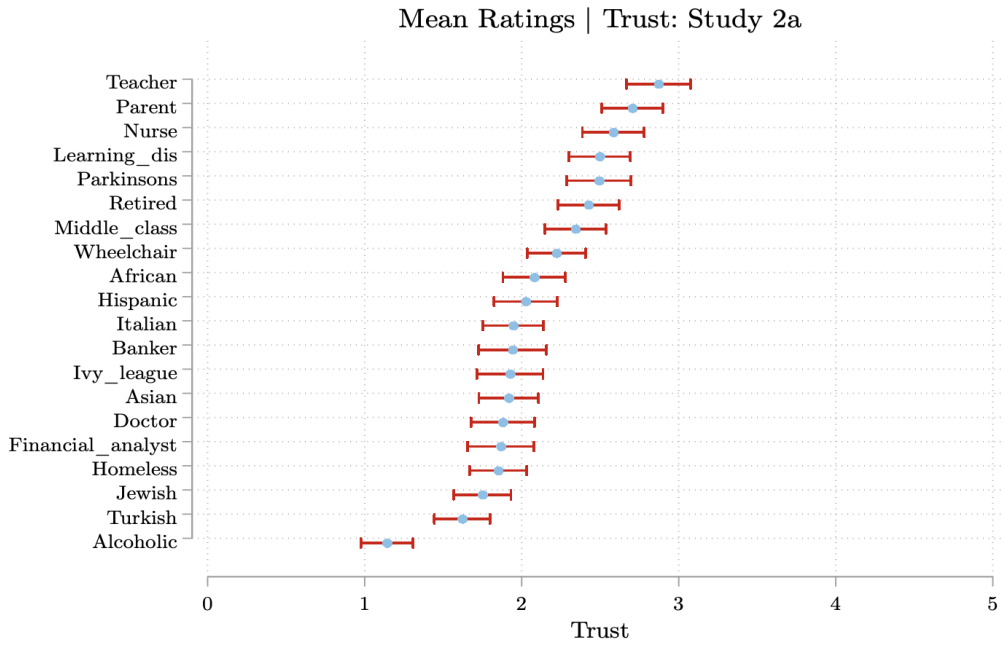

Mean Ratings | Warmth: Study 1a
Nurse
Doctor
Teacher
Parent
Retired
Wheelchair
Middle_class
Ivy_league
Financial_analyst
Parkinsons
Learning_dis
Asian
Jewish
Italian
Banker
African
Hispanic
Turkish
Homeless
Alcoholic
1
2
3
4
Warmth

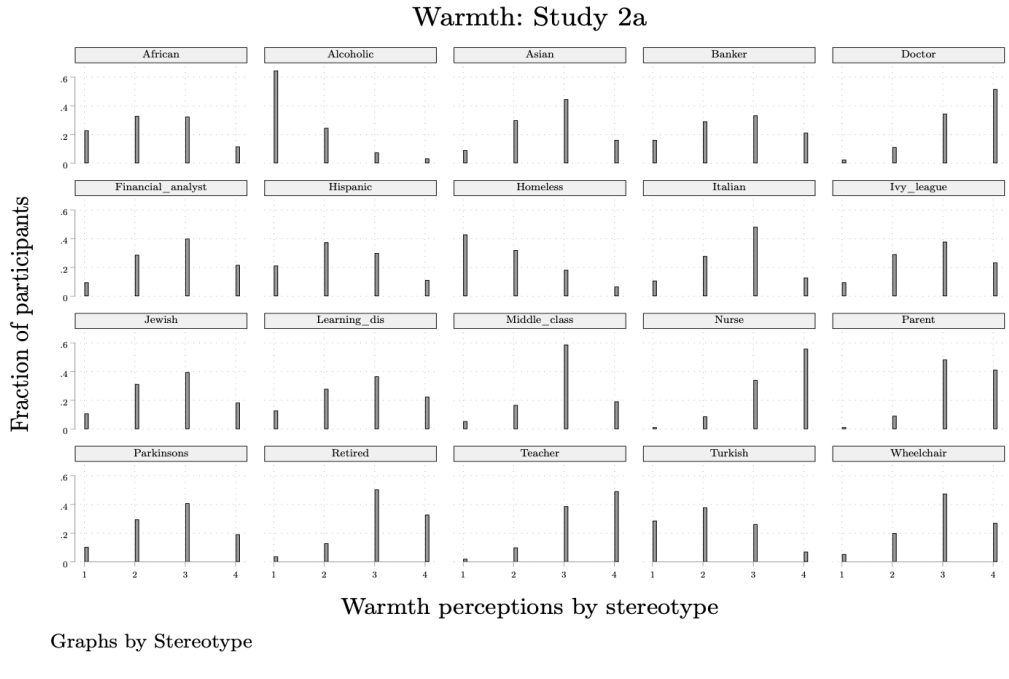

Competence: Study 1a
African
Alcoholic
Asian
Banker
Doctor
Financial_analyst
Hispanic
Homeless
Italian
Ivy_league
Jewish
Learning_dis
Middle_class
Nurse
Parent
Parkinsons
Retired
Teacher
Turkish
Wheelchair
Fraction of participants
Competence perceptions by stereotype
Graphs by Stereotype

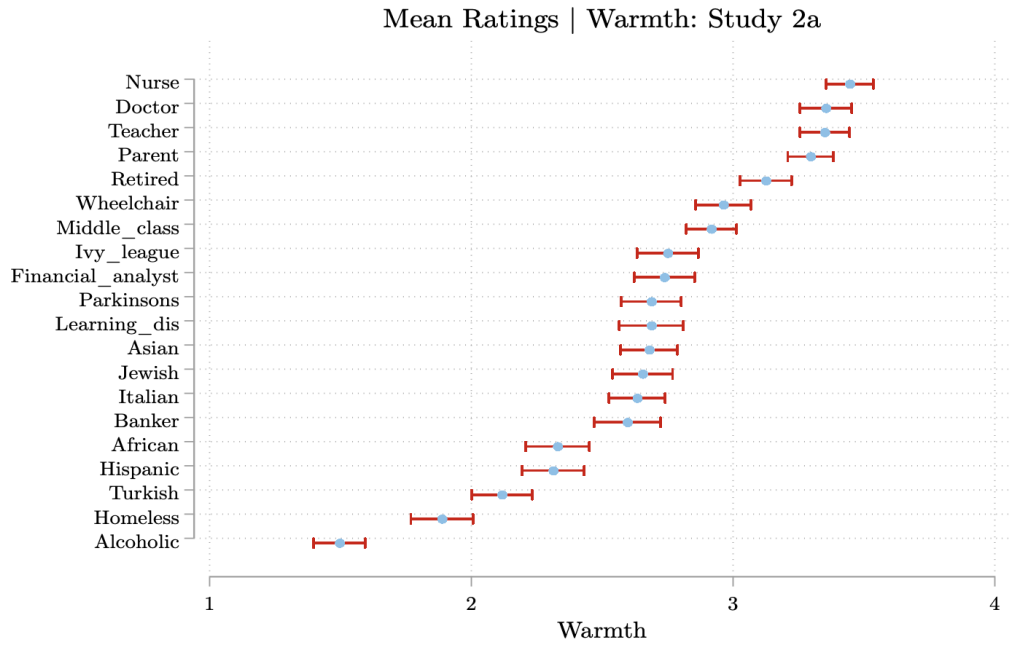

Mean Ratings | Competence: Study 1a
Doctor
Nurse
Teacher
Ivy_league
Parent
Financial_analyst
Banker
Middle_class
Retired
Asian
Jewish
Italian
Wheelchair
African
Hispanic
Parkinsons
Turkish
Learning_dis
Homeless
Alcoholic
1
2
3
4
Competence

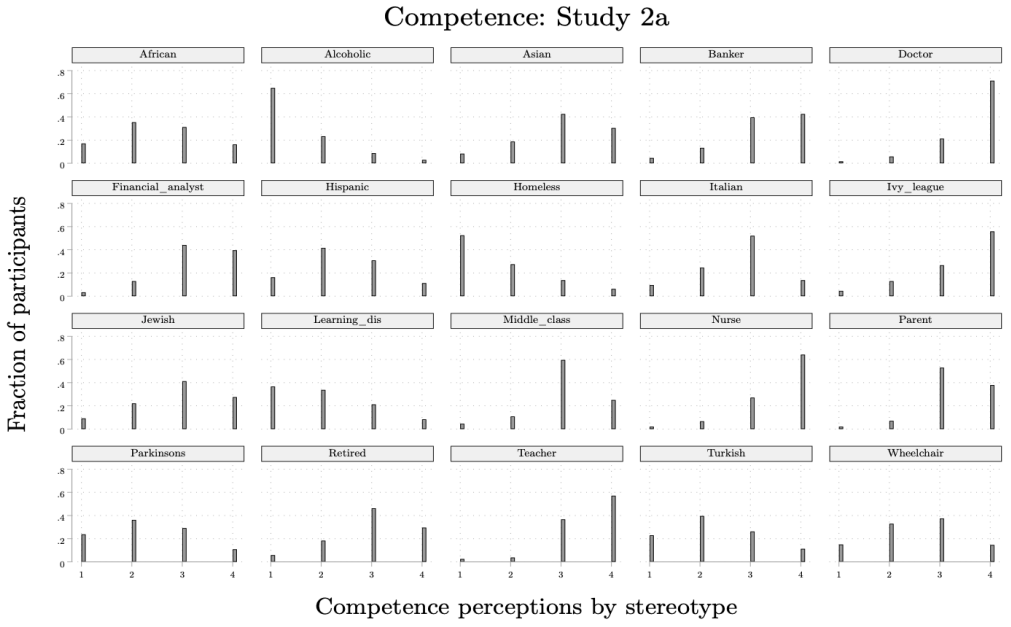
Study 2a

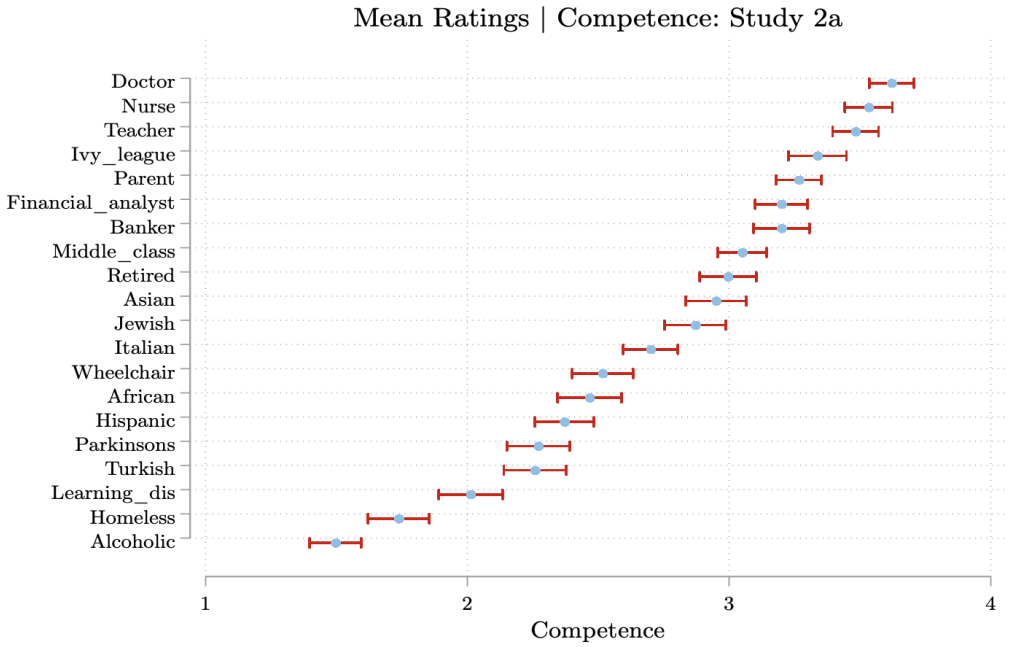

Trust: Study 2a
African
Alcoholic
Asian
Banker
Doctor
Financial_analyst
Hispanic
Homeless
Italian
Ivy_league
Jewish
Learning_dis
Middle_class
Nurse
Parent
Parkinsons
Retired
Teacher
Turkish
Wheelchair
Fraction of participants
Dollars given by stereotype
Graphs by Stereotype


Mean Ratings | Trust: Study 2a
Teacher
Parent
Nurse
Learning_dis
Parkinsons
Retired
Middle_class
Wheelchair
African
Hispanic
Italian
Banker
Ivy_league
Asian
Doctor
Financial_analyst
Homeless
Jewish
Turkish
Alcoholic
0
1
2
3
4
5
Trust

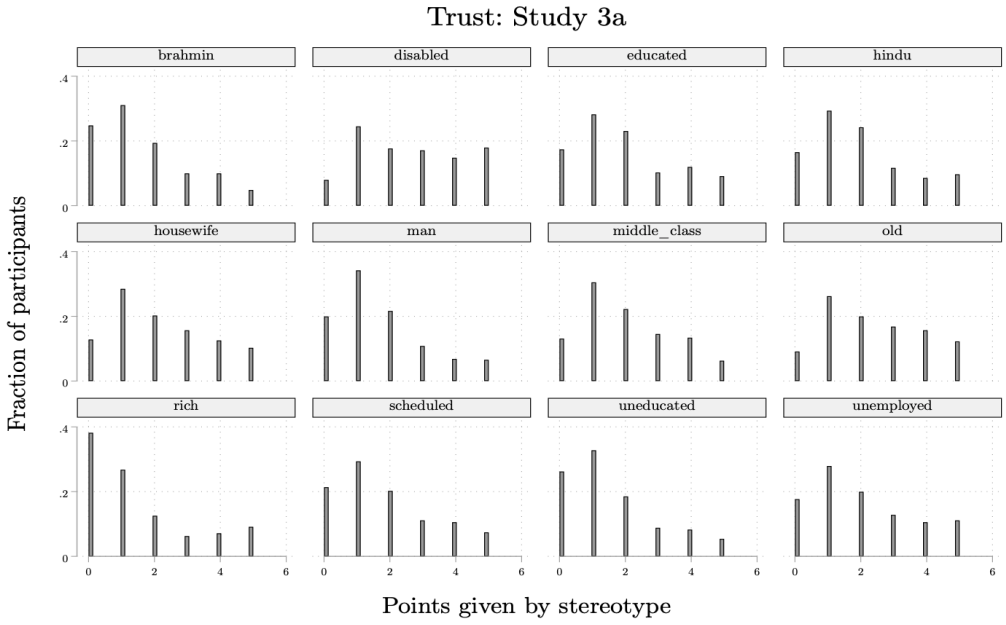

Warmth: Study 2a
African
Alcoholic
Asian
Banker
Doctor
Financial_analyst
Hispanic
Homeless
Italian
Ivy_league
Jewish
Learning_dis
Middle_class
Nurse
Parent
Parkinsons
Retired
Teacher
Turkish
Wheelchair
Fraction of participants
Warmth perceptions by stereotype
Graphs by Stereotype

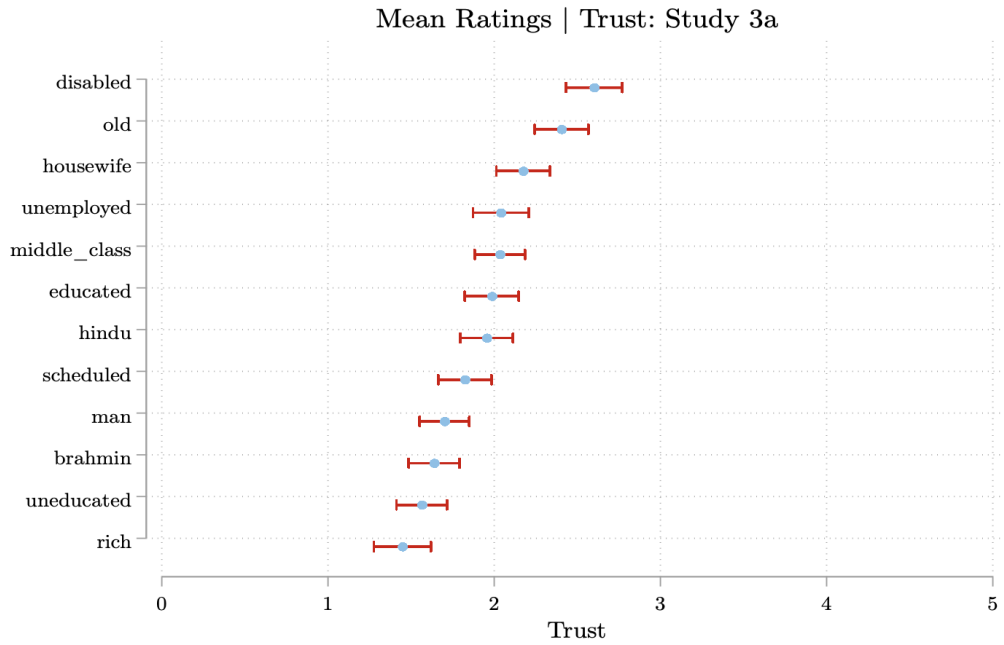

Mean Ratings | Warmth: Study 2a
Nurse
Doctor
Teacher
Parent
Retired
Wheelchair
Middle_class
Ivy_league
Financial_analyst
Parkinsons
Learning_dis
Asian
Jewish
Italian
Banker
African
Hispanic
Turkish
Homeless
Alcoholic
1
2
3
4
Warmth

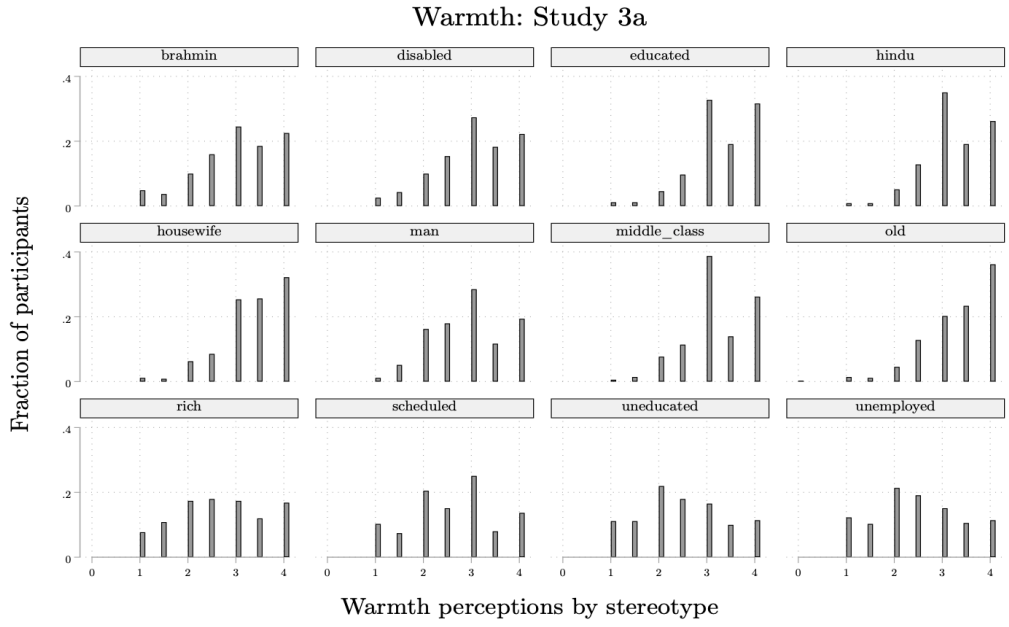

Competence: Study 2a
African
Alcoholic
Asian
Banker
Doctor
Financial_analyst
Hispanic
Homeless
Italian
Ivy_league
Jewish
Learning_dis
Middle_class
Nurse
Parent
Parkinsons
Retired
Teacher
Turkish
Wheelchair
Fraction of participants
Competence perceptions by stereotype
Graphs by Stereotype

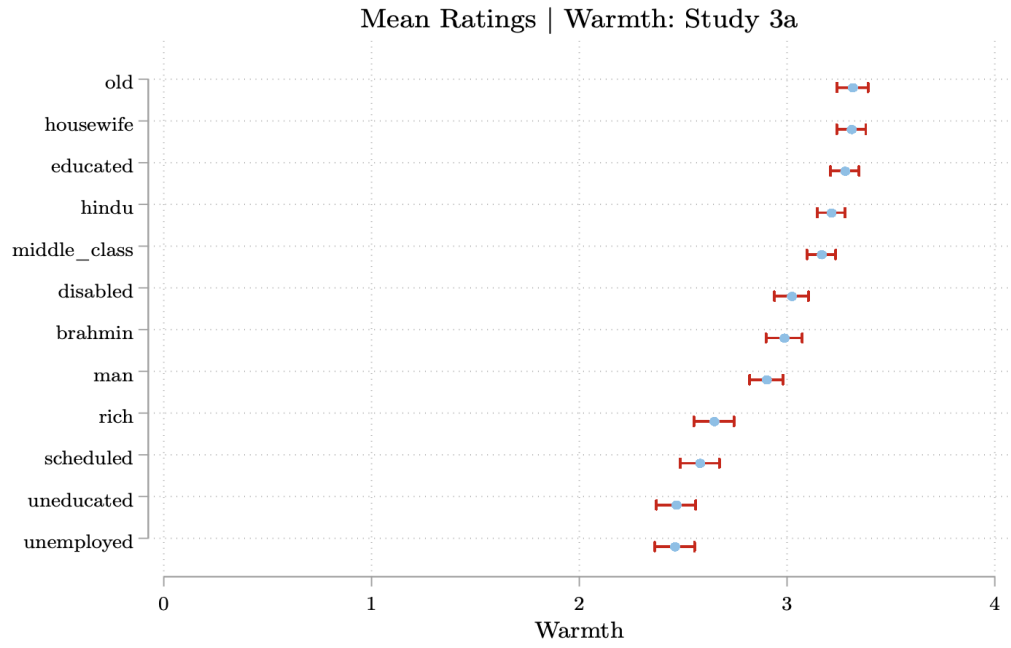

Mean Ratings | Competence: Study 2a
Doctor
Nurse
Teacher
Ivy_league
Parent
Financial_analyst
Banker
Middle_class
Retired
Asian
Jewish
Italian
Wheelchair
African
Hispanic
Parkinsons
Turkish
Learning_dis
Homeless
Alcoholic
1
2
3
4
Competence

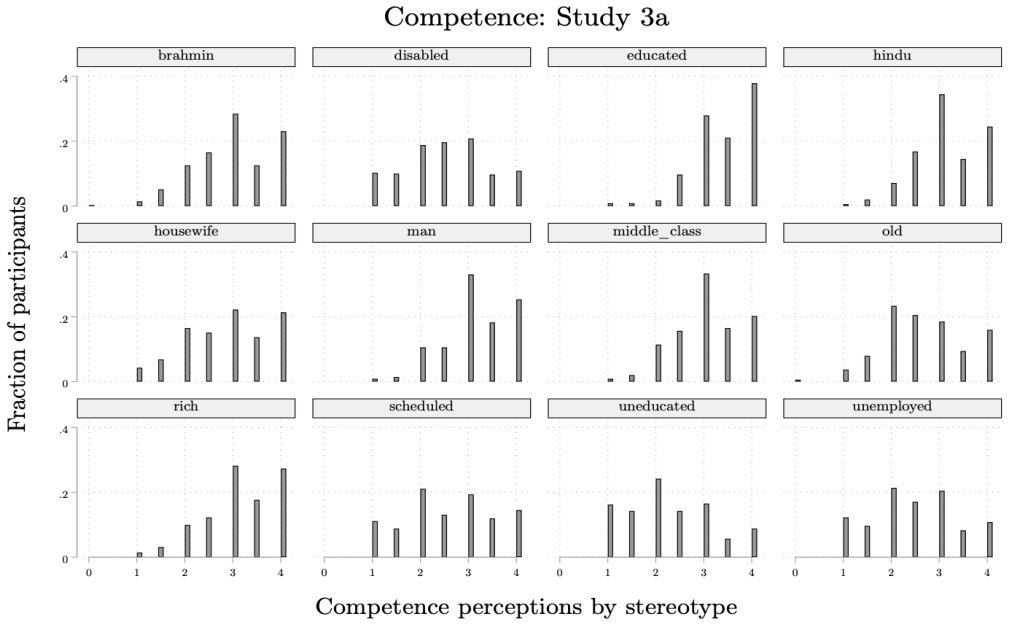
Study 3a

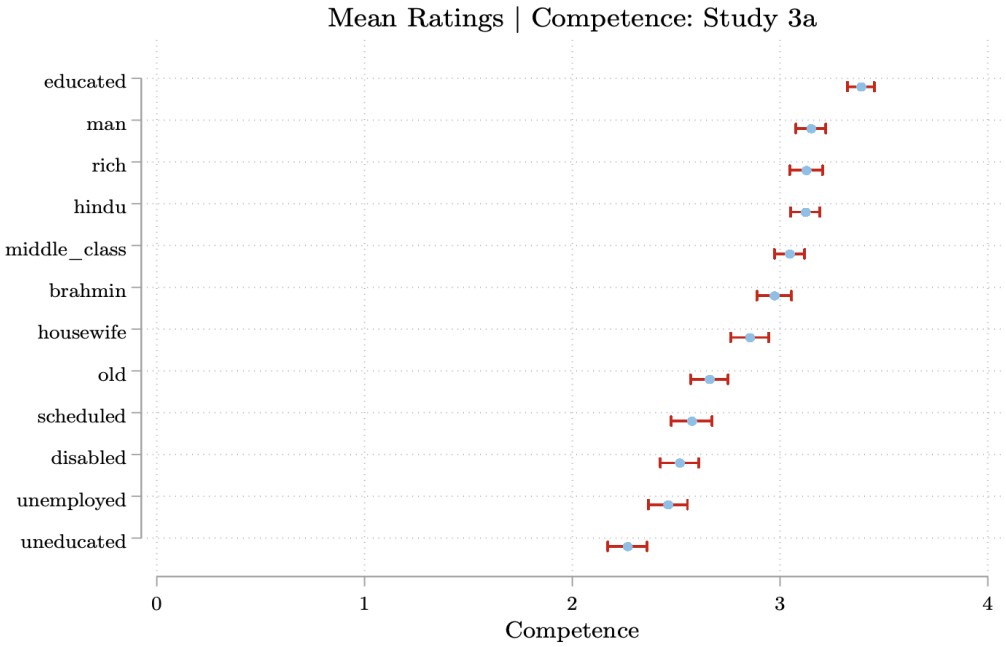

Trust: Study 3a
brahmin
disabled
educated
hindu
housewife
man
middle_class
old
rich
scheduled
uneducated
unemployed
Fraction of participants
Points given by stereotype
Graphs by Stereotype


Mean Ratings | Trust: Study 3a
disabled
old
housewife
unemployed
middle_class
educated
hindu
scheduled
man
brahmin
uneducated
rich
0
1
2
3
4
5
Trust

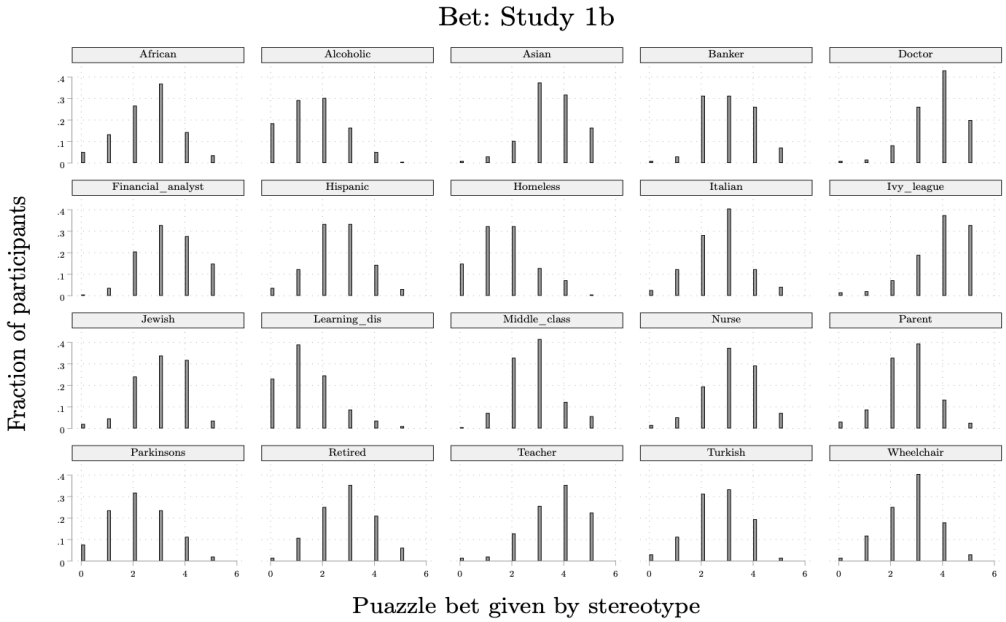

Warmth: Study 3a
brahmin
disabled
educated
hindu
housewife
man
middle_class
old
rich
scheduled
uneducated
unemployed
Fraction of participants
Warmth perceptions by stereotype
Graphs by Stereotype

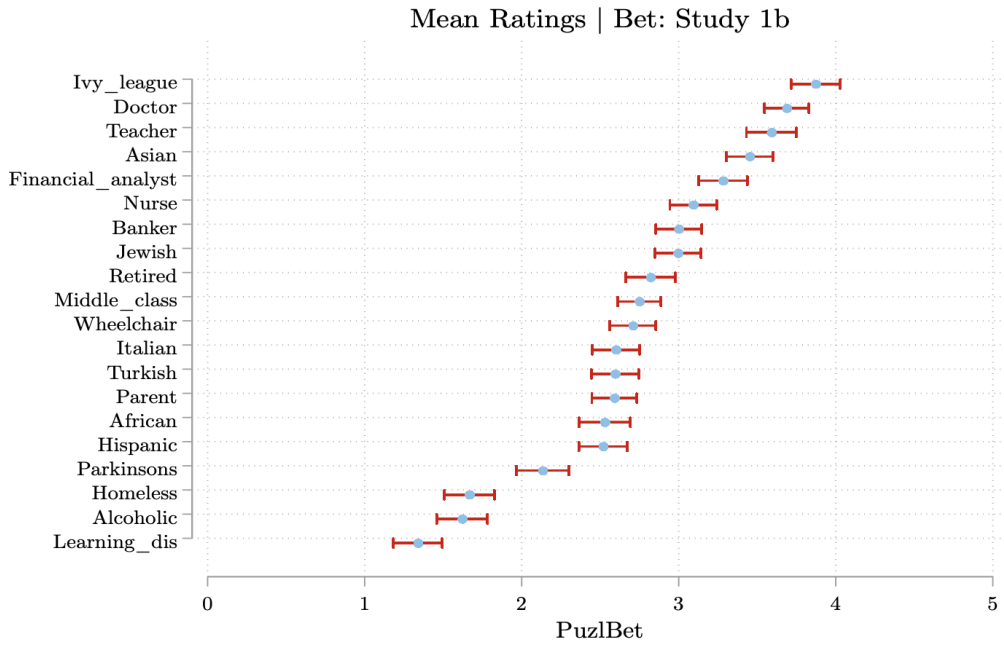

Mean Ratings | Warmth: Study 3a
old
housewife
educated
hindu
middle_class
disabled
brahmin
man
rich
scheduled
uneducated
unemployed
0
1
2
3
4
Warmth

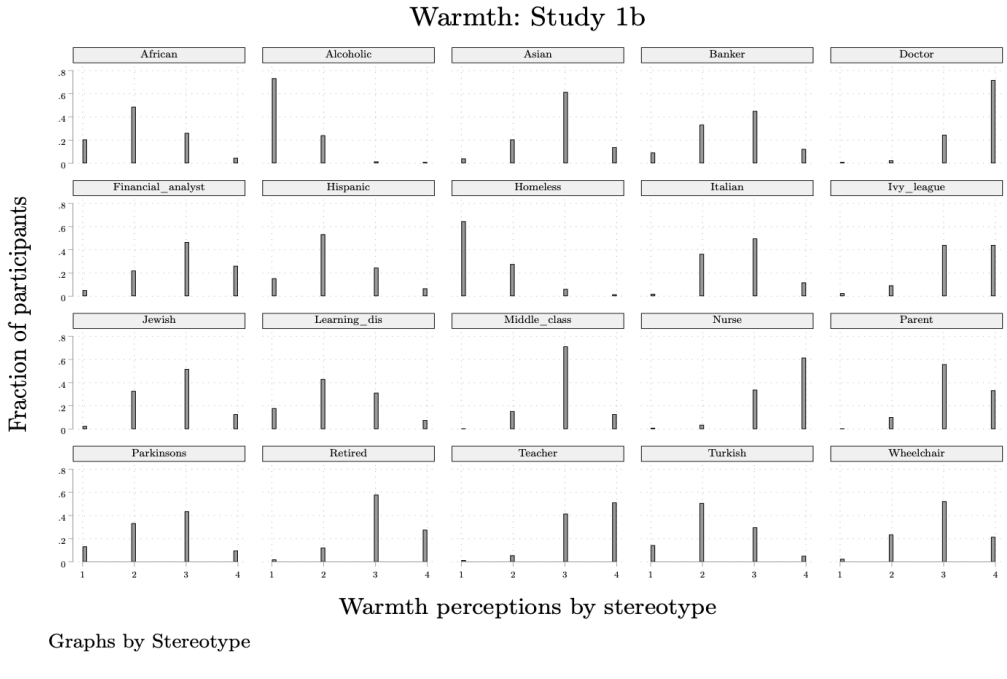

Competence: Study 3a
brahmin
disabled
educated
hindu
housewife
man
middle_class
old
rich
scheduled
uneducated
unemployed
Fraction of participants
Competence perceptions by stereotype
Graphs by Stereotype

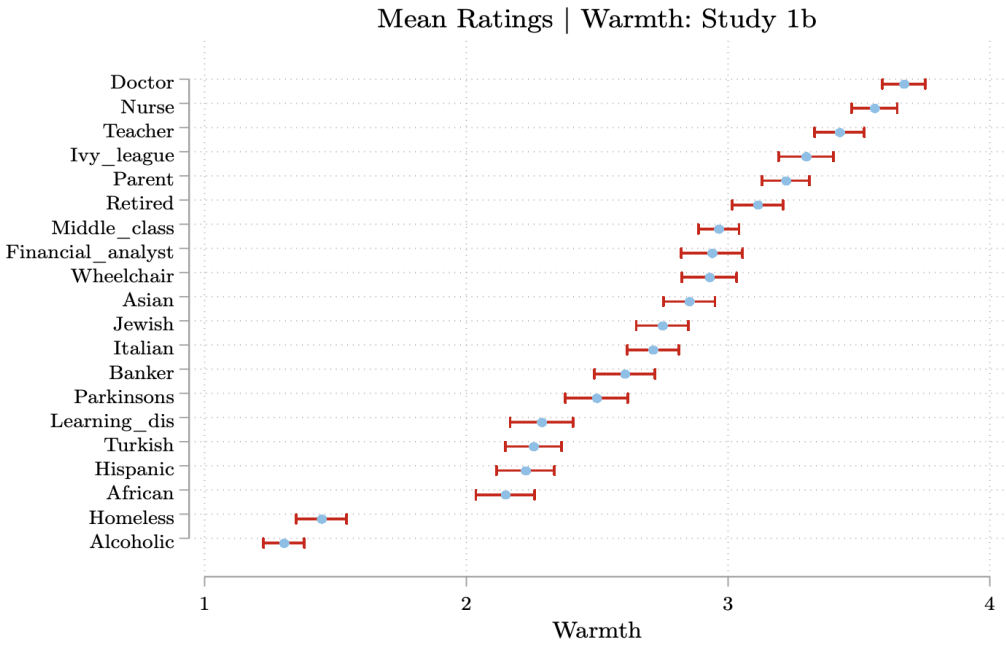

Mean Ratings | Competence: Study 3a
educated
man
rich
hindu
middle_class
brahmin
housewife
old
scheduled
disabled
unemployed
uneducated
0
1
2
3
4
Competence

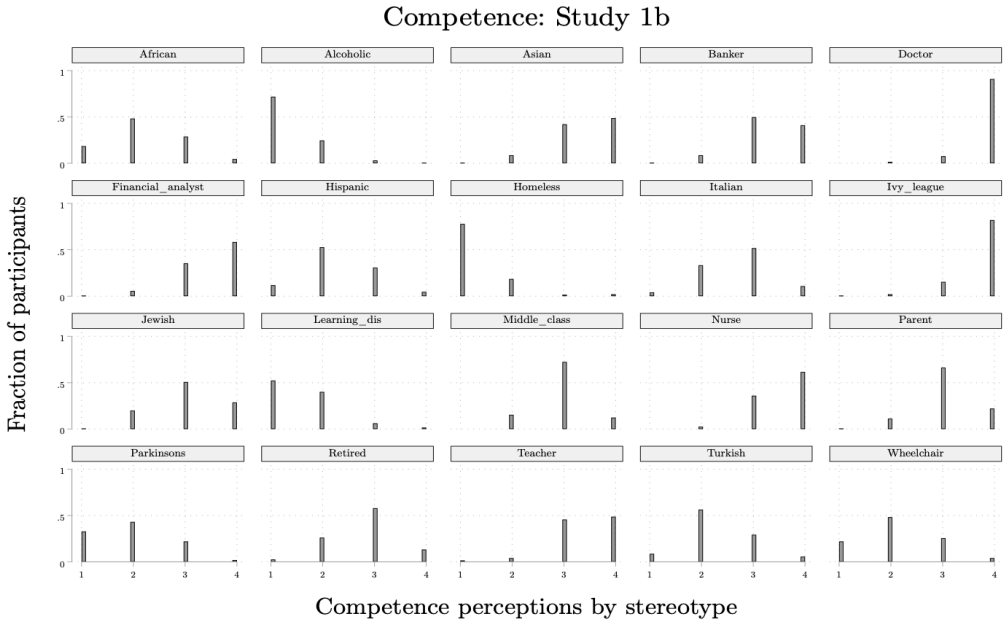
Study 1b

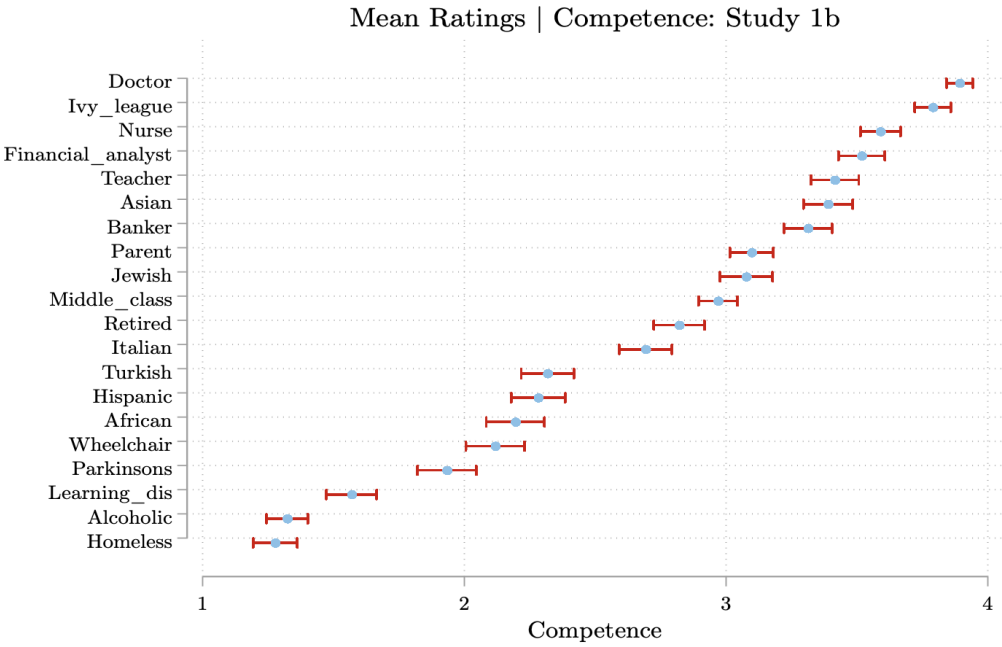

Bet: Study 1b
African
Alcoholic
Asian
Banker
Doctor
Financial_analyst
Hispanic
Homeless
Italian
Ivy_league
Jewish
Learning_dis
Middle_class
Nurse
Parent
Parkinsons
Retired
Teacher
Turkish
Wheelchair
Fraction of participants
Puazzle bet given by stereotype
Graphs by Stereotype


Mean Ratings | Bet: Study 1b
Ivy_league
Doctor
Teacher
Asian
Financial_analyst
Nurse
Banker
Jewish
Retired
Middle_class
Wheelchair
Italian
Turkish
Parent
African
Hispanic
Parkinsons
Homeless
Alcoholic
Learning_dis
0
1
2
3
4
5
PuzlBet

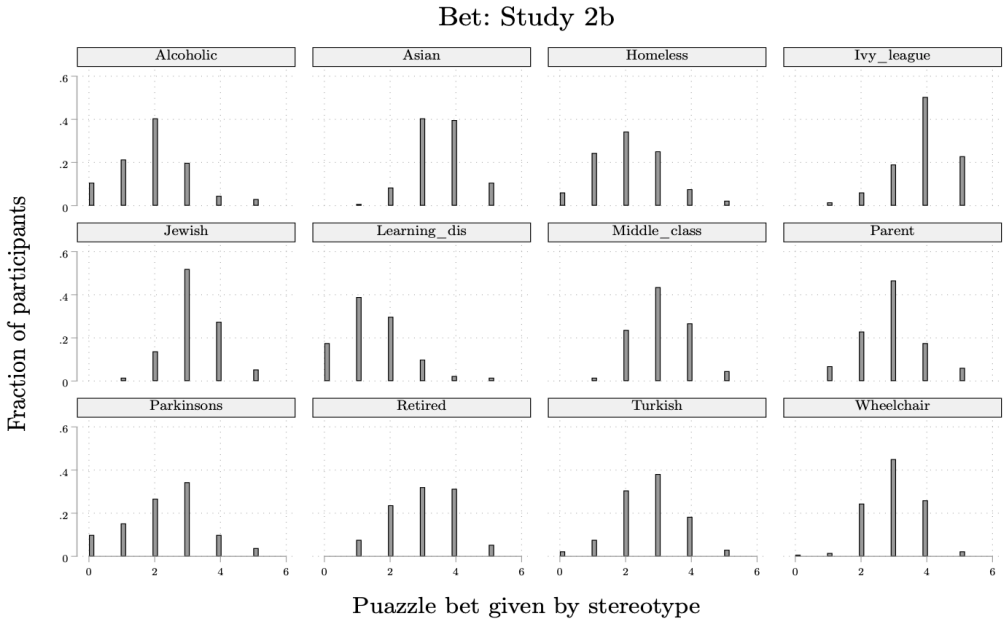

Warmth: Study 1b
African
Alcoholic
Asian
Banker
Doctor
Financial_analyst
Hispanic
Homeless
Italian
Ivy_league
Jewish
Learning_dis
Middle_class
Nurse
Parent
Parkinsons
Retired
Teacher
Turkish
Wheelchair
Fraction of participants
Warmth perceptions by stereotype
Graphs by Stereotype

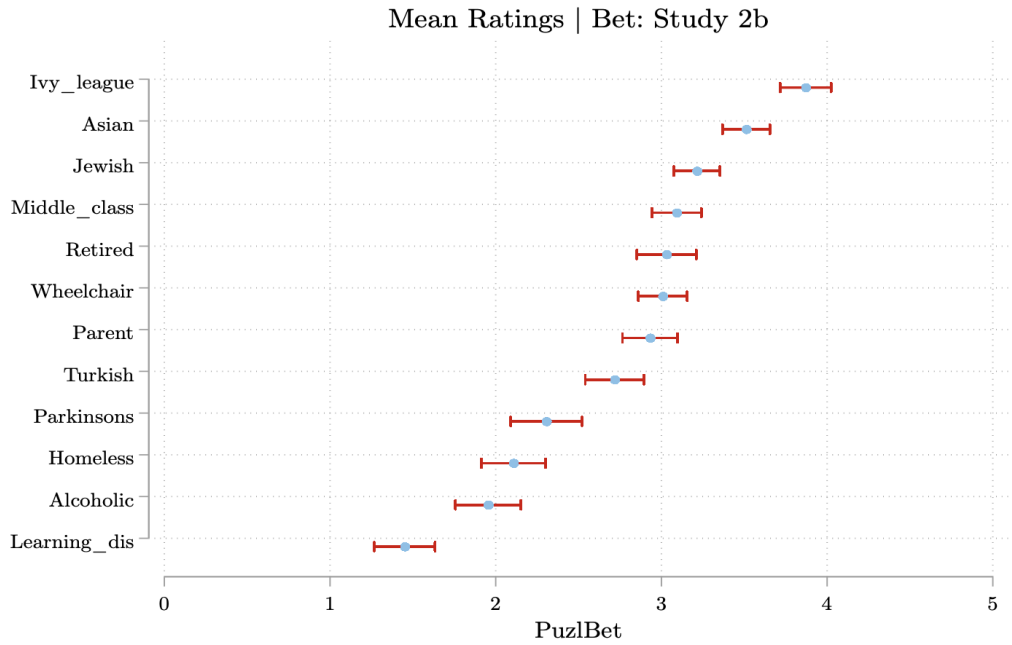

Mean Ratings | Warmth: Study 1b
Doctor
Nurse
Teacher
Ivy_league
Parent
Retired
Middle_class
Financial_analyst
Wheelchair
Asian
Jewish
Italian
Banker
Parkinsons
Learning_dis
Turkish
Hispanic
African
Homeless
Alcoholic
1
2
3
4
Warmth

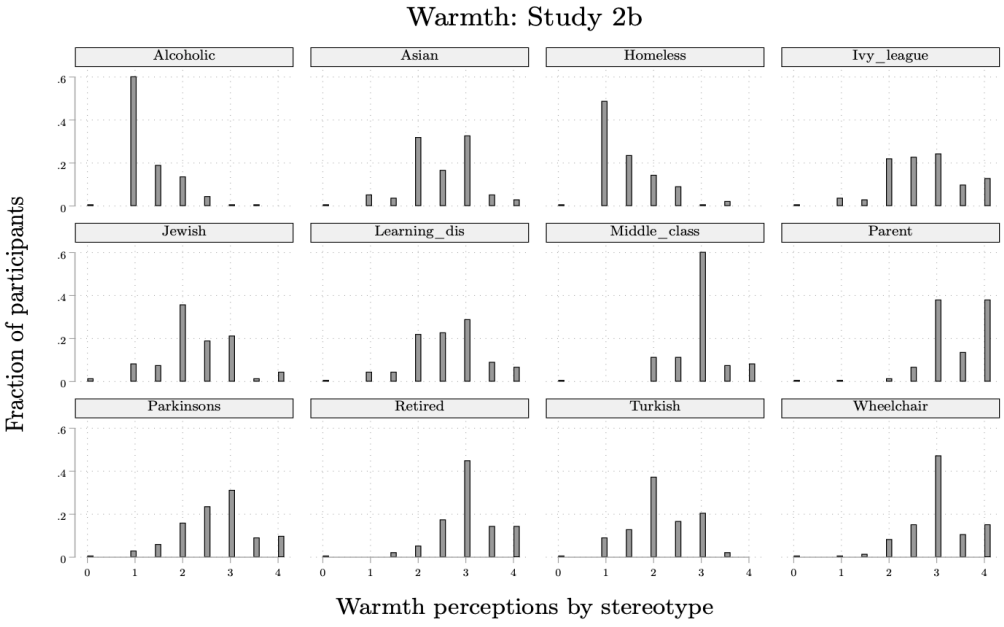

Competence: Study 1b
African
Alcoholic
Asian
Banker
Doctor
Financial_analyst
Hispanic
Homeless
Italian
Ivy_league
Jewish
Learning_dis
Middle_class
Nurse
Parent
Parkinsons
Retired
Teacher
Turkish
Wheelchair
Fraction of participants
Competence perceptions by stereotype
Graphs by Stereotype

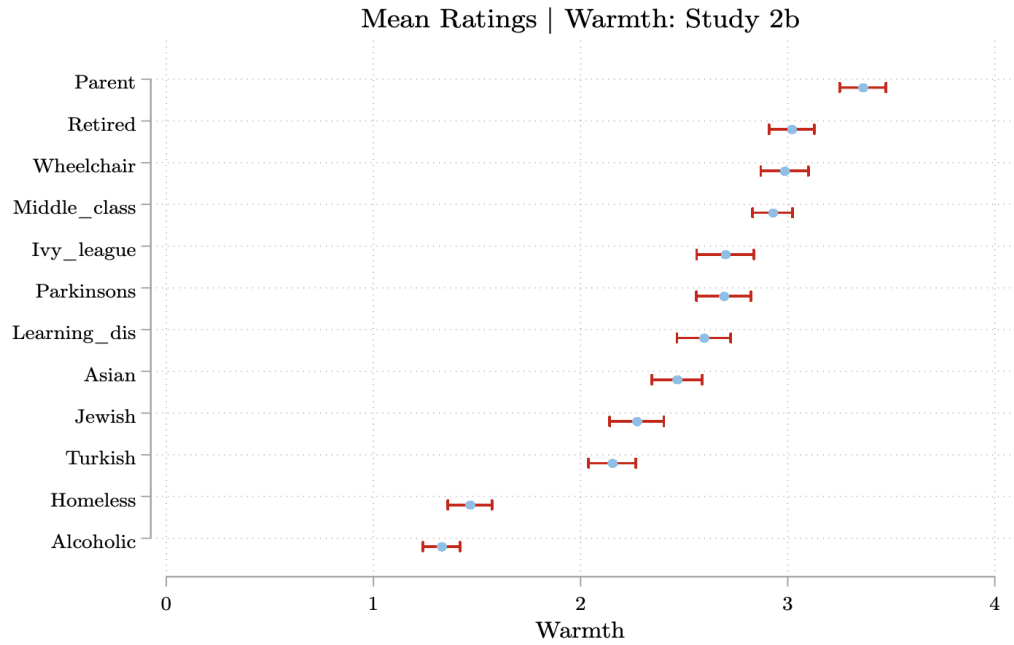

Mean Ratings | Competence: Study 1b
Doctor
Ivy_league
Nurse
Financial_analyst
Teacher
Asian
Banker
Parent
Jewish
Middle_class
Retired
Italian
Turkish
Hispanic
African
Wheelchair
Parkinsons
Learning_dis
Alcoholic
Homeless
1
2
3
4
Competence

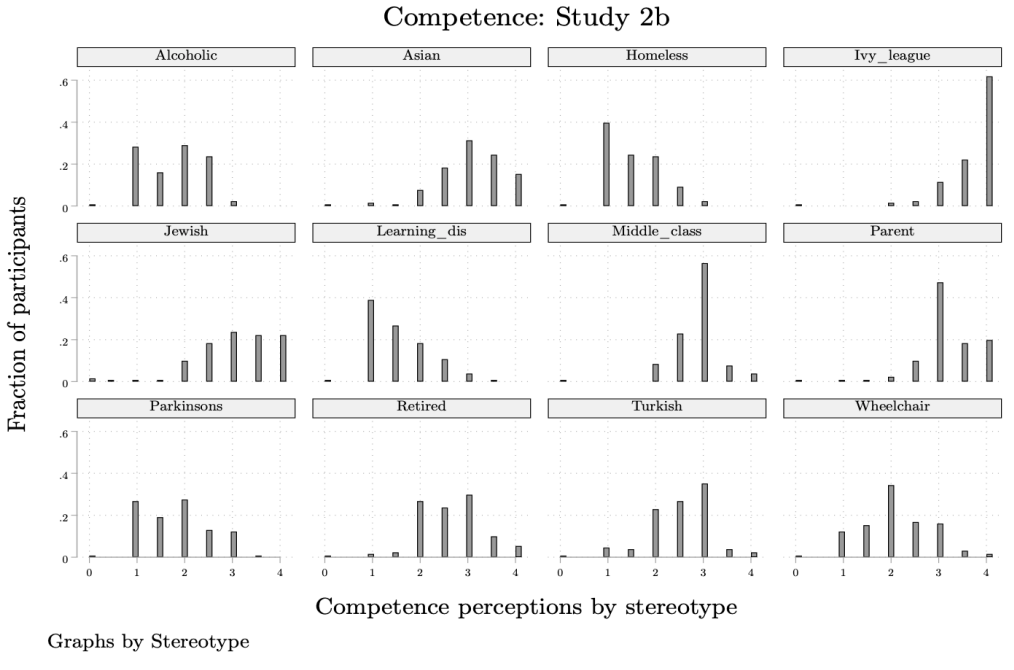
Study 2b

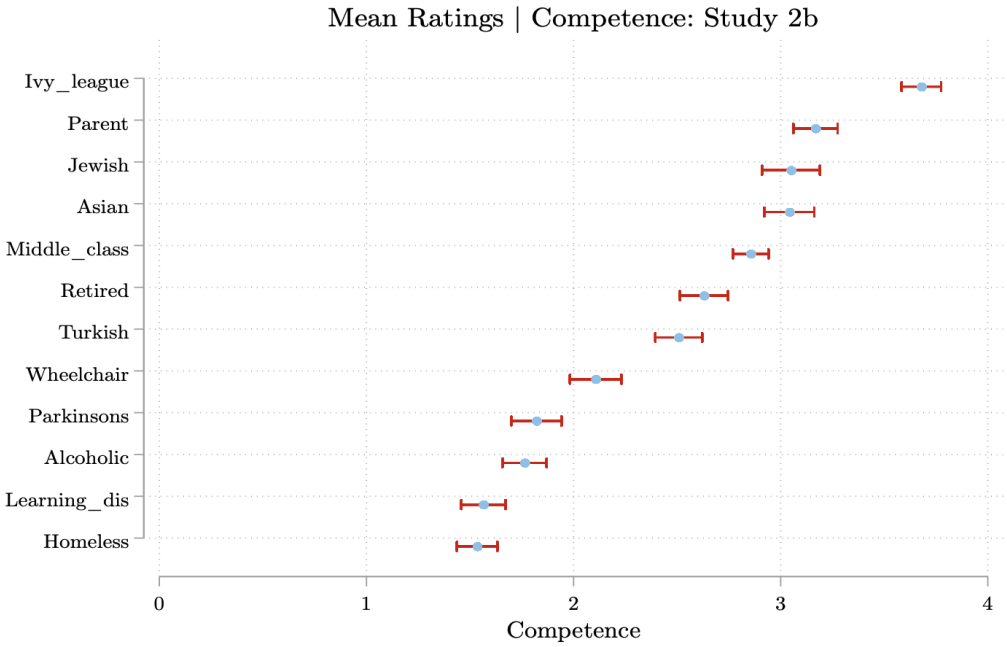

Bet: Study 2b
Alcoholic
Asian
Homeless
Ivy_league
Jewish
Learning_dis
Middle_class
Parent
Parkinsons
Retired
Turkish
Wheelchair
Fraction of participants
Puazzle bet given by stereotype
Graphs by Stereotype


Mean Ratings | Bet: Study 2b
Ivy_league
Asian
Jewish
Middle_class
Retired
Wheelchair
Parent
Turkish
Parkinsons
Homeless
Alcoholic
Learning_dis
0
1
2
3
4
5
PuzlBet

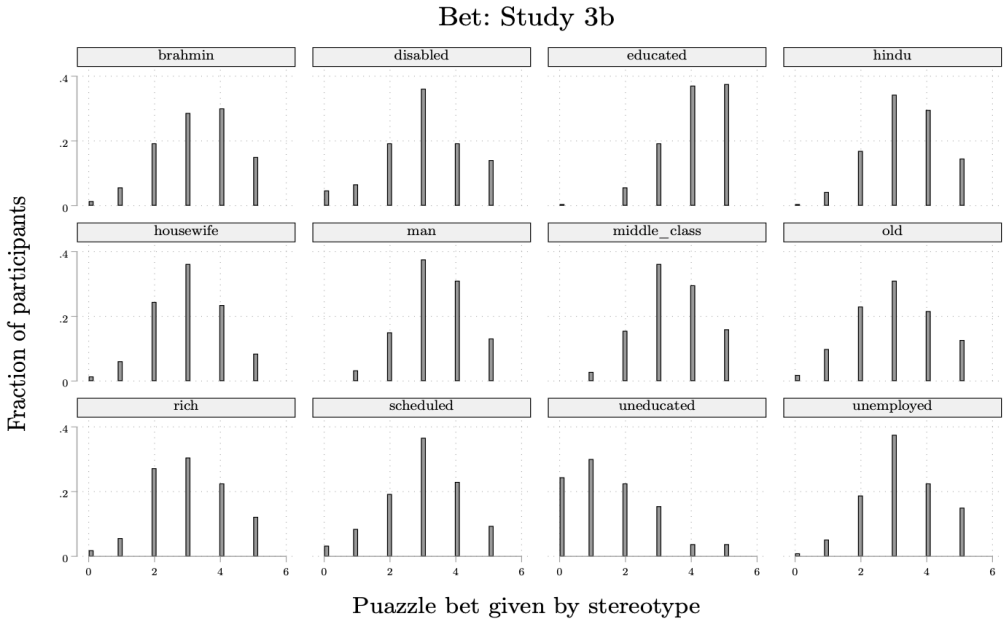

Warmth: Study 2b
Alcoholic
Asian
Homeless
Ivy_league
Jewish
Learning_dis
Middle_class
Parent
Parkinsons
Retired
Turkish
Wheelchair
Fraction of participants
Warmth perceptions by stereotype
Graphs by Stereotype

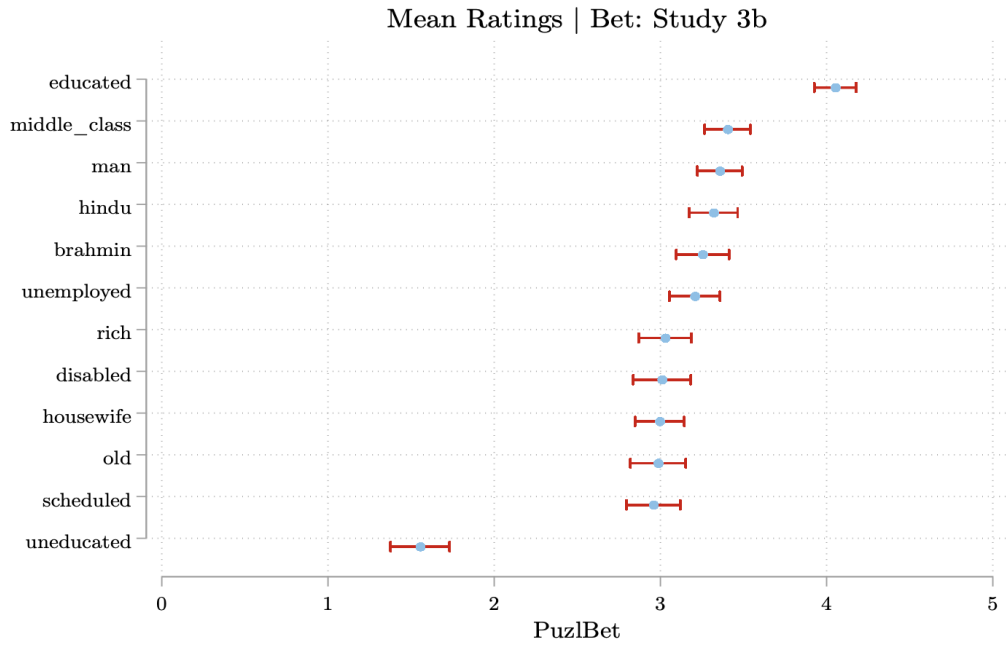

Mean Ratings | Warmth: Study 2b
Parent
Retired
Wheelchair
Middle_class
Ivy_league
Parkinsons
Learning_dis
Asian
Jewish
Turkish
Homeless
Alcoholic
0
1
2
3
4
Warmth

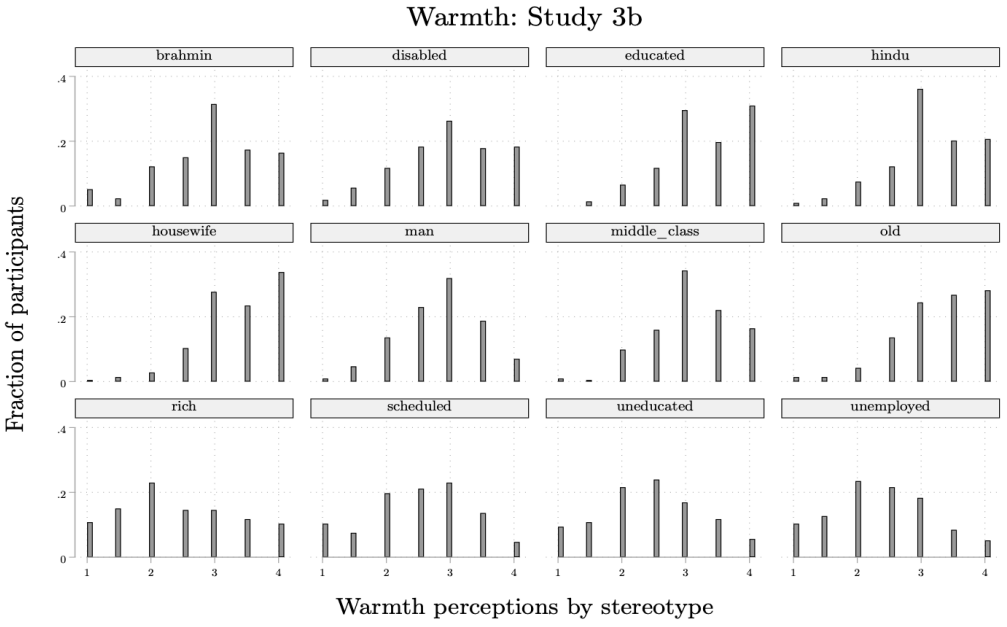

Competence: Study 2b
Alcoholic
Asian
Homeless
Ivy_league
Jewish
Learning_dis
Middle_class
Parent
Parkinsons
Retired
Turkish
Wheelchair
Fraction of participants
Competence perceptions by stereotype
Graphs by Stereotype

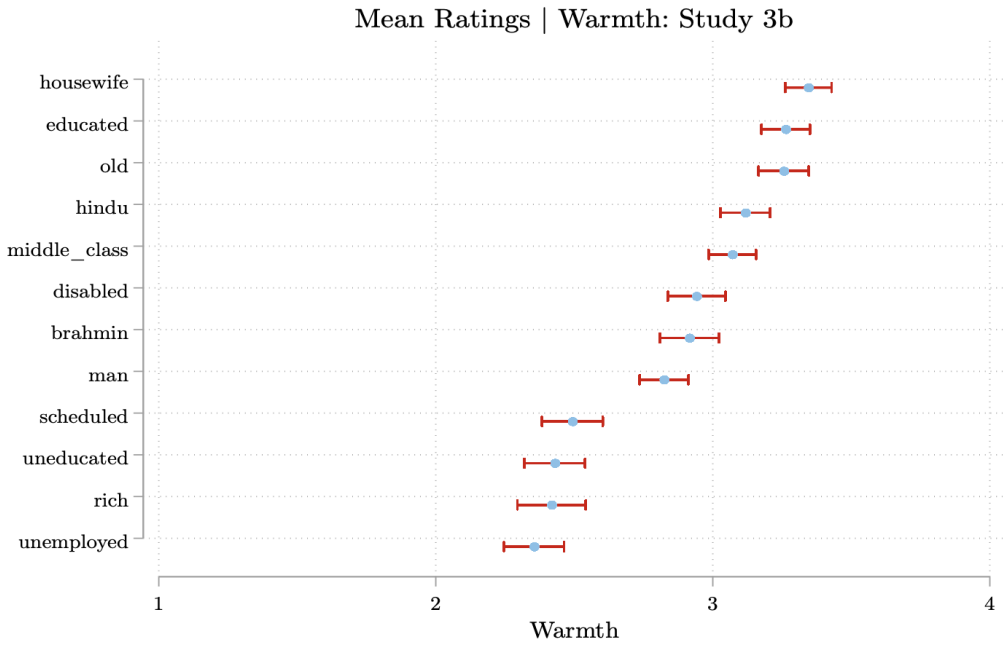

Mean Ratings | Competence: Study 2b
Ivy_league
Parent
Jewish
Asian
Middle_class
Retired
Turkish
Wheelchair
Parkinsons
Alcoholic
Learning_dis
Homeless
0
1
2
3
4
Competence

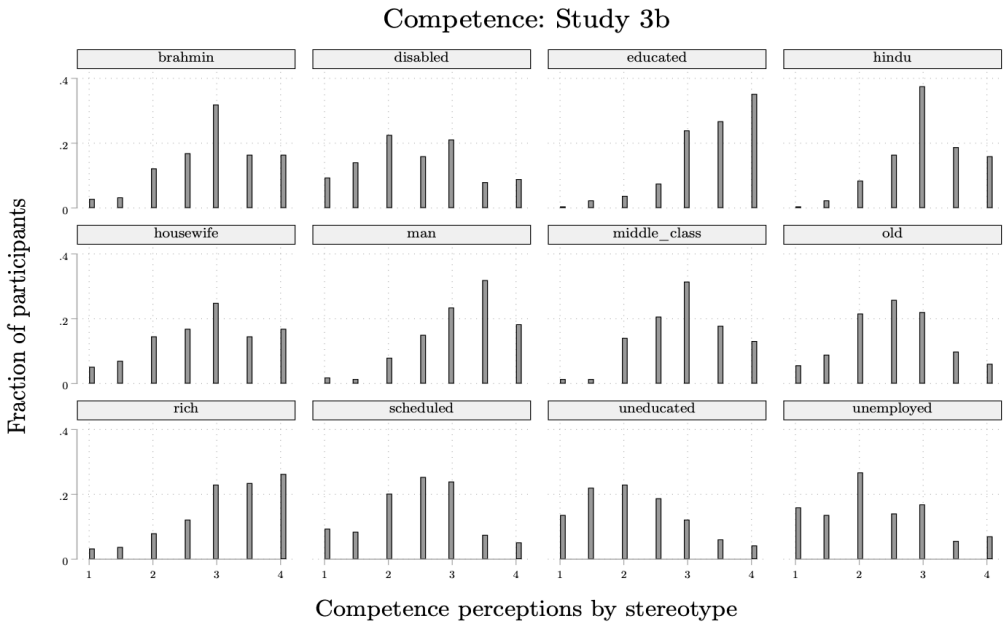
Study 3b

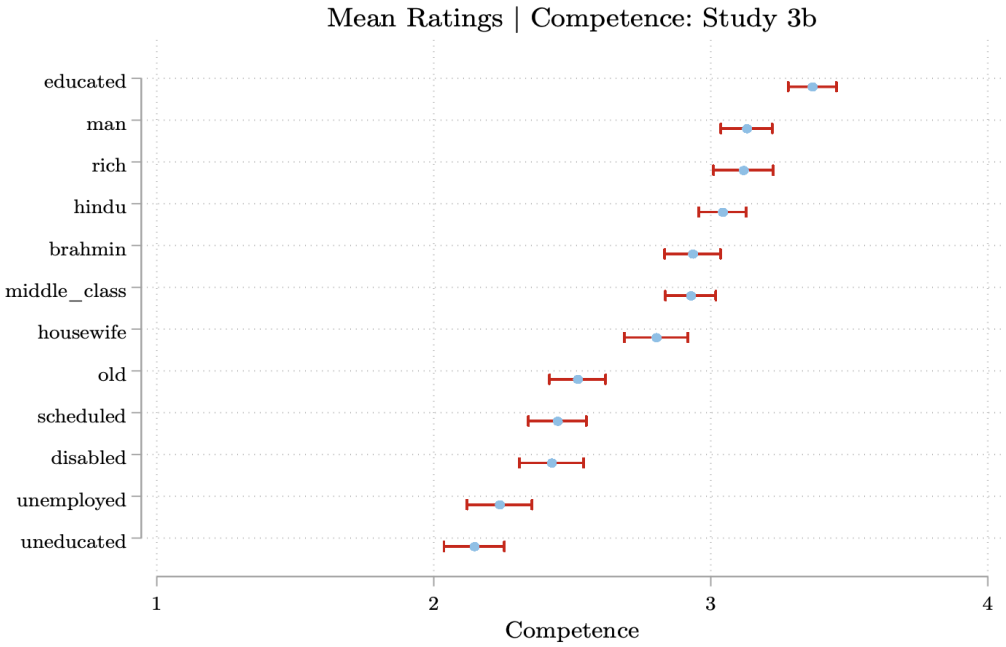

Bet: Study 3b
brahmin
disabled
educated
hindu
housewife
man
middle_class
old
rich
scheduled
uneducated
unemployed
Fraction of participants
Puazzle bet given by stereotype
Graphs by Stereotype


Mean Ratings | Bet: Study 3b
educated
middle_class
man
hindu
brahmin
unemployed
rich
disabled
housewife
old
scheduled
uneducated
0
1
2
3
4
5
PuzlBet


Warmth: Study 3b
brahmin
disabled
educated
hindu
housewife
man
middle_class
old
rich
scheduled
uneducated
unemployed
Fraction of participants
Warmth perceptions by stereotype
Graphs by Stereotype


Mean Ratings | Warmth: Study 3b
housewife
educated
old
hindu
middle_class
disabled
brahmin
man
scheduled
uneducated
rich
unemployed
1
2
3
4
Warmth


Competence: Study 3b
brahmin
disabled
educated
hindu
housewife
man
middle_class
old
rich
scheduled
uneducated
unemployed
Fraction of participants
Competence perceptions by stereotype
Graphs by Stereotype


Mean Ratings | Competence: Study 3b
educated
man
rich
hindu
brahmin
middle_class
housewife
old
scheduled
disabled
unemployed
uneducated
1
2
3
4
Competence

Chapter 3

The self

To function well and accomplish goals, people must manage threats to their sense of self-integrity: their mental model of themselves as morally adequate people who are capable of controlling important events in their lives (Steele, 1988; Bénabou and Tirole, 2002). People who face chronic threats to their esteem may experience such stigma that it "gets inside their heads," spoiling their motivation and locus of control. Strategies that have been developed in psychology to help people cope with chronic threats to self-integrity and implementation intention failures are self-affirmation and goal-setting, respectively (Steele, 1988; Cohen and Sherman, 2014; Locke and Latham, 1990). We test these strategies using a field experiment in Addis Ababa. We randomly allocate invitations to workshops on self-affirmation and goal-setting (and a control) to 1,766 unemployed youth. We have two main findings. First, the workshops increase both job-search and employment for men, but not for women – who appear to be constrained by gender norms. Second, the workshops are strikingly effective for men with low psychological agency (with a low measure of locus of control at baseline). For this sub-group, participating in the work-shops increases the likelihood of being employed by almost 30 percent, the days worked by around 36 percent and earnings by 34 percent. For these men we also observe major improvements in their locus of control measure.

3.1 Introduction

To function well and accomplish goals, people must manage their sense of self-integrity: their mental model of themselves as morally adequate and capable of controlling the important events in their lives (Steele, 1988; Bénabou and Tirole, 2002; Köszegi, 2006). A major potential threat to self-integrity is the inability to be productive, self-reliant, and contribute to community. Unemployment and underemployment undermine psychological wellbeing and predict poor mental health (Mousteri, Daly, and Delaney, 2018; Goldsmith, Veum, and Darity, 1997; Paul and Moser, 2009). The long run effects have been described as "psychological scarring" (Clark, Georgellis, and Sanfey, 2001). People are motivated to think of themselves as adequate but experiences such as unemployment and underemployment signal that it may not be true, causing cognitive dissonance (Festinger, 1957; Akerlof and Dickens, 1982). People who face chronic threats to the esteem of their identity may experience such stigma that it "gets inside their heads" and threatens to spoil their self-schema (Goffman,

1963). Sen (1999) and Nussbaum (2001) call this perversion of identity formation "adaptive preferences." Chronic threats to the esteem of one's identity can significantly undermine performance (Steele and Aronson, 1995; Hoff and Pandey, 2006).

This paper explores ways to mitigate the scarring effect of youth unemployment through workshops aimed at addressing two psychological risks of unemployment: threats to self-integrity and inhibited goal orientation. To test this these ideas, we conduct a field experiment in Addis Ababa with 1,766 out-of-work young people. We randomly assign participants to either a control group or one of two treatment arms. The first treatment group was invited to attend a three-hour workshop focused on self-affirmation. The second group was invited to a workshop that, in addition to self-affirmation, also focused on a goal-setting exercise. We follow all individuals fortnightly, through phone calls, for the subsequent four months, collecting data on economic indicators such as job search and employment as well as psychological indicators, such as locus of control and conscientiousness.

The first treatment is based on a strategy from psychology designed to help people cope with chronic threat to self-integrity is a practice called "self-affirmation" (Steele, 1988; Cohen and Sherman, 2014). It is based on the premise that negative thoughts and feelings caused by threats in the environment can be mitigated by the fact that people do not think of themselves in simply one way. Rather, people are concerned with developing a *global* mental model of themselves as morally and adaptively adequate (Steele, 1988). We can restore our sense of self-integrity by focusing on the values and aspects of our life that are going well, even if we do not resolve the threats we face (Sherman and Cohen, 2006). Self-affirmations are practices that emphasize people's moral and adaptive adequacy by expanding the mental model of the self and its resources to encompass roles or identities that sustain their sense of integrity. The most common approach to structure self-affirmations in the literature is for people to reflect upon their core values (McQueen and Klein, 2006). Core values are internalized standards that people use to self-evaluate (Rokeach, 1973). Self-affirmation exercizes have had positive effects on academic performance (Cohen et al., 2006), executive functions (Harris, Harris, and Miles, 2017), reducing health information avoidance (Howell, 2017) and self-control (Churchill et al., 2018).

Our second treatment combines self-affirmation with a session on goal-orientation. Managing self-image is not the only challenge that vulnerable groups face to achieving their goals. People must also construct a focused and realistic plan for meeting their objective, which may be challenging when competing priorities loom large. Goals represent focal points around which human action can be oriented (Fishbach and Ferguson, 2007). As Locke and Latham,

1990 write, "goals direct attention, effort, and action toward goal-relevant actions at the expense of non-relevant actions." The literature on goal orientation emphasizes an important nuance: fantasizing about a desired goal alone can lower the likelihood that individuals achieve goals (Oettingen and Mayer, 2002). However, when encouraged to contrast fantasies of goal attainment with the obstacles they faced, individuals become more likely to fulfil their goals in areas such as weight loss, exam performance, and social interactions. The combination of mental contrasting with implementation intentions (MCII) has been show to help people engage in self-regulatory strategies that support behavior change, increasing commitment towards the goal and the intensity of actions required to attain it (Oettingen, 2015).

The workshops are effective in increasing both job-search and employment for men, but not for women. We investigate why the effects vary by gender. We find suggestive evidence that the difference in the constraints that men and women face in accessing labour market opportunities at baseline. Women primarily report normative and structural constraints to searching for work: roughly half report childbearing or other family obligations. On the other hand, almost half of men report lack of hope or motivation as reasons for not searching for work. The workshops have large effects for men with low psychological agency (with a low measure of locus of control at baseline). For this sub-group, participating in the workshops increases the likelihood of being employed by almost 30 percent, the days worked by around 36 percent and earnings by 34 percent. For these men we also observe major improvements in their locus of control measure.

Ideas such as self-affirmation and goal-orientation have been tested in health and education domains in high-income countries, but their application in low-income countries is still emerging (Bank, 2015b). The results are informative for the design of active labour market policies in emerging economies that have rapidly growing urban environments with high levels of youth unemployment. First, psychological supports can help individuals to overcome psychological constraints that may be preventing them from entering the labour market. Second, these are most effective when targeted at those individuals with a low locus of control, and they can significantly increase their economic prospects. Third, the activation of women in the labour market requires addressing access to childcare as well as shifting the normative expectations that women should take primary responsibility for household obligations.

This paper is related to two strands of literature at the intersection of psychology and economics that focus on self-beliefs, goals, and aspirations. The first literature on the effect

of managing confidence in self-image on motivation, behaviour, and welfare (Loewenstein, Köszegi, and Murooka, 2019; Johnson and Mislin, 2011). Much of this research comes from behavioural economics and emphasizes the gap between subjective self-assessments and objective reality. One theme in this literature is the prevalence and adverse effects of *overconfident* self-assessments. These include imprecise expectations of future self-control (DellaVigna and Malmendier, 2006), decisions to make compensation based on relative performance (Dohmen and Falk, 2011), and CEO decisions to make new acquisitions (Malmendier and Tate, 2008).

A second literature, conversely, focuses on the role that positive self-evaluations play in normal, healthy functioning (Taylor and Brown, 1988). Much of this work comes from development economics and focuses on the psychology and aspirations of vulnerable groups in low-income settings (Garance and Debraj, 2017; Ray, 2006; Dalton, Ghosal, and Mani, 2016; Appadurai, 2004). To consider how social deprivation might affect individuals' psychology and motivation, the literature describes how "aspirations windows" are formed based on the zone of "similar," "attainable," reference points in individuals' environment (Ray, 2006). The values, behaviours, and accomplishments of reference points guide their' beliefs, preferences, and behaviour. Aspirations are conceived as a "capacity" that motivates people to strive (Appadurai, 2004). Aspirations can lead to inertia if the window is too low or frustration if it is too high. An emerging empirical literature provides evidence that role models that provide new reference points may indeed expand aspirations windows: shifting preferences and increasing productivity (Ghosal et al., 2020; Macours and Vakis, 2014; Riley et al., 2017; Tanguy et al., 2014).

3.2 Conceptual framework

To motivate the field experiment, we sketch an informal framework for interpreting the intended effects of the workshop. We propose to explore two mechanisms. First, the effect of self-integrity on perceived agency, which in turn effects behavior. Second, the effect of goal-orientation on conscientiousness, which in turn effects behavior.

Mechanism 1: Self-integrity and locus of control

Consider an unemployed young person deciding whether to search for work. They want a job but also know that their efforts may be in vain. An important consideration in determining whether to search for work is the degree of control they feel they have over the situation. This is captured by the locus of control construct (Rotter, 1966). If they feel that they

lack control, they will search less intensely. If they feel that they have control over their situation, they will search more intensely. Their sense of agency is determined not only by their search performance, but also by their sense of self-worth. If they have a higher sense of self-worth, setbacks and failures are interpreted alongside other self-evaluations. This provides a resilience that sustains perceptions of agency. In a context of poverty and unemployment, positive self-evaluations may be challenging to sustain. The workshop is intended to serve as an exogenous shock that refocuses their attention to aspects of their identity they value. This provides greater resilience, greater agency, and ultimately leads to greater search intensity.

Mechanism 2: Goal-orientation and conscientiousness

Just as before, an unemployed young person must decide whether to search for work. Another important consideration in determining whether to search for work is whether they have the capacity to be goal directed, to plan, and to be able to control impulses and delay gratification. These capacities are captured by the conscientiousness construct (Roberts et al., 2009), and will search more intensely. These abilities will be low if the value of finding work is vague or unspecified or if the obstacles that the job seeker is likely to face are unclear. Because they are unemployed and have given up on searching for work, the advantages of employment are distant and the path to gaining employment is unspecified. The workshop module on goal-orientation is intended to make their employment goals more salient and the pathway to achieving them more thought through.

3.3 Intervention and experiment

The study took place between April and November 2018 in Addis Ababa. Addis is the capital and largest city in Ethiopia, a country with over 109 million people. Ethiopia has one of the fastest rates of economic growth in Africa and although it still a predominantly rural country, it is rapidly urbanizing at a rate of about 5.4 percent a year (Bank, 2015a; Fund, 2019). This has been accompanied by a large scale shift over the past 15 years from from agriculture to services and industry, with agriculture's share of GDP declining from 55 percent to 35 percent between 2003 and 2018 (Bank, 2019). Despite the country's strong economic performance, labour market outcomes are surprisingly poor in urban areas, in particular for young people. In 2018, 25.3 percent of young people were employed, compared against the national average of 19.1 percent. Among young people, there is a large gender

gap in the unemployment rate. Over 30 percent of young women and 19 percent of young men were unemployed in 2018 (Bank, 2019).

A number of interventions have been tested in Ethiopia to understand what are the binding constraints to sustainable employment. Franklin (2018) investigated whether search frictions (transport costs) constrain job seekers using a field experiment. He found that transport subsidies increases the likelihood of finding permanent employment by 6 percentage points in the short run. In a related intervention, Abebe et al. (2021a) find long lasting positive effects on earnings, job satisfaction and employment duration of job application workshops that helped applicants to signal their skills to their potential employers. On the other hand, job fairs designed to reduce matching frictions have had little effect. Employers over-estimate the quality of talent while low-skilled job-seekers have higher reservation wages than firms are prepared to pay (Abebe et al., 2021b).

3.3.1 Target population and recruitment

We define young people or "youth" as people between the ages of 18 and 30.[1] Approximately one in three of the unemployed are not actively looking for a job. An important consideration for policymakers is why people are not searching and whether cost effective intervention can help. To develop a preliminary understanding of this, we conducted qualitative work with youth in Addis Ababa. The results suggested that many young people do not search because they lack motivation and belief in their ability to change their circumstances. To explore this hypothesis further, we conducted a randomized field experiment to investigate the effects of providing psychological supports to young people on their on employment outcomes.

Our goal was to recruit 1,800 low income young men and women who were functionally out of work and not enrolled in education. We worked with local Woreda administrators to identify records of young people in the area, but following a review determined that study participants would have to be recruited directly. We recruited participants from seven Woredas in two of the poorest sub-cities of Addis Ababa.

Enumerators conducted a listing exercise to assess eligibility and consent. We screened 2,500 people against three conditions. First, they had to be between 18-30 years of age. Second,

[1] The Urban Employment Unemployment Survey statistics are for young people between the ages of 15 and 29. We assume that the unemployment rate for young people between ages 18 to 30 isn't significantly different from the reported statistics.

they could not be enrolled or attending school of any kind. Third, they could not have been engaged in full-time paid work in the last week. 1,979 individuals were eligible and invited to participate in the program. The program was described as a 3-hour workshop on improving your mindset. Participants were given lunch as part of the workshop. Of those, 1,766 individuals agreed to participate and provided consent.[2] 1,979 individuals were eligible and invited to participate in the program. Of those, 1,766 individuals provided consent.

3.3.2 Experimental Design

Recruited participants were stratified by gender and randomly allocated to one of three arms: (i) a pure control (C); (ii) a self-affirmation workshop (SA); or (iii) a self-affirmation + goal setting (GS). A separate randomization took place for each Woreda, thus we implicitly blocked by Woreda. This was done sequentially Woreda by Woreda, so that recruitment occurred at different dates for each Woreda. Participants assigned to the treatment groups were given time slots for the corresponding workshops.

3.3.3 Interventions

The two treatment arms were designed on the basis of psychological interventions that have been found to be effective before in other contexts.Cohen and Sherman (2014) provide a summary of the literature on self-affirmation theory and Oettingen (2015) provides a summary of the literature on goal setting through mental contrasting/implementation intentions (MCII). To ensure that the workshops were suitable for the context, we worked closely on the curriculum with a local organization that specializes in youth counseling and this same organization implemented the workshops. On this basis, we added discussions on role models to help participants focus on goals and on locus of control to address the obvious external constraints that participants faced. All workshops were led by a facilitator trained in therapy, with the support of two volunteer assistants that worked in the respective Woreda office. They took place in a local community space such as the youth center or a nearby school. All workshops lasted approximately three hours.

[2] An important decision was whether to compensate youth for participating in the workshop. We determined not to pay participants for two reasons. First, it would not be feasible to pay participants in a scaled program run by the government (which was a scenario we wanted to mimic). Second, workshops were delivered in the same locality from which participants were recruited. Thus we did not anticipate participants to have to incur travel costs. Due to the timing, the workshops overlapped with lunch, so we determined to provide lunch. This was a service we expected the government could provide.

The self-affirmation workshop: The workshop applied insights from self-affirmation theory to encourage participants to develop a belief in their own capacity to affect change in their economic lives. Self-affirmation entails mentalizing aspects of one's self-identity that one is proud of and values. The activities for this were based on the procedures set out in the Supporting Online Materials by Cohen et al. (2006).[3] The workshop included modules in which the individual reflected on (i) the role models in their lives, (ii) how their role model relates to their values, and (iii) the external and internal constraints they face to getting a job.

The goal-setting workshop: The workshop included identical material on self-affirmation and was supplemented with an additional module on goal setting. The particular application of goal setting was informed by the theory of mental contrasting and implementation intentions by Gabriele Oettingen (for summary, see Oettingen (2015); for practical online examples see woopmylife.org). Through this module, participants went through four specific steps: (i) establishing a wish or goal; (ii) determining the aspects of their wish that they would value; (iii) picturing the obstacles to the goal; and (iv) determining a plan.

The interventions took place either in the morning or in the afternoon. To ensure that the effects were not driven by the time of day, we varied whether workshops took place in the morning and afternoon within each Woreda (e.g., if the self-affirmation workshop is in the morning on one day, it would be in the afternoon on the next day). The workshops were delivered by the same implementing organization but different facilitators.

Balance

We test for balance on outcome variables and covariates across treatments and control group. We find that the arms are balanced on virtually across all dimensions of interest (both covariates and outcomes), as can be seen in Table 3.1 and Table 3.2. For each variable of interest, we present the test of the null hypothesis that all experimental groups are balanced. We only reject this hypothesis for a 6 percent difference between the literacy rate of the control and the SA treatment group. Overall, we feel confident that the samples are balanced.[4]

[3] We note that strategies and activities for self-affirmation vary across studies. For example Hall, Zhao, and Shafir (2014) invite people to think of things they are proud of. We drew on Cohen et al. (2006) because this was the original study and because Cohen had conducted multiple studies on the intervention and authored a review on the subject in the *Annual Review of Psychology*. The methods we drew on can be found here.

[4] Note that the survey and the intervention was restricted to people who didn't have a full time job. People with part time jobs were included in the sample. Hence, the indicator for employment in the last 7 days is not zero. Moreover, there was a lag of about 2 weeks between the initial screening and baseline survey

TABLE 3.1: Balance on covariates: Full Sample

Variable	(1) Control Mean/SE	(2) Goal setting Mean/SE	(3) Self affirmation Mean/SE	(4) Total Mean/SE	T-test P-value (1)-(2)	(1)-(3)	(2)-(3)
Age	23.180	23.095	23.261	23.178	0.674	0.689	0.414
	(0.142)	(0.141)	(0.146)	(0.083)			
Female	0.579	0.588	0.581	0.583	0.761	0.937	0.824
	(0.020)	(0.020)	(0.021)	(0.012)			
Married	0.331	0.334	0.318	0.328	0.919	0.640	0.571
	(0.019)	(0.019)	(0.019)	(0.011)			
Born in Addis Ababa	0.671	0.639	0.670	0.660	0.251	0.971	0.271
	(0.019)	(0.020)	(0.020)	(0.011)			
Years of education	8.379	8.446	8.668	8.496	0.752	0.166	0.283
	(0.151)	(0.149)	(0.143)	(0.085)			
Training programs	0.166	0.187	0.166	0.173	0.343	0.989	0.341
	(0.015)	(0.016)	(0.015)	(0.009)			
Can read and write	0.932	0.937	0.957	0.942	0.718	0.062*	0.133
	(0.010)	(0.010)	(0.008)	(0.006)			
N	601	587	578	1766			

Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

TABLE 3.2: Balance on dependent variables: Full Sample

Variable	(1) Control N	Mean/SE	(2) Goal setting N	Mean/SE	(3) Self affirmation N	Mean/SE	(4) Total N	Mean/SE	T-test P-value (1)-(2)	(1)-(3)	(2)-(3)
Searched for work in last 7 days?	601	0.286 (0.018)	587	0.302 (0.019)	578	0.280 (0.019)	1766	0.289 (0.011)	0.562	0.822	0.425
Number of days spent searching (last week)	601	0.928 (0.069)	587	0.910 (0.066)	578	0.908 (0.068)	1766	0.916 (0.039)	0.844	0.836	0.988
Number of hours spent searching (last time searching)	601	0.614 (0.047)	587	0.634 (0.048)	578	0.640 (0.054)	1766	0.629 (0.029)	0.770	0.714	0.929
Worked last 7 days	601	0.183 (0.016)	587	0.162 (0.015)	578	0.161 (0.015)	1766	0.169 (0.009)	0.334	0.315	0.965
Days worked last week	601	0.539 (0.057)	587	0.496 (0.056)	578	0.452 (0.052)	1766	0.496 (0.032)	0.587	0.256	0.563
Number of hours worked, last day spent working	601	0.970 (0.100)	587	0.853 (0.092)	578	0.808 (0.092)	1766	0.878 (0.055)	0.391	0.233	0.726
Amount earned in last week	110	315.909 (31.056)	95	352.105 (39.839)	93	355.914 (44.219)	298	339.933 (21.931)	0.469	0.450	0.949

Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Compliance

Overall, around 58 percent of the sample assigned to the treatment status attended the workshop. We note that although the rate is low, it is broadly in line with other similar field experiments of this kind. Moreover, it was also likely affected by the fact that we did not provide financial incentives to attend the workshop due to the concern that it would not be a scalable policy component.[5] However, other reasons like the duration of the workshop, low interest or faith may be responsible for the low participation in the treatment.[6] Thus, the intent-to-treat effects that we report are likely a conservative estimates of the impact of the program. The compliance within the treatment allocation was also successful and with the exception of one person, all participants attended the workshop to which they were assigned. Similarly, as Table 3.11 shows, only 2 people allocated to the control attended a workshop. People who turned up to the workshop were more likely to have searched in the past two weeks.[7]

Phased Implementation

To limit or avoid equilibrium effects, we worked in 7 separate Woredas so that study participants represented only a small facilitators of the local labour supply. We also wanted to ensure that Woredas were receiving the same facilitators, so we moved from one Woreda to the next every two weeks. This is described in Table 3.10 in the appendix.

3.4 Data

We attempted to survey participants 7 times. We conducted a baseline survey in person prior to the intervention and 6 phone surveys following the intervention. We collected baseline data on basic demographic and socioeconomic characteristics (such as gender, education level, marital status, etc.), time use, employment and job search. In our sample, 58 percent are women, one third are married, and two-thirds were born in Addis Ababa. Almost all participants are literate and the average years of education is 8.4. This means that, on

during which some individuals may found jobs. Individuals may have under-reported their employment status hoping to take advantage of some welfare schemes.

[5] The workshops were conducted in the neighbourhood and were walking distance from all participants so we did not provide a transport subsidy.

[6] Note that the response rate for the follow-up surveys, which was incentivized by airtime of 15 Birr per call, was approximately 90 percent.

[7] For balance tests, see **??** and **??** in the appendix.

average, participants completed the primary cycle (level) and almost half a year of secondary level.[8]

Following the intervention, we conducted a total of six phone follow-up interviews approximately two weeks apart over a period of four months (16 weeks). We document the time elapsed between the intervention and each wave in Table 3.12. After the sixth follow-up took place for a particular individual, no further data was collected (despite the fact that in some cases the project continued to run in the other Woredas). As can be seen in Table 3.13, attrition stood at around 10 percent for any given wave, and there was no heterogeneity across treatments.

The phone surveys included questions on job search and economic activity, our two main outcomes of interest. We rotated modules on locus of control, conscientiousness, and time preferences through the waves. We capture these to identify the potential psychological mechanisms through which the workshop might affect job search and employment outcomes. The outcome measures collected in each wave are described in Table 3.14.

We use self-reported data. The primary concern with self-reported data in field experiments is that study participants may misreport data either intentionally or unintentionally. Study participants did not have an incentive to misreport, nor do we believe they may have mistakenly perceived there to be an incentive for misreporting due to the fact that there was no compensation for participating in the intervention. To reduce the likelihood that participants misremember their experiences, we asked concrete questions about the relatively recent past. Because the timing of the phone calls was arbitrary (participants could be called on any day of the week), we extrapolate that on average these results are representative.

3.4.1 Outcomes of Interest

We are interested in five main outcomes. Our primary outcomes are job search and economic activity. Our secondary outcome is earnings. We also test three psychological mechanisms: locus of control, conscientiousness, and time preferences. We construct summary indexes for both of the primary outcomes based on the procedure developed by Anderson (2008). The summary index is a weighted mean of the standardized outcomes (by the control group standard deviation), where the weights are designed to maximize the information contained

[8] Available at https://wenr.wes.org/wp-content/uploads/2018/11/WENR-1118-Country_Profile-Ethiopia-1.png.

in the index. When studying the primary outcomes, we use sharpened q-values to adjust for multiple hypothesis testing, following Benjamini, Krieger, and Yekutieli (2006).

Primary Outcomes

The primary aim of the interventions is to encourage participants to engage in the labour market, either by looking for a job, getting a job, or engaging in self-employment.

Job search: To capture study participants' job search intensity, we asked them whether or not they had searched for work in the past week. If they had searched, we asked them how many days they had spent searching and how many hours they had searched on the last day they had searched. At baseline, only about 28 percent of people were searching for work. Among those who were searching for work, spent on average 3 days searching. On a given day, they would typically search for 2 hours.

Work: To capture study participants' work search intensity, we asked them whether or not they had engaged in productive activity last week. This could include wage and self employment. If they had worked, we asked them how many days they had spent working and how many hours they had worked on the last day they had worked. About 17 percent of the sample had worked in the last week. Among those who worked, they had worked for 3 days in the past week and worked for about 5 hours on a given day.

Secondary Outcome

In a secondary analysis we will look at the earnings (both from wage employment and self-employment) of individuals in the past seven days. At baseline, participants reported earning about 340 Birr in the last week. At the time (November 2018), this was worth roughly $12 USD.

Mechanisms

The channel through which we hypothesize behavioural change takes place is psychological. To capture and investigate the psychological mechanisms, we look at three outcomes: (i) locus of control, (ii) conscientiousness, and (iii) time preferences.

1. *Locus of control*: Locus of control is a construct that measures the degree to which people believe that they have control over the outcome of events in their lives, as opposed to external forces beyond their control. It is conceptualized on a spectrum from internal (a belief that one has personal control over their life) to external (a

belief that one's life is controlled by factors that he or she cannot influence). We follow Caliendo, Cobb-Clark, and Uhlendorff (2015) to construct a summary index of locus of control, where a higher value represents that the individual beliefs she has more control over their life. Participants were asked their level of agreement with 9 statements (see Table 3.17 in the Appendix) on a 4 point scale (strongly disagree, disagree, agree, strongly agree), and the index was constructed by summing the values of responses in line with the psychological construct and subtracting those in conflict.[9]

2. *Conscientiousness*: We also measured Conscientiousness, a personality construct commonly used in psychology which captures the tendency to be goal directed, to plan, and to be able to control impulses and delay gratification (Roberts et al., 2009).[10] We applied the battery of ten questions used by Blattman and Dercon (2018) presented in Table 3.16 of the Appendix 1.[11] As with locus of control, a summary index was constructed by summing the values (out of a four point scale) from statements affirming behaviours predicted to be in line with the construct and subtracting statements affirming behaviours predicted to be in conflict with the construct (see signs in Table 3.16). A higher value of the index signals a higher degree of conscientiousness.

3. *Time preferences*: Similarly, we measured participants' time preferences. Time preferences (time discounting) is the relative value participants put on receiving a good (benefit) now compared to receiving it at a later date. To obtain individuals' discount rate (impatience), we follow Falk, Dohmen, and Huffman (2016). This is a method created for surveys that has been validated in the lab to be applied to surveys globally (Falk et al., 2018). Specifically, we used a staircase procedure (see Figure 3.3), on which a person lands at a particular time discounting level. Higher levels of this measure reflects a higher level of impatience.[12]

3.5 Empirical Strategy

To estimate the effects of providing psychological tools on individual level outcomes related to job search and employment, we use an ANCOVA specification, following McKenzie (2012). Our choice is based on the low auto-correlation of the outcome variables, and it takes

[9] The Chronbach's alpha on this is 0.55 which is below typical standards for internal validity. Weak internal validity is a common challenge associated with applying psychological tools in low-income countries (Laajaj et al., 2019).

[10] It is part of the Big Five model (Five Factor Model) of personality, a validated model that also includes extraversion, neuroticism, openness to experience, and agreeableness (see McAdams, 2008 chapter 5).

[11] The Chronbach's alpha on this is 0.79 which approximately the typical standard for internal validity, and fairs well compared to similar attempts to apply psychological constructs developed in WEIRD contexts to low income countries.

[12] We did not use incentives to elicit time preferences. Falk et al. (2018) do not find that this affects answers.

advantage of all the follow-up surveys collected.[13] A base specification pools both treatment groups as follows:

$$Y_{ir,t} = \beta \cdot T_i + \delta_t + \gamma_r + \lambda_{rt} + \tau_w + \alpha \cdot Y_{i,0} + \eta \cdot X_{i,0} + \mu_{i,t} \tag{3.1}$$

Here $Y_{ir,t}$ is the outcome of interest for individual i in Woreda (district) r at week t of the project (out of a total of 28 weeks). T_i is a dummy that indicates that the individual was invited to attend a workshop, capturing the intent to treatment effect (ITT), δ_t are time (week) fixed effects, γ_r are Woreda fixed effects, λ_{rt} are Woreda time-variant effects, τ_w are wave fixed effect (where $w = 1, ..., 6$ describing which follow-up was collected for individual i at time t) and $Y_{i,0}$ is the baseline measure of the outcome of interest. Additionally, following the recommendations of Bruhn and McKenzie (2009), we will control for a set of baseline covariates $X_{i,0}$.[14] Errors are clustered at the individual level.

Our preferred specification separately identifies the effect of the self-affirmation workshop (SA) and the goal-setting (GS) workshop, as follows:

$$Y_{ir,t} = \beta_1 \cdot SA_i + \beta_2 \cdot GS_i + \delta_t + \gamma_r + \lambda_{rt} + \tau_w + \alpha \cdot Y_{i,0} + \eta \cdot X_{i,0} + \mu_{i,t} \tag{3.2}$$

Where SA_i is a dummy that indicates participant was invited to the self-affirmation workshop, while GS_i is a dummy that indicates participant was invited to the goal setting workshop. We continue to cluster errors at the individual level. We test if the effect of the two treatment arms are statistically different. In addition, we explore how the impact of the intervention at the individual level varies along two dimensions: gender and locus of control.

3.6 Results

3.6.1 Average Intent-to-Treat Effects of Interventions

We first investigate the average treatment effect of the interventions on primary and secondary outcomes. For all tests, we use the regression specifications outlined in section 3.5. We use an intention to treat analysis. In other words, we assess the impact of *being invited*

[13] Job search intensity is known to vary throughout the unemployment spell and permanent stable employment is rare in Ethiopia.

[14] Our list of baseline controls include the following characteristics: age, gender, born in Addis Ababa, marital status, education level, literate and has received training in the past.

to the workshops, independent of whether the person actually attended. Since 58 percent of invited individuals attended the workshop, it is likely that this underestimates the effect of the treatment on those who were treated.

Table 3.3 presents the results for the primary outcomes. We find that the workshops have a significant positive effect on job search, but not on the secondary outcomes of employment and profit. Columns 2 and 3 present the separate treatment effects of the self-affirmation and goal-setting workshops. The results suggest that the SA workshop has a significant impact on job search at the 5 percent of confidence level. The ITT effects identified are of limited magnitude, and they are not robust to the multiple hypothesis testing adjustment (q-values in brackets).

Table 3.3: Full sample: Main outcomes

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search index	0.044	0.019	0.069**	-0.000	9620
	(0.114)	(0.549)	(0.034)		
	[0.227]	[0.549]	[0.204]		
Work index	0.050	0.059	0.041	0.000	9618
	(0.157)	(0.157)	(0.321)		
	[0.227]	[0.440]	[0.440]		
Earnings	9.736	10.785	8.662	92.112	9618
	(0.227)	(0.261)	(0.367)		
	[0.227]	[0.440]	[0.440]		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis and the a q-value in brackets. q-values, which adjust for multiple hypothesis testing, are obtained using the simes approach. ***: p<0.01, **: p<0.05, * : p<0.10.

Table 3.4: Full sample: Main outcomes, components

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search indicator	0.003	-0.011	0.017	0.574	9620
	(0.822)	(0.496)	(0.277)		
Job search days	0.119***	0.100*	0.138***	1.523	9620
	(0.007)	(0.054)	(0.008)		
Job search hours	0.066**	0.034	0.100**	1.138	9620
	(0.043)	(0.370)	(0.011)		
Work indicator	0.016	0.020	0.012	0.205	9618
	(0.237)	(0.211)	(0.445)		
Work days	0.106	0.138*	0.073	0.843	9618
	(0.116)	(0.086)	(0.347)		
Work hours	0.136	0.146	0.125	1.255	9618
	(0.167)	(0.204)	(0.278)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

TABLE 3.5: Full sample: Mechanisms

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Locus of control	0.140	0.181	0.098	-3.355	3193
	(0.264)	(0.214)	(0.505)		
Conscientiousness	0.257*	0.253*	0.261*	4.469	3223
	(0.053)	(0.096)	(0.088)		
Time preferences	-0.093	-0.237	0.054	7.101	4803
	(0.684)	(0.376)	(0.838)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

As discussed above, the GS workshop contained the same material of the SA workshop plus an additional goal-setting exercise. Thus, adding the additional content to the intervention appears to have dampened its efficacy. One explanation is that the goal-setting material undermined individuals' effort. However, given that the GS techniques employed have been effective in other areas, we do not believe this to be likely. Another explanation is that the GS workshop attempted to fit too much material into a restricted time frame (recall that the two workshops were the same length by design). This could have diluted the focus of the content and in extreme circumstances, cancelled the effects of the other modules. The effects of goal setting behaviour may not be tangibly visible in the short run and follow ups may be required for a longer period of time to observe effects. This is a potentially important lesson for the application of behavioural insights – while the marginal financial cost of adding materials is low, other costs, more challenging to observe and measure such as time and focus, may impact the effectiveness of policy design choices.

The effect on job search is explained by an increase in the intensity of job search rather than an increase at the extensive margin. Participants in the control group searched 1.5 days on average and about 1.1 hours. The number of days individuals searched for a job (during the last week) increased by almost 8 percent for those invited to the workshops, reaching 1.64 days (see Table 3.4)[15]. The time spent searching remains consistently low across all groups. On days that people search for jobs, they spend nearly 2 hours on, average, searching – a figure that barely changes across treatment groups. This perhaps reflects the nature of the search tasks. Within treatment arms, the effects are still being driven by changes at the intensive margin. The SA treatment increases the number of search days by 9 percent (Column 3). The GS treatment has a positive effect on the number of search days, but this goes away when integrated into the index. An invitation to the GS workshop increases the

[15]Note that statistics on days and hours spent working are calculated for the group as a whole – i.e., both participants who searched and did not search at the time of data collection.

days of employment by 16.37 percent but there is no significant effect of the SA workshop by itself.

In terms of mechanisms, the findings suggest that workshop increased conscientiousness by around 6 percent (see Table 3.5). The results are consistent for the two arms. In all cases, the impact is statistically significant at a 10 percent confidence level. Given the non-empirical and normative nature of conscientiousness and the small size of the point estimate, these results may be artefacts of demand effects.

3.6.2 Heterogeneity and Sub-Group Analysis

To better understand how the interventions worked, we conducted sub-group and heterogeneity analysis. Sub-group analysis enables us to transparently test whether the treatment had a causal effect on particular demographic subsets of the sample. Heterogeneity analysis enables us to assess the magnitude and statistical significance of differences between sub-groups. We focus here on gender and locus of control.[16]

First, consider impacts by gender. Table 3.6 shows results on the main outcomes for women. The intervention had no effect on any outcome for women. The results by item and psychological mechanism for both women and men are presented in section 3.8.4 along with the heterogeneity analysis. They also show no impacts for women. Table 3.7 shows results on the main outcomes for men. The self-affirmation treatment has a statistically significant effect on both search and work. As before, the results are not robust to the adjustment for multiple hypothesis testing. Table 3.24 shows that the effects are at both the extensive and intensive margin for search. The effects on work are at the intensive margin with a roughly 24 percent increase in work hours. Table 3.25 shows that there are also statistically significant effects on conscientiousness and locus of control. As shown in section 3.8.4, the heterogeneity in treatment impacts by gender are not statistically significant.

One explanation for the lack of impact for women is the distinct nature of the constraints faced by women: social norms and the role that women are expected to play by society. At baseline, almost 50 percent of women report that they are not searching for work because they are fulfilling family responsibilities (which includes looking after children/elderly, among other things), compared to just 19 percent of men, as seen in Figure 4.1 and Figure 3.2.

[16] In our pre-analysis plan, we pre-specified that we would look at gender, education, marital status, and self-reported income. Analysis based on locus of control should be considered exploratory as this was not pre-specified. The theoretical rationale for inclusion is that a consistent finding over over the past decade has been that interventions targeting mental models have generally been most effective for participants where the mental models has been maladaptive.

Table 3.6: Women: Main outcomes

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search index	0.010	-0.017	0.039	-0.125	5664
	(0.770)	(0.679)	(0.361)		
	[0.770]	[0.817]	[0.722]		
Work index	0.026	0.069	-0.018	-0.128	5664
	(0.534)	(0.164)	(0.708)		
	[0.770]	[0.722]	[0.817]		
Earnings	6.823	11.300	2.146	60.406	5664
	(0.410)	(0.258)	(0.817)		
	[0.770]	[0.722]	[0.817]		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis and the a q-value in brackets. q-values, which adjust for multiple hypothesis testing, are obtained using the simes approach. ***: p<0.01, **: p<0.05, * : p<0.10.

Table 3.7: Men: Main outcomes

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search index	0.091**	0.075	0.108**	0.177	3955
	(0.038)	(0.146)	(0.034)		
	[0.114]	[0.292]	[0.204]		
Work index	0.074	0.033	0.116*	0.182	3953
	(0.196)	(0.626)	(0.090)		
	[0.291]	[0.626]	[0.270]		
Earnings	15.416	9.576	21.355	136.832	3953
	(0.291)	(0.585)	(0.230)		
	[0.291]	[0.626]	[0.345]		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis and the a q-value in brackets. q-values, which adjust for multiple hypothesis testing, are obtained using the simes approach. ***: p<0.01, **: p<0.05, * : p<0.10.

Thus, to address the exclusion of women from the labour market, balancing household responsibilities between genders and childcare alternatives may be necessary.[17]

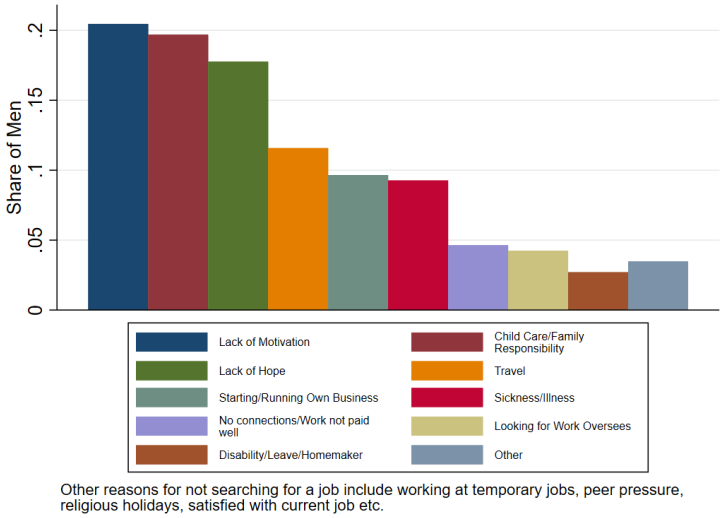


FIGURE 3.1: Reasons for Not Searching for Jobs, as Reported by Men

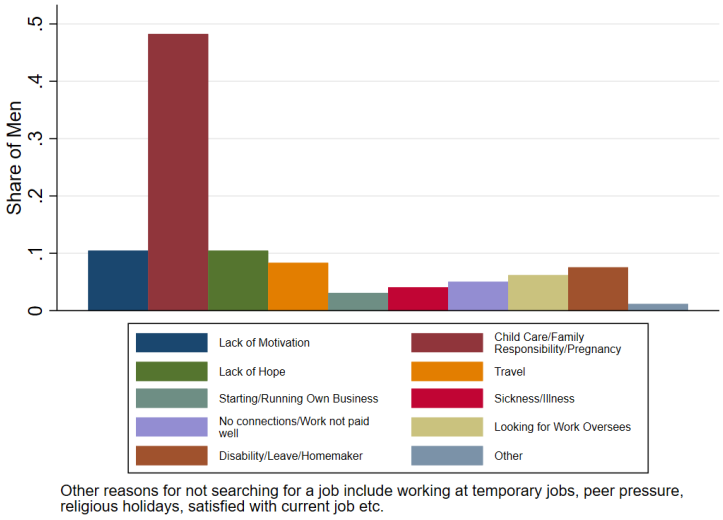


FIGURE 3.2: Reasons for Not Searching for Jobs, as Reported by Women

[17] The effect of the treatment is insignificant for women with and without children. This suggests that the added responsibility of taking care of children doesn't explain the lack of effect among women. Expectations related to other household responsibilities may also be at play.

In order to further understand the impact on men, we look at impacts based on locus of control at baseline. We report the main outcomes of the workshops for men with a low locus of control in Table 3.8.[18] For this sub-sample, we find that considerable positive effects on both employment and earnings. There is a large and statistically significant impact on employment. Similarly, earnings increase by a third for the combined treatment, mainly explained by the goal-setting workshop; a result which is robust to the multiple hypothesis testing adjustment. For this particular sub-group, the treatments have an effect on both the extensive margin of employment (employment status) and intensive margin (number of days worked), as shown in Table 3.36. In terms of mechanisms, the evidence suggests that both workshops had a positive effect on the locus of control of this men (combined effect of more than 20 percent), see Table 3.37. Note, however, that there are no effects on search.

TABLE 3.8: Men with low Locus of Control: Main outcomes

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search index	0.014	0.016	0.013	0.205	1769
	(0.824)	(0.834)	(0.862)		
	[0.824]	[0.862]	[0.862]		
Work index	0.230***	0.227**	0.235**	0.158	1769
	(0.007)	(0.023)	(0.027)		
	[0.021]	[0.081]	[0.081]		
Earnings	42.478**	43.862*	40.845	126.388	1769
	(0.033)	(0.065)	(0.115)		
	[0.049]	[0.130]	[0.173]		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis and the a q-value in brackets. q-values, which adjust for multiple hypothesis testing, are obtained using the simes approach. ***: $p<0.01$, **: $p<0.05$, * : $p<0.10$.

Among men with high locus of control, we find inverse results. There is a relatively large and statistically significant effect on job search. The effect on job search is driven by the SA treatment. One explanation for this is that men with a higher locus of control are more educated at baseline. While the treatment may increase job search, it is possible that men were not readily accepting jobs due to their high initial investment. Demand for labour in specialized industries could also be responsible for the insignificant effect on the work index for this subsample. The job search index for people who received an invitation is 1.2 times the control (see Table 3.8). Table 3.36 shows that the intensive margin of job search increases by 14.6 percent and that the effects are larger and more significant for the SA workshop than the GS workshop (17.3 vs 11.7 percent). Not surprisingly, for this group

[18] That is men with a locus of control measure in the bottom the 50 percent of the distribution among men at baseline.

the psychological dimension that improves is conscientiousness, mostly as a result of the SA intervention (see Table 3.32).

Table 3.9: Men with high Locus of Control: Main outcomes

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search index	0.143**	0.093	0.190***	0.157	2186
	(0.023)	(0.191)	(0.010)		
	[0.069]	[0.566]	[0.060]		
Work index	-0.008	-0.089	0.070	0.199	2184
	(0.921)	(0.318)	(0.431)		
	[0.988]	[0.566]	[0.566]		
Earnings	0.290	-13.828	13.607	144.169	2184
	(0.988)	(0.562)	(0.566)		
	[0.988]	[0.566]	[0.566]		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis and the a q-value in brackets. q-values, which adjust for multiple hypothesis testing, are obtained using the simes approach. ***: p<0.01, **: p<0.05, * : p<0.10.

3.7 Conclusion

Roughly a quarter of young people in urban Ethiopia were neither employed nor searching for work in 2018, despite a fast growing economy. As suggested by qualitative work, low motivation and a perceived helplessness over their fate is a constraint for young people. This study explores whether influencing vulnerable young people's mental models of themselves can affect their labour market outcomes and mindsets. We find that the supports have effects over a 4 month period on job search, employment and earnings. We note that these are relatively short term.

We find that interventions that target people's mental models, and particularly that seek to enhance self-affirmation or the belief that one has control over one's life, enhance labour market outcomes. In particular, being invited to the workshop increases job search efforts, mainly on the intensive margin, albeit the average intent-to-treat effects are small. However, this hides an important difference between men and women. Our results show that the workshops are effective for men, but they have *no effect for any women's outcomes*. For men, getting invited to the workshop increases both job search efforts. This gender gap is consistent with the set of constraints that men and women report in trying to access labour market opportunities.

Interestingly, we find that both for the entire sample and the male sample, the results are driven by the self-affirmation workshop, rather than the goal-setting. Given that, intentionally, both workshops had the same duration, we attribute this result to an attempt to fit too much material in a restricted time frame. In addition, goal-setting techniques function better when there are periodical reminders, a characteristic that was missing from our experimental design. Similarly, for both samples, the interventions seem to be operating through an increase in conscientiousness, or the desire to do tasks well and honour obligations, among participants. In the case of men, it also seems to be operating through an increase in the locus of control, the degree to which individuals believe they are in control of the outcomes in their lives.

When we analyze the effect of the intervention on the men with low locus of control (at baseline), we find it strikingly effective. Being invited to the workshop has now a large effect on economic activity/employment, around 60 percent higher than for the male sample, and on earnings, which increase by 30 percent with respect to the control group. Unsurprisingly, for these men we observe major improvements in the locus of control measure.

Our results provide important lessons for the design of active labour market policies in urban Ethiopia as well as other urban African contexts. First, interventions that aim to activate women in the labour market require to tackle the household responsibilities that often fell fully on women: housework and care-giving (for children and others). Without addressing these structural constraints, policies will likely be ineffective. Second, implementation matters in behavioral interventions. Sometimes, too much material can result counterproductive for participants, with negative implications for the effectiveness of the intervention. Finally, as with most policies, behaviorally informed intervention seems to work more effectively for those who need it the most, in this particular case, men with low psychological agency.

3.8 Appendix

3.8.1 Implementation and compliance

Table 3.10: Implementation period

Woreda	Date of implementation	N
Woreda 1 (Lideta 03)	17- 19 April, 2018	221
Woreda 2 (Kolfe13)	02-03 May, 2018	242
Woreda 3 (Kolfe 11)	16-27 May, 2018	265
Woreda 4 (Kolfe 08)	30-31 May, 2018	249
Woreda 5 (Kolfe 02)	13-14 June, 2018	231
Woreda 6 (Kolfe 03)	04-05 July, 2018	302
Woreda 7 (Lideta 04)	18-19 July, 2018	256

Table 3.11: Assignment and compliance

	Control	Treatment (GS+SA)	Goal-Setting	Self-Affirmation
Assigned to treatment	601	1165	587	578
Complied		677	330	347
(% that complied)		58%	56%	60%
Miss-compliance	2		0	1
(% of treatment share)	0%		0%	0%

3.8.2 Data Collection

Table 3.12: Time elapsed since intervention (by wave)

Wave	1	2	3	4	5	6
Weeks after workshop	2.5	4.6	6.9	9.3	11.4	13.9

Table 3.13: Response rates by treatment status and wave

	Control	%	Treatment 1	%	Treatment 2	%	Total
Baseline	601	100	587	100	578	100	1766
Wave 1	547	91	529	90.1	517	89.4	1593
Wave 2	554	92.2	539	91.8	522	90.3	1615
Wave 3	557	92.7	526	89.6	509	88.1	1592
Wave 4	545	90.7	530	90.3	527	91.2	1602
Wave 5	530	88.2	542	92.3	541	93.6	1613
Wave 6	552	91.8	525	89.4	534	92.4	1611
Total	3886		3778		3728		11392

Table 3.14: Outcome measures by wave

	Job Search	Economic Activity	Locus of Control	Conscientiousness	Time Preferences
Baseline	X	X	X	X	X
Wave 1	X	X	X		
Wave 2	X	X		X	
Wave 3	X	X			X
Wave 4	X	X	X		X
Wave 5	X	X		X	
Wave 6	X	X			X

Table 3.15: Balance on attrition

	(1) 0		(2) 1		(3) 2		(4) Total		T-test P-value		
Variable	N	Mean/SE	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(3)	(2)-(3)
Rate of response	3285	1.000 (0.000)	3191	1.000 (0.000)	3150	1.000 (0.000)	9626	1.000 (0.000)	N/A	N/A	N/A

Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 3.16: Conscientiousness Measures

Item	Obs	Sign	item-test corr.	item-rest corr.	interitem cov.	alpha
I am always prepared	1766	-	0.53	0.42	0.25	0.78
I get chores done right away	1766	-	0.29	0.1	0.27	0.82
I complete tasks successfully	1766	-	0.52	0.40	0.24	0.78
I waste my time	1766	+	0.54	0.36	0.23	0.79
I shirk my duties	1766	+	0.73	0.62	0.20	0.75
I leave things unfinished	1766	+	0.78	0.70	0.20	0.74
I don't put my mind on the task at hand	1766	+	0.77	0.67	0.20	0.75
I make a mess of things	1766	+	0.78	0.70	0.20	0.74
I need a push to get started	1766	+	0.64	0.48	0.21	0.77
Test scale (unstandardized mean)					0.22	0.79

3.8.3 Measurement

Table 3.17: Locus of Control Measures

Item	Obs	Sign	item-test corr.	item-rest corr.	interitem cov.	alpha
My life's direction depends on me.	1766	-	0.26	0.04	0.15	0.58
In comparison with others, I haven't achieved what I deserved to achieve.	1766	+	0.44	0.21	0.13	0.53
What can be achieved in life is mainly a result of fate or luck.	1766	+	0.42	0.17	0.13	0.55
If you are socially or politically active, you can influence social circumstance	1766	+	0.35	0.11	0.14	0.56
I often find that other people dictate my life.	1766	+	0.67	0.50	0.09	0.43
You must work hard to achieve success.	1766	-	0.29	0.17	0.14	0.54
When I encounter difficulties in life, I often doubt my abilities.	1766	+	0.67	0.50	0.09	0.44
The options that I have in life are determined by social circumstances.	1766	+	0.40	0.15	0.14	0.55
I don't have much control over what happens in my life.	1766	+	0.66	0.48	0.09	0.44
Test scale (Unstandardized mean)					0.12	0.55

Table 3.18: Primary outcome variables

Primary Outcomes	Family	Definition	Notes
		Job Search	
1	Job search	Summary index (weighted standardized) of the following sub-components.	
1.1	job search	Actively searched for a job last week	Dummy variable.
1.2	days search	Number of days searched for a job last 7 days.	Zero for those not working.
1.3	hours search	Number of hours searched for a job last time (day).	Zero for those not working. Winzorized at 99 percent.
		Economic activity	
2	Work	Summary index (weighted standardized) of the following sub-components.	
2.1	productive work	Engaged in productive activity last week (includes wage and self employment)	Dummy variable.
2.2	days worked 7days	Number of days engaged in productive activity last 7 days.	Zero for those not working.
2.3	hours per day	Number of hours engaged in productive activity last time (day).	Zero for those not working. Winzorized at 99 percent.
2.4	hours worked yesterday	Hours spent yesterday in productive activity (time use section)	Zero for those not working.

Table 3.19: Secondary outcome variable

Secondary Outcome	Family	Definition	Notes
1.1	day profit	Earnings in the last 7 days.	Takes the value of zero for those not working. Winzorized at 99 percent.

3.8.4 Heterogeneity

Gender

Table 3.20: Heterogeneity by gender specification 1

	Work Index	Job search index	Earnings
Any treatment	0.0692	0.0784	19.86
	(0.062)	(0.044)	(15.812)
Male	0.319***	0.247***	84.35***
	(0.066)	(0.048)	(14.887)
Female × Any treatment=1	-0.0661	-0.0593	-19.94
	(0.077)	(0.057)	(18.127)
Observations	9618	9620	9618

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.21: Heterogeneity by gender specification 2

	Work Index	Job search index	Earnings
Goal setting	0.0330	0.0623	13.73
	(0.072)	(0.052)	(18.707)
Self affirmation	0.106	0.0947	26.02
	(0.073)	(0.051)	(19.235)
Male	0.318***	0.247***	84.16***
	(0.066)	(0.048)	(14.888)
Female × Goal setting=1	0.0104	-0.0729	-8.385
	(0.091)	(0.066)	(21.488)
Female × Self affirmation=1	-0.144	-0.0449	-31.68
	(0.091)	(0.066)	(21.766)
Observations	9618	9620	9618

Standard errors errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Women

Table 3.22: Women: Main outcomes, components

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search indicator	-0.013	-0.025	-0.001	0.522	5664
	(0.472)	(0.236)	(0.962)		
Job search days	0.080	0.042	0.120*	1.293	5664
	(0.146)	(0.501)	(0.074)		
Job search hours	0.024	-0.013	0.063	1.008	5664
	(0.547)	(0.786)	(0.198)		
Work indicator	0.011	0.028	-0.007	0.145	5664
	(0.516)	(0.158)	(0.728)		
Work days	0.092	0.180*	0.000	0.673	5664
	(0.266)	(0.069)	(0.999)		
Work hours	0.019	0.130	-0.096	0.959	5664
	(0.869)	(0.337)	(0.466)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Table 3.23: Women: Mechanisms

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Locus of control	-0.164	-0.052	-0.280	-3.340	1878
	(0.318)	(0.786)	(0.144)		
Conscientiousness	0.172	0.262	0.081	4.587	1893
	(0.312)	(0.179)	(0.683)		
Time preferences	-0.103	-0.138	-0.068	9.128	2829
	(0.747)	(0.713)	(0.855)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Men

Table 3.24: Men: Main outcomes, components

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search indicator	0.025	0.010	0.041*	0.647	3955
	(0.217)	(0.663)	(0.088)		
Job search days	0.179**	0.193**	0.165*	1.846	3955
	(0.015)	(0.026)	(0.052)		
Job search hours	0.127**	0.106	0.150**	1.320	3955
	(0.021)	(0.102)	(0.022)		
Work indicator	0.023	0.008	0.037	0.289	3953
	(0.319)	(0.764)	(0.165)		
Work days	0.116	0.062	0.171	1.084	3953
	(0.275)	(0.621)	(0.175)		
Work hours	0.265*	0.129	0.404**	1.672	3953
	(0.098)	(0.494)	(0.038)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Table 3.25: Men: Mechanisms

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Locus of control	0.572***	0.522**	0.622***	-3.376	1315
	(0.002)	(0.017)	(0.005)		
Conscientiousness	0.326	0.188	0.469**	4.301	1330
	(0.109)	(0.422)	(0.045)		
Time preferences	-0.035	-0.323	0.261	4.276	1973
	(0.908)	(0.362)	(0.451)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Locus of control

TABLE 3.26: Heterogeneity by locus of control and gender specification 1

	Work Index	Job search index	Earnings
Any treatment	-0.128	0.0982	-17.67
	(0.127)	(0.094)	(28.844)
Low LoC	-0.0752	-0.0242	-16.10
	(0.076)	(0.057)	(14.174)
Male	0.302***	0.313***	77.76***
	(0.083)	(0.065)	(17.684)
Low LoC=0 × Any treatment=1	0.0608	0.0335	12.79
	(0.094)	(0.071)	(18.083)
Female × Any treatment=1	0.104	-0.0944	11.81
	(0.112)	(0.080)	(26.194)
Low LoC=0 × Male	0.0181	-0.114	9.116
	(0.121)	(0.090)	(26.544)
Low LoC=1 × Male × Any treatment=1	0.366*	-0.0918	68.85
	(0.154)	(0.113)	(35.471)
Observations	9618	9620	9618

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.27: Heterogeneity by locus of control and gender specification 2

	Work Index	Job search index	Earnings
Low LoC=0 × Male	0.0183	-0.114	9.149
	(0.121)	(0.090)	(26.555)
Low LoC=0 × Goal setting=1	-0.0339	-0.0885	1.533
	(0.148)	(0.112)	(35.408)
Low LoC=1 × Goal setting=0	0.0426	0.162	0.976
	(0.152)	(0.114)	(36.299)
Female × Goal setting=1	0.192	-0.0609	22.81
	(0.131)	(0.093)	(30.976)
Low LoC=1 × Male × Goal setting=1	0.386*	0.00360	67.40
	(0.181)	(0.133)	(42.576)
Low LoC=0 × Self affirmation=1	0.113	-0.00586	23.12
	(0.107)	(0.085)	(20.520)
Female × Self affirmation=1	0.0210	-0.124	1.643
	(0.132)	(0.095)	(31.108)
Low LoC=1 × Male × Self affirmation=1	0.353	-0.185	71.57
	(0.181)	(0.132)	(42.568)
Observations	9618	9620	9618

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Women with high locus of control

Table 3.28: Women with high Locus of Control: Main outcomes

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search index	0.020	0.012	0.027	-0.093	2673
	(0.709)	(0.842)	(0.670)		
	[0.709]	[0.842]	[0.804]		
Work index	0.079	0.099	0.058	-0.108	2673
	(0.210)	(0.166)	(0.446)		
	[0.315]	[0.624]	[0.669]		
Earnings	15.090	13.999	16.226	65.515	2673
	(0.206)	(0.312)	(0.263)		
	[0.315]	[0.624]	[0.624]		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis and the a q-value in brackets. q-values, which adjust for multiple hypothesis testing, are obtained using the simes approach. ***: p<0.01, **: p<0.05, * : p<0.10.

Table 3.29: Women with high Locus of Control: Main outcomes, components

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search indicator	-0.010	-0.010	-0.009	0.548	2673
	(0.724)	(0.753)	(0.775)		
Job search days	0.077	0.064	0.091	1.308	2673
	(0.335)	(0.490)	(0.357)		
Job search hours	0.050	0.032	0.068	1.033	2673
	(0.421)	(0.651)	(0.377)		
Work indicator	0.033	0.039	0.026	0.147	2673
	(0.184)	(0.170)	(0.373)		
Work days	0.174	0.214	0.133	0.731	2673
	(0.169)	(0.138)	(0.387)		
Work hours	0.163	0.221	0.103	1.047	2673
	(0.335)	(0.246)	(0.612)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Table 3.30: Women with high Locus of Control: Mechanisms

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Locus of control	0.083	0.289	-0.136	-2.564	884
	(0.711)	(0.241)	(0.629)		
Conscientiousness	0.456*	0.670**	0.233	4.878	896
	(0.053)	(0.011)	(0.404)		
Time preferences	-0.507	-0.542	-0.471	10.274	1334
	(0.270)	(0.323)	(0.377)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Men with high locus of control

Table 3.31: Men with high Locus of Control: Main outcomes, components

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search indicator	0.048*	0.026	0.068**	0.640	2186
	(0.100)	(0.436)	(0.040)		
Job search days	0.258***	0.207*	0.306***	1.771	2186
	(0.010)	(0.072)	(0.010)		
Job search hours	0.192**	0.116	0.263***	1.325	2186
	(0.017)	(0.191)	(0.007)		
Work indicator	-0.013	-0.043	0.016	0.289	2184
	(0.665)	(0.210)	(0.644)		
Work days	-0.018	-0.122	0.081	1.130	2184
	(0.902)	(0.471)	(0.634)		
Work hours	0.068	-0.185	0.307	1.774	2184
	(0.755)	(0.468)	(0.231)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Table 3.32: Men with high Locus of Control: Mechanisms

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Locus of control	0.352	0.403	0.305	-2.312	723
	(0.163)	(0.151)	(0.314)		
Conscientiousness	0.414	0.241	0.577*	4.931	728
	(0.114)	(0.426)	(0.051)		
Time preferences	-0.125	-0.240	-0.016	4.267	1095
	(0.765)	(0.630)	(0.973)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Women with low locus of control

Table 3.33: Women with low Locus of Control: Main outcomes

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search index	0.004	-0.048	0.056	-0.156	2989
	(0.930)	(0.393)	(0.319)		
	[0.930]	[0.590]	[0.590]		
Work index	-0.021	0.039	-0.080	-0.147	2989
	(0.713)	(0.573)	(0.178)		
	[0.930]	[0.688]	[0.590]		
Earnings	-2.470	5.637	-10.630	55.680	2989
	(0.826)	(0.697)	(0.348)		
	[0.930]	[0.697]	[0.590]		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis and the a q-value in brackets. q-values, which adjust for multiple hypothesis testing, are obtained using the simes approach. ***: p<0.01, **: p<0.05, * : p<0.10.

Table 3.34: Women with low Locus of Control: Main outcomes, components

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search indicator	-0.017	-0.045*	0.012	0.498	2989
	(0.492)	(0.100)	(0.666)		
Job search days	0.093	0.024	0.162*	1.280	2989
	(0.233)	(0.786)	(0.079)		
Job search hours	0.004	-0.048	0.057	0.986	2989
	(0.938)	(0.428)	(0.358)		
Work indicator	-0.008	0.017	-0.032	0.142	2989
	(0.731)	(0.532)	(0.172)		
Work days	0.013	0.140	-0.115	0.618	2989
	(0.900)	(0.294)	(0.285)		
Work hours	-0.101	0.053	-0.256	0.879	2989
	(0.516)	(0.781)	(0.121)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

TABLE 3.35: Women with low Locus of Control: Mechanisms

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Locus of control	-0.512**	-0.505*	-0.518**	-4.054	994
	(0.030)	(0.083)	(0.047)		
Conscientiousness	-0.020	0.052	-0.093	4.322	997
	(0.934)	(0.854)	(0.739)		
Time preferences	0.101	-0.077	0.280	8.066	1493
	(0.819)	(0.878)	(0.587)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Men with low locus of control

Table 3.36: Men with low Locus of Control: Main outcomes, components

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Job search indicator	-0.012	-0.025	0.004	0.658	1769
	(0.706)	(0.473)	(0.919)		
Job search days	0.049	0.104	-0.014	1.954	1769
	(0.663)	(0.434)	(0.914)		
Job search hours	0.063	0.073	0.051	1.311	1769
	(0.418)	(0.421)	(0.570)		
Work indicator	0.085**	0.085**	0.086**	0.290	1769
	(0.016)	(0.036)	(0.047)		
Work days	0.371**	0.362**	0.381**	1.018	1769
	(0.016)	(0.044)	(0.045)		
Work hours	0.687***	0.662**	0.716**	1.527	1769
	(0.003)	(0.015)	(0.014)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Table 3.37: Men with low Locus of Control: Mechanisms

	Any Treatment	Goal Setting	Self Affirmation	Control Mean	N
Locus of control	0.947***	0.725**	1.197***	-4.913	592
	(0.001)	(0.030)	(0.001)		
Conscientiousness	0.353	0.252	0.477	3.431	602
	(0.278)	(0.486)	(0.216)		
Time preferences	0.336	-0.182	0.921*	4.290	878
	(0.451)	(0.716)	(0.078)		

[a] This table presents the intent-to-treat estimates of the intervention. Below each coefficient estimate we report the s.e. in parenthesis. ***: p<0.01, **: p<0.05, * : p<0.10.

Education

Table 3.38: Heterogeneity by education specification 1

	Work Index	Job search index	Earnings
Any treatment	0.0550	0.0969*	15.25
	(0.050)	(0.039)	(11.083)
More education	-0.0279	-0.0247	-20.01
	(0.083)	(0.060)	(18.345)
More education=0 × Any treatment=1	-0.0606	-0.124*	-16.68
	(0.077)	(0.057)	(17.463)
Observations	9618	9620	9618

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.39: Heterogeneity by education specification 2

	Work Index	Job search index	Earnings
Goal setting	0.0372	0.0692	10.57
	(0.059)	(0.045)	(13.010)
Self affirmation	0.0727	0.124**	19.87
	(0.058)	(0.045)	(13.290)
More education	-0.0289	-0.0238	-20.18
	(0.082)	(0.060)	(18.330)
More education=0 × Goal setting=1	0.00479	-0.111	-3.202
	(0.091)	(0.066)	(20.767)
More education=0 × Self affirmation=1	-0.129	-0.134*	-30.71
	(0.091)	(0.067)	(20.802)
Observations	9618	9620	9618

Standard errors errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Socioeconomic status

Table 3.40: Heterogeneity by SES specification 1

	Work Index	Job search index	Earnings
Any treatment	0.0754	0.0897*	18.10
	(0.055)	(0.038)	(13.066)
SES	0.122*	-0.147**	40.46**
	(0.061)	(0.046)	(12.989)
SES=0 × Any treatment=1	-0.102	-0.110	-22.50
	(0.076)	(0.057)	(16.699)
Observations	9618	9620	9618

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.41: Heterogeneity by SES specification 2

	Work Index	Job search index	Earnings
Goal setting	0.0654	0.0484	13.34
	(0.064)	(0.043)	(15.108)
Self affirmation	0.0868	0.137**	23.55
	(0.065)	(0.045)	(16.205)
SES	0.122*	-0.147**	40.46**
	(0.061)	(0.046)	(12.990)
SES=0 × Goal setting=1	-0.0700	-0.0631	-13.35
	(0.089)	(0.066)	(19.604)
SES=0 × Self affirmation=1	-0.133	-0.163*	-31.89
	(0.088)	(0.067)	(20.046)
Observations	9618	9620	9618

Standard errors errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Marital status

Table 3.42: Heterogeneity by married specification 1

	Work Index	Job search index	Earnings
Any treatment	-0.0309	0.0292	-11.05
	(0.063)	(0.049)	(14.810)
Married	-0.282***	-0.123*	-51.27**
	(0.070)	(0.053)	(16.016)
Married=0 × Any treatment=1	0.0869	0.0177	27.43
	(0.079)	(0.060)	(18.203)
Observations	9618	9620	9618

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

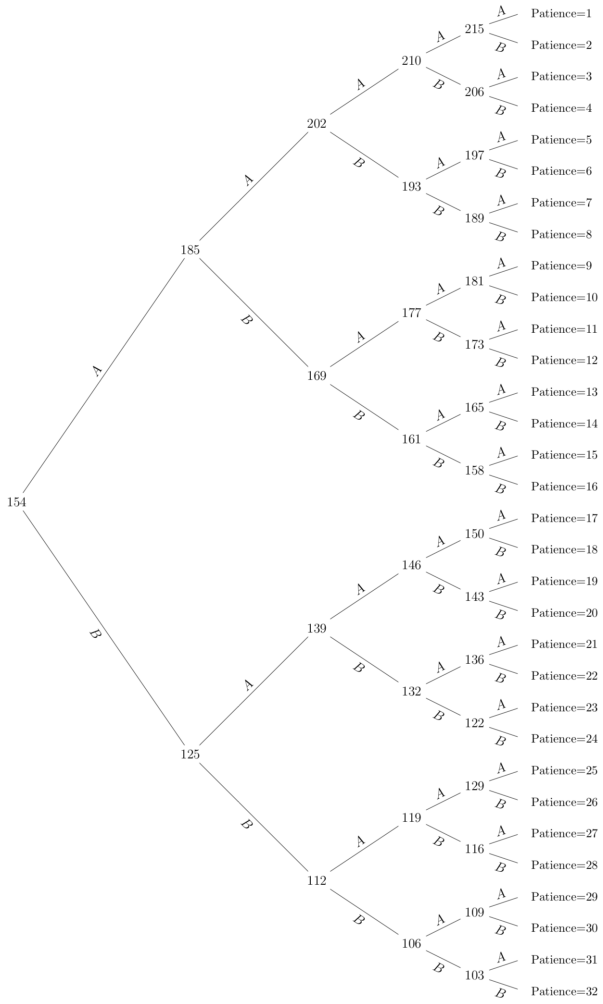
Table 3.43: Heterogeneity by married specification 2

	Work Index	Job search index	Earnings
Goal setting	0.00396	-0.0227	-4.900
	(0.075)	(0.057)	(17.810)
Self affirmation	-0.0684	0.0849	-17.65
	(0.070)	(0.058)	(16.819)
Married	-0.281***	-0.123*	-51.17**
	(0.070)	(0.053)	(16.031)
Married=0 × Goal setting=1	0.0517	0.0618	20.21
	(0.094)	(0.070)	(21.742)
Married=0 × Self affirmation=1	0.125	-0.0303	35.09
	(0.090)	(0.071)	(21.178)
Observations	9618	9620	9618

Standard errors errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

FIGURE 3.3: Time preference tree



Chapter 4

Technology

Vaccination is integral to public health. It not only protects the vaccinated individual against infection; it also provides externality benefits. Epidemiologists argue that mass vaccination is the only viable way to achieve herd immunity against COVID-19. However, the best available estimates indicate that only about 7 out of 10 people globally intend to get vaccinated against COVID-19 (Lazarus et al. 2020). A growing body of research emphasizes one reliable predictor of vaccine-hesitancy: lack of public trust. An important policy challenge, therefore, is for governments to find ways to communicate information about the risks and benefits of the vaccine in a way that is consistent with people's mental model of trustworthiness. This paper explores ways to communicate information about the risks and benefits of the COVID-19 vaccine. It compares information conveyed as objective knowledge with information embedded in story. In addition, it compares two strategies for building trust emphasizing good intentions (i.e., warmth) and expertise (i.e., competence). To test this these approaches, I conduct an online experiment with 1,042 vaccine-hesitant British participants. I oversample Black participants because they are five-times more likely to be hesitant to take the vaccine. I find that people who read the story are significantly more likely to favourably update mental models of vaccines and increase their trust in the government. I also find that story design matters. White participants only update in the condition where the doctor is conveyed as competent. On the other hand, Black participants only update in the condition where the doctor is conveyed as warm.

4.1 Introduction

Vaccination is integral to public health. It not only protects the vaccinated individual against infection; it also provides externality benefits. Epidemiologists argue that mass vaccination is the only viable way to achieve herd immunity against COVID-19 (Rasmussen, 2020; Anderson et al., 2020; Randolph and Barreiro, 2020). However, the best available estimates indicate that only about 7 out of 10 people globally intend to get vaccinated against COVID-19 (Lazarus et al., 2020). A growing body of research emphasizes one reliable predictor of vaccine-hesitancy: lack of public trust in the institutions responsible for creating and distributing the vaccine (Murphy et al., 2021; Lazarus et al., 2020; SAGE, 2020; Callaghan et al., 2020).[1] This is consistent with a long-standing literature that has identified distrust

[1]The common usage of the term "trustworthy" describes the intent of the actor. It captures the degree to which the target has your interests at heart. The literatures on trust – in philosophy, political science, and public health – also emphasize the importance of competence in trust. Thus, trustworthiness is theorized

as one of the main reasons that people give for their scepticism toward vaccines and their intention not to vaccinate (Larson et al., 2011; Larson et al., 2014; Larson et al., 2015; Organisation, 2017). An important policy challenge, therefore, is for governments to find ways to communicate information about the risks and benefits of the vaccine in a way that is consistent with peoples' mental models of the trustworthiness of vaccine providers.

This paper explores ways to communicate information about the risks and benefits of the COVID-19 vaccine. It compares information conveyed as objective knowledge with information embedded in story. In addition, it compares two strategies for building trust emphasizing good intentions (i.e., warmth) and expertise (i.e., competence). To test this these approaches, I conduct an online experiment with 1,042 vaccine-hesitant British participants. I oversample Black participants because they are five-times more likely to be hesitant to take the vaccine. I randomly assign participants to either a control group or one of two treatment arms. The first treatment group received information based closely on NHS guidance on the vaccine. The information was drawn from the UK's National Health Service (NHS) website. The second group received a story about a vaccine hesitant woman trying to decide whether to take the vaccine. The protagonist calls her doctor who provides the same information provided on the NHS website. Among those who receive the story treatment, I further randomized whether the doctor was portrayed as warm (i.e., has good intent) or competent (i.e., has expertise). To understand the effect of the treatments on participants' mental model of the COVID-19 vaccine, I ask participants (i) about their attitudes and intentions with respect to the vaccination and (ii) their attitudes toward the trustworthiness of the Government, NHS, doctors, and pharmaceutical companies. Crucially, I elicit people's mental models both before and after the informational intervention, and test individual-level updating.

The control condition is intended as a placebo. It contains an article about the history of games and the role they have played in human society. It was chosen because it is engaging and interesting but has nothing to do with vaccines or government. The first treatment contains information sourced directly from NHS website. It is presented in the form of questions and answers. For example, one page contained the question "What are the main side-effects associated with the vaccine?" and provided information about the main side effects sourced from the website. The goal of the treatment design is twofold. First, it was intended to capture "business as usual" – i.e., be an accurate representation of the government's main strategy for providing information about the vaccine. A second,

here as encompassing both positive intent and capacity of vaccine providers. In this study I focus on the government, the NHS, doctors, and pharmaceutical companies.

related, goal was to capture a counterfactual representation to narrative communication that addresses informational asymmetries but does not embed communication in a mental model. Thus, the information was intended to be perceived as abstract and authoritative knowledge.

The second treatment embedded the information from treatment one in the form of an explicitly fictional story. It describes a vaccine hesitant woman called Sarah deliberate over whether to get the COVID-19 vaccine. She calls her doctor for advice and the doctor provides the same guidance given in treatment one. The story ends without establishing whether the woman gets the vaccine. The goal of the treatment design is to demonstrate to participants (who themselves are all vaccine hesitant) that the communicator (i.e., information source) can articulate the perspective of a person who is hesitant about the vaccine.

Participants in Treatment 2 were randomly assigned to two versions of the story. In one version, the story emphasizes the doctor's competence – for example, by noting that he has a medical degree and PhD from the University of Birmingham and is regarded as one of the top experts on COVID-19 in the country. In a second version, the story emphasizes the doctor's warmth – for example emphasizing that he has a big family and gave up a highly paid job to stay serving his community. The goal of these treatments was to isolate the two dimensions of trust identified in the literature in philosophy, political science, and psychology: good intent and capability.

I find that people who read the story are significantly more likely to update their mental model of the COVID-19 vaccine and the trustworthiness of government. However, I find that there are only effects for women. Men do not respond to either the informational or the narrative treatment. I also find suggestive evidence that story design matters. White participants only update in the condition where the doctor is conveyed as competent. On the other hand, Black participants only update in the condition where the doctor is conveyed as competent warmth info provider.

This experiment has implications for both policy and research. In terms of policy implications, it provides evidence that stories may be a more effective strategy for communicating with vaccine-hesitant groups than conventional information provision. However, there are two important caveats. First, the design and structure of the story is likely to matter. The story was designed to signal the fact that the communicator understood the perspective of the audience. However, while I think it is unlikely that people would have responded if the story did not contain this feature, it was not experimentally tested. Second, the context of information provision matters. The information condition was not presented on the NHS

website, and it does not follow form this that Governments should replace official advice with stories. Rather, it indicates that stories may be an effective complement to conventional informational services. More evidence is needed to develop a systematic understanding of how (i.e., the features) and when (i.e., the contexts) stories are effective forms of communication.

The study is related to several strands of literature. First, this paper relates to the literature on vaccine hesitancy and the demand for vaccines (Larson, 2020; Larson et al., 2015; Figueiredo et al., 2020; Larson et al., 2011; Lazarus et al., 2020; MacDonald et al., 2015; Sood et al., 2021). Within this, there is relatively large public health literature on interventions aimed at encouraging vaccination (Sood et al., 2021; Leung et al., 2017; Alsan, Garrick, and Graziani, 2019; Shermohammed et al., 2021; Williams et al., 2013; Gottvall et al., 2010; Nyhan and Reifler, 2015; Nyhan et al., 2014). With respect to the COVID-19 vaccine, the evidence is mixed. In a recent study, Kerr et al. (2021) find that providing a British sample with detailed information about COVID-19 vaccines, including clinical trials results, had no effect on beliefs, attitudes, or intentions toward the vaccine. Davis, Golding, and McKay (2021) tested the effect of giving people information about the effectiveness and safety of COVID-19 vaccines increases vaccine intentions by 0.39 standard deviations. In a study conducted in mid-2021, Chang et al. (2021) look at whether incentives, information, or nudges encouraged vaccine-hesitant people to take the COVID-19 vaccine. They found that messages increased vaccination intentions, but did not increase vaccination rates among participants (nor did any other). This paper contributes to this literature by adding evidence on the COVID-19 after they have become available en masse.

The study is also related to the edutainment and narrative persuasion and edutainment literature (Hinyard and Kreuter, 2007; Kreuter et al., 2007; Orozco-Olvera, Shen, and Cluver, 2019; Green, Strange, and Brock, 2002). This literature finds that two key mechanisms through which stories influence attitudinal change and lead people to update their beliefs are transportation (i.e., when participants become immersed in the story) (Green and Brock, 2000) and identification (Graaf et al., 2012). Stories have been used to promote policy goals in areas such as sexual health (Banerjee, Ferrara, and Orozco-Olvera, 2019), poverty alleviation (Tanguy et al., 2014), and peace-building (Paluck, 2009). This paper contributes to this literature testing the mechanisms of impact through precise variations in the design of the messaging, comparing narrative forms with non-narrative and varying messenger to be either warm or competent.[2]

[2]While narrative and non-narrative communication have been compared before, there have been no studies to my knowledge that construct the messages with the goal of testing the messages with strictly equivalent information.

Finally, this paper is related literature is trust of medical institutions and services among Black people. As discussed later in the paper, medical trust has been a social dilemma for thousands of years. Through the 20th century, medical researchers engaged in highly abusive practices with Black participants. In one study, researchers recruited Black men who had Syphilis and told them they would receive free medical care for the disease. This was a lie. The goal of the study was to understand the progression of untreated Syphilis. As a results, over 100 men died. The study was made public in 1972. Alsan and Wanamaker (2018) estimates that among men who were 45 at the time of disclosure, life expectancy fell by up to 1.5 years. Consistent with these results, Lowes and Montero (2021) find that in areas of Central Africa in which French colonial governments organized health campaigns that entailed coercive medical practices with severe and sometimes deadly side effects, vaccination rates and trust in medicine are lower. This paper adds to this literature by identifying targeted communication strategies of rebuilding trust.

4.2 Intervention and experiment

The study took place in January 2022 in the UK. In the UK, roughly 4 in 5 people over the age of 12 have received both shots of the vaccine and 96 percent of all adults self-reported being vaccinated (ONS).[3] Although vaccine hesitancy has declined, only 44 percent of previously vaccine-hesitant adults are now vaccinated against COVID-19 as of September 2021 (ONS). Non-White ethnic minorities are about one third less likely than Whites to be vaccinated (ONS) and vaccine hesitancy is around five times higher among Black or Black British adults compared with White adults.[4] Around 1 in 5 (21 percent) Black or Black British adults report vaccine hesitancy, compared with 4 percent of White adults (ONS). Research emphasizes lack of public trust as a major driver of hesitancy related to covd-19 vaccine (Murphy et al., 2021; Lazarus et al., 2020; SAGE, 2020; Callaghan et al., 2020).

4.2.1 Target population and recruitment

The study takes place with an online sample of adults living in Britain who are vaccine hesitant. I define people as being vaccine hesitant if they have either (i) not yet had the COVID-19 vaccine or (ii) do not intend to get a shot of the COVID-19 vaccine, despite not having their full schedule. Within my sample, 92.8 percent have not had a COVID-19 vaccine, 6.5 percent have had one shot, and 0.7 percent have had 2 shots. At the time of the study, the government was advising people to get a booster vaccine.

[3]UK Government Official Statistics, available at: https://coronavirus.data.gov.uk/details/vaccinations
[4]UK Government Official. Statistics, available at: https://coronavirus.data.gov.uk/details/vaccinations

Because of disparities in vaccine uptake, I over-sampled Black participants. The reason for this was twofold. First, available evidence from the Office of National Statistics indicates that about round 1 in 5 (21 percentage) Black adults in Britain report vaccine hesitancy, compared with 4 percent of White adults. In the first wave of the pandemic, men of Black African ethnicity were 3.7 times more likely to die than White men, while women of Black African ethnicity were 2.6 times more likely. The rates are similar but lower for men of Black Caribbean ethnicity. Second, despite the increased risk, there is relatively little systematic quantitative evidence on Black people's perspective on vaccination. This is in part because Black people are a relatively small proportion of the population and their response rates in surveys tend to be lower. For example, a frequently cited paper in *Nature Communications* on psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in the UK and Ireland had 38 Black participants out of a sample of 2025 (Murphy et al., 2021). This constitutes 1.9 percent of the sample, whereas Black people make up about 3 percent of the population as of the 2011 census. Together, this means that the scientific evidence is most scarce for the group that is most at risk from COVID-19 and least well served by the vaccine.

All participants were recruited from the Prolific platform and were paid £1.25 for taking the survey. In order to recruit participants, I employed two methods. First, I invited participants using the pre-screening tool on Prolific. Asian, Black, Mixed, and White people on Prolific who answered "No" to the question "Have you received a coronavirus (COVID-19) vaccination?" were invited through this process. However, only 37 Black participants were identifiable through this pre-screening. To compliment this, I screened Black and Mixed participants in a rapid survey to identify people who were vaccine hesitant. Participants were asked whether they have had the vaccine for COVID-19 asked to indicate how many shots, if any, they have had. They were also asked "How many shots of the vaccine do you think you'll get of the COVID-19 vaccine in the coming year?" Participants indicated that they hadn't had the vaccine or who had not received all three vaccines were included. An additional 326 participants were included through this process.

The study was conducted in January 2022. I pre-registered the study on aspredicted.org prior to the data collection, specifying the empirical specifications, and hypotheses. [5]

[5] https://aspredicted.org/RST_6Y7. Note that I made a modification in the early stages of data collection to clarify that the outcome measure was updated attitudes rather than attitudes at endline. This entailed a small modification to the specification, where the pre-intervention attitudes is subtracted from the post-intervention attitudes, rather than used a control in the regression. Upon consultation with my supervisor, I report the original specification. The choice of specification had no impact on the statistical significance of results.

4.2.2 Experimental design

Participants began the experiment by answering questions about (i) their mental models of the vaccine (ii) their mental models of vaccine producers. They then were randomly allocated to their treatment condition. Following their intervention, they were asked the same questions as before. Finally, they were asked a small number of demographic questions and invited to visit the NHS website.

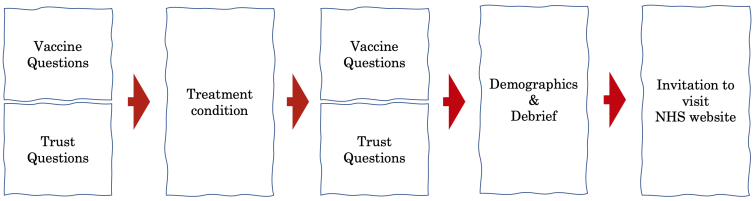


FIGURE 4.1: Experimental procedure

An important feature of the experimental design was that participants are asked the same questions in short succession. The rationale for this was to be able to capture attitudinal updating at the individual level. This allows for greater empirical precision in estimating the effects of the intervention – enabling us to pin down changes in attitudes in addition to post-treatment average differences attitudes. A further advantage is that it enables richer theoretical investigation – for example, to see where in the distribution of attitudes people were more likely to move from. There are concerns with this approach, however. The first is that it might lead to demand effects. The second is that participants might feel that the survey was malfunctioning if they are asking the exact same questions as before. In order to address these issues, I prepared people for the new questions and made it clear that there were no right or wrong answers. The bridge text is documented below:

> In this section, I am keen to understand whether you found the piece useful and whether it helped you think through issues related to getting the vaccine. Some questions are repeats of previous questions. I appreciate you taking the time to carefully fill them out. Like before, there are no right or wrong answers. The goals is just to understand how you see things.

4.2.3 Pre-treatment attitudes and demographics

Table 4.1 presents participants pre-treatment attitudes. The sample is balanced across conditions terms of pre-treatment attitudes, with the exception of attitudes about the trustworthiness of government which is slightly higher for the control condition. Note that participants ratings are very low. For example, participants' average response on government trustworthiness is 1.8 on a scale of 1-5. When asked to rate how trustworthy and competent the government is when it comes to the vaccine, over half selected 1 on both cases. For both measures, only one percent indicated 5. The sample is also balanced on demographics with the exception of gender (Table 4.2), though this is controlled for in all specifications.

TABLE 4.1: Pre-intervention attitudes

Variable	(1) Control Mean/SE	(2) Information Mean/SE	(3) Stories Mean/SE	(4) Total Mean/SE	T-test P-value (1)-(2)	(1)-(3)	(2)-(3)
vaccine_effective_pre z_vaccine_effective_pre_raw vaccine_effective_pre	0.020	-0.023	0.004	0.000	0.573	0.836	0.722
	(0.054)	(0.053)	(0.054)	(0.031)			
vaccine_safe_pre z_vaccine_safe_pre_raw vaccine_safe_pre	0.063	-0.014	-0.049	0.000	0.314	0.138	0.639
	(0.054)	(0.054)	(0.053)	(0.031)			
call_doctor_pre z_call_doctor_pre_raw call_doctor_pre	-0.031	0.045	-0.014	-0.000	0.308	0.823	0.437
	(0.052)	(0.053)	(0.055)	(0.031)			
vaccine_coming_year_pre z_vaccine_coming_year_pre_raw vaccine_coming_year_pre	0.018	-0.002	-0.015	0.000	0.791	0.662	0.866
	(0.053)	(0.054)	(0.053)	(0.031)			
gov_warmth_pre z_gov_warmth_pre_raw gov_warmth_pre	0.086	-0.024	-0.063	0.000	0.154	0.051*	0.594
	(0.058)	(0.052)	(0.050)	(0.031)			
gov_competence_pre z_gov_competence_pre_raw gov_competence_pre	0.067	-0.025	-0.043	-0.000	0.225	0.155	0.810
	(0.057)	(0.051)	(0.053)	(0.031)			
doctors_warmth_pre z_doctors_warmth_pre_raw doctors_warmth_pre	0.044	-0.038	-0.005	-0.000	0.274	0.515	0.663
	(0.053)	(0.053)	(0.054)	(0.031)			
doctors_competence_pre z_doctors_competence_pre_raw doctors_competence_pre	0.031	0.018	-0.050	0.000	0.858	0.279	0.365
	(0.054)	(0.054)	(0.052)	(0.031)			
pharma_warmth_pre z_pharma_warmth_pre_raw pharma_warmth_pre	-0.002	-0.005	0.007	-0.000	0.961	0.906	0.865
	(0.055)	(0.053)	(0.053)	(0.031)			
pharma_competence_pre z_pharma_competence_pre_raw pharma_competence_pre	-0.018	0.003	0.015	-0.000	0.785	0.672	0.869
	(0.054)	(0.051)	(0.056)	(0.031)			
NHS_warmth_pre z_NHS_warmth_pre_raw NHS_warmth_pre	0.082	-0.042	-0.040	0.000	0.096*	0.108	0.972
	(0.052)	(0.053)	(0.055)	(0.031)			
NHS_competence_pre z_NHS_competence_pre_raw NHS_competence_pre	0.081	-0.044	-0.037	0.000	0.096*	0.112	0.937
	(0.051)	(0.055)	(0.054)	(0.031)			
N	350	355	345	1050			

Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

TABLE 4.2: Demographics

	(1)	(2)	(3)	(4)	T-test P-value		
	Control	Information	Stories	Total			
Variable	Mean/SE	Mean/SE	Mean/SE	Mean/SE	(1)-(2)	(1)-(3)	(2)-(3)
Age	33.417	33.068	32.933	33.140	0.640	0.531	0.860
	(0.536)	(0.520)	(0.555)	(0.310)			
Percent women	0.691	0.755	0.684	0.710	0.060*	0.834	0.037**
	(0.025)	(0.023)	(0.025)	(0.014)			
College graduate	0.600	0.532	0.586	0.572	0.070*	0.698	0.158
	(0.026)	(0.027)	(0.027)	(0.015)			
Left wing	0.283	0.254	0.258	0.265	0.380	0.461	0.893
	(0.024)	(0.023)	(0.024)	(0.014)			
Right wing	0.283	0.254	0.258	0.265	0.380	0.461	0.893
	(0.024)	(0.023)	(0.024)	(0.014)			
Black	0.234	0.206	0.252	0.230	0.359	0.583	0.143
	(0.023)	(0.021)	(0.023)	(0.013)			
White	0.460	0.465	0.467	0.464	0.899	0.860	0.960
	(0.027)	(0.027)	(0.027)	(0.015)			
Asian	0.049	0.045	0.038	0.044	0.826	0.481	0.624
	(0.012)	(0.011)	(0.010)	(0.006)			
Mixed	0.091	0.118	0.078	0.096	0.245	0.534	0.076*
	(0.015)	(0.017)	(0.014)	(0.009)			
N	350	355	345	1050			

Notes: The value displayed for t-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

4.2.4 Interventions

Recruited participants were randomly allocated to one of three conditions: (i) a control; an information condition; or a story condition. Among those who were in the story condition, approximately half were allocated to the warmth condition and approximately half were allocated to the competence condition.

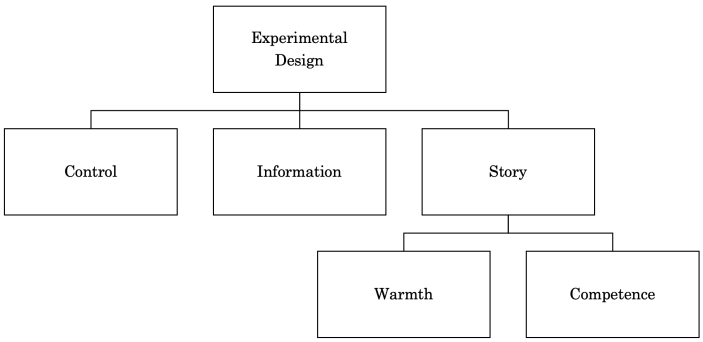


FIGURE 4.2: Experimental design

Information Treatment

The full text of the information is provided in the appendix. I summarize the questions it dealt with in here and provide an example of a response:

- What are the main side-effects associated with the vaccine?
- Can the Oxford AstraZeneca vaccine cause blood clots?
- How effective are the COVID-19 vaccines?
- How do we know that the vaccine is safe?
- Are there long-term risks associated with the vaccine?

FIGURE 4.3: Sample page from informational treatment

What are the main side-effects associated with the vaccine?

The vaccines can cause some side effects, but not everyone gets them.
The side effects are usually mild and should not last longer than a week, such as:

- a sore arm from the injection
- feeling tired
- a headache
- feeling achy
- feeling or being sick

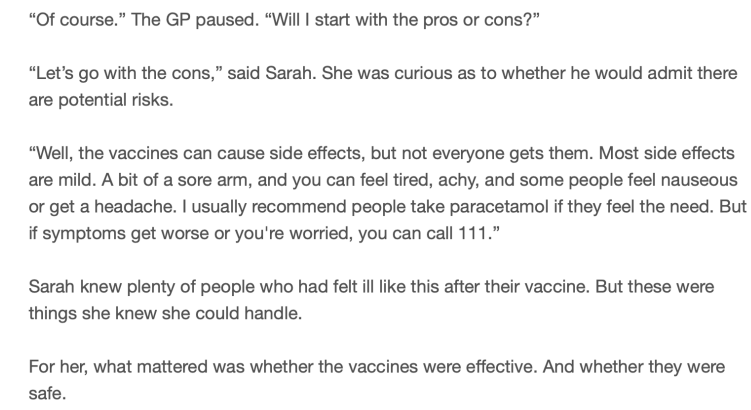
You can take painkillers such as paracetamol if you need to. If your symptoms get worse or you're worried, call 111.

Narrative Treatment

The second treatment embedded the information from treatment one in the form of an explicitly fictional story. The story is set in Birmingham in December 2021. It describes a vaccine hesitant woman called Sarah deliberate over whether to get the COVID-19 vaccine. Although she has been eligible for months, she has not signed up to get it. When her cousin gets admitted to hospital with COVID-19, she decided to call her doctor for advice on the vaccine. In her conversation with the doctor, he provides the same guidance given in

treatment one. Although the conversation is good natured, we learn that Sarah is sceptical about the arguments supporting the safety of the vaccine and is concerned that she will be stereotyped for her views. The story ends without establishing whether Sarah decides to get the vaccine. The goal of the treatment design is to demonstrate to participants (who themselves are all vaccine hesitant) that the communicator (i.e., information source) can articulate the perspective of a person who is hesitant about the vaccine. This signals that they understand and respect the point of view of the target group.

I share below how information is placed within the story. Sarah has just stated that she plans to make up her own mind but that she would like to hear from a doctor she knows.



"Of course." The GP paused. "Will I start with the pros or cons?"

"Let's go with the cons," said Sarah. She was curious as to whether he would admit there are potential risks.

"Well, the vaccines can cause side effects, but not everyone gets them. Most side effects are mild. A bit of a sore arm, and you can feel tired, achy, and some people feel nauseous or get a headache. I usually recommend people take paracetamol if they feel the need. But if symptoms get worse or you're worried, you can call 111."

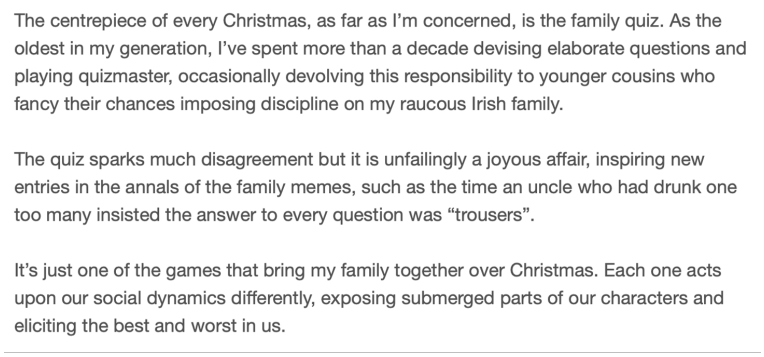
Sarah knew plenty of people who had felt ill like this after their vaccine. But these were things she knew she could handle.

For her, what mattered was whether the vaccines were effective. And whether they were safe.

FIGURE 4.4: Opening page of control condition

Control Condition

The control condition contained no information related to the vaccine. It was, however, engaging. The story is in the appendix. However, the introduction is included below for reference:



The centrepiece of every Christmas, as far as I'm concerned, is the family quiz. As the oldest in my generation, I've spent more than a decade devising elaborate questions and playing quizmaster, occasionally devolving this responsibility to younger cousins who fancy their chances imposing discipline on my raucous Irish family.

The quiz sparks much disagreement but it is unfailingly a joyous affair, inspiring new entries in the annals of the family memes, such as the time an uncle who had drunk one too many insisted the answer to every question was "trousers".

It's just one of the games that bring my family together over Christmas. Each one acts upon our social dynamics differently, exposing submerged parts of our characters and eliciting the best and worst in us.

FIGURE 4.5: Example of information embedded in story

4.3 Data and Outcomes of interest

Data was collected via Qualtrics and included two main outcomes of interest: mental models of the vaccine and vaccine providers. I construct summary indexes for both of the primary outcomes based on the procedure developed by Anderson (2008). The summary index is a weighted mean of the standardized outcomes (by the control group standard deviation), where the weights are designed to maximize the information contained in the index. In addition, I will present results on specific items to understand the precise dynamics of change. All individual items have been standardized to have a zero sample mean and unit sample variance.

Primary outcome 1: Mental model of the vaccine

Participant's mental model of the vaccine is captured using five measures. The first four measures are attitudes and intentions related to the vaccine. I elicit attitudes on a five-point Likert scale rather than eliciting subjective beliefs. The rationale for this was premised on the assumption that participants' decisions with respect to vaccination were unlikely to be guided by expectations about the likelihood of rare *specific* events. Instead, this assumes that participants' decision-making is more likely to be influenced by attitudinal judgments about the safety and effectiveness of the vaccine. The fifth variable is a dummy variable, indicating whether the participant chooses to click on the NHS website. Note that participants were first asked whether they would like "a link with more information about the vaccine?" If they responded no, the survey ended. If they responded yes, they were shown the link. This design measure was taken to be transparent with participants and to reduce the likelihood that people mindlessly click on the link.

1. Confidence in the effectiveness of COVID-19 vaccines
2. Confidence in the safety of COVID-19 vaccines
3. Likelihood of getting a shot this year
4. Likelihood of calling today
5. Participant clicks link to NHS website

Primary outcome 2: Mental model of vaccine providers

Trust in the vaccine providers was captured for four groups related to the distribution of the vaccine: the NHS, the government, doctors, and pharmaceutical companies. All variables were measured on a five-point Likert scale.

Participants were asked:

- How trustworthy do you think the [Group] are when it comes to vaccines?
- How competent do you think the [Group] are when it comes to vaccines?

4.3.1 Other exploratory variables

In addition to the main variables, I asked how useful and interesting participants found the articles they read. I also asked some questions to participants who received the story (described in Appendix 4.8.3 of the Appendix)

4.4 Empirical approach

4.4.1 Hypotheses

I pre-registered[6] the following hypotheses:

H1: Receiving information[7] on the vaccine will lead people to update their attitudes, intentions, and behaviours (mental models) in favour of vaccination

H2: Receiving information on the vaccine in the form of a story will lead people to update their attitudes, intentions, and behaviours (mental models) in favour of vaccination

H3: Receiving information on the vaccine in the form of a story will lead people to update their attitudes, intentions, and behaviours (mental models) in favour of vaccination more than receiving information on the vaccine in the form

H4: Within the story, emphasizing the warmth of the information provider (the doctor) will elicit more positive attitudes, intentions, and behaviours toward the vaccine than emphasizing the competence of the information provider.[8]

4.4.2 Identification strategy

All analyses will be conducted using linear regressions (ANCOVA). For all specifications except one, the outcome of interest is post-treatment attitudes and intentions. The one exception to this is the behavioural measure: whether participants click on the NHS link.

In the first specification, I pool the narrative treatments:

$$Y_{i,t} = \beta_0 + \beta_1 \cdot Information + \beta_2 \cdot Story + \chi \cdot Y_{i,t-1} + \gamma + \mu_i \tag{4.1}$$

[6] https://aspredicted.org/RST_6Y7
[7] Information sourced from NHS. Main webpage: https://www.nhs.uk/conditions/coronavirus-COVID-19/coronavirus-vaccination/coronavirus-vaccine
[8] Warmth and competence operationalised on the basis of the stereotype content model (Fiske et al. 2002).

Y_i is the variable of interest for participant i. $Information$ is a dummy variable indicating that the participant was allocated to the information condition. $beta_1$ is the coefficient of interest for for the information condition. $Story$ is a dummy variable indicating that the participant was allocated to the story condition. $beta_2$ is the coefficient of interest for the story condition. $Y_{i,t-1}$ is the pre-treatment outcome of interest. γ is a vector of covariates that includes gender and age. μ is the error term. This regression will be used to answer H1, H2, and H3. **H1** will be tested with $beta_1$ as the statistic of interest. **H2** will be tested with $beta_2$ as the statistic of interest. **H3** will be tested with $beta_2$ - $beta_1$ as the statistic of interest.

In the second specification, I separate the narrative treatments:

$$Y_{i,t} = \beta_0 + \beta_1 \cdot Information + \beta_2 \cdot Warmth + \beta_3 \cdot Competence + \chi \cdot Y_{i,t-1} + \gamma + \mu_i \tag{4.2}$$

In this specification, the regression remains the same as before except that the variables related to the story condition are broken up into two. However, $Warmth$ is a dummy variable indicating that the participant was allocated to the story condition that emphasized the doctor's warmth. $beta_2$ is the coefficient of interest for the Warmth condition. In addition, $Competence$ is a dummy variable indicating that the participant was allocated to the Competence condition that emphasized the doctor's competence. $beta_3$ is the coefficient of interest for the Competence condition. This specification enables us to test the effect of the warmth version of the story and the competence version of the story respectively.

A third specification is used to answer H4:

$$Y_{i,t} = \beta_0 + \beta_1 \cdot Warmth + \chi \cdot Y_{i,t-1} + \gamma + \mu_i \tag{4.3}$$

For this specification, only the sample who received the story are included. The outcome variables and covariates are the same as in previous regressions. $Warmth$ is a dummy variable indicating that the participant was allocated to the story condition that emphasized the doctor's warmth. $beta_1$ is the coefficient of interest for the Warmth condition. **H4** will be tested with $beta_1$.

4.5 Results

This section presents the main results of the study. I sequentially run each hypothesis, then explore variation in the hypotheses by the sub-groups outlined in the pre-registration. The results are extensively documented in regressions tables in the Appendix.

Treatment effects of the informational condition

The first rows of Table 4.3 and Table 4.4 present the results for the effects of the informational treatment on participants' mental models of the vaccine and vaccine providers. It shows that the informational treatment does not lead people to update their mental model of the vaccine (thus rejecting H1). Participants favourably update their attitude toward effectiveness of the vaccine, an effect that is significant at the ten percent level.

Table 4.3: Vaccine outcomes: Specification 1 (full sample)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Story Effect	0.113***	0.127***	0.141***	0.0531*	0.132**	0.0177
	(0.026)	(0.035)	(0.040)	(0.026)	(0.045)	(0.017)
Information Effect	0.0115	0.0466	0.0856*	-0.00201	0.000599	-0.0122
	(0.026)	(0.035)	(0.040)	(0.026)	(0.045)	(0.017)
Observations	1042	1042	1042	1042	1042	1042

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.4: Trust outcomes: Specification 1 (full sample)

	Index	NHS	Doctors	Government	Pharmaceuticals
Story Effect	0.0730*	0.0973	0.136	0.210**	-0.110
	(0.029)	(0.081)	(0.085)	(0.072)	(0.080)
Information Effect	0.0405	0.128	-0.0426	0.0991	0.0626
	(0.029)	(0.081)	(0.084)	(0.072)	(0.080)
N	1042	1042	1042	1042	1042

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Treatment effects of the story condition

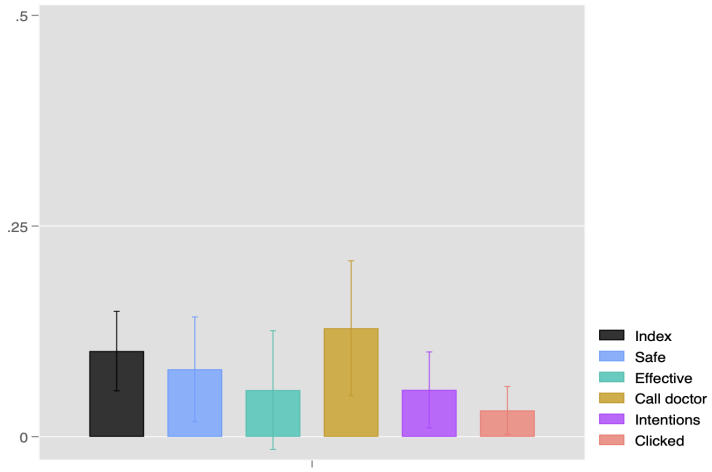
The second rows of Table 4.3 and Table 4.4 presents the results for the effects of the story treatment on mental models of the vaccines and vaccine providers. It shows that the story treatment does lead people to update their mental model of the vaccine as a whole (thus finding in favour of H2). Participants favourably update their attitudes toward the safety and effectiveness of the vaccine, indicate that they would be more likely to call their doctor to discuss the vaccine, and are more likely to get the vaccine in the coming year. They

were no more likely to click on the NHS information at the end of the survey however. I find a statistically significant effect on trust, an effect that is driven by a large effect on the trustworthiness of government.

Story boost

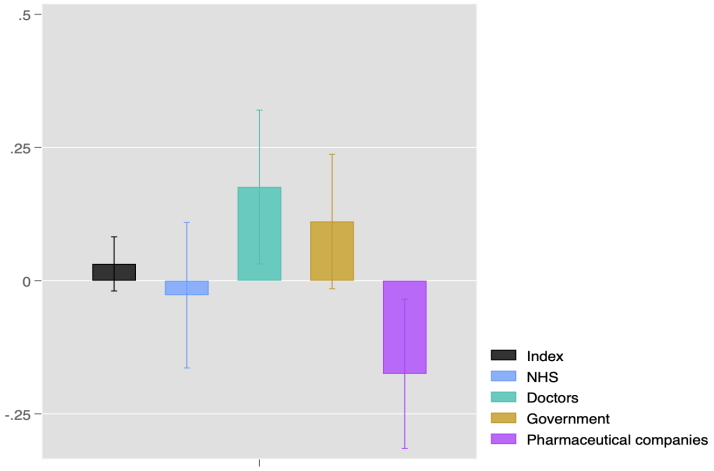
Figure 4.6 and Figure 4.7 test for differences in the rate of updating between the information story condition. As the figure shows, participants are more likely to favourably update their mental models of vaccines and trust in the vaccine producers (supporting H3). The coefficient for all vaccine measures is positive and statistically significant at the ten percent level with the exception of attitude toward the effectiveness of the vaccine. Participants increase their trust in doctors and their increase in trust of the government is at the threshold of significance. However, they downgrade their trust in pharmaceutical companies.

FIGURE 4.6: Comparison of stories with information: Vaccine outcomes



Note: Coefficient on outcome variables for OLS regression testing the effect of being in the story compared with the counterfactual information condition. Confidence intervals at 90 percent. Specification controls for age and gender.

FIGURE 4.7: Comparison of stories with information: Trust



Note: Coefficient on outcome variables for OLS regression testing the effect of being in the story compared with the counterfactual information condition. Confidence intervals at 90 percent. Specification controls for age and gender.

Finally, I look at variation in the design of the story. The stories are included in full in Appendix 4.8.4, with variations highlighted in the boxes. The results for these regressions are presented in Table 4.5 and Table 4.6. As can be seen, within the full sample, participants do not change their mental models of the vaccines or vaccine providers.

TABLE 4.5: Vaccine outcomes (full sample)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Warmth v Competence	0.0268	0.0919	0.0566	0.0819	0.0272	-0.0252
	(0.047)	(0.058)	(0.065)	(0.043)	(0.080)	(0.028)
Observations	344	344	344	344	344	344

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 4.6: Trust outcomes (full sample)

	Index	NHS	Doctors	Government	Pharmaceuticals
Warmth v Competence	-0.0323	0.210	-0.0120	-0.122	0.101
	(0.047)	(0.118)	(0.133)	(0.119)	(0.121)
N	344	344	344	344	344

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.5.1 Sub-group analysis

I focus here on analysing sub-groups by gender and race. First, consider gender. Women make up the majority of the sample. The regression tables for women are documented in

Appendix 4.8.1 of the appendix. For women, I find statistically significant effects for the story condition on participants' mental model of vaccines. Women who read the story were more likely to update their attitudes in favour of the safety and effectiveness of the vaccine and are more likely to click the link, though they do not report greater intention to get vaccinated. The story version had mixed impacts. Emphasizing the warmth had larger influences on attitudes toward safety, while emphasizing competence influenced behavioural measures (intention to take the vaccine, clicking on the link). The effects on men are documented in Appendix 4.8.1 of the appendix. I find that neither the informational treatment or the story treatment lead men to update their mental models of the vaccine or vaccine producers. However, although the index is not statistically significant, men report greater intentions to get vaccinate in the coming year and are more likely to click on the NHS link, effects that are significant at the ten percent level. Neither intervention affects how men update their mental models of vaccine providers.

Table 4.7: Vaccine outcomes: Specification 2 (Women)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Warmth Effect	0.129***	0.194***	0.178***	0.0876***	0.107	0.0373
	(0.037)	(0.051)	(0.057)	(0.033)	(0.065)	(0.025)
Competence Effect	0.104***	0.0955*	0.116**	0.00674	0.0988	0.0488**
	(0.037)	(0.050)	(0.056)	(0.032)	(0.064)	(0.025)
Information Effect	0.00674	0.0461	0.0889**	0.0179	-0.0128	-0.00808
	(0.029)	(0.040)	(0.045)	(0.026)	(0.051)	(0.020)
Observations	746	746	746	746	746	746

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, consider impacts by race.[9] The regression tables for Black participants are documented in Appendix 4.8.1 of the appendix. For Black participants, the story treatment, but not the informational treatment, has a statistically significant effect on the vaccine and trust indices. The results are driven by the warmth condition, where the effects are large and statistically significant on all measures except attitudes toward the effectiveness of the vaccine. Although the vaccine index is statistically significant for the competence condition, none of measures that from which the index is constructed are statistically significant.

The regression tables for White participants are documented in Appendix 4.8.1 of the appendix. For White participants, neither the informational nor the story treatment have a statistically significant effect. However, story version matters. White participants have the inverse response to Black participants. The effect of the story emphasizing competence are

[9]The sample size for Asian or mixed-race participants is too small to analyse impacts.

TABLE 4.8: Vaccine outcomes: Specification 2 (Black participants)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Warmth Effect	0.231***	0.250**	0.246**	0.296***	0.386***	0.0503
	(0.071)	(0.099)	(0.110)	(0.069)	(0.126)	(0.059)
Competence Effect	0.0658	-0.0489	-0.0225	0.0151	0.00602	0.126**
	(0.068)	(0.096)	(0.104)	(0.066)	(0.119)	(0.057)
Information Effect	0.126**	0.149*	0.000538	0.0129	0.0823	0.0758
	(0.059)	(0.083)	(0.091)	(0.057)	(0.105)	(0.050)
Observations	240	240	240	240	240	240

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

statistically significant on all measures except their likelihood of calling their doctor. Although the point estimates on these distinct effects are large, the tables in Appendix 4.8.2 show that the effects are not statistically significant.

TABLE 4.9: Vaccine outcomes: Specification 2 (White participants)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Warmth Effect	0.0321	0.0694	0.0672	0.000487	0.00411	-0.00593
	(0.043)	(0.059)	(0.067)	(0.041)	(0.073)	(0.025)
Competence Effect	0.102**	0.113*	0.189***	0.00195	0.163**	-0.00754
	(0.043)	(0.059)	(0.067)	(0.040)	(0.072)	(0.025)
Information Effect	-0.0499	-0.00431	0.117**	-0.0200	-0.0847	-0.0192
	(0.035)	(0.048)	(0.054)	(0.033)	(0.059)	(0.021)
Observations	485	485	485	485	485	485

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.6 Discussion

Much of the policy discourse and literature on COVID-19 vaccines has focused on attempts to nudge (Milkman et al., 2021), incentivize (Chang et al., 2021), or compel people to take the vaccine. While these strategies are generally at odds with guidance from vaccine research conducted over the past two decades (Larson, 2020), efforts to encourage vaccination can also be considered within the broader context in which medical professions have traditionally sought to establish trust in health contexts. The relationship between patients and medical practitioners presents a special case for trust because of the patient's vulnerability (Hall et al., 2001). There are substantive informational asymmetries between the principal (patient) and the agent (the medical professional), and the fact that the patient must entrust the medical professional with sensitive personal information to facilitate diagnoses, the causal effects of treatments are often opaque, and the downside risk can be deathly for the patient.

The case of the vaccine is no different. The externality benefits of vaccination imply that the incentives of a government that aims to promote social welfare may not be aligned with all individuals' private interest.

Since antiquity, medical ethics has sought to engender genuine trust and trustworthiness in the integrity of the medical profession through the Hippocratic Oath (Miles, 2005). Indeed, through to the 1800s, the good character of doctors played an important role in medical practice. For example, one of America's preeminent physicians during the time of the Revolution, Benjamin Rushe, delivered lectures "On the Virtues and Vices of Physicians" and "On the Duty of Patients to their Physician." The patients' duties included "prompt, strict, and universal" obeyance of their doctor's prescriptions (Jonsen, 2000). Beginning in the early 1800s the medical profession has sought to establish greater trust through ethical codes of practice. Thomas Percival, an English physician and philosopher, published a Code of Medical Ethics in 1803 that has become a foundational document in medical ethics. This inspired the American Medical Association to produce a similar Code of Ethics in 1847. These codes of practice are continually updated and the vast majority of medical schools in the United States, Canada, and the United Kingdom maintain oath taking as part of their graduations (Green, 2017; Scheinman, Fleming, and Niotis, 2018).

Efforts to foster genuine trust have long been complimented through institutional arrangements that establish incentives for people to behave trustworthily (Bal, 2009). For example: the Code of Hammurabi, an ancient Babylonian legal text, stated that: "If a physician performs major surgery with a bronze lancet upon a gentleman and thus causes the gentleman's death, or opens a gentleman's temple with a bronze lancet and thus blinds the gentleman's eye, they shall cut off his hand" (Roth, Hoffner, and Michalowski, 1997). Jonsen (2000) shows that attempts to establish ethics boards were focused on establishing greater confidence in the competence and integrity of the medical profession. However, these steps to improve the trustworthiness of medical professionals have often been undermined by high-profile cases that highlighted stories in which the trust placed in scientific and medical professionals was abused.

Because both the degree of uncertainty and the stakes are often so high in health decisions, the question of trust may not ever be resolved through institutional means alone. A rich literature in economics emphasizes the problem of incomplete contracts: no institutional design, private contract, or network can completely erode the need for genuine trust (Hart, 2017). As discussed in earlier chapters in earlier chapters in this DPhil, people navigate the limitations of institutions using mental models. People are "good-enough" perceivers of

others (Fiske 1992, 1993; Fiske and Taylor 2013). They make pragmatic social judgments on a continuum: initially judging others based on automatic categorizations (stereotypes), then – depending on their goals and the situation (Erber and Fiske 1984) – allocating attention to make individuating assessments of the other (Fiske and Neuberg 1990). Trustworthiness is the most important trait people seek out in interdependent contexts (Cottrell, Neuberg, and Li 2007).

Thus, the question arises: how can governments signal trustworthiness. The intervention tested in this paper, stories, on one interpretation is an archetypal case of cheap talk – it is costless for the government to communicate through narrative. However, the story also signalled to the audience (who were vaccine-hesitant) that the story's author understood their perspective. By articulating the perspective of a vaccine hesitant participant, the narrative condition signalled that the communicator could interpret their perspective. Given the degree of polarization with respect to vaccination, this may not represent cheap talk to people who are hesitant about the vaccine.

4.7 Conclusion

In this paper, I present evidence that communicating about the costs and benefits of the COVID-19 vaccine in the form of story leads people to positively update their mental model of the vaccine and trust in the parties associated with making and distributing the vaccine. The increases in trust are largely driven by effects on trust in government. While none of the groups for which effects were investigated responded to the informational treatment, demographic groups respond differently to the stories. The story influenced women's mental models but not men.

I randomly varied the content of the story in order to emphasize either the warmth or competence of the doctor giving guidance on the vaccine. When the doctor is portrayed as competent as opposed to warm, White participants are more likely update their mental models in favour of vaccination. Conversely, when the doctor is portrayed as warm as opposed to competent, Black participants are more likely update their mental models in favour of vaccination. These results are suggestive (the difference in effects are not statistically significant, though the point estimates are sizeable). They indicate that when communicating with Black communities, signalling good intent may be more important than signalling expertise.

4.8 Appendix

4.8.1 Regression tables

Table 4.10: Vaccine outcomes: Specification 2 (full sample)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Warmth Effect	0.127***	0.170***	0.165***	0.0956***	0.142**	0.00540
	(0.032)	(0.042)	(0.049)	(0.031)	(0.055)	(0.021)
Competence Effect	0.100***	0.0836*	0.116**	0.0101	0.122**	0.0301
	(0.032)	(0.043)	(0.049)	(0.032)	(0.055)	(0.021)
Information Effect	0.0114	0.0464	0.0855**	-0.00213	0.000588	-0.0122
	(0.026)	(0.035)	(0.040)	(0.026)	(0.045)	(0.017)
Observations	1042	1042	1042	1042	1042	1042

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.11: Trust outcomes: Specification 2 (full sample)

	Index	NHS	Doctors	Government	Pharmaceuticals
Warmth Effect	0.0577	0.198**	0.131	0.154*	-0.0542
	(0.035)	(0.099)	(0.104)	(0.088)	(0.098)
Competence Effect	0.0886**	-0.00431	0.141	0.268***	-0.166*
	(0.036)	(0.100)	(0.104)	(0.089)	(0.099)
Information Effect	0.0406	0.128	-0.0426	0.0995	0.0624
	(0.029)	(0.081)	(0.085)	(0.072)	(0.080)
N	1042	1042	1042	1042	1042

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Black participants

Table 4.12: Vaccine outcomes: Specification 1 (Black participants)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Story Effect	0.142*	0.0923	0.101	0.147*	0.181	0.0908
	(0.057)	(0.080)	(0.087)	(0.056)	(0.102)	(0.047)
Information Effect	0.126*	0.154	0.00405	0.0149	0.0832	0.0757
	(0.060)	(0.084)	(0.091)	(0.059)	(0.106)	(0.050)
Observations	240	240	240	240	240	240

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.13: Trust outcomes: Specification 1 (Black participants)

	Index	NHS	Doctors	Government	Pharmaceuticals
Story Effect	0.159*	0.139	0.408*	0.455**	-0.171
	(0.065)	(0.163)	(0.189)	(0.171)	(0.178)
Information Effect	0.0356	-0.199	0.00275	0.148	0.0405
	(0.068)	(0.170)	(0.197)	(0.178)	(0.186)
N	240	240	240	240	240

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.14: Trust outcomes: Specification 2 (Black participants)

	Index	NHS	Doctors	Government	Pharmaceuticals
Warmth Effect	0.172**	0.330	0.533**	0.451**	-0.0835
	(0.082)	(0.202)	(0.234)	(0.214)	(0.224)
Competence Effect	0.146*	-0.0330	0.296	0.458**	-0.246
	(0.078)	(0.195)	(0.226)	(0.204)	(0.213)
Information Effect	0.0355	-0.203	-0.000228	0.148	0.0399
	(0.068)	(0.170)	(0.197)	(0.178)	(0.186)
N	240	240	240	240	240

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

White participants

Table 4.15: Vaccine outcomes: Specification 1 (White participants)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Story Effect	0.0676	0.0913	0.129*	0.00123	0.0847	-0.00675
	(0.035)	(0.048)	(0.055)	(0.033)	(0.059)	(0.021)
Information Effect	-0.0501	-0.00444	0.116*	-0.0200	-0.0852	-0.0192
	(0.035)	(0.048)	(0.054)	(0.033)	(0.059)	(0.021)
Observations	485	485	485	485	485	485

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.16: Trust outcomes: Specification 1 (White participants)

	Index	NHS	Doctors	Government	Pharmaceuticals
Story Effect	0.0227	-0.00617	-0.0215	0.0302	-0.0858
	(0.040)	(0.111)	(0.118)	(0.103)	(0.107)
Information Effect	0.0257	0.187	-0.187	0.0168	0.0925
	(0.040)	(0.111)	(0.118)	(0.102)	(0.106)
N	485	485	485	485	485

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.17: Trust outcomes: Specification 2 (White participants)

	Index	NHS	Doctors	Government	Pharmaceuticals
Warmth Effect	-0.0304	0.0289	-0.0829	-0.156	-0.0478
	(0.050)	(0.137)	(0.146)	(0.125)	(0.132)
Competence Effect	0.0745	-0.0402	0.0383	0.212*	-0.123
	(0.049)	(0.136)	(0.144)	(0.124)	(0.130)
Information Effect	0.0260	0.187*	-0.186	0.0181	0.0922
	(0.040)	(0.111)	(0.118)	(0.101)	(0.106)
N	485	485	485	485	485

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Women

Table 4.18: Vaccine outcomes: Specification 1 (Women)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Story Effect	0.116***	0.144***	0.146**	0.0462	0.103	0.0432*
	(0.030)	(0.041)	(0.046)	(0.026)	(0.053)	(0.020)
Information Effect	0.00677	0.0462	0.0890*	0.0180	-0.0128	-0.00810
	(0.029)	(0.040)	(0.045)	(0.026)	(0.051)	(0.020)
Observations	746	746	746	746	746	746

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.19: Trust outcomes: Specification 1 (Women)

	Index	NHS	Doctors	Government	Pharmaceuticals
Story Effect	0.0633	0.0987	0.0884	0.225**	-0.155
	(0.033)	(0.099)	(0.103)	(0.084)	(0.090)
Information Effect	0.0277	0.212*	-0.0311	0.0708	0.0324
	(0.032)	(0.096)	(0.100)	(0.081)	(0.088)
N	746	746	746	746	746

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.20: Trust outcomes: Specification 2 (Women)

	Index	NHS	Doctors	Government	Pharmaceuticals
Warmth Effect	0.0557	0.197	0.102	0.136	-0.0138
	(0.040)	(0.123)	(0.128)	(0.103)	(0.112)
Competence Effect	0.0705*	0.00496	0.0755	0.310***	-0.289***
	(0.040)	(0.121)	(0.126)	(0.102)	(0.110)
Information Effect	0.0278	0.211**	-0.0311	0.0711	0.0323
	(0.032)	(0.096)	(0.100)	(0.081)	(0.087)
N	746	746	746	746	746

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Men

Table 4.21: Vaccine outcomes: Specification 1 (Men)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Story Effect	0.105*	0.0915	0.129	0.0669	0.198*	-0.0393
	(0.051)	(0.066)	(0.081)	(0.060)	(0.087)	(0.032)
Information Effect	0.0249	0.0514	0.0712	-0.0593	0.0313	-0.0197
	(0.054)	(0.070)	(0.086)	(0.064)	(0.093)	(0.034)
Observations	296	296	296	296	296	296

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.22: Vaccine outcomes: Specification 2 (Men)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Warmth Effect	0.118*	0.122	0.142	0.105	0.215**	-0.0595
	(0.061)	(0.078)	(0.096)	(0.071)	(0.104)	(0.038)
Competence Effect	0.0908	0.0560	0.114	0.0233	0.179	-0.0157
	(0.064)	(0.082)	(0.101)	(0.075)	(0.109)	(0.040)
Information Effect	0.0249	0.0512	0.0711	-0.0591	0.0313	-0.0197
	(0.054)	(0.070)	(0.086)	(0.064)	(0.093)	(0.034)
Observations	296	296	296	296	296	296

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.23: Trust outcomes: Specification 1 (Men)

	Index	NHS	Doctors	Government	Pharmaceuticals
Story Effect	0.0965	0.0981	0.244	0.184	-0.00898
	(0.060)	(0.140)	(0.149)	(0.143)	(0.167)
Information Effect	0.0680	-0.106	-0.0955	0.177	0.138
	(0.064)	(0.149)	(0.158)	(0.152)	(0.178)
N	296	296	296	296	296

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.24: Trust outcomes: Specification 2 (Men)

	Index	NHS	Doctors	Government	Pharmaceuticals
Warmth Effect	0.0665	0.201	0.203	0.200	-0.130
	(0.071)	(0.166)	(0.178)	(0.170)	(0.199)
Competence Effect	0.131*	-0.0201	0.291	0.166	0.132
	(0.075)	(0.174)	(0.187)	(0.179)	(0.209)
Information Effect	0.0682	-0.105	-0.0955	0.177	0.138
	(0.064)	(0.148)	(0.159)	(0.153)	(0.178)
N	296	296	296	296	296

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.8.2 Tables and Figures

Warmth versus competence

Black participants

Table 4.25: Vaccine outcomes (Black participants)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Warmth v Competence	0.155	0.331*	0.241	0.289**	0.411*	-0.0616
	(0.099)	(0.130)	(0.144)	(0.094)	(0.181)	(0.076)
Observations	86	86	86	86	86	86

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.26: Trust outcomes (Black participants)

	Index	NHS	Doctors	Government	Pharmaceuticals
Warmth v Competence	0.0135	0.415	0.329	-0.0644	0.131
	(0.105)	(0.244)	(0.322)	(0.282)	(0.254)
N	86	86	86	86	86

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

White participants

Table 4.27: Vaccine outcomes (White participants)

	Index	Safe	Effective	Call Doctor	Intentions	Clicked
Warmth v Competence	-0.0679	-0.0345	-0.128	-0.0129	-0.153	0.000761
	(0.065)	(0.080)	(0.091)	(0.055)	(0.109)	(0.031)
Observations	161	161	161	161	161	161

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.28: Trust outcomes (White participants)

	Index	NHS	Doctors	Government	Pharmaceuticals
Warmth v Competence	-0.102	0.0785	-0.123	-0.346*	0.0473
	(0.068)	(0.176)	(0.183)	(0.166)	(0.169)
N	161	161	161	161	161

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.8.3 Survey instrument

Vaccine Questions

- Rating it on a scale of 1-5, how likely would you be to call your doctor to talk about the covid vaccine?
- Rating it on a scale of 1-5, how confident are you that the COVID-19 vaccines are effective?
- Rating it on a scale of 1-5, how confident are you that the COVID-19 vaccines are safe?
- Rating it on a scale of 1-5, How likely would you be to get the COVID-19 vaccine in the coming year?

Trust Questions

- Rating it on a scale of 1-5, How *trustworthy* do you think the **Government** are when it comes to vaccines?
- Rating it on a scale of 1-5, How *competent* do you think the **Government** are when it comes to vaccines?
- Rating it on a scale of 1-5, How *trustworthy* do you think the **NHS** is when it comes to vaccines?
- Rating it on a scale of 1-5, How *competent* do you think the **NHS** is when it comes to vaccines?
- Rating it on a scale of 1-5, How *trustworthy* do you think the **pharmaceutical** companies are when it comes to vaccines?
- Rating it on a scale of 1-5, How *competent* do you think the **pharmaceutical** companies are when it comes to vaccines?
- Rating it on a scale of 1-5, How *trustworthy* do you think **doctors** are when it comes to vaccines?
- Rating it on a scale of 1-5, How *competent* do you think **doctors** are when it comes to vaccines?

Other questions

Story feedback

Participants were asked for their level of agreement with respect to two lists of statements. In each case, they were given the option to strongly disagree, moderately disagree, moderately agree, or strongly agree.

1. To what extent do you agree or disagree with the following statements:
    - I enjoyed the story
    - The story was believable
    - I was mentally involved in the story while reading it
    - I could picture myself in the scene of the events described in the story
    - The story affected me emotionally
2. To what extent do you agree or disagree with the following statements:
    - I could identify with Sarah
    - Sarah was likeable
    - Sarah was smart
    - Doctor George seemed competent
    - Doctor George seemed trustworthy
    - Doctor George seemed warm
    - Doctor George seemed intelligent

Response to treatments

- Rating it on a scale of 1-5, how useful did you find the piece?
- Rating it on a scale of 1-5, how interesting did you find the piece?
- What is your gender?
- How old are you?
- How would you describe your highest level of education?
- On economic policy matters, where do you see yourself on the left/right spectrum?
- Do you feel that this survey was biased?

4.8.4 Treatments

Informational treatment

What are the main side-effects associated with the vaccine?

The vaccines can cause some side effects, but not everyone gets them. The side effects are usually mild and should not last longer than a week, such as: a sore arm from the injection feeling tired a headache feeling achy feeling or being sick

You can take painkillers such as paracetamol if you need to. If your symptoms get worse or you're worried, call 111.

Can the Oxford AstraZeneca vaccine cause blood clots?

There is an extremely rare blood clotting problem affecting a small number of people who had the Oxford/AstraZeneca vaccine.

For people aged 40 or over and those with other health conditions, the benefits of being vaccinated with the Oxford/AstraZeneca vaccine outweigh any risk of clotting problems.

For people under 40 without other health conditions, it's preferable for you to have the Pfizer/BioNTech or Moderna vaccine instead of the Oxford/AstraZeneca vaccine.

How effective are the COVID-19 vaccines?

Research has shown the vaccines help: reduce your risk of getting seriously ill or dying from COVID-19 reduce your risk of catching or spreading COVID-19 protect against COVID-19 variants

Recent analysis by Public Health England found that the vaccines have prevented between 26,000 and 28,000 deaths in England alone and between 6.4 and 7.9 million infections.

You can still get COVID-19 if you are vaccinated but studies show that you are about 5 or 10 times more likely to get COVID-19 and pass it on if you are not fully vaccinated.

How do we know that the vaccine is safe?

The vaccines would not have been approved if they weren't effective and safe. They have met strict standards of safety, quality and effectiveness set out by the independent Medicines and Healthcare products Regulatory Agency (MHRA) in the UK.

The vaccines were developed quickly because everybody – the regulators, the scientists, the volunteers, the public funding agencies – made it their top priority when the pandemic broke.

All of the vaccines have met the same standards of any other vaccine.

Are there long-term risks associated with the vaccine?

The side effects are monitored closely by the regulator. So far, millions of people have gotten the vaccine and no long-term complications have been detected.

It is worth noting that an estimated 1.1 million people in the UK have self-reported symptoms of "long covid," which you can get from the coronavirus.

Story treatment

Birmingham, December 2021.

At the beginning of the pandemic, Sarah shared the fear of many around her. The huge disruption to life, and the loneliness, and the uncertainty – COVID-19 had exposed how flimsy everything was. But as time went on and the first year rolled into the next, the energy required to remain in a state of worry was too much, and Sarah settled into indifference.

When she became eligible for the vaccine, Sarah couldn't summon any strong feeling about whether or not she would get it. It didn't seem as though the vaccine could prevent someone from contracting the virus, and Sarah didn't think that she was at risk of serious illness in any case. On the day that her mother phoned to tell her that her cousin had covid, Sarah didn't think much of it. He was only two years older than her and had always been in good health. He would beat it, she knew, and would be feeling better in a week or so.

The next phone call, two days later, came as a shock: her cousin had been admitted to hospital, and quickly transferred to intensive care. After hanging up, Sarah stood with her phone in her hand for a long time. She looked out the window at the bird feeder on her balcony. A house sparrow clung to the wire. There had been more birds around, she thought, since this all started.

There was a new feeling dawning on Sarah now, a feeling of conflict. She wondered – if she caught the virus, she could get as sick as her cousin, but she had worried that the vaccine could make her very sick, too. She didn't trust the way the pharmaceutical companies were pushing the vaccine, and it almost made her want to take her chances with the virus. She tried to think of someone that she could trust. Outside, the sparrow delicately extracted a seed from the feeder. Sarah wondered if the bird had a nest somewhere, hatchlings to feed. Her daughter Nathalie would be home from school within the hour.

Table 4.29: Variation 1

Warmth	Competence
Sarah decided to call her GP, Doctor George. He had cared for Nathalie when she had a high temperature as a baby. She remembered how he had kept up a running commentary of what he was doing, in a sing song voice to comfort the child, and to reassure Sarah. He had a big family of his own, Sarah had seen, in a framed picture on his desk.	Sarah decided to call her GP, Doctor George. He had treated Nathalie when she had a high temperature as a baby. She remembered how he had known immediately what was wrong and began to work on her with silent efficiency. Sarah recalled the framed qualifications from the University of Birmingham lining the walls of the GP's office. He had a medical degree and a PhD.
He spoke to her like a friend, an equal, and had her laughing by the end of the appointment, even though she had arrived in tears with worry.	He spoke with a calm authority as he quickly took Nathalie's vitals and prescribed her a medicine that brought her temperature down within minutes.
Her mother had gone to Doctor George for years. She told her that Doctor George had been offered a highly paid role working in management at a Hospital but turned it down to stay serving his community. He was someone, Sarah figured, whose intentions she could trust.	Her mother told her that Doctor George was one of the top experts on covid in the country. Sarah had heard him on the radio a few times talking about the pandemic. He was someone, Sarah figured, whose expertise she could trust.

After a few minutes on hold, the receptionist patched the call through to Doctor George. At the sound of his voice, Sarah felt self-conscious. She started to wonder whether it was a good idea. Would he just be judgmental? Would he stereotype her for not having already gotten it?

She explained everything in careful detail - about her worry, then indifference, about her cousin and her fear that her daughter could be left without a mother. Ultimately, she said, she just didn't trust the people promoting the vaccine. As she spoke, Sarah could almost feel the GP listening to her through the phone.

Table 4.30: Variation 2

Warmth	Competence
"I appreciate you calling me, Sarah. I suppose to start: I just want to make clear – it's completely up to you whether you get the vaccine or not. I can only offer advice based on what I believe to be best for you."	"I appreciate you calling me, Sarah. I suppose to start: I just want to make clear – it's completely up to you whether you get the vaccine or not. I can only offer advice based on my knowledge of the research."

"Really?" Sarah said. "I thought you'd tell me I had to get it."

"No. Again, it's your choice. But you sound torn. What is it you want to know?"

"Well I just want good information. I'm going to make my own mind up. I'm not going to be forced one way or the other. But I want to know what you think of the vaccine. I want to hear it from a doctor I actually know."

"Of course." The GP paused. "Will I start with the pros or cons?"

"Let's go with the cons," said Sarah. She was curious as to whether he would admit there are potential risks.

"Well, the vaccines can cause side effects, but not everyone gets them. Most side effects are mild. A bit of a sore arm, and you can feel tired, achy, and some people feel nauseous or get a headache. I usually recommend people take paracetamol if they feel the need. But if symptoms get worse or you're worried, you can call 111."

Sarah knew plenty of people who had felt ill like this after their vaccine. But these were things she knew she could handle.

For her, what mattered was whether the vaccines were effective. And whether they were safe.

"What about the blood clots people have gotten from the vaccine? It seems like women are particularly at risk."

"There is an extremely rare blood clotting problem affecting a small number of people who had the AstraZeneca vaccine," said the doctor. "For people over 40, the benefits outweigh the risks. For people under 40, like yourself, I recommend you get the Pfizer or Moderna vaccine instead."

Sarah decided to ask about the benefits the vaccine was supposed to provide.

“Alright then, now let’s do pros”

“Well this is what we know about all of the vaccines that have been approved in the UK. They reduce your likelihood of getting seriously ill or dying if you do get covid. To give you a sense of this, a recent study by Public Health England found that the vaccines have prevented about 27,000 deaths in England alone and about 7 million infections.”

“For me,” said the doctor, “that personal assurance that it helps avoid getting really sick or dying is the main benefit. But vaccines also make it less likely that you’ll catch it or spread it to people around you. They also protect against COVID-19 variants.”

"But you can still get it!" said Sarah.

"Yes, you can. But the studies show that you are about 5 or 10 times more likely to get and pass it on if you don’t get fully vaccinated."

“How do we know all this, though? The studies on the vaccine were rushed out.”

Table 4.31: Variation 3

Warmth	Competence
“A lot of people have this concern, and it’s really understandable. The fact is, these vaccines would not have been approved if they weren’t effective and safe. They have met strict standards of safety, quality and effectiveness set out by the independent MHRA – that’s the Medicines and Healthcare products Regulatory Agency in the UK.”	“A lot of people have this concern. The best way to think about this is to look at the science. The fact is, these vaccines would not have been approved if they weren’t effective and safe. They have met strict standards of safety, quality and effectiveness set out by the independent MHRA – that’s the Medicines and Healthcare products Regulatory Agency in the UK.”

“They were able to do it so quickly because everybody – the regulators, the scientists, the volunteers, the public funding agencies – made it their top priority when the pandemic broke. All of the vaccines have met the same standards of any other vaccine.”

“And those studies can’t tell anything about the long-term effects.”

“Well side effects are monitored closely by MHRA, the regulator. So far, millions of people have gotten the vaccine and no long-term complications have been detected.”

“On the other hand, about 1.1 million are estimated to have experienced ‘long covid’.”

Although they had only been talking for minutes, Sarah was exhausted. She felt like she needed to wrap the conversation up.

"Thanks for all this. Where can I find out more information?"

"Not at all," said the doctor. "You were right to call. I'll text you a link with more information."

As she hung up, Sarah heard her daughter's key in the lock. Outside, the sparrow opened its wings, but Sarah had turned away from the window to greet her daughter, and didn't see it take flight.

Her phone lit up. Doctor George had sent a link with the phone.

Control condition

The centrepiece of every Christmas, as far as I'm concerned, is the family quiz. As the oldest in my generation, I've spent more than a decade devising elaborate questions and playing quizmaster, occasionally devolving this responsibility to younger cousins who fancy their chances imposing discipline on my raucous Irish family.

The quiz sparks much disagreement but it is unfailingly a joyous affair, inspiring new entries in the annals of the family memes, such as the time an uncle who had drunk one too many insisted the answer to every question was "trousers".

It's just one of the games that bring my family together over Christmas. Each one acts upon our social dynamics differently, exposing submerged parts of our characters and eliciting the best and worst in us.

Games have played a role in human society across history and cultures stretching back millennia. When archaeologists found unusual quantities of sheep ankle bones during excavations in the ancient Mediterranean, they deduced they must have been used as a game that required players to balance the bones on their hands, throw them in the air then catch as many as possible. This game, now called Knucklebones, was already ancient when it was written about by Plato, Sophocles and Herodotus.

More recently, the board game creators of the late 18th-century were also mapmakers whose boards propagated alongside the British empire. Before the advent of mass production, such games passed along as a kind of folk technology, with people playing the games in communal spaces and then recreating their own versions at home.

"Games reflect the culture in which they exist," says Nic Ricketts, curator of board games at the Strong National Museum of Play in upstate New York. "During every major war, war games became popular in one way or another."

So we can trace social change through the games that people play. One striking example is The Game of Life, created in 1860 as a moralist fable that reflected puritanical American society, encouraging "honour", "industry" and the achievement of "happy old age". The evolution of a different form of capitalism becomes apparent in the 1960 version of the game, which is all about reaching "Millionaire Acres" and avoiding the "Poor Farm", where players are eliminated.

My favourite genre, known as Eurogames, originated in the 1960s but hit the mainstream with Catan in 1995. These games are focused not on fighting or moving towards a goal

on the board, but rather trading and building. Nobody is eliminated before the end of a Eurogame, meaning it is more inclusive (and fun) to play.

Such games point towards a recent development, a rise in collaborative play. After 5,000 years of competitive games, team sports aside, it is only in the past few decades that humans have decided to play on the same side. The game Pandemic was launched in 2008 but hit the mainstream recently, asking players to dash around a Risk-style world map working together to fight a deadly contagion (though recently it has, for me, hit a little too close to home).

At home, games don't just bridge the generation gap — they can turn that gap to players' advantage, by offering opportunities to educate. Parents and grandparents who play board games with children find that they help improve reflexes and mental agility, and that they promote healthy social interaction, from learning to be a good winner or loser to understanding the complexities of cheating.

They also confer a thrilling sense of autonomy and agency to younger players. "Board games change the power balance for kids," says Tristan Donovan, writer of the book It's All a Game: A Short History of Board Games. "Instead of my parents deciding what happens to me, I could be the person winning this game, because it's really down to the dice."

We continue to play games into adulthood not just because they promote neuroplasticity in older players, helping the brain to continue learning new systems, but because of the opportunity for togetherness that they offer. Those who live in houseshares or gather with their extended family at Christmas often find themselves thrown together with people who may not share their political beliefs or world views. For them, games can act as a bridge to restore a sense of goodwill and empathy. McGonigal goes a step further, arguing that games can actually make us better people: "They're like a healing balm for other areas of life," she says. "When we play games we allow ourselves the pleasure of going from scratch to success and mastery. This provokes a range of positive mindsets like curiosity, creativity, flexibility and, most importantly, self-efficacy — the experience of getting better at something through your own efforts and attention."

I've often observed games fulfilling a different emotional need, allowing us to express unpleasant emotions that would usually be unwelcome. In a game we can let out all our built-up reserves of competition, stinginess or cut-throat egotism.

Bibliography

Abebe, Girum et al. (May 2021a). "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City". In: *The Review of Economic Studies* 88 (3), pp. 1279–1310. ISSN: 0034-6527. DOI: 10.1093/restud/rdaa057.

Abebe, Girum et al. (2021b). *Job Fairs: Matching Firms and Workers in a Field Experiment in Ethiopia*.

Abele, Andrea E. et al. (2020). "Navigating the Social World: Toward an Integrated Framework for Evaluating Self, Individuals, and Groups". In: *Psychological Review* 128 (2), pp. 290–314. ISSN: 19391471. DOI: 10.1037/rev0000262.

Acemoglu, Daron and James A Robinson (2012). *Why nations fail: The origins of power, prosperity, and poverty*. Currency. ISBN: 0307719227.

Akerlof, George A. and William T. Dickens (1982). "The Economic Consequences of Cognitive Dissonance". In: *The American Economic Review* 72 (3), pp. 307–319. ISSN: 0036-8075. URL: http://www.ncbi.nlm.nih.gov/pubmed/24473658.

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and Identity*". In: *Quarterly Journal of Economics* 115 (3), pp. 715–753. ISSN: 0033-5533. DOI: 10.1162/003355300554881. URL: https://academic.oup.com/qje/article-lookup/doi/10.1162/003355300554881.

Akerlof, George A and Robert J Shiller (2015). *Phishing for phools: The economics of manipulation and deception*. Princeton University Press. ISBN: 1400873266.

Allport, Gordon Willard (1954). *The Nature of Prejudice*. Addison-wesley Reading, MA.

Alrababa'h, Ala' et al. (2021). "Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes". In: *American Political Science Review* 115 (4), pp. 1111–1128. ISSN: 0003-0554. DOI: DOI:10.1017/S0003055421000423. URL: https://www.cambridge.org/core/article/can-exposure-to-celebrities-reduce-prejudice-the-effect-of-mohamed-salah-on-islamophobic-behaviors-and-attitudes/A1DA34F9F5BCE905850AC8FBAC78BE58.

Alsan, Marcella, Owen Garrick, and Grant Graziani (2019). "Does diversity matter for health? Experimental evidence from Oakland". In: *American Economic Review* 109 (12), pp. 4071–4111. ISSN: 19447981. DOI: 10.1257/aer.20181446.

Alsan, Marcella and Marianne Wanamaker (Feb. 2018). "Tuskegee and the health of black men". In: *Quarterly Journal of Economics* 133 (1), pp. 407–455. ISSN: 15314650. DOI: 10.1093/qje/qjx029.

Anderson, Michael L (2008). "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". In: *Journal of the American statistical Association* 103 (484). Williams, Jones, and Tukey 1999, pp. 1481–1495. ISSN: 0162-1459.

Anderson, Roy M. et al. (2020). "Challenges in creating herd immunity to SARS-CoV-2 infection by mass vaccination". In: *The Lancet* 396 (10263), pp. 1614–1616. ISSN: 1474547X. DOI: 10.1016/S0140-6736(20)32318-7.

Andre, Peter et al. (2019). "Subjective Models of the Macroeconomy: Evidence from Experts and a Representative Sample". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3355356.

Andrews, Matt, Lant Pritchett, and Michael Woolcock (2013). "Escaping capability traps through problem driven iterative adaptation (PDIA)". In: *World Development* 51, pp. 234–244. ISSN: 0305-750X.

Appadurai, Arjun (2004). *The Capacity to Aspire: Culture and the Terms of Recognition*. Ed. by Vijayendra Rao and Michael Walton.

Arrow, Kenneth (1973). "The Theory of Discrimination". In: *Discrimination in Labor Markets*. Ed. by Orley Ashenfelter and Albert Rees, pp. 3–33. ISSN: 09680004. DOI: 0026-30379/84/2502--0039$01.50/0.

Arrow, Kenneth J (1974). *The Limits of Organization*. WW Norton and Company. ISBN: 0393355799.

Bal, B. Sonny (2009). "An introduction to medical malpractice in the United States". In: *Clinical Orthopaedics and Related Research* 467 (2), pp. 339–347. ISSN: 15281132. DOI: 10.1007/s11999-008-0636-2.

Banaji, Mahzarin R and Anthony G Greenwald (2016). *Blindspot: Hidden biases of good people*. Bantam. ISBN: 0345528433.

Bandura, Albert (1977). *Social learning theory*. Vol. 1. Englewood cliffs Prentice Hall.

– (2001). *Social cognitive theory: An agentic perspective*. Vol. 52, pp. 1–26.

Banerjee, Abhijit, Eliana La Ferrara, and Victor H Orozco-Olvera (2019). "The Entertaining Way to Behavioral Change: Fighting HIV with MTV". In: *NBER Working Paper* (26096).

Bank, The World (2011). *World Development Report 2011*, p. 416. ISBN: 9780821384398. DOI: 10.1596/978-0-8213-8439-8. URL: http://elibrary.worldbank.org/content/book/9780821384398.

Bank, World (2015a). *Ethiopia Urbanization Review: Urban Institutions for a Middle-Income Ethiopia*. World Bank.

– (2015b). *World Development Report 2015: Mind, Society, and Behavior*. World Bank.

– (2019). "Ethiopia Economic Update 7 SPECIAL TOPIC: Poverty and Household Welfare in Ethiopia". In.

Banuri, Sheheryar, Stefan Dercon, and Varun Gauri (2019). "Biased policy professionals". In: *The World Bank Economic Review* 33 (2), pp. 310–327. ISSN: 0258-6770.

Bargh, John A. and Tanya L. Chartrand (1999). "The unbearable automaticity of being." In: *American Psychologist* 54 (7), pp. 462–479. ISSN: 0003-066X. DOI: 10.1037/0003-066X.54.7.462. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.54.7.462.

Barkow, Jerome, Leda Cosmides, and John Tooby (1992). *The adapted mind*. Vol. 14, pp. 397–402. ISBN: 0195101073. DOI: 10.1016/0162-3095(93)90028-G.

Barrett, H Clark (2014). *The shape of thought: How mental adaptations evolve*. Oxford University Press. ISBN: 0199348324.

Bartlett, Frederic Charles (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press. ISBN: 0521483565.

Baumeister, Roy F, Kathleen D Vohs, and David C Funder (2007). "Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?" In: *Perspectives on Psychological Science* 2 (4), pp. 396–403. ISSN: 1745-6916.

Beaman, Lori, Jeremy Magruder, and Jonathan Robinson (2014). "Minding small change among small firms in Kenya". In: *Journal of Development Economics* 108, pp. 69–86. ISSN: 03043878. DOI: 10.1016/j.jdeveco.2013.12.010. URL: http://dx.doi.org/10.1016/j.jdeveco.2013.12.010.

Becker, Gary (1957). *The Economics of Discrimination*. University of Chicago press. ISBN: 0226041042.

– (1968). "Crime and Punishment: An Economic Approach". In: *The economic dimensions of crime*, pp. 13–68. URL: http://www.nber.org/chapters/c3625.pdf.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli (2006). "Adaptive linear step-up procedures that control the false discovery rate". In: *Biometrika* 93 (3), pp. 491–507. ISSN: 1464-3510.

Bertrand, Marianne and Esther Duflo (2017). *Field experiments on discrimination*.

Bertrand, Marianne and Sendhil Mullainathan (2001). "Do people mean what they say? Implications for subjective survey data". In: *American Economic Review* 91 (2), pp. 67–72. ISSN: 0002-8282.

Bertrand, Marianne and Sendhil Mullainathan (2004). "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination". In: *American Economic Review* 94 (4), pp. 991–1013. ISSN: 0002-8282. DOI: 10.1257/0002828042002561. URL: http://pubs.aeaweb.org/doi/10.1257/0002828042002561.

Bettinger, Eric P and Rachel B Baker (2014). "The effects of student coaching: An evaluation of a randomized experiment in student advising". In: *Educational Evaluation and Policy Analysis* 36 (1), pp. 3–19. ISSN: 0162-3737.

Bicchieri, Cristina (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press. ISBN: 1139447149.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992). "A Theory of Fads , Fashion , Custom , and Cultural Change as Informational Cascades Sushil Bikhchandani , David Hirshleifer ," in: *Journal of Political Economy* 100 (5), pp. 992–1026.

Blattman, Christopher and Stefan Dercon (2018). "The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia". In: *American Economic Journal: Applied Economics* 10 (3), pp. 1–38. ISSN: 1945-7782. URL: http://www.aeaweb.org/articles?id=10.1257/app.20170173.

Blattman, Christopher, Julian C Jamison, and Margaret Sheridan (2017). "Reducing Crime and Violence: Experimental Evidence on Adult Noncognitive Investments in Liberia". In: *American Economic Review* 107 (4), pp. 1165–1206. ISSN: 1556-5068. DOI: 10.2139/ssrn.2594868. URL: http://www.nber.org/papers/w21204.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini (Nov. 2021). "Identity, Beliefs, and Political Conflict". In: *Quarterly Journal of Economics* 136 (4), pp. 2371–2411. ISSN: 15314650. DOI: 10.1093/qje/qjab034.

Bordalo, Pedro et al. (2016). "Stereotypes". In: *Quarterly Journal of Economics* (October), pp. 1753–1794. DOI: 10.1093/qje/qjw029.Advance.

Bourdieu, Pierre (1986). *The forms of capital*. Ed. by J Richardson.

Bransford, John D. and Marcia K. Johnson (1972). "Contextual prerequisites for understanding: Some investigations of comprehension and recall". In: *Journal of Verbal Learning and Verbal Behavior* 11 (6), pp. 717–726. ISSN: 00225371. DOI: 10.1016/S0022-5371(72)80006-9.

Brewer, William F and Glenn V Nakamura (1984). "The nature and functions of schemas". In: *Center for the Study of Reading Technical Report; no. 325*.

Bruhn, Miriam and David McKenzie (2009). "In Pursuit of Balance: Randomization in Practice in Development Field Experiments". In: *American Economic Journal: Applied Economics* 1 (4), pp. 200–232. ISSN: 1945-7782. DOI: 10.1257/app.1.4.200.

Bruner, Jerome S (1957). "On perceptual readiness." In: *Psychological review* 64 (2), p. 123. ISSN: 1939-1471.

Bruner, Jerome S, Jacqueline J Goodnow, and George A Austin (1956). *A Study of Thinking*. Vol. 14. New York: John Wiley and Sons, Inc, p. 330.

Bénabou, Roland and Jean Tirole (2002). "Self-confidence and personal motivation". In: *Quarterly Journal of Economics* 117 (3), pp. 871–915. ISSN: 00335533. DOI: 10.1162/003355302760193913.

Caliendo, Marco, Deborah A Cobb-Clark, and Arne Uhlendorff (2015). "Locus of control and job search strategies". In: *Review of Economics and Statistics* 97 (1), pp. 88–103. ISSN: 0034-6535.

Callaghan, Timothy et al. (2020). *Correlates and Disparities of COVID-19 Vaccine Hesitancy*. DOI: 10.2139/ssrn.3667971.

Camerer, Colin and George Loewenstein (2004). *Behavioral Economics: Past, Present, Future in*. Ed. by Colin Camerer, George Loewensteing, and Matthew Rabin.

Carey, Susan (2004). "Bootstrapping and the origin of concepts beings". In: *Daedalus* Wonter (1), pp. 59–68. ISSN: 0011-5266.

– (2009). *The Origin of Concepts*. Oxford University Press, p. 608. ISBN: 9780195367638. DOI: 10.1093/acprof:oso/9780195367638.001.0001. URL: https://oxford.universitypressscholarship.com/10.1093/acprof:oso/9780195367638.001.0001/acprof-9780195367638.

Caria, A Stefano et al. (2021). "An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan *". In: DOI: 10.1257/rct.3870-2.2. URL: https://doi.org/10.1257/rct.3870-2.2..

Castelfranchi, Christiano and Rino Falcone (2010). *Trust Theory: A socio-cognitive and computational model*. Vol. 18. John Wiley and Sons. ISBN: 0470519843.

Cerulo, Karen A (2002). *Culture in mind: Toward a sociology of culture and cognition*. Psychology Press. ISBN: 041592944X.

Chang, Tom et al. (2021). *Financial incentives and other nudges do not increase covid-19 vaccinations among the vaccine hesitant*.

Charlesworth, Tessa E.S. and Mahzarin R. Banaji (Feb. 2019). "Patterns of Implicit and Explicit Attitudes: I. Long-Term Change and Stability From 2007 to 2016". In: *Psychological Science* 30 (2), pp. 174–192. ISSN: 14679280. DOI: 10.1177/0956797618813087.

Chomsky, Noam (1980). "Rules and representations". In: *Behavioral and Brain Sciences* 3 (1), pp. 31–32. ISSN: 14691825. DOI: 10.1017/S0140525X00001680.

Churchill, Susan et al. (2018). "Self-affirmation improves self-control over snacking among participants low in eating self-efficacy". In: *Appetite* 123, pp. 264–268. ISSN: 10958304.

DOI: 10.1016/j.appet.2017.12.028. URL: https://doi.org/10.1016/j.appet.2017.12.028.

Clark, Andrew E., Yannis Georgellis, and Peter Sanfey (2001). "Scarring: The psychological impact of past unemployment". In: *Economica* 68 (270), pp. 221–241. ISSN: 00130427. DOI: 10.1111/1468-0335.00243.

Cohen, Geoffrey L. and David K. Sherman (2014). "The Psychology of Change: Self-Affirmation and Social Psychological Intervention". In: *Annual Review of Psychology* 65 (1), pp. 333–371. ISSN: 0066-4308. DOI: 10.1146/annurev-psych-010213-115137. URL: http://www.annualreviews.org/doi/10.1146/annurev-psych-010213-115137.

Cohen, Geoffrey L. et al. (2006). "Reducing the racial achievement gap: A social-psychological intervention". In: *Science* 313 (5791), pp. 1307–1310. ISSN: 00368075. DOI: 10.1126/science.1128317.

Coleman, James S (1994). *Foundations of Social Theory*. Harvard university press. ISBN: 0674312260.

Collier, Paul (2011). *Wars, Guns and Votes: Democracy in dangerous places*. Random House. ISBN: 144645018X.

Collins, Allan and Elizabeth Loftus (1975). "A spreading activation theory of semantic memory". In: *Psychological Review* 82, pp. 407–428.

Collins, Allan M. and M. Ross Quillian (1969). "Retrieval time from semantic memory". In: *Journal of Verbal Learning and Verbal Behavior* 8 (2), pp. 240–247. ISSN: 00225371. DOI: 10.1016/S0022-5371(69)80069-1.

Craik, Kenneth James Williams (1943). *The nature of explanation*. Vol. 445. CUP Archive. ISBN: 0521094453.

Cuddy, Amy J C, Michael I Norton, and Susan T Fiske (2005). "This old stereotype: The pervasiveness and persistence of the elderly stereotype". In: *Journal of social issues* 61 (2), pp. 267–285. ISSN: 0022-4537.

Cuddy, Amy J.C., Susan T. Fiske, and Peter Glick (2007). "The BIAS Map: Behaviors From Intergroup Affect and Stereotypes". In: *Journal of Personality and Social Psychology* 92 (4), pp. 631–648. ISSN: 00223514. DOI: 10.1037/0022-3514.92.4.631.

Dalton, Patricio S., Sayantan Ghosal, and Anandi Mani (2016). "Poverty and Aspirations Failure". In: *Economic Journal* 126 (590), pp. 165–188. ISSN: 14680297. DOI: 10.1111/ecoj.12210.

d'Andrade, Roy G (1995). *The development of cognitive anthropology*. Cambridge University Press. ISBN: 0521459761.

Davis, Colin J., Matt Golding, and Ryan McKay (2021). "Efficacy information influences intention to take COVID-19 vaccine". In: *British Journal of Health Psychology*. ISSN: 20448287. DOI: 10.1111/bjhp.12546.

Deese, James (1959). "On the prediction of occurrence of particular verbal intrusions in immediate recall." In: *Journal of experimental psychology* 58 (1), p. 17. ISSN: 0022-1015.

DellaVigna, Stefano and Ulrike Malmendier (2006). "Paying not to go to the gym". In: *American Economic Review* 96 (3), pp. 694–719. ISSN: 00028282. DOI: 10.1257/aer.96.3.694.

Denzau, Arthur T and Douglass C North (1994). "Shared mental models: ideologies and institutions". In: *Kyklos* 47 (1), pp. 3–31. ISSN: 1467-6435.

Desimone, Robert and John Duncan (1995). "Neural mechanisms of selective visual attention". In: *Annual review of neuroscience* 18 (1), pp. 193–222. ISSN: 0147-006X.

DiMaggio, Paul (1997). "Culture and cognition". In: *Annual review of sociology* 23 (1), pp. 263–287. ISSN: 0360-0572.

– (2012). "Sociological perspectives on the face-to-face enactment of class distinction". In: *Facing social class: How societal rank influences interaction*, pp. 15–38.

Dohmen, Thomas and Armin Falk (2011). "Performance pay and multidimensional sorting: Productivity, preferences, and gender". In: *American economic review* 101 (2), pp. 556–590. ISSN: 0002-8282.

Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt (2011). "The framing of games and the psychology of play". In: *Games and Economic Behavior* 73 (2), pp. 459–478. ISSN: 08998256. DOI: 10.1016/j.geb.2011.02.003. URL: http://dx.doi.org/10.1016/j.geb.2011.02.003.

Dunbar, Robin I.M. (1998). "The social brain hypothesis". In: *Evolutionary Anthropology* 6 (5), pp. 178–190. ISSN: 10601538. DOI: 10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8.

Durante, Federica and Susan T. Fiske (2017). "How social-class stereotypes maintain inequality". In: *Current Opinion in Psychology* 18, pp. 43–48. ISSN: 2352250X. DOI: 10.1016/j.copsyc.2017.07.033. URL: http://dx.doi.org/10.1016/j.copsyc.2017.07.033.

Durante, Federica, Courtney Bearns Tablante, and Susan T. Fiske (2017). "Poor but Warm, Rich but Cold (and Competent): Social Classes in the Stereotype Content Model". In: *Journal of Social Issues* 73 (1), pp. 138–157. ISSN: 15404560. DOI: 10.1111/josi.12208.

Durante, Federica et al. (2013). "Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap". In: *British Journal of Social Psychology* 52 (4), pp. 726–746. ISSN: 01446665. DOI: 10.1111/bjso.12005.

Dweck, Carol S and Ellen L Leggett (1988). "A social-cognitive approach to motivation and personality." In: *Psychological review* 95 (2), p. 256. ISSN: 1939-1471.

Eliaz, Kfir and Ran Spiegler (Dec. 2020). "A Model of Competing Narratives". In: *American Economic Review* 110 (12), pp. 3786–3816. ISSN: 19447981. DOI: 10.1257/AER.20191099.

Enke, Benjamin (2020). *Moral Values and Voting*.

Er-rafiy, Abdelatif and Markus Brauer (2013). "Modifying perceived variability: Four laboratory and field experiments show the effectiveness of a ready-to-be-used prejudice intervention". In: *Journal of Applied Social Psychology* 43 (4), pp. 840–853. ISSN: 15591816. DOI: 10.1111/jasp.12010.

Erber, Ralph and Susan T Fiske (1984). "Outcome dependency and attention to inconsistent information." In: *Journal of personality and social psychology* 47 (4), p. 709. ISSN: 1939-1315.

Erev, Ido and Greg Barron (2005). "On adaptation, maximization, and reinforcement learning among cognitive strategies." In: *Psychological review* 112 (4), p. 912. ISSN: 1939-1471.

Erev, Ido and Alvin E Roth (1998). "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria". In: *American economic review*, pp. 848–881. ISSN: 0002-8282.

Falk, Armin, Thomas Dohmen, and David Huffman (2016). "The preference survey module: A validated instrument for measuring risk, time, and social preferences". In: *IZA Discussion Paper* No. 9674 (9674).

Falk, Armin et al. (2018). "Global evidence on economic preferences". In: *The Quarterly Journal of Economics* 133 (4), pp. 1645–1692. ISSN: 0033-5533.

Fehr, Ernst and Jean Robert Tyran (2005). "Individual irrationality and aggregate outcomes". In: *Journal of Economic Perspectives* 19 (4), pp. 43–66. ISSN: 08953309. DOI: 10.1257/089533005775196651.

Festinger, Leon (1957). *A theory of cognitive dissonance*. Row, Petereson, and Company. ISBN: 9781912282432. DOI: 10.4324/9781912282432.

Figueiredo, Alexandre de et al. (2020). "Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study". In: *The Lancet* 396 (10255), pp. 898–908. ISSN: 1474547X. DOI: 10.1016/S0140-6736(20)31558-0. URL: http://dx.doi.org/10.1016/S0140-6736(20)31558-0.

Finetti, Bruno De (1937). "La prévision: ses lois logiques, ses sources subjectives". In: vol. 7, pp. 1–68.

Fishbach, Ayelet and Melissa J Ferguson (2007). *The goal construct in social psychology*. Ed. by A W Kruglanski and E T Higgins.

Fiske, Susan and Shelley Taylor (2021). *Social Cognition: From Brains to Culture*. Sage.

Fiske, Susan T. (1992). "Thinking is for doing: Portraits of social cognition from Daguerreotype to laserphoto." In: *Journal of Personality and Social Psychology* 63 (6), pp. 877–889. ISSN: 0022-3514. DOI: 10.1037//0022-3514.63.6.877.

– (2012). "Journey to the edges: Social structures and neural maps of inter-group processes". In: *British Journal of Social Psychology* 51 (1), pp. 1–12. ISSN: 01446665. DOI: 10.1111/j.2044-8309.2011.02092.x.

– (2018). "Stereotype Content: Warmth and Competence Endure". In: *Current Directions in Psychological Science* 27 (2), pp. 67–73. ISSN: 14678721. DOI: 10.1177/0963721417738825.

Fiske, Susan T. and Cydney Dupree (2014). "Gaining trust as well as respect in communicating to motivated audiences about science topics". In: *Proceedings of the National Academy of Sciences of the United States of America* 111, pp. 13593–13597. ISSN: 10916490. DOI: 10.1073/pnas.1317505111.

Fiske, Susan T and Patricia W Linville (1980). "What does the schema concept buy us?" In: *Personality and social psychology bulletin* 6 (4), pp. 543–557. ISSN: 0146-1672.

Fiske, Susan T. and Steven L. Neuberg (1990). "A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation". In: *Advances in Experimental Social Psychology* 23 (C), pp. 1–74. ISSN: 00652601. DOI: 10.1016/S0065-2601(08)60317-2.

Fiske, Susan T and Shelley E Taylor (2020). *Social Cognition: From brains to culture*. New York, NY: Sage. ISBN: 0070211914.

Fiske, Susan T. et al. (2002). "A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition". In: *Journal of Personality and Social Psychology* 82 (6), pp. 878–902. ISSN: 00223514. DOI: 10.1037//0022-3514.82.6.878.

Foucault, Michel (1978). *The history of sexuality: An introduction*. Vintage. ISBN: 0679724699.

– (2003). *Madness and civilization*. Routledge. ISBN: 0203164695.

– (2014). *On the Government of the Living: Lectures at the Collège de France, 1979-1980*. Springer. ISBN: 1137491825.

– (2019). *Power: the essential works of Michel Foucault 1954-1984*. Penguin UK. ISBN: 0141991372.

Franklin, Simon (2018). "Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies". In: *Economic Journal* 128 (614), pp. 2353–2379. ISSN: 14680297. DOI: 10.1111/ecoj.12509.

Frye, Margaret (2012). "Bright futures in Malawi's new dawn: Educational aspirations as assertions of identity". In: *American Journal of Sociology* 117 (6). Statements about the future as moral claims rather than as rational choices.

four elements are understood by young women in Malawi to jointly produce educational success:
– Ambitious career goals,
– Sustained effort,
– Unflagging optimism, and
– Resistance to (mostly sexual) temptations.

Resources:
– 40 in-depth interviews with female secondary school students
– archival sources, including school curricula, newspaper articles, and printed materials from nongovernmental organizations.

identity as an individual's ongoing narrative account
of who she is at present, modeled on the future self that she imagines
becoming.88, pp. 1565–1624. ISSN: 00029602. DOI: 10.1086/664542.

– (2017). *Cultural Meanings and the Aggregation of Actions: The Case of Sex and Schooling in Malawi*. Vol. 82, pp. 945–976. ISBN: 0003122417. DOI: 10.1177/0003122417720466.

Fujiwara, Thomas (2015). "Voting technology, political responsiveness, and infant health: Evidence from Brazil". In: *Econometrica* 83 (2), pp. 423–464. ISSN: 0012-9682. DOI: 10.3982/ECTA11520. URL: https://www.econometricsociety.org/doi/10.3982/ECTA11520.

Fukuyama, Francis (1995). *Trust: The social virtues and the creation of prosperity*. Vol. 99. Free press New York.

Fund, International Monetary (Oct. 2019). *World Economic Outlook: Global Manufacturing Downturn, Rising Trade Barriers*. International Monetary Fund.

Gabaix, Xavier (2014). "A sparsity-based model of bounded rationality". In: *The Quarterly Journal of Economics* 129 (4), pp. 1661–1710. ISSN: 0033-5533.

– (2018). *Behavioral Inattention*. DOI: 10.3386/w24096. URL: http://www.nber.org/papers/w24096.pdf.

Gambetta, Diego (2000). "Can we trust trust". In: *Trust: Making and breaking cooperative relations* 13, pp. 213–237.

Garance, Genicot and Ray Debraj (Mar. 2017). "Aspirations and Inequality". In: *Econometrica* 85 (2). From Duplicate 1 (Aspirations and Inequality - Garance, Genicot; Debraj, Ray)

doi: 10.3982/ECTA13865, pp. 489–519. ISSN: 0012-9682. DOI: 10.3982/ECTA13865. URL: https://doi.org/10.3982/ECTA13865.

Gardner, Howard (1987). *The mind's new science: A history of the cognitive revolution*. Basic books. ISBN: 0465046355.

Gennaioli, Nicola and Andrei Shleifer (2010). "What comes to mind". In: *The Quarterly journal of economics* 125 (4), pp. 1399–1433. ISSN: 1531-4650.

Gentner, Dedre and Albert L Stevens (1983). *Mental models*. Psychology Press. ISBN: 1317769406.

Ghosal, Sayantan et al. (2020). "Sex Workers, Stigma and Self-Image: Evidence from Kolkata Brothels". In: *The Review of Economics and Statistics* Advanced publication, pp. 1–45. ISSN: 0034-6535. DOI: 10.1162/rest_a_01013.

Gigerenzer, Gerd (2000). *Adaptive thinking: Rationality in the real world*. Oxford University Press, USA. ISBN: 0195136225.

Gilboa, Asaf and Hannah Marlatte (2017). "Neurobiology of Schemas and Schema-Mediated Memory". In: *Trends in Cognitive Sciences* 21 (8), pp. 618–631. ISSN: 1879307X. DOI: 10.1016/j.tics.2017.04.013. URL: http://dx.doi.org/10.1016/j.tics.2017.04.013.

Gilovich, Thomas, Dale Griffin, and Daniel Kahneman (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press. ISBN: 0521796792.

Glaeser, Edward L. et al. (2000). "Measuring Trust". In: *Quarterly Journal of Economics* 115 (3), pp. 811–846. ISSN: 00335533. DOI: 10.1162/003355300554926.

Gneezy, Uri and Aldo Rustichini (2000). "A fine is a price". In: *Journal of Legal Studies* 29 (1 PART I), p. 1. ISSN: 00472530. DOI: 10.1086/468061.

Goffman, Erving (1963). *Stigma: Notes on the Management of Spoiled Identity*. Penguin.

Goldin, Claudia and Cecilia Rouse (2000). "Orchestrating impartiality: The impact of "blind" auditions on female musicians". In: *American Economic Review* 90 (4), pp. 715–741. ISSN: 00028282. DOI: 10.1257/aer.90.4.715.

Goldsmith, Arthur H, Jonathan R Veum, and William Darity (1997). "Unemployment, joblessness, psychological well-being and self-esteem: Theory and evidence". In: *The Journal of Socio-Economics* 26 (2), pp. 133–158. ISSN: 1053-5357. DOI: https://doi.org/10.1016/S1053-5357(97)90030-5. URL: http://www.sciencedirect.com/science/article/pii/S1053535797900305.

Gopnik, Alison and Andrew N Meltzoff (1997). *Words, thoughts, and theories*. Mit Press. ISBN: 0262571269.

Gottvall, M. et al. (Aug. 2010). "Knowledge of human papillomavirus among high school students can be increased by an educational intervention". In: *International Journal of STD and AIDS* 21 (8), pp. 558–562. ISSN: 09564624. DOI: 10.1258/ijsa.2010.010063.

Graaf, Anneke de et al. (2012). "Identification as a Mechanism of Narrative Persuasion". In: *Communication Research* 39 (6), pp. 802–823. ISSN: 00936502. DOI: 10.1177/0093650211408594.

Graesser, Arthur C., Sallie E. Gordon, and John D. Sawyer (1979). "Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis". In: *Journal of Verbal Learning and Verbal Behavior* 18 (3), pp. 319–332. ISSN: 00225371. DOI: 10.1016/S0022-5371(79)90182-8.

Green, Ben (2017). "Use of the Hippocratic or other professional oaths in UK medical schools in 2017: Practice, perception of benefit and principlism". In: *BMC Research Notes* 10 (1), pp. 1–4. ISSN: 17560500. DOI: 10.1186/s13104-017-3114-7. URL: https://doi.org/10.1186/s13104-017-3114-7.

Green, Melanie, Jeffrey Strange, and Timothy Brock (2002). "Narrative impact : social and cognitive foundations". In: URL: http://uwashington.worldcat.org/title/narrative-impact-social-and-cognitive-foundations/oclc/2830049626152&referer=brief_results.

Green, Melanie C. and Timothy C. Brock (2000). "The role of transportation in the persuasiveness of public narratives." In: *Journal of Personality and Social Psychology* 79 (5), pp. 701–721. ISSN: 1939-1315. DOI: 10.1037/0022-3514.79.5.701. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.79.5.701.

Hall, Crystal C., Jiaying Zhao, and Eldar Shafir (2014). "Self-Affirmation Among the Poor: Cognitive and Behavioral Implications". In: *Psychological Science*. ISSN: 14679280. DOI: 10.1177/0956797613510949.

Hall, Mark A. et al. (2001). "Trust in Physicians and Medical Institutions: What Is It, Can It Be Measured, and Does It Matter?" In: *Milbank Quarterly* 79 (4), pp. 613–639. ISSN: 0887378X. DOI: 10.1111/1468-0009.00223.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein (2014). "Learning through noticing: Theory and Evidence from a Field Experiment". In: *Quarterly Journal of Economics* 129 (3), pp. 1311–1354. DOI: 10.2307/26372574. URL: https://www.jstor.org/stable/10.2307/26372574?seq=1&cid=pdf-.

Harris, Peter R. Philine S., Peter R. Philine S. Harris, and Eleanor Miles (2017). "Self-affirmation improves performance on tasks related to executive functioning". In: *Journal of Experimental Social Psychology* 70, pp. 281–285. ISSN: 10960465. DOI: 10.1016/j.jesp.2016.11.011. URL: http://dx.doi.org/10.1016/j.jesp.2016.11.011.

Hart, Oliver (2017). "Incomplete contracts and control". In: *American Economic Review* 107 (7), pp. 1731–1752. ISSN: 00028282. DOI: 10.1257/aer.107.7.1731.

Hayek, Friedrich August (1945). "The use of knowledge in society". In: *The American economic review* 35 (4), pp. 519–530. ISSN: 0002-8282.

Heller, Sara B. et al. (2017). "Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago". In: *Quarterly Journal of Economics* 132 (1), pp. 1–54. ISSN: 15314650. DOI: 10.1093/qje/qjw033.

Henrich, Joseph (2017). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press. ISBN: 0691178437.

Henrich, Joseph et al. (Dec. 2005). *"Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies*. DOI: 10.1017/S0140525X05000142.

Herrmann, Esther et al. (2007). "Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis". In: *science* 317 (5843), pp. 1360–1366. ISSN: 0036-8075.

Hertwig, Ralph et al. (2004). "Decisions from experience and the effect of rare events in risky choice". In: *Psychological science* 15 (8), pp. 534–539. ISSN: 0956-7976.

Heyes, Cecilia (2018). *Cognitive gadgets*. Harvard University Press. ISBN: 067498515X.

Heyes, Cecilia M (1994). "Social learning in animals: categories and mechanisms". In: *Biological Reviews* 69 (2), pp. 207–231. ISSN: 1464-7931.

Hinyard, Leslie J. and Matthew W. Kreuter (Oct. 2007). "Using narrative communication as a tool for health behavior change: A conceptual, theoretical, and empirical overview". In: *Health Education and Behavior* 34 (5), pp. 777–792. ISSN: 10901981. DOI: 10.1177/1090198106291963.

Hobbes, Thomas (1651). *Leviathan*. Ed. by William George Pogson Smith. Oxford: Clarendon Press, 1909.

Hoff, Karla and Priyanka Pandey (2006). "Discrimination, Social Identity, and Durable Inequalities". In: *American Economic Review* 96 (2), pp. 206–211. ISSN: 0002-8282.

– (2014). "Making up people—The effect of identity on performance in a modernizing society". In: *Journal of Development Economics* 106, pp. 118–131. ISSN: 0304-3878.

Hoff, Karla and Joseph E. Stiglitz (2016). "Striving for balance in economics: Towards a theory of the social determination of behavior". In: *Journal of Economic Behavior and Organization* 126, pp. 25–57. ISSN: 01672681. DOI: 10.1016/j.jebo.2016.01.005. URL: http://dx.doi.org/10.1016/j.jebo.2016.01.005.

Hoff, Karla and James Walsh (2018). "The whys of social exclusion: Insights from behavioral economics". In: *World Bank Research Observer* 33 (1), pp. 1–33. ISSN: 15646971. DOI: 10.1093/wbro/lkx010.

– (2021). *The Third Function of Law Is to Transform Cultural Categories*. Ed. by Kaushik Basu and Robert Hockett.

Honig, Dan (2018). *Navigation by judgment: Why and when top-down management of foreign aid doesn't work*. Oxford University Press. ISBN: 0190672471.

Honig, Dan (Oct. 2020). "Information, power, and location: World Bank staff decentralization and aid project success". In: vol. 33. Blackwell Publishing Ltd, pp. 749–769. DOI: 10.1111/gove.12493.

Howell, Andrew J. (2017). "Self-Affirmation Theory and the Science of Well-Being". In: *Journal of Happiness Studies*.

Inoue, Sana and Tetsuro Matsuzawa (2007). "Working memory of numerals in chimpanzees". In: *Current Biology* 17 (23), R1004–R1005. ISSN: 0960-9822.

Iyer, Lakshmi et al. (2012). "The power of political voice: women's political representation and crime in India". In: *American Economic Journal: Applied Economics* 4 (4), pp. 165–193. ISSN: 1945-7782.

Jenkins, Adrianna C. et al. (2018). "Predicting human behavior toward members of different social groups". In: *Proceedings of the National Academy of Sciences of the United States of America* 115 (39), pp. 9696–9701. ISSN: 10916490. DOI: 10.1073/pnas.1719452115.

Johnson, Noel D and Alexandra A Mislin (2011). "Trust games: A meta-analysis". In: *Journal of Economic Psychology* 32 (5), pp. 865–889. ISSN: 0167-4870.

Johnson, Samuel and David Tuckett (2018). "Narrative Decision-Making in Investment Choices: How Investors Use News About Company Performance". In: *SSRN Electronic Journal* (September). ISSN: 1556-5068. DOI: 10.2139/ssrn.3037463.

Johnson-Laird, Philip N (2004). "The history of mental models". In: *Psychology of reasoning: Theoretical and historical perspectives*, pp. 179–212.

– (2010). "Mental models and human reasoning". In: *Proceedings of the National Academy of Sciences* 107 (43), pp. 18243–18250. ISSN: 0027-8424.

Johnson-Laird, Philip Nicholas (1980). "Mental models in cognitive science". In: *Cognitive science* 4 (1), pp. 71–115. ISSN: 0364-0213.

Johnston, William A and Veronica J Dark (1986). "Selective attention". In: *Annual review of psychology* 37 (1), pp. 43–75. ISSN: 0066-4308.

Jonsen, Albert R (2000). *A short history of medical ethics*. Oxford University Press. ISBN: 0195134559.

Jr, Robert S Wyer (2007). *Principles of Mental Representation*. Ed. by Arie W. Kruglanski and E. Tory Higgins.

Kandel, Eric R (2007). *In search of memory: The emergence of a new science of mind*. WW Norton and Company. ISBN: 0393070514.

Kant, Immanuel (1781). *Critique of Pure Reason*. Cambridge University Press. ISBN: 9780521657297. DOI: DOI:10.1017/CBO9780511804649. URL: https://www.cambridge.org/core/books/critique-of-pure-reason/259C2355B74458963EC285F53337AAF0.

Katz, Daniel and Kenneth Braly (1933). "Racial stereotypes of one hundred college students." In: *The Journal of Abnormal and Social Psychology* 28 (3), p. 280. ISSN: 0096-851X.

Kerr, John R. et al. (Apr. 2021). "Effect of information about COVID-19 vaccine effectiveness and side effects on behavioural intentions: Two online experiments". In: *Vaccines* 9 (4). ISSN: 2076393X. DOI: 10.3390/vaccines9040379.

Kesteren, Marlieke T.R. Van et al. (2012). "How schema and novelty augment memory formation". In: *Trends in Neurosciences* 35 (4), pp. 211–219. ISSN: 01662236. DOI: 10.1016/j.tins.2012.02.001.

Keyes, Katherine M, Mark L Hatzenbuehler, and Deborah S Hasin (2011). "Stressful life experiences, alcohol consumption, and alcohol use disorders: the epidemiologic evidence for four main types of stressors". In: *Psychopharmacology* 218 (1), pp. 1–17. ISSN: 1432-2072.

King, Mervyn and John Kay (2020). *Radical uncertainty: Decision-making for an unknowable future*. Hachette UK. ISBN: 140871258X.

Kitayama, Shinobu and Jiyoung Park (2010). "Cultural neuroscience of the self: Understanding the social grounding of the brain". In: *Social Cognitive and Affective Neuroscience* 5 (2-3), pp. 111–129. ISSN: 17495016. DOI: 10.1093/scan/nsq052.

Knight, Frank Hyneman (1921). *Risk, uncertainty and profit*. Vol. 31. Houghton Mifflin.

Kreuter, Matthew W et al. (June 2007). "Narrative communication in cancer prevention and control: a framework to guide research and application." In: *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine* 33 (3), pp. 221–235. ISSN: 0883-6612 (Print). DOI: 10.1007/bf02879904.

Kuhn, Thomas S (1970). *The structure of scientific revolutions*. Vol. 111. Chicago University of Chicago Press.

Köszegi, Botond (2006). "Ego utility, overconfidence, and task choice". In: *Journal of the European Economic Association* 4 (4), pp. 673–707. ISSN: 15424766. DOI: 10.1162/JEEA.2006.4.4.673.

Laajaj, Rachid et al. (2019). "Challenges to capture the big five personality traits in non-WEIRD populations". In: *Science Advances* 5 (7). ISSN: 23752548. DOI: 10.1126/sciadv.aaw5226.

Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor (2012). "Heuristic thinking and limited attention in the car market". In: *American Economic Review* 102 (5), pp. 2206–2236. ISSN: 00028282. DOI: 10.1257/aer.102.5.2206.

Lai, Calvin K et al. (2016). "Reducing implicit racial preferences: II. Intervention effectiveness across time." In: *Journal of Experimental Psychology: General* 145 (8), p. 1001. ISSN: 1939-2222.

Lakoff, George (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago press. ISBN: 0226471012.

Lampinen, James M, Susann M Copeland, and Jeffrey S Neuschatz (2001). "Recollections of things schematic: Room schemas revisited." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27 (5), p. 1211. ISSN: 1939-1285.

Larson, Heidi J (2020). *Stuck: How Vaccine Rumors Start–and Why They Don't Go Away*. Oxford University Press. ISBN: 0190077255.

Larson, Heidi J. et al. (2011). "Addressing the vaccine confidence gap". In: *The Lancet* 378 (9790), pp. 526–535. ISSN: 1474547X. DOI: 10.1016/S0140-6736(11)60678-8.

Larson, Heidi J. et al. (2014). "Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature, 2007-2012". In: *Vaccine* 32 (19), pp. 2150–2159. ISSN: 18732518. DOI: 10.1016/j.vaccine.2014.01.081. URL: http://dx.doi.org/10.1016/j.vaccine.2014.01.081.

Larson, Heidi J. et al. (2015). "Measuring vaccine confidence: Introducing a global Vaccine Confidence Index". In: *PLoS Currents* 7 (OUTBREAKS). ISSN: 21573999. DOI: 10.1371/currents.outbreaks.ce0f6177bc97332602a8e3fe7d7f7cc4.

Lazarus, Jeffrey V. et al. (2020). "A global survey of potential acceptance of a COVID-19 vaccine". In: *Nature Medicine*. ISSN: 1546170X. DOI: 10.1038/s41591-020-1124-9. URL: http://dx.doi.org/10.1038/s41591-020-1124-9.

Leung, Ka Chun et al. (Oct. 2017). *Impact of patient education on influenza vaccine uptake among community-dwelling elderly: A randomized controlled trial*. DOI: 10.1093/her/cyx053.

Liberman, Varda, Steven M. Samuels, and Lee Ross (Sept. 2004). "The name of the game: Predictive power of reputations versus situational labels in determining Prisoner's Dilemma game moves". In: *Personality and Social Psychology Bulletin* 30 (9), pp. 1175–1185. ISSN: 01461672. DOI: 10.1177/0146167204264004.

Locke, Edwin A and Gary P Latham (1990). *A theory of goal setting and task performance*. Prentice-Hall, Inc. ISBN: 0139131388.

Loewenstein, George, Botond Köszegi, and Takeshi Murooka (2019). *Fragile Self-Esteem*.

Lowes, Sara and Eduardo Montero (2021). "The legacy of colonial medicine in Central Africa". In: *American Economic Review* 111 (4), pp. 1284–1314. ISSN: 0002-8282.

Lowrey, Annie (2018). *Give people money: How a universal basic income would end poverty, revolutionize work, and remake the world*. Broadway books. ISBN: 1524758779.

MacDonald, Noni E. et al. (2015). "Vaccine hesitancy: Definition, scope and determinants". In: *Vaccine* 33 (34), pp. 4161–4164. ISSN: 18732518. DOI: 10.1016/j.vaccine.2015.04.036.

Macours, Karen and Renos Vakis (2014). "Changing households' investment behaviour through social interactions with local leaders: Evidence from a randomised transfer programme". In: *The Economic Journal* 124 (576), pp. 607–633. ISSN: 1468-0297.

Malmendier, Ulrike and Geoffrey Tate (2008). "Who makes acquisitions? CEO overconfidence and the market's reaction". In: *Journal of Financial Economics* 89 (1), pp. 20–43. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2007.07.002.

Mandler, George (2011). *A history of modern experimental psychology: From James and Wundt to cognitive science*. MIT press. ISBN: 0262263882.

Mandler, Jean Matter (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Psychology Press. ISBN: 1317768590.

Manski, Charles F. (2004). "Measuring expectations". In: *Econometrica* 72 (5), pp. 1329–1376. ISSN: 00129682. DOI: 10.1111/j.1468-0262.2004.00537.x.

Markus, Hazel (1977). "Self-schemata and processing information about the self." In: *Journal of personality and social psychology* 35 (2), p. 63. ISSN: 1939-1315.

Markus, Hazel Rose and Shinobu Kitayama (2010). "Cultures and selves: A cycle of mutual constitution". In: *Perspectives on psychological science* 5 (4), pp. 420–430. ISSN: 1745-6916.

Massey, Douglas S (2007). *Categorically Unequal: The American stratification system*. Russell Sage Foundation. ISBN: 1610443802.

Massey, Douglas S and Garvey Lundy (2001). "Use of Black English and racial discrimination in urban housing markets: New methods and findings". In: *Urban affairs review* 36 (4), pp. 452–469. ISSN: 1078-0874.

Mayer, R. C., J. H. Davis, and F. D. Schoorman (1995). "An Integrative Model of Organizational Trust". In: *Academy of Management Review* 20 (3), pp. 709–734. ISSN: 0363-7425. DOI: 10.5465/amr.1995.9508080335.

McAdams, Dan P (2008). *The person: An introduction to the science of personality psychology*. John Wiley and Sons. ISBN: 0470129131.

McGuigan, Nicola, Jenny Makinson, and Andrew Whiten (2011). "From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children". In: *British Journal of Psychology* 102 (1), pp. 1–18. ISSN: 20448325. DOI: 10.1348/000712610X493115.

McKenzie, David (2012). "Beyond baseline and follow-up: The case for more T in experiments". In: *Journal of development Economics* 99 (2), pp. 210–221. ISSN: 0304-3878.

McQueen, Amy and William M.P. P Klein (2006). "Experimental manipulations of self-affirmation: A systematic review". In: *Self and Identity* 5 (4), pp. 289–354. ISSN: 15298868. DOI: 10.1080/15298860600805325.

Medin, Douglas L. and Marguerite M. Schaffer (1978). "Context theory of classification learning". In: *Psychological Review* 85 (3), pp. 207–238. ISSN: 0033295X. DOI: 10.1037/0033-295X.85.3.207.

Mendoza-Denton, Rodolfo et al. (2001). "Person× Situation interactionism in self-encoding (Iam... when...): Implications for affect regulation and social information processing." In: *Journal of Personality and Social Psychology* 80 (4), p. 533. ISSN: 1939-1315.

Mesoudi, Alex (2011). *Cultural evolution*. University of Chicago Press. ISBN: 0226520455.

Michalopoulos, Stelios and Melanie Meng Xue (Jan. 2021). "Folklore". In: *The Quarterly Journal of Economics*. ISSN: 0033-5533. DOI: 10.1093/qje/qjab003. URL: https://doi.org/10.1093/qje/qjab003.

Miles, Steven H (2005). *The Hippocratic Oath and the ethics of medicine*. Oxford University Press. ISBN: 0195188209.

Milkman, Katherine L et al. (2021). "A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment". In: *Proceedings of the National Academy of Science* 118 (20). DOI: 10.1073/pnas.2101165118/-/DCSupplemental. URL: https://www.reuters.com/article/health-coronavirus-.

Minsky, Marvin (1975). *A Framework for Representing Knowledge*. Ed. by P. Winston.

Mousteri, Victoria, Michael Daly, and Liam Delaney (2018). "The scarring effect of unemployment on psychological well-being across Europe". In: *Social Science Research* 72, pp. 146–169. ISSN: 0049-089X.

Mukand, Sharun and Dani Rodrik (2018). *The political economy of ideas: On ideas versus interests in policymaking*. URL: http://www.nber.org/papers/w24467.

Murphy, Jamie et al. (2021). "Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom". In: *Nature Communications* 12 (1), pp. 1–15. ISSN: 20411723. DOI: 10.1038/s41467-020-20226-9. URL: http://dx.doi.org/10.1038/s41467-020-20226-9.

Neisser, Ulric and Robert Becklen (1975). "Selective looking: Attending to visually specified events". In: *Cognitive psychology* 7 (4), pp. 480–494. ISSN: 0010-0285.

Neuberg, Steven L and Susan T Fiske (1987). "Motivational influences on impression formation: outcome dependency, accuracy-driven attention, and individuating processes." In: *Journal of personality and social psychology* 53 (3), p. 431. ISSN: 1939-1315.

Nisbett, Richard E and Timothy D Wilson (1977). "Telling more than we can know: Verbal reports on mental processes." In: *Psychological review* 84 (3), p. 231. ISSN: 1939-1471.

North, Douglass C (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press. ISBN: 9780521394161. DOI: DOI:10.1017/CBO9780511808678.

Nussbaum, Martha C (2001). "Symposium on Amartya Sen's philosophy: 5 Adaptive preferences and women's options". In: *Economics and Philosophy* 17 (1), pp. 67–88. ISSN: 1474-0028.

– (2011). *Creating capabilities*. Harvard University Press. ISBN: 0674050541.

Nyhan, Brendan and Jason Reifler (2015). "Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information". In: *Vaccine* 33 (3), pp. 459–464. ISSN: 18732518. DOI: 10.1016/j.vaccine.2014.11.017. URL: http://dx.doi.org/10.1016/j.vaccine.2014.11.017.

Nyhan, Brendan et al. (2014). "Effective messages in vaccine promotion: A randomized trial". In: *Pediatrics* 133 (4). ISSN: 10984275. DOI: 10.1542/peds.2013-2365.

Oettingen, Gabriele (2015). *Rethinking positive thinking: Inside the new science of motivation*. Current. ISBN: 1617230235.

Oettingen, Gabriele and Doris Mayer (2002). "The motivating function of thinking about the future: Expectations versus fantasies". In: *Journal of Personality and Social Psychology* 83 (5), pp. 1198–1212. ISSN: 00223514. DOI: 10.1037//0022-3514.83.5.1198.

Organisation, World Health (2017). *Vaccination and trust*, p. 35. URL: http://www.euro.who.int/en/health-topics/disease-prevention/vaccines-and-immunization/publications/2017/vaccination-and-trust-2017.

Orozco-Olvera, Victor, Fuyuan Shen, and Lucie Cluver (2019). "The effectiveness of using entertainment education narratives to promote safer sexual behaviors of youth: A meta-analysis, 1985-2017." In: *PloS one* 14 (2), e0209969. ISSN: 1932-6203 (Electronic). DOI: 10.1371/journal.pone.0209969.

Ortony, Andrew (1978). "Remembering , Understanding , and Representation". In: *Cognitive Science* 69, pp. 53–69.

Ostrom, Elinor (1990). *Governing the Commons*. ISBN: 978-3-531-90400-9. DOI: 10.1007/978-3-531-90400-9_93. URL: http://dx.doi.org/10.1007/978-3-531-90400-9_93.

– (2009). *Understanding institutional diversity*. Princeton university press. ISBN: 1400831733.

Ostrom, Elinor and Michael Cox (2010). "Moving beyond panaceas: A multi-tiered diagnostic approach for social-ecological analysis". In: *Environmental Conservation* 37 (4), pp. 451–463. ISSN: 03768929. DOI: 10.1017/S0376892910000834.

Pager, Devah, Bruce Western, and Bart Bonokowski (2006). "Discrimination in a Low Wage Labor Market: A Field Experiment". In: *American Sociological Review* 74 (5), pp. 777–789. ISSN: 0003-1224. DOI: 10.1177/000312240907400505.Discrimination.

Paluck, Elizabeth Levy (2009). "Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda." In: *Journal of Personality and Social Psychology* 96 (3), pp. 574–587. ISSN: 1939-1315. DOI: 10.1037/a0011989.

Paul, Karsten I. and Klaus Moser (2009). "Unemployment impairs mental health: Meta-analyses". In: *Journal of Vocational Behavior* 74 (3), pp. 264–282. ISSN: 00018791. DOI: 10.1016/j.jvb.2009.01.001. URL: http://dx.doi.org/10.1016/j.jvb.2009.01.001.

Paunesku, David et al. (2015). "Mind-Set Interventions Are a Scalable Treatment for Academic Underachievement". In: *Psychological science* 26 (6), pp. 784–793. ISSN: 14679280. DOI: 10.1177/0956797615571017.

Petty, Richard E and Jon A Krosnick (2014). *Attitude strength: Antecedents and consequences*. Psychology Press. ISBN: 1317782364.

Pezdek, Kathy et al. (1989). "Memory for Real-World Scenes: The Role of Consistency With Schema Expectation". In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15 (4), pp. 587–595. ISSN: 02787393. DOI: 10.1037/0278-7393.15.4.587.

Phelps, Edmund (1972). "The Statistical Theory of Racism and Sexism". In: *The American Economic Review* 62 (4), pp. 659–661.

Pinker, Steven (1993). *The language instinct: How the mind creates language*. Penguin UK. ISBN: 0141929685.

– (1997). "How the mind works." In: *NY: Norton*.

Putnam, Robert D (1993). *Making Democracy Work: Civic traditions in modern Italy*. Princeton university press. ISBN: 140082074X.

Quillian, Lincoln et al. (2017). "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time". In: *Proceedings of the National Academy of Sciences* 114 (41), pp. 10870–10875. ISSN: 0027-8424.

Ramsey, Frank P (1926). *Truth and probability*.

Randolph, Haley E. and Luis B. Barreiro (2020). "Herd Immunity: Understanding COVID-19". In: *Immunity* 52 (5), pp. 737–741. ISSN: 10974180. DOI: 10.1016/j.immuni.2020.04.012. URL: https://doi.org/10.1016/j.immuni.2020.04.012.

Rasmussen, Angela L. (2020). "Vaccination Is the Only Acceptable Path to Herd Immunity". In: *Med* 1 (1), pp. 21–23. ISSN: 26666340. DOI: 10.1016/j.medj.2020.12.004. URL: https://doi.org/10.1016/j.medj.2020.12.004.

Ray, Debraj (2006). "Aspirations, poverty, and economic change". In: *Understanding poverty* 409421.

Reid, Vincent M. et al. (2017). "The Human Fetus Preferentially Engages with Face-like Visual Stimuli". In: *Current Biology* 27 (12), 1825–1828.e3. ISSN: 09609822. DOI: 10.1016/j.cub.2017.05.044. URL: http://dx.doi.org/10.1016/j.cub.2017.05.044.

Rendell, Luke et al. (2011). "Cognitive culture: Theoretical and empirical insights into social learning strategies". In: *Trends in Cognitive Sciences* 15 (2), pp. 68–76. ISSN: 13646613. DOI: 10.1016/j.tics.2010.12.002. URL: http://dx.doi.org/10.1016/j.tics.2010.12.002.

Riley, Amy Henderson et al. (Jan. 2017). "Encoded Exposure and Social Norms in Entertainment-Education." In: *Journal of health communication* 22 (1), pp. 66–74. ISSN: 1087-0415 (Electronic). DOI: 10.1080/10810730.2016.1250843.

Roberts, Brent W et al. (2009). *Conscientiousness*. Ed. by M. Leary and R. Hoyle.

Rodrik, Dani (2010). "Diagnostics before Prescription". In: *Journal of Economic Perspectives* 24 (3), pp. 33–44. ISSN: 0895-3309. DOI: 10.1257/jep.24.3.33.

– (2014). "When Ideas Trump Interests". In: *Journal of Economic Perspectives* 28 (1), pp. 189–208. ISSN: 0895-3309. DOI: 10.1257/jep.28.1.189. URL: file:///C:/Users/nmaloberti/Downloads/DownloadtoMendeley/s9.pdf.

Roediger, Henry L. and Kathleen B. McDermott (1995). "Creating False Memories: Remembering Words Not Presented in Lists". In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21 (4), pp. 803–814. ISSN: 02787393. DOI: 10.1037/0278-7393.21.4.803.

Rokeach, Milton (1973). *The nature of human values.* Free press. ISBN: 0029267501.

Rosch, Eleanor (1999). "Principles of categorization". In: *Concepts: core readings* 189, pp. 312–322.

Roth, Alvin E and Ido Erev (1995). "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term". In: *Games and economic behavior* 8 (1), pp. 164–212. ISSN: 0899-8256.

Roth, Martha Tobi, Harry A Hoffner, and Piotr Michalowski (1997). *Law collections from Mesopotamia and Asia minor*. Vol. 134. Scholars Press Atlanta.

Rotter, Julian B (1966). "Generalized expectancies for internal versus external control of reinforcement." In: *Psychological monographs: General and applied* 80 (1), p. 1. ISSN: 0096-9753.

Rubinstein, Ariel (2012). *Economic fables*. Open book publishers. ISBN: 1906924775.

Rumelhart, David E (1980). "Schemata: The building blocks of cognition". In: *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence and education* 11 (1), pp. 33–58.

Rumelhart, David E and Andrew Ortony (1977). *The representation of knowledge in memory*. Ed. by R. C. Anderson, R. J. Spiro, and W. E. Montague.

SAGE (2020). *Factors influencing COVID-19 vaccine uptake among minority ethnic groups*. Presented at Sage73, pp. 1–15.

Savage, Leonard J (1954). *The foundations of statistics*. Courier Corporation. ISBN: 0486623491.

Scheinman, Steven J., Patrick Fleming, and Kellyann Niotis (2018). "Oath taking at U.S. And Canadian medical school ceremonies: Historical perspectives, current practices, and future considerations". In: *Academic Medicine* 93 (9), pp. 1301–1306. ISSN: 1938808X. DOI: 10.1097/ACM.0000000000002097.

Schwartzstein, Joshua and Adi Sunderam (2021). "Using models to persuade". In: *American Economic Review* 111 (1), pp. 276–323. ISSN: 19447981. DOI: 10.1257/aer.20191074.

Scott, James C (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press. ISBN: 0300078153.

Sen, Amartya (1999). *Development as Freedom*. DOI: 10.1215/0961754X-9-2-350.

Sherman, David K and Geoffrey L Cohen (2006). "The psychology of self-defense: Self-affirmation theory". In: *Advances in experimental social psychology* 38, pp. 183–242. ISSN: 0065-2601.

Shermohammed, Maheen et al. (Feb. 2021). "Informing patients that they are at high risk for serious complications of viral infection increases vaccination rates." In: *medRxiv : the preprint server for health sciences*. DOI: 10.1101/2021.02.20.21252015. URL: http://www.ncbi.nlm.nih.gov/pubmed/33655258http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7924279.

Shiller, Robert J (2017). "Narrative economics". In: *American Economic Review* 107 (4), pp. 967–1004. ISSN: 0002-8282.

Simon, Herbert A (1955). "A behavioral model of rational choice". In: *The quarterly journal of economics* 69 (1), pp. 99–118. ISSN: 1531-4650.

Simons, Daniel J. and Christopher F. Chabris (1999). "Gorillas in our midst: Sustained inattentional blindness for dynamic events". In: *Perception* 28 (9), pp. 1059–1074. ISSN: 03010066. DOI: 10.1068/p281059.

Smith, Eliot R and Sarah Queller (2001). "Mental representations". In: *Blackwell handbook of social psychology: Intraindividual processes*, pp. 111–133.

Sood, Neeraj et al. (2021). "Shifting the Demand for Vaccines: A Review of Strategies". In: *Annu. Rev. Public Health*. DOI: 10.1146/annurev-publhealth. URL: https://doi.org/10.1146/annurev-publhealth-.

Spelke, Elizabeth S. and Katherine D. Kinzler (2007). "Core knowledge". In: *Developmental Science* 10 (1), pp. 89–96. ISSN: 1363755X. DOI: 10.1111/j.1467-7687.2007.00569.x.

Stadler, Michael A., Henry L. Roediger, and Kathleen B. McDermott (1999). "Norms for word lists that create false memories". In: *Memory and Cognition* 27 (3), pp. 494–500. ISSN: 0090502X. DOI: 10.3758/BF03211543.

Steele, Claude and Joshua Aronson (1995). "Stereotype Threat and the Intellectual Test Performance of African Americans". In: *Journal of Personality and Social Psychology* 69 (5), pp. 797–811. ISSN: 0013-189X. DOI: 10.4324/9780203463932.

Steele, Claude M (1988). *The psychology of self-affirmation: Sustaining the integrity of the self*.

Strauss, Claudia and Naomi Quinn (1997). *A cognitive theory of cultural meaning*. Cambridge University Press. ISBN: 052159541X.

Swidler, Ann (1986). "Culture in action: Symbols and strategies". In: *American sociological review*, pp. 273–286. ISSN: 0003-1224.

Tabellini, Guido (2010). "Culture and institutions: Economic development in the regions of Europe". In: *Journal of the European Economic Association*. URL: https://academic.oup.com/jeea/article/8/4/677/2295864.

Tanguy, Bernard et al. (2014). *The future in mind: Aspirations and forward-looking behaviour in rural Ethiopia*.

Taylor, Shelley E. and Jonathon D. Brown (1988). "Illusion and Well-Being: A Social Psychological Perspective on Mental Health". In: *Psychological Bulletin* 103 (2), pp. 193–210. ISSN: 00332909. DOI: 10.1037/0033-2909.103.2.193.

Tomasello, Michael (1999). *The Cultural Origins of Human Cognition*. Harvard university press. ISBN: 0674044371.

Tomasello, Michael, Ann Cale Kruger, and Hilary Horn Ratner (1993). "Cultural learning". In: *Behavioral and brain sciences* 16 (3), pp. 495–511. ISSN: 1469-1825.

Tuckett, David and Milena Nikolic (2017). "The role of conviction and narrative in decision-making under radical uncertainty". In: *Theory and Psychology* 27 (4), pp. 501–523. ISSN: 14617447. DOI: 10.1177/0959354317713158.

Tversky, Amos and Daniel Kahneman (1973). "Availability: A heuristic for judging frequency and probability". In: *Cognitive psychology* 5 (2), pp. 207–232. ISSN: 0010-0285.

– (1974). "Judgment under uncertainty: Heuristics and biases". In: *Science* 185 (4157), pp. 1124–1131. ISSN: 0036-8075.

Walsh, James (2013). "The Case for Quotas". In: *Women's Policy Journal of Harvard* 10, p. 104. ISSN: 1534-0473.

Whiten, Andrew et al. (2009). "Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1528), pp. 2417–2428. ISSN: 14712970. DOI: 10.1098/rstb.2009.0069.

Williams, S Elizabeth et al. (2013). "A Randomized Trial to Increase Acceptance of Childhood Vaccines by Vaccine-Hesitant Parents: A Pilot Study". In: *American Pediatrics* 13, pp. 475–480. URL: http://www.vvrp.info/.

Wilson, Timothy D. and Elizabeth W. Dunn (2004). "Self-knowledge: Its limits, value, and potential for improvement". In: *Annual Review of Psychology* 55, pp. 493–518. ISSN: 00664308. DOI: 10.1146/annurev.psych.55.090902.141954.

Yeager, David S et al. (2016). "Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school." In: *Journal of educational psychology* 108 (3), p. 374. ISSN: 1939-2176.

Zerubavel, Eviatar (2009). *Social mindscapes: An invitation to cognitive sociology*. Harvard University Press. ISBN: 0674045432.