Essays in Behavioral Economics

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Peter Andre

aus Mainz

2022

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Armin Falk
Zweitreferent:	Prof. Dr. Thomas Dohmen
Tag der mündlichen Prüfung:	14. Januar 2022

Acknowledgements

Among the myriad alternative paths that the universe could have taken, there are few in which this dissertation would exist. My own contribution appears negligible if one considers the many circumstances and events that have made this work possible: a society that devotes considerable resources to the education of its members, the pioneers who transformed economics into a profession and Bonn into a place where it is fun and exciting to be its apprentice, the insights and inspiration I found in the illuminating work of others, and, of course, the invaluable support of many institutions, colleagues, friends, and my family. Dissertations in economics do not enjoy a large readership. This may hold even more for their acknowledgment section. Those who – for whatever reason – made it to these lines, be aware that what follows could well be the most important sentence among the many I have written in the last few years: My deepest gratitude to all those who made this possible!

I want to thank my advisor Armin Falk for the many conversations we had. Each a lesson in economics, academia, and life (often at the same time). Each has shaped my understanding of the world. Likewise, I want to thank Thomas Dohmen, my second supervisor. Their guidance, support, and honest, encouraging (and sometimes challenging) feedback have been integral for me and the work in this dissertation.

But even two supervisors of Armin's and Thomas's caliber can take you only so far. It arguably takes a department to "raise" a Ph.D. student, and I am indebted to many faculty members and colleagues, among them Sebastian Kube, Florian Zimmermann, and Chris Roth. Chris has been a mentor and friend for me. He helped me navigate the academic landscape, and, throughout the years, we have discussed almost all my research ideas.

I also want to thank my co-authors, Teodora Boneva, Felix Chopra, Armin Falk, Carlo Pizzinelli, Chris Roth, and Johannes Wohlfart, for the collaboration, the great time and insightful discussions, and the challenges we overcame together. Felix and I were in the same Ph.D. cohort and became office mates, sparring partners, co-authors, and friends.

I am also grateful for the support of the Bonn Graduate School of Economics, the briq Institute on Behavior & Inequality, the Institute for Applied Microeconomics, the BonnEconLab (Holger Gerhardt!), the Collaborative Research Center Transregio

224, ECONtribute, and the Joachim-Herz foundation. The support of Markus Antony and Stefanie Sauter at briq was integral to many of my Ph.D. projects.

Fortunately, research is not a solitary endeavor but a joint effort of thousands of interconnected people all around the world. I was fortunate to meet many great people, colleagues, and friends at Bonn and the rest of the world. The full list would likely surpass in length even the appendix of chapter 2, but I want to mention: Susann Adloff (we traveled to Papua New Guinea), Zvonimir Bašić (we used to discuss image concerns for hours each week), and Luca Henkel (we love taking walks by the Rhine).

Finally, I want to thank my wife, Maren, my parents, sisters, and friends for their love, faith, and support. Their support was never conditional on this dissertation, but this dissertation would not have been possible without them.

Contents

List of Figures

List of Tables

Introduction

We all aspire to understand the social and economic world around us. Our way of living, our decisions, desires, and attitudes are strongly shaped by how we understand the economy and our socioeconomic communities. How, for example, would the economy react to a change in government spending? Should we be worried about high inflation rates when central banks flood the markets with money? How individuals subjectively answer these questions matters when they decide how much money to save, whether to search for a new job or to expect low or high inflation rates in the future (Woodford, 2013; Shiller, 2017; Gennaioli and Shleifer, 2018). People's perceptions of the social world are similarly consequential. Which social rules apply in my community? Would others shun me if I take a plane for a short private weekend trip and ignore the ensuing CO_2 emissions? Or, on the contrary, would they regard me as a hillbilly if I stay at home and do not travel the world? Humans share a deep desire to be respected members of their community. Hence, what they perceive to be their community's social norms strongly affects their behavior and attitudes (Fehr and Fischbacher, 2004; Bicchieri, 2006).

At the same time, we all desire to make the social and economic world a better and fairer place. Questions of fairness and justice are ubiquitous in political debate, the workplace, and our private relationships with family and friends. Which inequalities are unfair? Should policies address climate change? Should people stop eating meat to reduce their carbon footprint? Again, how individuals answer these questions matters because it shapes their demand for economic policies and the standards to which they hold their fellow citizens (Alesina and Giuliano, 2015; Bursztyn and Jensen, 2017).

This dissertation revolves around these two themes: individuals' perceptions of reality and their beliefs about what characterizes a good, fair, and just world. Both are powerful determinants of people's behavior and attitudes. For economic analysis, they matter not only because they influence economic behavior but also because they influence which economic policies individuals support and thus often set the political constraints that policymakers face in practice. The dissertation presents four independent research papers that can be viewed as concrete examples for the two broad themes of the dissertation. Below, I briefly summarize each chapter.

Chapter 1: Shallow Meritocracy. Chapter 1 focuses on people's fairness views. It starts from the observation that meritocracies aspire to reward effort and hard work but promise not to judge individuals by the circumstances they were born into. The choice to work hard is, however, often shaped by circumstances. The chapter investigates whether people's merit and fairness judgments are sensitive to this endogeneity of choice. Do they hold others responsible for their choices even when these choices have been shaped by external circumstances?

The study proceeds in four steps. First, I isolate and identify the effect of interest in an incentivized choice experiment with a representative sample from the US population. Study participants judge how much money two workers deserve for the effort they exerted. In the treatment condition, unequal circumstances strongly discourage one of the workers from working hard. Nonetheless, I find that individuals hold the disadvantaged worker fully responsible for his choice. In the second step, additional follow-up experiments explore the behavioral mechanism underpinning this result. I find that participants neglect the endogeneity of choices even though they understand that choices are strongly influenced by circumstances. Instead, in light of an uncertain counterfactual state – what would have happened on a level playing field – participants base their merit judgments on the only reliable evidence they have: observed effort levels. In the third step, a structural model integrates the findings into a preference framework. Finally, a vignette study showcases the relevance of the experimental findings in labor market and career choice scenarios.

While meritocratic fairness promises that the family, neighborhood, and circumstances one is born into should not matter, the findings of this study suggest that meritocratic fairness is likely to be "*shallow*" in practice. People neglect that external circumstances also influence the choices that agents make and hold them fully responsible for these choices. Thus, choices can "launder" unequal circumstances and legitimize the ensuing inequality.

Chapter 2: Subjective Models of the Macroeconomy. Chapter 2 investigates people's mental "models" of how the macroeconomy works. The study measures beliefs about the effects of hypothetical macroeconomic shocks on unemployment and inflation. It finds that beliefs are widely dispersed. This even holds for beliefs about the directional effects of shocks. Moreover, there are large differences in the average beliefs between households and experts.

Part of this disagreement arises from selective retrieval of different propagation channels. Respondents think about different channels through which the shocks affect the economy. For instance, households tend to think about supply-side channels even for shocks that are traditionally viewed as demand-side shocks. The propagation channels that are on top of their minds affect their predictions. The study confirms this causally by exogenously shifting households' attention to either supply-side or demand-side channels. The chapter also shows that households with different personal experiences recall different propagation channels for the shocks. These findings offer a new perspective on the widely documented disagreement in macroeconomic expectations.

Chapter 3: Fighting Climate Change: The Role of Norms, Preferences, and Moral Values. Chapter 3 documents that people in the US underestimate the prevalence of climate-friendly behaviors and norms among their fellow citizens. Moreover, providing respondents with correct information causally raises their willingness to fight climate change, measured through an incentivized donation decision, as well as individual support for climate policies. The effects are strongest for individuals who are skeptical about the existence and threat of global warming. The study also explores which other behavioral determinants shape individual willingness to fight climate change and finds that economic preferences, such as patience and altruism, and universal moral values positively predict climate preferences.

This chapter demonstrates that misperceptions of climate norms prevail in the US and can form a dangerous obstacle to climate action. However, at the same time, they can provide a unique opportunity to promote and accelerate climate-friendly behavior. A simple, easily scalable, and cost-effective intervention can correct these misperceptions and encourage climate-friendly behavior. This intervention is particularly effective for climate change skeptics, who are commonly difficult to reach but crucial for building a broad alliance against climate change. The results suggest that social norms should play a pivotal role in the policy response to climate change.

Chapter 4: What's Worth Knowing? The final Chapter 4 raises the question of what is worth knowing and worth studying in economics. Since there is no clear, scientific, or objective response to this question, researchers are forced to retreat to their gut feeling, instincts, and personal value judgments (Weber, 1919). The fourth chapter aims to document these judgments and analyzes the views of almost 10,000 academic economists from all fields and ranks of the profession. It asks which topics economics should work on and which research objectives it should pursue. The chapter describes three main results of the survey. First, economists' opinions are substantially heterogeneous. Second, most researchers are dissatisfied with economics' current research topics and objectives. Third, on average, respondents think economic research should become more policy-relevant, multidisciplinary, risky and disruptive, and pursue more diverse topics. The results, thus, suggest that economics as a field does not appreciate and work on what economists collectively prefer.

On the one hand, the chapter can be read as an application of the dissertation's two main themes to the production of economic research. Perceptions of the status quo and views about "what economics should do" are likely to be relevant for researchers' choice of research topics, referees' publication recommendations, and the design of research institutions and academic incentive systems. On the other hand, the chapter can be viewed as a personal attempt to come to grips with the question of what is worth knowing and studying. This question haunts many Ph.D. students, and, yes, it also haunted me during my Ph.D. studies. Which research projects should I start? Which topics matter? The four chapters of this dissertation document what I eventually found worth studying. Chapter 4 also sheds light on economists' opinions from all around the world.

People's perceptions of reality and their fairness views shape individual behavior and support for policies. The chapters of this dissertation illustrate that they matter in many domains of economic analysis. While the importance of studying determinants of individual behavior is widely acknowledged in economics, the lack of systematic research on people's policy views still surprises me. Economists invest a lot of time and energy into analyzing and determining optimal policies, yet, in many cases, the public decides which policies are eventually implemented. Many economists would profess that the public's understanding of economics is intriguing and imaginative but often detached from established economic knowledge. Economic policies need to be designed and communicated in light of these political constraints. This is another argument for why it is crucial to understand people's political perceptions and preferences. For example, we are just beginning to understand lay models of the economy. Likewise, climate change raises novel and largely unexplored questions of how the public thinks about intergenerational fairness. I hope to explore these issues and the themes of this dissertation in future work.

References

Alesina, Alberto, and Paola Giuliano. 2015. "Culture and Institutions." *Journal of Economic Literature* 53 (4): 898–944. [1]

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York: Cambridge University Press. [1]

Bursztyn, Leonardo, and Robert Jensen. 2017. "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure." *Annual Review of Economics* 9: 131–53. [1]

Fehr, Ernst, and Urs Fischbacher. 2004. "Social norms and human cooperation." *Trends in Cognitive Sciences* 8 (4): 185–90. [1]

Gennaioli, Nicola, and Andrei Shleifer. 2018. *A Crisis of Beliefs: Investor Psychology and Financial Fragility*. Princeton: Princeton University Press. [1]

Shiller, Robert J. 2017. "Narrative Economics." *American Economic Review* 107 (4): 967–1004. [1]

Weber, Max. 1919. "Wissenschaft als Beruf." In *Geistige Arbeit als Beruf*. München, and Leipzig: Duncker & Humblot. [3]

Woodford, Michael. 2013. "Macroeconomic Analysis Without the Rational Expectations Hypothesis." *Annual Review of Economics* 5: 303–46. [1]

Chapter 1

Shallow Meritocracy: An Experiment on Fairness Views

Abstract: Meritocracies aspire to reward effort and hard work but promise not to judge individuals by the circumstances they were born into. The choice to work hard is, however, often shaped by circumstances. This study investigates whether people's merit judgments are sensitive to this endogeneity of choice. In a series of incentivized experiments with a large, representative US sample, study participants judge how much money two workers deserve for the effort they exerted. In the treatment condition, unequal circumstances strongly discourage one of the workers from working hard. Nonetheless, I find that individuals hold the disadvantaged worker fully responsible for his choice. They do so, even though they understand that choices are strongly influenced by circumstances. Additional experiments identify the cause of this neglect. In light of an uncertain counterfactual state – what would have happened on a level playing field – participants base their merit judgments on the only reliable evidence they possess: observed effort levels. I confirm these patterns in a structural model of merit views and a vignette study with real-world scenarios.

Acknowledgements: I thank Felix Chopra, Thomas Dohmen, Armin Falk, Thomas Graeber, Leander Heldring, Luca Henkel, Paul Hufe, Ingo Isphording, Fabian Kosse, Yucheng Liang, Matt Lowe, Wladislaw Mill, Franz Ostrizek, Christopher Roth, Sebastian Schaube, Andreas Stegmann, Florian Zimmermann, and participants at various conferences for helpful comments and discussions. **Funding**: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866. Funding by the Deutsche Forschungsgemeinschaft (DFG) through CRC TR 224 (Project A01) is gratefully acknowledged. Supported by the Reinhard-Selten Scholarship (German Association for Experimental Economic Research). Supported by the Joachim Herz Foundation. **Ethics approval**: The study obtained ethics approval from the German Association for Experimental Economic Research (#HyegJqzx, 12/11/2019). **Research transparency**: The study was pre-registratered at the AEA RCT Registry (#AEARCTR-0005811). Data and code will be made available. I declare no competing interests. See also Appendix 1.D on research transparency. **Instructions**: The full experimental instructions of all studies are available at https://osf.io/xj7vc/.

1.1 Introduction

The notion of meritocratic fairness is at the heart of Western political and economic culture. It shapes which inequalities we consider to be fair, which redistributive policies we implement, and how we design our welfare states (Alesina and Glaeser, 2004; Alesina and Angeletos, 2005; Cappelen, Falch, and Tungodden, 2020; Sandel, 2020). In essence, meritocratic fairness means that people should be rewarded in proportion to their merit. Besides talent and skill, the choice to work hard and exert effort is considered a central determinant of merit. By contrast, external circumstances outside the individual's control, such as parental background, race, or sex, are not legitimate sources of merit (Konow, 2000; Alesina and Angeletos, 2005; Almås, Cappelen, and Tungodden, 2020; Cappelen, Falch, and Tungodden, 2020). Meritocratic fairness thus distinguishes between effort choices (relevant for merit) and external circumstances (irrelevant for merit). However, this distinction is clouded by a fundamental feature of reality: Agents' choices are endogenous to and shaped by their circumstances, opportunities, and incentives. For instance, a person growing up with few opportunities and incentives to work hard might respond by exerting little effort. Likewise, minorities that experience discrimination might be discouraged from working hard. Indeed, empirical studies have linked effort, career, and schooling choices to gender norms, racial inequality, and the socio-economic environment (e.g., Carrell, Page, and West, 2010; Bursztyn, Fujiwara, and Pallais, 2017; Glover, Pallais, and Pariente, 2017; Falk, Kosse, and Pinger, 2020; Altmejd, Barrios-Fernández, Drlje, Goodman, Hurwitz, et al., 2021). Moreover, the fact that adverse environments often encourage detrimental decision-making is considered a key cause of poverty (e.g., Bertrand, Mullainathan, and Shafir, 2004; Haushofer and Fehr, 2014).

Any meritocracy thus needs to take a stance on how choices that are shaped by external circumstances should be rewarded. Should choices be evaluated in light of or irrespective of their circumstances? This study explores the prevailing concept of meritocratic fairness and investigates how people reward choices in a series of online experiments with a large, representative US sample of about 4,000 respondents. The study proceeds in four steps. First, I isolate and identify the effect of interest, which requires the control of an incentivized choice experiment. I find that merit judgments completely neglect the endogeneity of choices. In the second step, additional follow-up experiments explore the behavioral mechanism underpinning this result. Third, a structural model integrates the findings into a preference framework. Finally, a vignette study showcases the relevance of the experimental findings in labor market and career choice scenarios.

In the main experiment, participants ("spectators", $n = 653$) judge how much money other people ("workers") deserve for their effort in a piece-work job. The workers initially earn a randomly assigned piece-rate (their circumstances). They know that their piece-rate can either be high ($0.50) or low ($0.10) with 50%

chance each. Chance determines that one worker receives the high rate, whereas the other worker receives the low rate. Workers decide freely how many tasks they want to complete (their effort choice). Unsurprisingly, workers work much harder and complete roughly three times as many tasks for the higher piece-rate (the endogeneity of choices). In the second step, each spectator is assigned to one pair of workers and informed about their task and circumstances. Spectators decide which final reward each worker deserves. In multiple scenarios, they can redistribute the earnings between the two workers, conditional on workers' effort choices. These merit judgments are the central outcome variable of the study.

The experiment exogenously varies in which circumstances workers make their effort choices. In the *control* condition, the workers do not know their realized piece-rates yet. They only know their odds to obtain a high or low piece-rate, which are identical for both workers. Hence, their effort choices are directly comparable because their choices are made in the same environment and subject to the same situational influence – a level playing field. By contrast, in the *treatment* condition, workers immediately learn about their realized piece-rates. Workers with a high piece-rate are encouraged and advantaged by these circumstances, whereas workers with a low piece-rate are discouraged and disadvantaged. Thus, in the treatment condition, but not in the control condition, the endogeneity of choices differentially (dis)advantages the workers. I compare spectators' merit judgments across the two conditions. Do merit judgments reward the same effort choices equally across conditions, thereby ignoring the external circumstances in which workers make their decisions? Alternatively, do spectators compensate the disadvantaged workers in the treatment condition for the fact that they are discouraged from working hard?

The results show that participants' merit judgments are completely insensitive to the endogeneity of choices. The spectators strongly redistribute payments to reward workers for higher effort, but they do so equally in both conditions. They neglect that the disadvantaged worker is discouraged from working hard in the treatment condition but not in the control condition. The average reward share of the disadvantaged worker is even (insignificantly) 0.49 percentage points (pp) lower in the treatment than in the control condition. The large sample size allows me to rule out even minor increases in the reward of the disadvantaged worker (0.8 pp of total payoff). The results thus provide strong evidence for the absence of a meaningful effect. Spectators hold workers responsible for their choices, even if these choices are endogenous and shaped by external situational influence over which the workers have no control.

Why do spectators neglect the endogeneity of workers' choices? To shed light on the behavioral mechanism behind this finding, I run tailored follow-up experiments. I start by investigating whether spectators underestimate the power of situational influence, in line with the well-known *fundamental attribution error* (Ross, 1977). I measure incentivized beliefs about how strongly the piece-rates influence workers' effort choices. However, spectators even slightly overestimate the piece-rate effect, so that its neglect cannot simply be attributed to biased beliefs. Of course, this does not rule out that the endogeneity of choices escapes spectators' *attention* while rewarding the workers. In the second step, I therefore implement an attention intervention ($n = 274$) in which I draw spectators' attention to the effect of situational influence just before their merit judgments. However, merit judgments remain insensitive to the endogeneity of choices even then.

Compensating for disadvantageous situational influence also raises the question of what the two workers to whom a spectator is assigned would have done in identical circumstances. This *counterfactual* is unknown and uncertain even for spectators who accurately anticipate the average piece-rate effect. Therefore, I test for the role of counterfactual reasoning in another experiment ($n = 945$) in which I provide a subset of spectators with accurate information about what the disadvantaged worker would have done in the advantaged environment. I find that, on average, spectators' merit judgments react strongly to the counterfactual effort choice of the disadvantaged worker. Once the counterfactual is revealed to them, they take the endogeneity of choices into account and compensate workers who are disadvantaged by external situational influence. This suggests that the uncertainty of the counterfactual – what would have happened on a level playing field – explains why merit judgments are insensitive to the endogeneity of choices. When the counterfactual is unknown, spectators simply base their merit judgments on the only clear and reliable evidence they have, namely the observed effort choices. This results in a "burden of the doubt" for the disadvantaged worker.

The average results discussed earlier conceal that merit judgments are vastly heterogeneous. In the next step, I therefore estimate a structural model of merit views to assess the prevalence of different merit views in the population. The model builds on a simple theoretical framework that I sketch in the introductory Section 1.2 and thus brings the study's argument full circle. I distinguish between four distinct merit views: comparable choice meritocrats, actual choice meritocrats, libertarians, and egalitarians. "Comparable choice meritocrats" hold workers accountable for the counterfactual effort choices that workers would make in identical, comparable circumstances, but – in line with the reduced-form results – potentially discount this counterfactual when it is unknown and uncertain. "Actual choice meritocrats" reward workers proportional to their actual effect choices, even if these choices are endogenous to external circumstances. "Libertarians" accept any inequality and do not redistribute. Lastly, "egalitarians" think that the workers always deserve equal payments. The estimated model classifies 26% of participants as comparable choice meritocrats. In line with the reduced-form results, I estimate that they fully neglect situational influence when the counterfactual is uncertain. Meanwhile, 37% of participants are classified as actual choice meritocrats, 23% as libertarians, and 14% as egalitarians. The results show that people hold fundamentally different merit views. Importantly, they also reveal that, even in a (counterfactual) world where counterfactual choices were known, only about 26% of individuals would compensate for disadvantageous situational influence. The prevailing meritocratic fairness ideal ignores the endogeneity of choices.

Although the controlled experimental environment comes with the crucial advantage that the effect of interest is clearly and credibly identified, it also comes at a cost: It differs from many real-life settings that characterize the debate about merit, choices, and circumstances. To mitigate this concern, I run a vignette study ($n = 1{,}222$) showing that the insensitivity of merit judgments to the endogeneity of choices can also be observed in labor market and career choice scenarios. For instance, participants do not compensate a black employee who chooses not to work hard for a promotion but faces racial discrimination and has no chance of being promoted anyway. Likewise, they do not compensate a person who shows hardly any effort in his or her life but grew up in a discouraging environment with few opportunities and incentives to work hard. In both cases, the choice not to work hard legitimizes a highly unequal outcome, irrespective of the disadvantageous external situational influence.

Discussion. The pros and cons of meritocracy have been the subject of a heated public debate (Young, 1958; Greenfield, 2011; Frank, 2016; Markovits, 2019; Sandel, 2020). Meritocratic fairness promises that the family, neighborhood, and circumstances one is born into should not matter – a popular notion that closely connects to the prominent ideas of equal opportunity and the American dream. However, the findings of this study suggest that meritocratic fairness is likely to be "*shallow*". Even though meritocratic fairness holds that individuals should not be judged by their external circumstances, people neglect that these external circumstances also influence the choices that agents make and hold them fully responsible for these choices. Thus, choices "launder" unequal circumstances and legitimize the ensuing inequality.

In practice, not only effort but also valuable talents and abilities, such as cognitive skills, are viewed as meritorious and worthy of reward. These talents, skills, and personality traits are also shaped by external circumstances, in particular, during early childhood (e.g., Heckman, 2006; Putnam, 2016; Alan and Ertac, 2018; Kosse, Deckers, Pinger, Schildberg-Hörisch, and Falk, 2019). Hence, while this study focuses on the endogeneity of *choices*, an analogous question arises for the endogeneity of *skills*. The former is the starting point of this study because it is the simpler, more transparent, and relatable channel. Because individuals ignore the endogeneity of choices – an effect they should be well familiar with –, I expect that a similar neglect also arises for the endogeneity of *skills*.

Of course, holding others responsible for their actual choices (or skills) may simply be a practical necessity of living together. The results of the study thus connect to an old theme in the philosophy of responsibility (Eshleman, 2016; Nelkin, 2019), but the study neither can nor aims to settle this normative debate. Instead, it documents which merit views people endorse in practice.

These views on fairness matter because they characterize the society in which we live. Ultimately, the neglect of endogeneity is likely to shape which policies voters demand. “Shallow meritocrats” endorse *predistribution* policies that level the playing field and equate circumstances ex-ante. Yet, they are reluctant to compensate others for unequal circumstances via *redistribution* after unequal choices have been made. This could explain why ex-post policies such as affirmative action are considered controversial and suggests that policymakers who want to mobilize support for advancing equality of opportunity should emphasize ex-ante, predistributive policies.

Related literature. The study builds on and contributes to several strands of the literature. The fairness views of the general population have long been a focus of economic research because they are recognized as an important determinant of welfare systems and a defining feature of political culture (Alesina and Glaeser, 2004; Alesina and Angeletos, 2005; Giuliano and Spilimbergo, 2013; Kuziemko, Norton, Saez, and Stantcheva, 2015; Alesina, Stantcheva, and Teso, 2018; Andreoni, Aydin, Barton, Bernheim, and Naecker, 2020; Fisman, Kuziemko, and Vannutelli, 2020; Stantcheva, 2021). Past research documents that the idea of merit is at the center of fairness and inequality acceptance. Merit is associated with choices such as to work hard or to take risks. Unequal rewards derived from unequally meritorious choices are typically considered fair and legitimate (Cappelen, Hole, Sorensen, and Tungodden, 2007; Cappelen, Sørensen, and Tungodden, 2010; Krawczyk, 2010; Cappelen, Konow, Sørensen, and Tungodden, 2013; Mollerstrom, Reme, and Sørensen, 2015; Akbaş, Ariely, and Yuksel, 2019; Almås, Cappelen, and Tungodden, 2020). Small differences in merit sometimes justify large reward inequalities (Bartling, Cappelen, Ekström, Sørensen, and Tungodden, 2018; Cappelen, Moene, Skjelbred, and Tungodden, 2020). Moreover, Cappelen, Fest, Sørensen, and Tungodden (2020) show that even degenerate choices can have meritorious character. Participants in their study reward “choices” even when the agents have no real choice and can only decide between two identical alternatives. Thus, merit judgments seem to be all about choice. By contrast, luck and circumstances outside the agents’ control are commonly rejected as a legitimate source of merit. However, how do merit judgments deal with the ubiquitous endogeneity of choices to external circumstances? This study is the first to address this question and provide an in-depth analysis of the underlying behavioral mechanisms.

The finding that people are held responsible for their choices even if these choices are the product of external circumstances also relates to the literature on moral responsibility and moral luck (Nagel, 1979; Baron and Hershey, 1988; Bartling and Fischbacher, 2012; Gurdal, Miller, and Rustichini, 2013; Brownback and Kuhn, 2019; Falk, Neuber, and Szech, 2020). Individuals are often approved or disapproved not only for their choices but also the consequences of their choices, even if these are accidental, unintended, and the product of chance. Here, I show that individuals can be held responsible for external luck not only if it shapes the consequences of their decisions but also if it directly shapes the decision they make.

This study also connects to a recent literature on inference in economics (e.g., Enke and Zimmermann, 2017; Benjamin, 2019; Han, Liu, and Loewenstein, 2020; Graeber, 2021; Liang, 2021). In particular, individuals often struggle with complex decisions in uncertain and contingent environments (Esponda and Vespa, 2014; Esponda and Vespa, 2019; Martínez-Marquina, Niederle, and Vespa, 2019) – a key element of counterfactual reasoning. However, counterfactual reasoning itself remains relatively unexplored in economics, even though cognitive scientists have long since acknowledged its centrality to causal reasoning and inference (Kahneman and Miller, 1986; Roese, 1997; Sloman, 2005; Byrne, 2016; Lagnado and Gerstenberg, 2017). This study illustrates that counterfactual reasoning is a potent mechanism. The inherent uncertainty of the counterfactual strongly affects individuals' choices even though they accurately anticipate the expected counterfactual.

The remainder of the paper is structured as follows. Section 1.2 sets the stage by discussing a simple conceptual framework of merit views, Section 1.3 describes the main experimental design, and Section 1.4 presents the main results. Section 1.5 examines their behavioral foundations, Section 1.6 structurally estimates the model of fairness views, and Section 1.7 reports the vignette study. Finally, Section 1.8 concludes the paper.

1.2 Conceptual framework

To fix ideas, I introduce a simple theoretical framework that directly maps into the experimental design. Two workers, $k \in \{A, B\}$, independently choose how much effort E_k they exert, given their external circumstances, namely their returns to effort π_k. As in the experiment, the workers' returns to effort are externally determined by a lottery. Worker A randomly receives a high piece-rate, whereas worker B randomly receives a low piece-rate. The workers have convex effort costs $\frac{1}{2}(E_k - \theta_k)^2$, where θ_k is their diligence or taste for hard work. Hence, worker k maximizes $\pi_k E_k - \frac{1}{2}(E_k - \theta_k)^2$, chooses the optimal effort level

$$E_k^* \;=\; \theta_k + \pi_k,$$

and earns $P_k = \pi_k E_k^*$. The optimal choice E_k^* can be decomposed into an "internal" cause (θ_k) and an "external" cause (π_k). Thus, conditional on their types θ, worker A works harder due to their higher returns to effort. Worker A (high piece-rate) is advantaged, whereas worker B (low piece-rate) is disadvantaged by external situational influence.[1]

1. I abstract from income effects on labor provision (i.e., worker's utility is linear in money) because income effects will arguably be absent in the experimental application. The structural as-

How is workers' merit in this setting evaluated? Suppose that a neutral third person observes this situation. In line with the literature on fairness preferences, I refer to the third party as "spectator" because the spectator's own monetary payoff is not at stake. The spectator (hereafter referred to as "she" or "her") observes the workers' (referred to as "he" or "him") circumstances, the share of the total payment that the disadvantaged worker B receives $p = \frac{P_B}{P_A+P_B}$, and the share of total work that he conducts $e = \frac{E_B}{E_A+E_B}$. Without loss of generality, I focus on the disadvantaged worker B because he will be at the center of the later analysis. Moreover, I focus on his payment and effort shares (denoted by lower case letters) because they can easily be compared across situations. The spectator can redistribute the workers' earnings to implement the reward share r of worker B that she prefers. Redistribution comes at no cost.[2] I assume that spectator i maximizes the utility function

$$U(r_i) \;=\; -\frac{1}{2}\left[r_i - m_i(e,s)\right]^2$$

where $m_i(e,s)$ denotes i's merit view, that is, her view about which reward the disadvantaged worker deserves for providing the effort share e in the external situation s. Thus, the spectator wants to implement the reward share r_i that she thinks is merited by worker B in situation s.

$$r_i^* \;=\; m_i(e,s)$$

This set-up combines several features that are well-suited to characterize merit judgments. First, it focuses on choices about effort and hard work that play a major role in the debate about merit. Second, it deals with *relative* merit judgments, that is the merit of worker B compared to worker A. After all, "high" or "low" effort and "high" or "low" rewards can most easily be distinguished in comparison. Third, rewards are assigned via redistribution to mirror the fact that society's fairness views are often implemented via redistributive schemes that intervene into naturally arising market outcomes.

I consider four distinct merit views.

Actual choice meritocrat: $m_i(e,s) = e$
For "actual choice meritocrats", choice is the only relevant criterion of merit. They hold people fully responsible for their choices, even if these choices are endogenous to external circumstances. In the worker setting sketched earlier, actual choice meritocrats hold that the disadvantaged worker B deserves a payment share equal to

---

sumptions on the effort cost function C are made for illustrative purposes only. The argument in this paper depends mainly on $\partial E_k^* / \partial \pi_k > 0$.

2. For simplicity, I abstract from the frequently studied fairness-efficiency trade-off. Existing research shows that fairness concerns often dominate efficiency concerns (Almås, Cappelen, and Tungodden, 2020).

his effort share. For instance, he deserves 25% of the payment if he completed 25% of the tasks; he deserves 75% of the payment if he completed 75% of the tasks.

Comparable choice meritocrat: $m_i(e,s) = \hat{E}_i c(e,s)$
"Comparable choice meritocrats" do not hold individuals responsible for external causes of choice (π) but only for internal (θ).[3] To subtract any external influence on choice, a comparable choice meritocrat asks, "What would the two workers have done in an identical, comparable situation?" Merit is derived from these counterfactual, comparable effort choices. Accordingly, comparable choice meritocrats think that the disadvantaged worker B deserves a payment share equal to the counterfactual effort share c that he would have provided had he been in the same advantaged circumstances as worker A.[4] Since comparable choice meritocratism requires an inference about the counterfactual, biased counterfactual reasoning could lead to a discrepancy between the perceived counterfactual effort share $\hat{E}_i c(e,s)$ and the actual but unknown counterfactual effort share $c(e,s)$.

Egalitarian: $m_i(e,s) = 50\%$
The workers always deserve equal payment shares (as in Almås, Cappelen, and Tungodden, 2020).

Libertarian: $m_i(e,s) = p$
Any pre-existing earning share p is regarded as legitimate and accepted (as in Almås, Cappelen, and Tungodden, 2020).

In sum, actual choice meritocrats equate merit and effort even if external circumstances shape the effort choices. By contrast, comparable choice meritocrats identify merit with counterfactual effort choices in identical, comparable circumstances. Because the counterfactual is uncertain, their merit judgments also depend on their inference and counterfactual reasoning. The other two types, egalitarians and libertarians, do not condition merit on choice. They respectively accept no or any form of unequal rewards and play only a minor role in the context of this study.

Conceptually, there are intriguing *normative* arguments for both actual choice and comparable choice meritocratism. For instance, incentives to behave well could deteriorate if individuals are not fully accountable for their actual choices. Moreover, workers already bore the costs of their working decisions. Why should a lazy worker be rewarded for the hard work he would have done (but did not do) in a counterfactual environment? On the other hand, it seems inconsistent to claim that external

3. These internal causes of choice, such as type or preference differences, can often be attributed to differential external circumstances as well – be it nature or nurture (Heckman, 2006; Cesarini, Dawes, Johannesson, Lichtenstein, and Wallace, 2009; Dohmen, Falk, Huffman, and Sunde, 2012; Kosse et al., 2019). Ultimately, one could hence even ask whether these differences can justify merit differences. However, this question is outside the scope of this paper.

4. In principle, comparable choice meritocrats could also base their merit judgments on counterfactual effort choices in another environment, for example, the low piece-rate situation. Relatedly, Roemer (1993) takes an individual's relative ranking in the effort distribution conditional on circumstances, $f(E_k^*|\pi_k)$ as a comparable measure of merit. These details affect neither the qualitative argument here nor the qualitative interpretation of later treatment effects.

circumstances should not influence merit judgments, while their external influence on choice does.

Here, however, the research question is of *positive* nature: Which merit judgments does the general population make? First, are they sensitive to the endogeneity of choices? Second if not, are they insensitive because comparable choice meritocrats are absent from the population or because they misinfer what would have happened without situational influence and fail to apply their merit view? The main experiment sets out to investigate the first question in an environment that mirrors the simple model sketched above. Tailored mechanism experiments follow to explore the second question.

1.3 Experimental design

Studying how the endogeneity of effort choices shapes merit judgments requires a setting where choices are central to merit and merit judgments can be measured in an incentivized way. And it requires experimental conditions that exogenously vary the situational influence of external circumstances on choices. Below, I describe how I tailor the experimental design to meet both requirements.

1.3.1 Setting: Redistribution task

I create an experimentally controlled situation of inequality between *workers* and observe how study participants (*spectators*) redistribute money between the workers, conditional on workers' effort choices. Spectators decide what each worker deserves and thereby judge which merit originates from the workers' choices. The set-up is in line with the framework sketched in Section 1.2.

Workers. I hire US workers on Amazon's online labor market *Mechanical Turk* for a crowd-working job in which they collect email address data for another research project. In each task, a worker is given the name of a person, searches for the person's website, identifies their email address, and enters it in a data collection form. Typically, it takes about two minutes to complete one task. The crowd-working job requires no special qualification but demands effort and concentration, ensuring that hard work determines success rather than skill. Each worker k earns a piece-rate π_k and can freely choose how many tasks E_k to complete. Workers know that a lottery determines their piece-rate, which can either be high ($0.50) or low ($0.10). A worker's initial payment is $\pi_k E_k$. Workers know that someone else might influence their payment, but they neither know when, why, nor how this happens, nor who is involved in this process. This guarantees that workers cannot distort their effort decisions in anticipation of a later redistribution stage. Each worker additionally receives a fixed remuneration of $1. The full instructions for the workers are available online (https://osf.io/xj7vc/).

For the redistribution stage, workers are assigned to pairs. I will refer to the two workers in a pair as workers A and B. I focus on pairs where worker A receives a high piece-rate of $0.50 and worker B receives a low piece-rate of $0.10.[5] Inequality between the two workers is likely to prevail – either due to differences in effort E_k or the piece-rate π_k. Whereas effort E_k is a choice variable, the piece-rate π_k is outside the control of workers but is likely to shape the workers' effort choices. Indeed, workers complete, on average, more than three times as many tasks (mean: 16.8 tasks) for a high piece-rate of $0.50 than for a low piece-rate of $0.10 (mean: 5.0 tasks, see Appendix 1.C), rendering the setting well-suited to study how merit judgments react to situational influence.

Spectators. I invite adults from the general US population to participate in the online experiment. Each study participant ("spectator") is assigned to a pair of workers and informed about the workers' task, situation, choices, and earnings. In particular, spectators know that a lottery determines the workers' piece-rate. Spectators then determine the final earnings of both workers and judge which percentage share of the total performance-based earnings each worker deserves. That is, they can redistribute the earnings between both workers.[6] Redistribution comes at no cost. Spectators know that their decision is strictly anonymous and that workers are unaware of the redistribution stage. Appendix 1.E provides the main instructions for spectators, and the full instructions are available online (https://osf.io/xj7vc/).

The redistribution decisions of spectators, neutral third-parties who have no monetary stake in the distribution of funds, commonly serve as a measure of fairness behavior and views (e.g., Mollerstrom, Reme, and Sørensen, 2015; Cassar and Klein, 2019; Almås, Cappelen, and Tungodden, 2020; Andreoni et al., 2020). They mirror the fact that society's fairness views are often implemented via redistributive schemes that intervene into naturally arising market outcomes – a feature that I want to capture in the experiment. I implement the merit judgments of 100 randomly selected spectators so that spectators' decisions are (probabilistically) incentivized. After all, their decisions can have real and meaningful consequences for the workers.[7]

5. In the experiment, I randomly vary whether worker A or worker B is the worker with the advantageous, high piece-rate. Here, I recode all responses as if worker A was the advantaged worker to ease analysis and exposition. Reassuringly, Table 1.B.5 shows that spectators' redistributive behavior is insensitive to whether worker A or worker B is advantaged. Moreover, sometimes both workers of a pair receive a piece-rate of $0.10 or both receive a piece-rate of $0.50. These worker pairs are used in additional experiments that I will introduce later.

6. Spectators cannot redistribute the fixed remuneration of $1 but only the performance-based rewards.

7. Charness, Gneezy, and Halladay (2016) review the advantages and disadvantages of implementing the decisions of a subset of participants versus those of all participants. The literature documents little difference between both methods.

Table 1.3.1. Overview of effort scenarios, experimental conditions, and studies

(A) Effort scenarios (presented in random order)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Effort share of worker B: *e*	**0%**	**10%**	**30%**	**50%**	**70%**	**90%**	**100%**
Effort of worker A	50	45	35	25	15	5	0
Effort of worker B	0	5	15	25	35	45	50
Payment of worker A	\$25.00	\$22.50	\$17.50	\$12.50	\$7.50	\$2.50	\$0.00
(Share)	(100%)	(98%)	(92%)	(83%)	(68%)	(36%)	(0%)
Payment of worker B	\$0.00	\$0.50	\$1.50	\$2.50	\$3.50	\$4.50	\$5.00
(Share)	(0%)	(2%)	(8%)	(17%)	(32%)	(64%)	(100%)

Contingent response method: Each spectator faces eight effort scenarios. The seven scenarios above are hypothetical. An eighth effort scenario (not shown) is real. Spectators do not know which scenario is real and have to take each of their decisions seriously.

(B) Experimental conditions (between-subject)

	Control condition		***Treatment condition***	
Worker	**A**	**B**	**A**	**B**
Constant across conditions				
Realized π	\$0.50	\$0.10	\$0.50	\$0.10
Effort choices	*Depends on effort scenario*			
Payment	*Results from effort scenario and realized π*			
Varies across conditions				
Expected π	\$0.50 or \$0.10 with 50% each	\$0.50 or \$0.10 with 50% each	\$0.50	\$0.10

(C) All experimental studies (for later reference)

Study	**Section**	**Description**
Main study	1.3, 1.4	Varies whether endogeneity of choices (dis)advantages workers.
Attention study	1.5.2	Shifts attention towards endogeneity of choices.
Counterfactual study	1.5.3, 1.6	Reveals what would have happened in equal circumstances.
Vignette study	1.7	Explores merit judgments in exemplary real-world scenarios.
Robustness		
“Equal rates” study	1.4	Replicates main study, but workers receive same piece-rate.
“Disappointment” study	1.4	Explores motive to compensate workers for disappointment.
“Equal rates” attention study	1.5.2	“Equal rates” version of the attention study (see above).

Notes: Panel A presents an overview of all effort scenarios. Panel B summarizes and compares the experimental conditions. Panel C lists all experimental studies that I present in this paper. Only the main study is introduced in this section. The details of all other studies will be introduced in later sections.

To elicit spectators' merit judgments for various effort choices, I employ a contingent response method. Each spectator decides whether and how to redistribute the earnings in eight different effort scenarios. Each scenario describes how many tasks worker A and how many tasks worker B completed. The first seven scenarios are hypothetical, presented in random order, and selected to represent various effort shares of worker B (denoted by $e = \frac{e_B}{e_A + e_B}$). Panel A of Table 1.3.1 summarizes these effort scenarios. For example, in Scenario 1, worker A does all the work and completes 50 tasks, whereas worker B completes no task at all ($e = 0\%$). In Scenario 4, both workers complete 25 tasks ($e = 50\%$). Moreover, in Scenario 7, worker A completes 0 tasks and worker B completes 50 tasks ($e = 100\%$). The other scenarios present intermediate cases. The eighth scenario is real and describes how many tasks the two workers actually complete. Spectators' decisions in this scenario determine the workers' final payoff. However, spectators are not told which scenario is real and hence have to take each of their decisions seriously.[8] Effort choices in the real scenario vary across experimental conditions (introduced in the following) due to the incentive effects of the conditions. Thus, the real scenario does not allow a consistent comparison across treatments. To circumvent this problem, I only analyze the merit judgments in the first seven scenarios. The contingent response method is central for the identification because it allows analyzing merit judgments for the same effort scenario and effort choices across the treatment and control conditions.

1.3.2 Experimental conditions: Varying situational influence

In a between-subject design, I exogenously vary whether workers' effort choices are affected by situational influence. For this purpose, I manipulate *when* the workers learn about the realized piece-rate of their lottery and inform spectators about this. Panel B of Table 1.3.1 provides an overview of both conditions.

Control: Both workers do not know their realized piece-rate while making their effort choices. They are aware that their piece-rates might either be \$0.50 or \$0.10 with equal chance. They learn about their realized piece-rate (\$0.50 for worker A and \$0.10 for worker B) only after completing their work.

Treatment: Both workers are informed about their realized piece-rate already before they decide how much effort they exert. Thus, worker A knows about his high rate of \$0.50 and worker B about his low rate of \$0.10 when they decide how many tasks they complete.

8. Indeed, only a few spectators can distinguish the hypothetical scenarios from the real one, even after they saw all scenarios and made all of their redistribution decisions. When I ask them to guess which of the scenarios is real, 46% respond that they do not know. Among the others, only 16% guess correctly. Thus, the recognition rate is only slightly higher than what would be expected under random guessing (12.5%). The results are robust to excluding respondents who recognize the real scenario (see Appendix 1.B.2).

The experimental conditions vary whether the two workers in a pair optimize against identical or different piece-rate expectations. In the control condition, both workers face the same expected circumstances and respond to the same environment so that their effort choices are comparable. If one worker completes more tasks, this directly signals his higher taste for hard work. In the treatment condition, the workers face different circumstances and their effort choices are differentially affected by situational influence. The high piece-rate encourages worker A to work more, whereas the low piece-rate discourages worker B. Thus, if the advantaged worker A completes more tasks, this may reflect his higher taste for work or the advantageous situational influence. Do spectators account for this? By comparing spectators' redistributive behavior across treatment and control, I test whether and how the endogeneity of choices shapes merit judgments.

The contingent response method allows me to study merit judgments and their sensitivity to situational influence in seven different effort scenarios. Each scenario describes how much effort each worker exerts and how much money they initially earn. The scenarios are identical across the treatment and control conditions, but their interpretation changes. For instance, two workers who complete 25 tasks each (Scenario 4) show identical diligence in the control condition. However, in the treatment condition, working on 25 tasks for a \$0.50 piece-rate signals a much lower taste for hard work than working on 25 tasks for a \$0.10 piece-rate. As another example, if worker A completes 50 tasks and worker B does nothing (Scenario 7), worker A clearly signals higher diligence in the control condition. The situation is less clear in the treatment condition because the effort choices can be partially attributed to unequal circumstances.

For actual choice meritocrats, the difference between the treatment and control conditions is irrelevant. Their merit judgments depend solely on workers' actual effort choices which are identical across both conditions. But comparable choice meritocrats who recognize that worker B is disadvantaged by the endogeneity of choices and would work harder for a high piece-rate should compensate him with a higher reward share.

1.3.3 Experimental procedures

Workers. I recruited 336 workers on Amazon Mechanical Turk in May and June 2020 to participate in the crowd-working job. On average, the workers complete 12 tasks and earn about \$5.40, but both figures vary across experimental conditions. I form 100 pairs with 200 of those workers and use them to incentivize spectators' redistribution decisions.[9]

9. I ran the main experimental conditions together with robustness and mechanism experiments with a total of 1,855 participants. The additional conditions will be introduced later. The workers were recruited jointly for all experimental conditions. Appendix 1.A provides an overview. Workers who were not selected for the redistribution stage received their original performance-based payments.

Table 1.3.2. Comparison of the sample to the American Community Survey

Variable	ACS (2019)	Sample
Gender		
Female	51%	51%
Age		
18-34	30%	30%
35-54	32%	33%
55+	38%	37%
Household net income		
Below 50k	37%	40%
50k-100k	31%	34%
Above 100k	31%	27%
Education		
Bachelor's degree or more	31%	43%
Region		
Northeast	17%	21%
Midwest	21%	21%
South	38%	36%
West	24%	22%
Sample size	2,059,945	653

Notes: Column 1 presents data from the American Community Survey (ACS) 2019. Column 2 presents data from the representative online sample.

Spectators. I recruit a sample of 653 participants in collaboration with Lucid, an online panel provider which is frequently used in social science research (Coppock and McClellan, 2019, Haaland et al., forthcoming). The sample excludes participants who do not complete the first seven redistribution decisions or speed through the experimental instructions (see Appendix 1.A). The sampling plan and the exclusion criteria were pre-registered (see Appendix 1.D). The participants are broadly representative of the US adult population in terms of gender, age, region, income, and education. Table 1.3.2 displays summary statistics from the sample and compares them to the data obtained from the American Community Survey 2019. The sample follows the characteristics of the American population closely, except perhaps for education: 43% of the sample possess an undergraduate degree, compared to about 31% of the US population. Respondents were randomly assigned to either the treatment ($n = 329$) or the control ($n = 324$) condition. Appendix Table 1.A.2 shows that the covariates are balanced across experimental conditions.

The experiment took place online in June 2020. Most participants spent 10 to 30 minutes to complete the experiment (15% and 85% percentile), with a median response duration of 16 minutes. The experiment is structured as follows. First, the participants answer a series of demographic questions, which monitor the sampling process. Inattentive participants are screened out in an attention check. Detailed instructions on the workers' situation and the redistribution decisions follow. The experimental treatment-control variation is introduced only at the end of the instructions. This guarantees that the instructions about the workers' task and the redistribution decisions are understood and interpreted identically across conditions. Then, a quiz tests whether participants understand the key aspects of the experiment and corrects them if necessary. Subsequently, participants make their redistribution decisions. Each redistribution decision screen also contains a tabular summary of the workers' situation, including their expected and realized piece-rates, to ensure that this information is salient in the moment of decision-making. Finally, I ask a series of follow-up questions to collect additional demographic variables and probe for possible mechanisms.

1.3.4 Additional experiments

I run a series of additional experiments to explore the robustness of the results and shed light on its behavioral mechanisms. The details will be introduced in later sections. For later reference, Panel C of Table 1.3.1 provides an overview and brief description of all experiments.

1.4 Main result

I start by studying spectators' merit judgments in the control treatment. Here, workers' effort choices are comparable because they are made in an identical environment: Both workers expect either a \$0.50 or \$0.10 piece-rate (each with 50%). Only after completing their work, worker A learns that he randomly receives the high piece-rate of \$0.50, whereas worker B learns that he earns \$0.10 per completed task. Do spectators compensate worker B for the bad luck of a low piece-rate? Figure 1.4.1 visualizes the share of the total earnings that a spectator assigns to the disadvantaged worker. Panel A displays the mean share, averaged across all seven scenarios, and Panel B presents the results in each of the seven effort scenarios. The results show that spectators indeed counterbalance the bad luck of a low piece-rate. They strongly redistribute money from worker A (high piece-rate) to worker B (low piece-rate). Averaged across scenarios, worker B receives 44.1% of the total earnings (red bar), which is much higher than the share he would receive without redistribution (31.9%, gray line). In fact, many participants reward worker B proportionally to his effort share. They implement the payment shares that would have occurred if both workers had earned an identical rate (Appendix Figure 1.B.1). Thus, in the control condition where both workers react to the same environment, merit derives mostly from effort choices.[10]

10. Deviations from effort-proportional rewards indicate traces of libertarian and egalitarian redistributive behavior. For instance, in effort Scenario 4 where worker B contributes exactly half of the

This sets the stage for my main research question. Do spectators take the endogeneity of effort choices into account? In the treatment condition, workers learn about their realized piece-rates already before they make their effort choice. Consequently, worker B is disadvantaged as he endogenously reacts to a discouragingly low piece-rate of $0.10. By contrast, worker A is encouraged by a high piece-rate of $0.50. Do spectators assign a higher reward share to worker B in the treatment than in the control condition to compensate him for this disadvantageous situational influence?

The results show that merit judgments are fully insensitive to the endogeneity of choices. Figure 1.4.1 shows that the payment shares are indistinguishable between the treatment and the control condition. Worker B receives on average 43.6% of the total earnings in the treatment condition and 44.1% in the control condition (Panel A). Hence, spectators do not compensate worker B for the disadvantageous situational influence in the treatment condition. They even assign an (insignificant) 0.49 pp lower share to him ($p = 0.464$; see Table 1.4.1). Panel B shows that this conclusion holds for all seven scenarios. Whether worker A or B completes more tasks, or both work equally hard, spectators do not counterbalance the effect of external situational influence. None of the seven treatment-control comparisons detects a significant difference, nor does a highly powered joint F-test that tests the null hypothesis that treatment differences are zero in all seven effort scenarios ($p = 0.668$).[11]

This null result does not reflect a noisy estimate but rather constitutes a precisely estimated null finding. Averaged across scenarios, the 95% confidence interval of the treatment effect ranges from −1.8 to 0.8 pp. This means that I can reject even tiny effect sizes with high statistical confidence, namely that workers who are disadvantaged by situational influence receive a compensation of more than 0.8 pp of the total payment. The results thus provide strong evidence for the absence of a meaningful effect.[12]

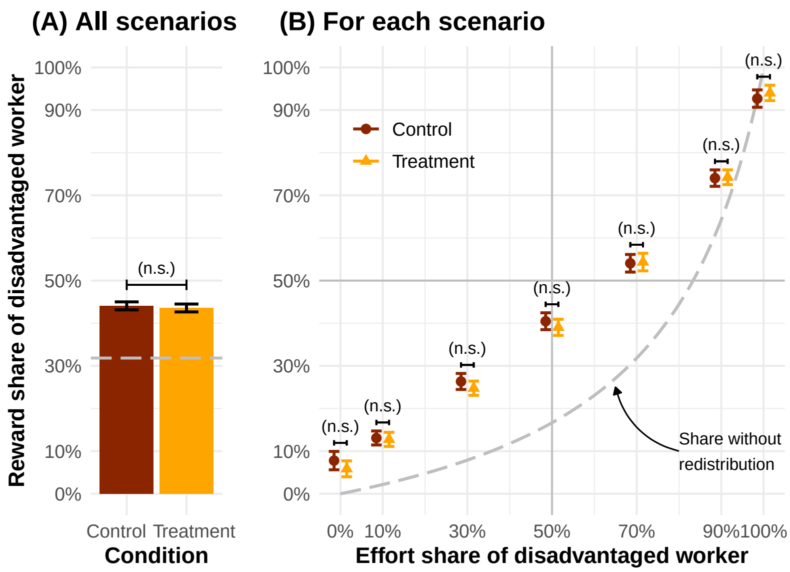
An average null effect might still conceal meaningful treatment effects for parts of the population. I therefore test for heterogeneous treatment effects. In the first step, I test for heterogeneity alongside six pre-registered covariates: gender, educa-

---

tasks, worker B receives a mean payment share of 40.5% rather than an equal 50.0% share. This is due to “libertarian” spectators who never redistribute and always accept the pre-existing reward share of 17% (see Figure 1.B.1). By contrast, in effort Scenario 1 where worker B completes no task at all, he still receives an average reward share of 7.8%. This is due to “egalitarian” spectators who always implement equal shares irrespective of the workers’ effort decisions (see Figure 1.B.1).

11. The F-test is derived from a regression of worker B’s payment share r_{is} on a treatment dummy interacted with a dummy for each scenario s and scenario fixed effects. It tests the null hypotheses that the treatment effects are zero in all seven effort scenarios. Standard errors are clustered on the participant level.

12. Precisely estimated null results are very informative from a Bayesian learning perspective – often even more informative than rejections of a null hypothesis (Abadie, 2020).



Notes: Results from the main study. Panel A displays the mean reward share assigned to the disadvantaged worker B in both experimental conditions, averaged across all seven effort scenarios, with 95% confidence intervals. Panel B plots the mean reward share in each effort scenario with 95% confidence intervals. The gray dashed line shows the default share, that is, which payment share worker B would receive if spectators do not redistribute. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, (n.s.) $p \geq 0.10$.

Figure 1.4.1. Average reward share of disadvantaged worker with 95% CI

tion, party affiliation, income, empathy, and internal locus of control. I assess empathy with four survey questions that measure perspective-taking and empathetic concern adopted from Davis (1983) and locus of control with a streamlined four-item scale developed in Kovaleva (2012). An internal locus of control measures whether a person attributes successes and failures to his or her own action and abilities instead of attributing them to luck, fate, or the actions of others. None of these variables significantly moderates the treatment effect (see Table 1.B.3).[13] In the second step, I apply the model-free approach of Ding, Feller, and Miratrix (2016) that tests whether *any* significant treatment heterogeneity exists. The method relies on randomization inference and basically tests whether the treatment distribution of the outcome variable is identical to the control distribution shifted by the average treatment effect. No significant heterogeneity in treatment effects is detected ($p = 0.446$), which corroborates my main result.

> ***Result***: *Individual merit judgments fully neglect the endogeneity of choices. People reward others for their effort, even if effort decisions are endogenous to external circumstances.*

13. Moreover, none of the variables is significantly associated with merit judgments in the baseline control condition.

Table 1.4.1. Treatment effects on average reward share of disadvantaged worker

	Mean reward share of disadvantaged worker (in %)				
	Main	Robust: No quiz mistakes	Robust: Decisions 1-3	Robust: High duration	Robust: With controls
	(1)	(2)	(3)	(4)	(5)
Treatment	−0.493	−1.002	−0.135	0.160	−0.353
	(0.673)	(0.827)	(1.335)	(0.785)	(0.684)
Constant	44.068***	44.792***	43.652***	43.479***	47.264***
	(0.480)	(0.573)	(0.915)	(0.553)	(4.569)
Controls	–	–	–	–	✓
Observations	653	395	653	471	634
R^2	0.001	0.004	0.000	0.000	0.004

Notes: Results from the main study, ordinary least squares (OLS) regressions, robust standard errors in parentheses. The outcome variable is the reward share (in %) a spectator assigns to the disadvantaged worker B, averaged across all seven effort scenarios. The independent variable is a treatment indicator. Column 1 presents the main specification. Columns 2-5 present different robustness specifications: Column (2) excludes respondents who initially answer at least one quiz question incorrectly, Column (3) considers only the first three decisions of each participant, Column (4) excludes the 25% respondents with the lowest response duration, and Column (5) includes controls (indicators for female gender, college degree, and being Republican, as well as log income, and age). *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

Robustness

I replicate the results in multiple robustness checks. In the first set of robustness tests, I ensure that the findings are not driven by a misunderstanding of the instructions, survey-taking fatigue, or inattentive participants – all of which would increase survey noise and thus could potentially conceal treatment effects. In Column 2 of Table 1.4.1, I exclude participants who initially answer one of the control questions incorrectly which could indicate a lack of understanding. In Column 3, I restrict the analysis to the first three redistribution decisions each participant makes, which would arguably be less affected by survey fatigue. In Column 4, I exclude the 25% of participants with the lowest response duration to drop participants who might "speed through" the survey. All three specifications replicate the main results. Moreover, I obtain virtually identical results if I control for respondents' demographic background (Column 5).

Second, one might be concerned that the direct effect of the piece-rates on earnings is too salient and crowds out attention to situational influence. For example, a disadvantaged worker who completes 15 tasks and earns only \$1.50 would have earned \$7.50 with a high piece-rate. Spectators might primarily think about this difference and thereby overlook that the worker would also have worked much harder (e.g., complete 35 tasks for a payment of \$17.50). In other words, spectators might primarily focus on the fact that, for the same effort choice, the disadvantaged worker would have earned more with a higher piece-rate and simply forget that a higher piece-rate would also have changed his effort choice. However, evidence from an additional experiment (**robustness study: equal rates**, $n = 661$) does not support this explanation. The experiment relies on a between-subject treatment-control variation which is analogous to the main study but keeps the realized piece-rate of both workers constant.[14] As before, both workers have identical expectations about their piece-rate ($0.10 or $0.50 with an equal chance) in the control condition. In the treatment condition, worker A expects to earn either $0.50 or $0.90, whereas worker B expects to earn only $0.10 or $0.50. Thus, worker A is advantaged by situational influence and encouraged to work hard, whereas worker B is disadvantaged and discouraged from working hard. However, in both conditions, chance determines that both workers earn the same rate of $0.50, so that their initial earnings are fully proportional to their effort.[15] Consequently, there is no direct piece-rate effect on payments that could distract spectators. Turning to the results, I detect no significant difference in merit judgments across the two conditions. Spectators accept that earnings move proportionally with effort in both conditions. They reward effort – irrespective of whether or not it is shaped by situational influence. This independent robustness experiment thus fully replicates the main results. Again, the null result is obtained with high precision. The 95% confidence interval rules out even small treatment effects (above 0.9 pp), and I observe a null effect in each of the seven effort scenarios (Table 1.B.1, Panel B).

A third potential concern is that a compensation for disappointment confounds the null effect. Worker B receives bad news upon learning that he only earns a low piece-rate, and the timing of bad news could matter. In the control condition, worker B receives this information only after he stopped working which could lead to larger disappointment. For instance, workers who completed ten tasks hoping for a $0.50 piece-rate might be more disappointed to learn that they earn only $0.10 per task (control condition) than workers who learn about their $0.10 piece-rate already before they complete the ten tasks (treatment condition). If spectators share this concern, they might want to assign a higher payment share to worker B in the *control* condition to compensate him for the higher disappointment. Any such effect would

14. I ran the "equal rates" experiment in parallel to the main study in June 2020. The study protocol closely follows the main experiment. As before, the sample broadly represents the US population, and treatment assignment is balanced across covariates (see Appendix 1.A). The results are robust to excluding potentially inattentive responses (misunderstanding of the instructions, survey-taking fatigue, and "speeders"; see Appendix 1.B.2).

15. Workers who receive a $0.90 piece-rate are not used for this robustness study and receive their payments without a redistribution stage. Workers with a $0.10 piece-rate are used in a second "equal rates" control condition in which both workers earn $0.10. To maximize statistical power, I present results in which I pool the $0.50 and the $0.10 control conditions, but the results are virtually identical if I only use the $0.50 control condition described in the main text (see Appendix 1.B.2).

run opposite to the main treatment effect and could therefore conceal its existence. Admittedly, the explanation seems unlikely because, to account for the results, the two effects would need to offset each other in all seven effort scenarios. Nevertheless, I design an additional experiment (**robustness study: disappointment**, $n = 606$) that rules out this confounding channel.[16] I replicate the main design with one crucial exception: Workers do not make a choice. Instead, all workers have to complete exactly ten tasks. Since no choice is involved, choices are not endogenous, situational influence on effort choices does not exist, and there is no reason to compensate for it. However, the motive to compensate for the timing of bad news is still present. If it matters, spectators should compensate worker B with a higher payment share in the control condition. The results reveal a negligible and insignificant treatment difference (Table 1.B.4). On average, spectators assign a 2.2 pp higher reward share to worker B in the control than in the treatment condition – an effect size that could not even conceal a minor treatment effect.

Lastly, one could argue that the spectators attempt to draw inferences about the workers' life situations outside the experiment. For instance, a worker who completes 25 tasks for a \$0.10 piece-rate (treatment) might not only be more diligent than a worker who completes the same amount of tasks for better piece-rate prospects of either \$0.10 or \$0.50 (control). He might also assign a higher marginal value to money or have lower marginal opportunity costs of time. Spectators could interpret this as a sign of neediness and assign a higher payment share to the disadvantaged worker B in treatment than control. Any such argument predicts the existence of a treatment effect and is thus firmly rejected by the null result.

1.5 Mechanism

This section investigates why individuals' merit judgments are insensitive to the endogeneity of effort choices. The theoretical framework of Section 1.2 suggests two explanations. On the one hand, the endogeneity of choices could simply be irrelevant for merit views. Spectators' fairness preferences might hold that merit should be solely grounded on actual effort choices ("actual choice meritocratism"). On the other hand, spectators might actually prefer to correct for situational influence ("comparable choice meritocratism"), but they struggle to do so because they fail to infer what would have happened in identical, comparable circumstances. Here, I explore three behavioral obstacles that could impair spectators' inference:

16. I ran the "disappointment" experiment in February 2021 with a convenience sample of US adults recruited with the help of the survey company Lucid. Treatment assignment is balanced across covariates (see Appendix 1.A). The results are robust to the use of post-stratification weights (see Table 1.B.4).

the fundamental attribution error, a lack of attention, and the uncertainty of the counterfactual.[17]

1.5.1 Fundamental attribution error

Do spectators understand that circumstances affect choices, that is, that workers' effort strongly react to the piece-rate workers earn? It may well be the case that spectators overly attribute choices to the decision-maker and underestimate the role of circumstances. Such an inferential error would be in line with the so-called fundamental attribution error, namely the notion that individuals underestimate situational influence on human decisions (Ross, 1977). To shed light on this mechanism, the main study elicits participants' beliefs about how workers' effort choices react to the piece-rate. Spectators learn that workers complete on average five tasks for a \$0.10 piece-rate and estimate how many tasks workers complete on average for a \$0.50 piece-rate. Their responses are incentivized: One out of ten participants earns a \$5 Amazon gift card if her response is at most one task away from the true value.

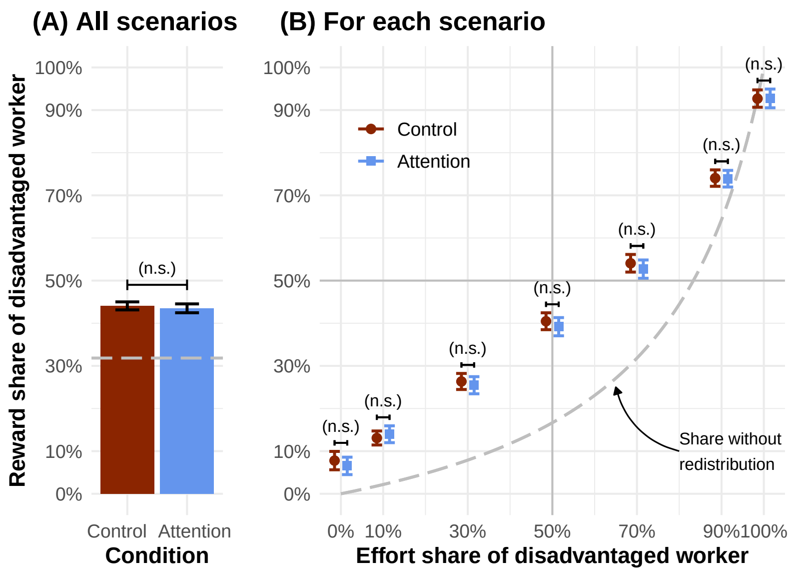
The findings do not support that a fundamental attribution error is driving the neglect of situational influence. Participants believe that workers complete 3.46 times as many tasks for a rate of \$0.50 than for a rate \$0.10. Thus, the perceived incentive effect is even slightly larger (though not significantly so) than the observed effect of 3.33 ($p = 0.749$, t-test).

1.5.2 Attention

Are spectators aware of the endogeneity of effort choices while making their merit judgments? Once asked explicitly about it, participants acknowledge that situational influence exists, but it might still escape their attention while they make their merit judgments. Attention (or a lack thereof) is a powerful explanation of behavior in many other domains (e.g., Chetty, Looney, and Kroft, 2009; Taubinsky and Rees-Jones, 2018; Gabaix, 2019; Andre, Pizzinelli, Roth, and Wohlfart, 2021). To test for this mechanism, I ran an additional experimental condition that draws participants' attention to the endogeneity of effort choices just before their merit judgments ($n = 274$).[18]

> **Attention**: I explicitly inform spectators that "the piece-rates strongly influence the number of tasks a worker completes." Spectators learn how large this incentive effect is on average and read two typical comments by workers that explain why

17. Cappelen, Mollerstrom, Reme, and Tungodden (2019) also study fairness views in an uncertain environment but their mechanism can only play a negligible role in my setting. They show that individuals do not want to risk rewarding the wrong person and hence prefer more equal rewards when it is unclear who merits the higher reward. However, in my setting, it is clear for comparable choice meritocrats that worker B merits a (weakly) higher reward in the treatment than in the control condition. It remains only unclear how much higher the reward should be. "Risk-averse" comparable



Notes: Results from the attention study. Panel A displays the mean reward share assigned to the disadvantaged worker B in both experimental conditions, averaged across all seven effort scenarios, with 95% confidence intervals. Panel B plots the mean reward share in each effort scenario with 95% confidence intervals. The gray dashed line shows the default share, that is, which payment share worker B would receive if spectators do not redistribute. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, (n.s.) $p \geq 0.10$.

Figure 1.5.1. Attention study: Average reward share of disadv. worker with 95% CI

this is the case. For example, the comment of a typical disadvantaged worker with a \$0.10 rate is: "For the amount of time that goes into these tasks, the compensation is simply just not sufficient." Participants have to spend at least 20 seconds on this information page, whose key message is repeated on the next page and tested for in the subsequent quiz.

Combining a qualitative statement, quantitative information, and workers' first-hand comments on their own experiences ensures that situational influence is salient to spectators while making their merit judgments. If a lack of attention to situational influence explains its neglect, spectators should compensate the disadvantaged worker with a higher reward share in the attention condition compared to the baseline control condition.

This is not the case. Participants who are informed about and focused on situational influence still do not compensate the disadvantaged workers. As before, the null effect is precisely estimated and present in each of the seven effort scenarios

choice meritocrats would still want to compensate the disadvantaged worker when the counterfactual is uncertain to ensure their reward decision is close to the expected fair merit judgment.

18. I ran this experiment in parallel to the main conditions in June 2020. The study protocol closely follows the main experiment. As before, the sample broadly represents the US population, and treatment assignment is balanced across covariates (see Appendix 1.A).

(see Figure 1.5.1). Aggregated across scenarios, the mean payment share of worker B is 43.5% in the attention condition versus 44.1% in the control condition. The 95% interval of their difference allows me to rule out even tiny treatment effects of 0.8 pp (see Table 1.B.1, Panel C).[19] Hence, a lack of attention to the endogeneity of effort choices also does not explain the results.

1.5.3 Uncertainty of the counterfactual

Compensating worker B for the disadvantageous situational influence he is exposed to does not only require an understanding and an awareness of the average piece-rate effect. It also raises the concrete question of what the two workers to whom a spectator has been assigned would have done in identical circumstances. How many tasks would worker B have completed had he also earned a high piece-rate of $0.50? Such a counterfactual benchmark would underlie the reward decision of a comparable choice meritocrat, who believes that external situational influence cannot justify merit and hence would want to correct for it.[20] However, this counterfactual is unknown and uncertain even for spectators who accurately anticipate the average piece-rate effect. Recent research shows that people struggle with complex decisions in uncertain and contingent environments, rendering this a promising explanation for why spectators' merit judgments neglect the endogeneity of choices (Esponda and Vespa, 2019; Martínez-Marquina, Niederle, and Vespa, 2019).

I devise a new mechanism experiment in which some spectators are explicitly informed about worker B's counterfactual effort choice, thereby removing any uncertainty about the counterfactual state (counterfactual study, $n = 945$).[21] For this purpose, I recruit new workers and elicit their effort choice for both the high and the low piece-rate. Workers commit to how many tasks they would complete for both piece-rates, are then randomly assigned to one piece-rate, and subsequently have to follow-up on their commitment. Importantly, this technique measures worker's counterfactual effort choice in an incentivized way. Thus, I know how many tasks the workers (would) complete for both piece-rates. Spectators are informed about this procedure. As before, they make merit judgments in eight scenarios of which

19. The results are robust to excluding potentially inattentive responses (misunderstanding of the instructions, survey-taking fatigue, "speeders"; see Appendix 1.B.2). I also replicate the results in an analogous extension of the robustness experiment with equal piece-rates (**attention: equal rates**, $n = 267$, see Table 1.B.1, Panel D).

20. As discussed in Section 1.2, this benchmark is not unique. For instance, a comparable choice meritocrat might also ask what both workers would have done for a low piece-rate of $0.10 or in another common piece-rate environment.

21. I ran this experiment in January 2021. The study protocol closely follows the main experiment. As before, the sample broadly represents the US population, and treatment assignment is balanced across covariates (see Appendix 1.A). The results are robust to excluding potentially inattentive responses (misunderstanding of the instructions, survey-taking fatigue, "speeders"; see Appendix 1.B.2).

seven are hypothetical and allow me to freely vary the counterfactual effort choice of worker B (contingent response method). Spectators do not know which of the eight scenarios is real so that all of their decisions are probabilistically incentivized. The first three scenarios are taken from the main experiment and are presented in random order. Here, the advantaged worker A completes more tasks than the disadvantaged worker B, that is, 50 to 0 tasks ($e = 0\%$), 45 to 5 tasks ($e = 10\%$), or 35 to 15 tasks ($e = 30\%$).[22] The next four scenarios are randomly generated and will be used in Section 1.6. Spectators are randomized into one of three experimental conditions. The conditions vary whether and what spectators learn about what the disadvantaged worker would have done in the advantaged environment. Table 1.5.1 provides an overview of all effort scenarios and experimental conditions.

No information (short: None): No information about worker B's counterfactual effort choice is provided. The condition thus replicates the main treatment condition and serves as a baseline condition in this experiment.

Low counterfactual (short: Low): Spectators are informed about worker B's counterfactual effort choice for a high piece-rate. In the "low counterfactual" condition, worker B would not change his effort provision and thus would not exert more effort for a higher piece-rate. This also means that worker B's effort choice is not shaped by situational influence.

High counterfactual (short: High): This condition provides information about worker B's counterfactual effort choice, too. Here, however, worker B would complete as many tasks as worker A for a high piece-rate. Situational influence thus exists and strongly affects worker B's choice. Workers A and B (would) make the same choices in the advantaged environment; hence, this information also implies that they share the same taste for hard work.

Figure 1.5.2 presents the results (see also Table 1.B.2). First, it reveals that the average reward for worker B is very similar in the "no information" condition and the "low counterfactual" condition.[23] Thus, in the baseline condition with unknown counterfactual, spectators reward worker B as if they knew that his counterfactual effort choice would be no different. This suggests that spectators in the baseline condition base their merit judgments on the assumption that choices have not been shaped by situational influence. They focus on observable effort choices, the only reliable evidence they have, akin to a "burden of the doubt" for the disadvantaged worker.

22. In the other scenarios of the main experiment, the disadvantaged worker completes the same or a larger number of tasks than the advantaged worker. These scenarios are not compatible with the "high counterfactual" condition and therefore not included.

23. If at all, spectators are even slightly more generous toward worker B in the "low counterfactual" condition. This difference is significant in the scenario where worker B has an effort share of 30%.

Table 1.5.1. Experimental conditions in the counterfactual study

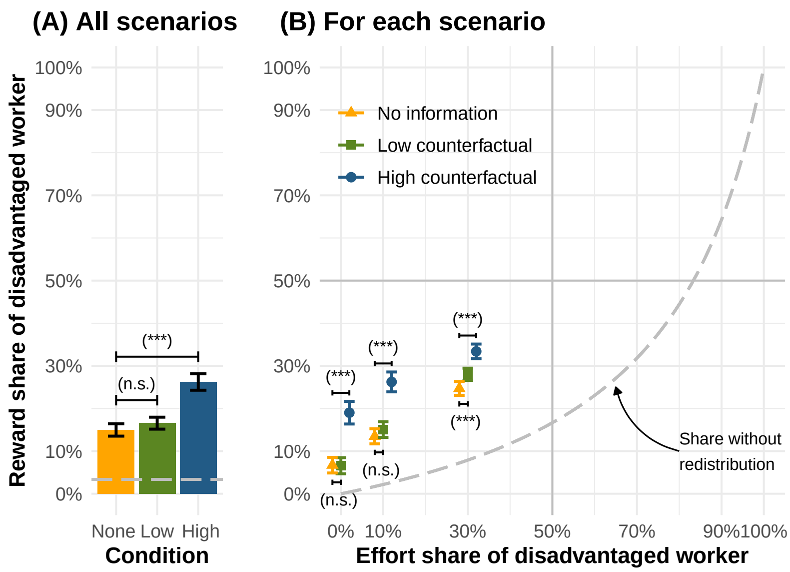
	(1)	(2)	(3)	(4)-(7)
	Actual effort share of worker B			
Effort scenario	0%	10%	30%	Random*
	Counterfactual effort share of worker B, by experimental condition			
No information	–	–	–	–
Low counterfactual	0%	10%	30%	Random*
High counterfactual	50%	50%	50%	Random*

*Effort choices: E_A is uniformly randomly drawn from the integers between 0 and 50. E_B ranges from 0 to 25. Counterfactual effort choice of worker B: C_B equals $E_B + X$ where X ranges from 0 to 25.
Notes: This table presents an overview of all seven effort scenarios and the experimental conditions in the counterfactual study. A contingent response method is used: Each spectator faces eight effort scenarios. The seven scenarios above are hypothetical. An eighth effort scenario (not shown) is real. Spectators do not know which scenario is real and have to take each of their decisions seriously. Scenarios (1) to (3) provide the reduced-form evidence analyzed in this section. They are presented in random order to spectators. Data from scenarios (4) to (7) are used in Section 1.6 to structurally estimate a model of merit views.

Second, a comparison of the "low counterfactual" and "high counterfactual" conditions exposes that, once known, the counterfactual choice of worker B matters substantially for spectators' merit judgments. Spectators distribute on average a 9.7 pp higher payment share to worker B when they know that he would have worked as hard as worker A, had he earned a high piece-rate. This effect is driven by a subset of spectators who distribute the payment equally once they know that both workers would have worked equally hard for a high piece-rate. About 32% of spectators implement equality in the "high counterfactual" condition, whereas only 7% do so in the "low counterfactual" and "no information" condition respectively (see also Figure 1.B.2).[24]

In short, spectators care about the counterfactual effort choice of worker B. Once known, their merit judgments take situational influence into account and compensate workers who are disadvantaged by external circumstances. This effect is driven by about one-quarter of participants, whereas the remaining participants do not adjust their reward behavior to the counterfactual information. However, *all* participants fully neglect the effect of situational influence when no information on the counterfactual choice is provided. This suggests that, in the presence of an un-

24. Could the large effect of the "high counterfactual" treatment be partially driven by an experimenter demand effect? Respondents might interpret the counterfactual information as a hint from the experimenter to make use of the information. However, the null result in the attention experiment renders such an explanation unlikely. Here, the scope for demand effects seems to be higher. Respondents receive two pages of information which strongly emphasize the endogeneity of choices. Nonetheless, I do not find a treatment effect, suggesting that demand effects are not an empirically important factor in the experimental context of this study.



Notes: Results from the counterfactual study, decisions 1-3. Panel A displays the mean reward share assigned to the disadvantaged worker B in each experimental condition, averaged across all three effort scenarios, with 95% confidence intervals. Panel B plots the mean reward share in each effort scenario with 95% confidence intervals. The gray dashed line shows the default share, that is, which payment share worker B would receive if spectators do not redistribute. I test for differences between the "High counterfactual" and the "No information" condition (upper test) and between the "Low counterfactual" and the "No information" condition (lower test). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, (n.s.) $p \geq 0.10$.

Figure 1.5.2. Counterfactual study: Avg. reward share of disadv. worker with 95% CI

known, uncertain counterfactual, spectators base their merit judgments on the only clear and reliable evidence they have, namely observed effort choices.

> ***Result**: Once the counterfactual is revealed, spectators on average compensate workers for disadvantageous situational influence. The uncertainty of the counterfactual state is thus responsible for the main finding that merit judgments neglect the endogeneity of effort choices.*

In light of the model discussed in Section 1.2, this means that comparable choice meritocrats exist but do not apply their merit view when the counterfactual effort choice under equal circumstances is uncertain and unknown. The next section organizes this and other reduced-form findings in a structured framework.

1.6 A structural model of heterogeneous merit views

Data from all experiments reveal that individuals endorse distinct fairness types. Typically, the distribution of merit judgments exhibits discrete spikes that coincide with the model of merit views introduced in Section 1.2 (see Figures 1.B.1 and

1.B.2). In this section, I structurally estimate the model to gauge the prevalence of these fairness views in the population.

1.6.1 Model and estimation

I assume that each participant rewards the disadvantaged worker B according to $m_i(e,s) + \varepsilon_{is}$. $m_i(e,s)$ is her merit view conditional on worker B's effort share e in situation s, and $\varepsilon_{is} \sim_{iid} N(0, \sigma^2)$ is a normally distributed response error. The model assumes that the population is separated into four distinct fairness types.

Actual choice meritocrats reward workers based on actual effort shares, $m_i(e,s) = e$, irrespective of whether effort choices are endogenous to external situational influence.

Comparable choice meritocrats reward workers based on (counterfactual) effort shares under equally advantaged, comparable circumstances, $m_i(e,s) = \hat{E}_i c(e,s)$, and thus compensate for situational influence. When the counterfactual $c(e,s)$ is known and revealed to the spectators, we have $\hat{E}_i c(e,s) = c(e,s)$. When the counterfactual is uncertain, I assume that comparable choice meritocrats accurately anticipate the expected counterfactual effort share $Ec(e,s)$ but "discount" it and put more weight on the observed effort share e.

$$\hat{E}_i c(e,s) \;=\; \rho Ec(e,s) + (1-\rho)e \quad \text{where } 0 \;\leq\; \rho \;\leq\; 1$$

Both assumptions are in line with the reduced-form results. The discounting of the expected counterfactual could be interpreted as a probabilistic failure to engage in counterfactual reasoning (with probability $1-\rho$) or has a preference to base merit judgments on verifiable information (with weighting factor $1-\rho$).[25]

Egalitarians always implement equality: $m_i(e,s) = 50\%$.

Libertarians fully accept any pre-existing inequality p: $m_i(e,s) = p$.

I use the merit judgments made in Scenarios 4 to 7 of the counterfactual study to estimate the model. These scenarios randomly vary the effort share of both workers and, in the counterfactual conditions, the counterfactual effort share of worker B (see Table 1.5.1, Scenarios 4-7). They cover a rich variety of cases and are hence ideally suited to estimate how common different merit views are. Moreover, this procedure allows me to explore the replicability of my reduced-form findings, which do not depend on data from Scenarios 4 to 7. I estimate six parameters, namely the population shares of each preference type together with the discount parameter ρ and the standard deviation of the response error σ.

The parameters are identified by the within-subject variation in effort scenarios and the between-subject variation in experimental conditions. For example, the

25. I calibrate $Ec(e,s)$ to the worker data. Appendix 1.B.4 shows that the results of the model are insensitive to two different calibration approaches.

share of egalitarians is reflected in the number of individuals who equalize payments in all effort scenarios. Likewise, the share of comparable choice meritocrats becomes evident in the conditions where the counterfactual is known. Here, the share influences how many respondents are willing to redistribute payments according to counterfactual effort shares. In turn, the discount parameter ρ can be identified in the condition where the counterfactual is uncertain and the merit judgments of comparable choice meritocrats crucially hinge on the discounting of the expected counterfactual.

I employ a constrained maximum likelihood procedure. Appendix 1.B.4 presents the technical details of the estimation procedure and shows that the results are robust to a series of sensitivity checks, such as a specification with trembling-hand response error or an exclusion of participants who initially failed a control question. I also confirm the numerical stability of the maximum likelihood estimator in Monte Carlo experiments.

1.6.2 Results

The model estimates that 37% of the population are actual choice meritocrats, while 26% are comparable choice meritocrats. Libertarians and egalitarians have a population share of 23% and 14%, respectively (see Table 1.6.1). Thus, a large majority of participants, namely 63%, endorse a meritocratic fairness ideal.[26] However, most meritocrats are actual choice meritocrats (about 60% of all meritocrats). They ignore that workers' choices are shaped by unequal situational influence, even if they know what would have happened in equal circumstances. Only a few individuals are comparable choice meritocrats and prefer to take the endogeneity of choices into account. For them, I estimate a ρ of 0.00 which means that even they fully discount counterfactual choices if the counterfactual is uncertain.[27]

The estimated model mirrors the reduced-form results. For instance, a ρ of 0.00 explains why merit judgments are entirely insensitive to situational influence in the conditions where the counterfactual is unknown. Likewise, the model estimates a share of comparable choice meritocrats of 26% which aligns with the observation that a quarter of respondents is responsible for the treatment effect in the counterfactual experiment (see Section 1.5.3). To give another example, the estimated libertarian share of 23% is broadly consistent with the fact that, depending on the

26. The estimated share of meritocrats is much higher than in Almås, Cappelen, and Tungodden (2020) who classify 37.5% of the US population as meritocrats. In their setting, spectators receive only coarse, binary information about effort choices, namely which of two workers is more productive. Merit presumably plays an even larger role in my setting because the piece-rate task provides a clear and fine-grained measure of effort.

27. The estimate for ρ is on the boundary. Standard inference in constrained maximum likelihood models can become unreliable if one of the parameters is on or near the boundary (Schoenberg, 1997). In Appendix 1.B.4, I run simulation experiments to show that the inference is nevertheless reliable.

Table 1.6.1. Results of the structural estimation

	Estimate	95% confidence interval
Population shares		
Actual choice meritocrats	36.7%	[33.0% – 40.3%]
Comparable choice meritocrats	26.2%	[22.8% – 29.6%]
Libertarians	23.0%	[20.2% – 25.7%]
Egalitarians	14.2%	–
Counterfactual discount parameter		
ρ	0.00	[0.00 – 0.09]
Error term and sample		
σ noise	9.27	[9.06 – 9.49]
Respondents	945	
Decisions	3777	

Notes: Results from the counterfactual study, decisions 4-7, maximum likelihood estimation of the structural model of merit views. The estimates indicate the population shares of different fairness views and the uncertainty discount parameter ρ. No confidence interval is reported for the share of egalitarians because their share is deduced from the other estimates. See Appendix 1.B.4 for further details.

effort scenario, 18% to 29% of respondents accept the pre-existing inequality (see Figures 1.B.1 and 1.B.2).

Does the composition of fairness types or the uncertainty discount parameter ρ vary across different parts of the population? To answer this question, I re-estimate the model and allow its parameters to vary across two separate groups of the population (see Appendix 1.B.4). I compare female versus male respondents, above-median versus below-median respondents, respondents with versus without a college degree, and Republicans versus Democrats. I detect no significant differences across groups. In particular, I estimate a ρ of 0.00 in each group, which suggests that the neglect of uncertain counterfactual states is a fundamental feature of merit judgments (see Table 1.B.7).

Taken together, the results show that people endorse fundamentally different merit views. Crucially, even if the counterfactual choice were known (arguably a rare if not "counterfactual" situation in the real world), only about 26% of individuals would compensate for situational influence. Thus, the prevailing fairness ideals ignore the endogeneity of choices.

> ***Result****: A structural model of merit views classifies only 26% of individuals as comparable choice meritocrats who want to correct for the endogeneity of choices. Replicating earlier results, the model also estimates that even comparable choice meritocrats fully neglect the endogeneity of choices when the counterfactual is uncertain.*

1.7 Vignette study with real-world scenarios

The controlled set-up of the online experiment has many advantages. In particular, it measures merit judgments in situations with real consequences, and it allows for an exogenous variation of external situational influence. However, its stylized environment – two crowd-workers, working for a randomly assigned piece-rate, earning up to $25 – also comes at a cost: It differs from many real-life settings that characterize the debate about meritocracy.

In this section, I therefore explore whether merit judgments are also insensitive to the endogeneity of choices in three real-world scenarios. I report results from an additional vignette study ($n = 1{,}222$) which sheds light on the following three questions, chosen as common and important practical examples of merit judgments: Are minorities compensated for the detrimental choices they might make because they are discriminated? Is a person growing up with few opportunities and incentives to exert effort blamed for being idle? And is an entrepreneur rewarded for taking the risk of founding a company if he inherited a fortune so generous that it made founding easy and substantially reduced any risk involved? The study was run in February 2021 in collaboration with the survey company Lucid. Respondents were recruited from the general US population.[28]

1.7.1 Vignettes

Each vignette describes a simple hypothetical scenario with two people that are exposed to unequal situational influence. The person disadvantaged by situational influence earns much less money due to the detrimental choice he makes. Below, I outline each vignette.[29]

> **Discrimination vignette**: A white and a black employee compete for a promotion which comes with a one-time bonus of $10,000. However, their boss is notorious for being racist, and he never promotes black employees. The white employee works hard to win the promotion, the black person does not, and the white employee is promoted.

28. The study was conducted in two waves. Wave 1 was collected together with the *robustness study: disappointment*. Here, every respondent faced two randomly selected vignettes. Wave 2 was launched shortly thereafter, and respondents faced all vignettes in random order. I exclude respondents who speed through the survey and complete the vignettes with an average response time of less than one minute. The results are robust to both stricter and more lenient exclusion criteria (see Table 1.B.8). Table 1.A.1 shows that the sample does not fully match the characteristics of the general population. Among others, the sample contains more females, more older respondents, and more respondents with a low income. However, the results are robust to the use of survey weights that correct for these imbalances (see Appendix 1.B.5).

29. The full wording of the vignettes is presented in Appendix 1.F. The vignette survey also contained a fourth vignette on criminal behavior which requires a tailored analysis and discussion and is not reported here for brevity (but see Appendix 1.B.5).

Poverty vignette: In this vignette, the advantaged person grew up in a rich family, went to good schools, and was taught that "you can go as far as your hard work takes you." The disadvantaged person grew up in a poor family, went to poor-quality schools, and was always told that "the poor stay poor, and the rich get richer." Whereas the advantaged person always worked hard in his life and, as a consequence, earns \$125,000 a year, the disadvantaged person never worked hard and earns only \$25,000 a year.

Start-up vignette: The vignette portrays two passionate software developers who always dreamed of founding a software start-up. The advantaged person inherited a considerable fortune that provided him with enough money to found and fail several times without any risk of financial ruin. By contrast, the disadvantaged person would have struggled to gather enough money to launch even a first start-up and would have been broke if his first attempt had failed. The advantaged person decided to take the risk and founded his own software start-up. He earns \$200,000 a year today. The disadvantaged person decided to work as a software developer for a local company. He earns \$50,000 a year today.

Analogous to the main experiment, respondents can specify how much money each person deserves by hypothetically redistributing the income between the two people. If their merit judgments are sensitive to situational influence, they should compensate the disadvantaged person for the adverse situational influence that shaped his choice. Redistribution toward the disadvantaged person could, however, also be explained by other fairness motives. In particular, respondents might assign more money to the disadvantaged person simply because they prefer a more equal outcome. Or they want to compensate the disadvantaged person for living in worse circumstances, for example, for not inheriting any money in the start-up vignette.

To identify the sensitivity of merit judgments to situational influence, I introduce a between-subject variation that is analogous to the counterfactual study of Section 1.5.3. Respondents are randomized into one of three treatments. The treatments vary whether and what spectators learn about what the disadvantaged person would have done in the advantaged environment.

Baseline: The vignettes describe only the actual decisions of both persons.

Low counterfactual: Each vignette states that the disadvantaged person would not have made a different choice if he had been in the advantaged situation. Hence, his choice was not shaped by his circumstances.

High counterfactual: Here, the disadvantaged person would have made the same choice as the advantaged person if he had been in the advantaged situation. Hence, his choice was strongly shaped by his circumstances.

1.7.2 Results

Table 1.7.1 summarizes the results. Once more, I find that merit judgments neglect the endogeneity of effort choices. First, the neglect of situational influence already

Table 1.7.1. Merit judgments in the vignette study

(A) Share of respondents redistributing towards the disadvantaged worker

	Binary indicator for compensation			
	Discrimination	Poverty	Start-up	Pooled
	(1)	(2)	(3)	(4)
Low counterfactual	0.015	−0.001	−0.026	−0.004
	(0.041)	(0.041)	(0.040)	(0.029)
High counterfactual	0.230***	0.090**	0.059	0.126***
	(0.040)	(0.040)	(0.039)	(0.029)
Constant	0.424***	0.547***	0.630***	
	(0.028)	(0.028)	(0.028)	
Vignette FE	–	–	–	✓
Observations	889	887	888	2,664
R^2	0.044	0.008	0.005	0.587

(B) Mean reward share of disadvantaged person

	Reward share of disadv. person (in %)			
	Discrimination	Poverty	Start-up	Pooled
	(1)	(2)	(3)	(4)
Low counterfactual	0.133	−2.387*	−2.391	−1.539
	(1.658)	(1.197)	(1.413)	(1.085)
High counterfactual	13.590***	4.003***	2.867*	6.795***
	(1.797)	(1.277)	(1.463)	(1.177)
Constant	13.994***	24.208***	33.497***	
	(1.182)	(0.874)	(1.044)	
Initial reward share	0.00	17.00	20.00	
Vignette FE	–	–	–	✓
Observations	889	887	888	2,664
R^2	0.082	0.029	0.015	0.683

Notes: Results from the vignette study, OLS regressions, robust standards (Columns 1-3) and standard errors clustered on the respondent level (Column 4) in parentheses. The dependent variable in Panel A is a binary indicator for whether a respondent compensates the disadvantaged person by redistributing money toward him. The dependent variable in Panel B is the reward share assigned to the disadvantaged person. The independent variables are treatment dummies. Columns 1-3 report results from different vignettes, and Column 4 displays the pooled results. In each panel, p-values of the coefficients in Columns 1-3 are adjusted for multiple hypothesis, using the Benjamini-Hochberg adjustment. *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

induces little redistribution toward the disadvantaged person in the baseline condition. For instance, in the discrimination vignette, only 42% of respondents assign a positive reward share to the discriminated black employee (Column 1, Panel A), and, on average, he receives only 14% of the total pay-off (Column 2, Panel B). Most respondents accept that he comes away empty-handed. His choice not to work hard legitimizes the highly unequal outcome. In the poverty vignette, 55% of respondents are willing to compensate the person who grew up in poverty, but he is still assigned only 24% of the total earnings (only 7 pp more than he would receive without redistribution).

Next, I study the difference in merit judgments between the baseline and the "low counterfactual" condition. In the baseline condition, situational influence is present (though uncertain), whereas it is verifiably absent in the "low counterfactual" condition. If, as in the main experiment, baseline merit judgments are insensitive to situational influence, they should be similar across the baseline and the "low counterfactual" condition. Indeed, the reward decisions are virtually identical in both conditions. Pooled across vignettes, only 0.4 pp more respondents redistribute money toward the disadvantaged person in baseline than in "low counterfactual" (Column 4, Panel A). Likewise, the average reward share of the disadvantaged is only 1.5 pp higher in the baseline condition (Column 4, Panel B). Both effects are statistically insignificant.

In stark contrast, the "high counterfactual" condition increases the share of respondents who redistribute money toward the disadvantaged person by 12.6 pp and raises his mean reward share by 6.8 pp across vignettes. The results are mainly driven by the discrimination and the poverty vignette, whereas they are more muted in the start-up vignette. For instance, in the discrimination vignette, 23 pp more respondents are willing to assign a positive reward share to the black employee once they know that he would have worked equally hard had his boss given him a fair chance. Likewise, the fraction of respondents who compensate the disadvantaged person increases by 9 pp in the poverty vignette. Respondents thus only integrate situational influence in their merit judgments once the counterfactual is known but ignore it if the counterfactual is uncertain.

Taken together, the results suggest that merit judgments are insensitive to situational influence not only in the controlled experimental setting but that the same phenomenon is to be expected in many important real-life domains of a meritocracy.

Result*: Merit judgments neglect the endogeneity of choices also in important real-world scenarios.*

1.8 Concluding remarks

The idea of meritocracy has become central in Western politics where it has shaped the public debate, the economic culture, and social reforms. Meritocracy promises that the family, neighborhood, and circumstances one is born into should not matter. This promise is popular and closely connects to the prominent ideas of equal opportunity and the American dream. However, the findings of this study suggest that, in practice, meritocratic fairness is likely to be "*shallow*". Even though it claims that individuals should not be judged by their external circumstances, people ignore that these external circumstances also influence the choices that agents make.

In a series of experiments with about 4,000 participants from the general US population, I document that individuals reward and penalize workers for their effort choices, even if their choices are strongly endogenous to and shaped by external circumstances. I experimentally identify the uncertainty of the counterfactual – what the disadvantaged person would have done in advantaged circumstances – as the cause of the neglect. Only once the uncertainty of the counterfactual is resolved and participants know what would have happened on a level playing field, about a quarter of respondents start to compensate for the disadvantageous endogeneity.

The uncertainty of the counterfactual state is often an inevitable feature of reality, and so is, this suggests, the neglect of endogeneity. Therefore, it seems likely that the neglect is common also outside the US and extends to other determinants of merit, such as cognitive skills, personality traits, or educational achievements, which are also highly endogenous to and shaped by circumstances (e.g., Heckman, 2006; Putnam, 2016; Alan and Ertac, 2018; Kosse et al., 2019).

A structurally estimated model of merit views reveals that the prevailing fairness ideal ignores the endogeneity of choices, even when the counterfactual state is known. Most participants endorse *actual choice meritocratism*: They reward and hold workers responsible for their observable effort choices, irrespective of whether choices are endogenous to external circumstances.

Of course, holding others responsible for their actual choices may simply be a practical necessity of living together. Any fairness principle must also be evaluated in terms of its prospective incentive effects. Actual choice meritocratism provides clear guidance to both agents and spectators. By contrast, comparable choice meritocratism could create a complicated signaling game where disadvantaged agents try to signal high counterfactual effort choices strategically, while spectators anticipate this behavior and face even greater difficulties in inferring the counterfactual. This may explain why actual choice meritocratism is more popular in the US population and why even comparable choice meritocrats account for the endogeneity of choices only if they have access to reliable information about the counterfactual state.

The structure of merit judgments is likely to affect which policies voters demand. In particular, "shallow meritocrats" may accept the consequences of unequal opportunities, even though they oppose unequal opportunities themselves. Once unequal opportunities led to unequally meritorious choices, these choices can justify the resulting inequality. Consequently, meritocrats endorse *predistribution* policies that level the playing field and equate circumstances ex-ante. By contrast, they are more reluctant to compensate others for unequal circumstances via *redistribution* after unequal choices have been made. In practice, a policymaker is therefore likely to face much larger support for predistributive than for redistributive policies. This may also explain why many affirmative action policies are considered controversial and often depicted as undermining the merit principle (Harrison, Kravitz, Mayer, Leslie, and Lev-Arey, 2006), even though they attempt to correct for the unequal opportunities that agents faced in producing merit. Simply put, in a meritocracy, choices can launder circumstances and legitimize the ensuing inequality.

References

Abadie, Alberto. 2020. "Statistical Nonsignificance in Empirical Economics." *American Economic Review: Insights* 2 (2): 193–208. [21]

Akbaş, Merve, Dan Ariely, and Sevgi Yuksel. 2019. "When is inequality fair? An experiment on the effect of procedural justice and agency." *Journal of Economic Behavior and Organization* 161: 114–27. [10]

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865–911. [9, 39]

Alesina, Alberto, and George-Marios Angeletos. 2005. "Fairness and Redistribution." *American Economic Review* 95 (4): 960–80. [6, 10]

Alesina, Alberto, and Edward Glaeser. 2004. *Fighting Poverty in the US and Europe: A World of Difference.* Oxford University Press. [6, 10]

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. "Intergenerational Mobility and Preferences for Redistribution." *American Economic Review* 108 (2): 521–54. [10]

Almås, Ingvild, Alexander Cappelen, and Bertil Tungodden. 2020. "Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-seeking than Scandinavians?" *Journal of Political Economy* 128 (5): 1753–88. [6, 10, 12, 13, 15, 33]

Altmejd, Adam, Andrés Barrios-Fernández, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson, and Jonathan Smith. 2021. "O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries." *Quarterly Journal of Economics* 136 (3): 1831–86. [6]

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. 2021. "Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples." *Working Paper*, [26]

Andreoni, James, Deniz Aydin, Blake Allen Barton, B. Douglas Bernheim, and Jeffrey Naecker. 2020. "When Fair Isn't Fair: Understanding Choice Reversals Involving Social Preferences." *Journal of Political Economy* 128 (5): 1673–711. [10, 15]

Baron, Jonathan, and John C. Hershey. 1988. "Outcome Bias in Decision Evaluation." *Journal of Personality and Social Psychology* 54 (4): 569–79. [10]

Bartling, Björn, Alexander W. Cappelen, Mathias Ekström, Erik Ø. Sørensen, and Bertil Tungodden. 2018. "Fairness in Winner-Take-All Markets." *Working Paper*, [10]

Bartling, Björn, and Urs Fischbacher. 2012. "Shifting the Blame: On Delegation and Responsibility." *Review of Economic Studies* 79 (1): 67–87. [10]

Benjamin, Daniel J. 2019. "Errors in probabilistic reasoning and judgmental biases." In *Handbook of Behavioral Economics: Applications and Foundations 2*. Edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson. North-Holland. Chapter 2. [11]

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir. 2004. "A Behavioral-Economics View of Poverty." *American Economics Reveiw* 94 (2): 419–23. [6]

Brownback, Andy, and Michael A. Kuhn. 2019. "Understanding outcome bias." *Games and Economic Behavior* 117: 342–60. [10]

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107 (11): 3288–319. [6]

Byrne, Ruth M.J. 2016. "Counterfactual Thought." *Annual Review of Psychology* 67: 135–57. [11]

Cappelen, Alexander W., Ranveig Falch, and Bertil Tungodden. 2020. "Fair and Unfair Income Inequality." In *Handbook of Labor, Human Resources and Population Economics*. Edited by K. F. Zimmermann. Springer, 1–25. [6]

Cappelen, Alexander W., Sebastian Fest, Erik Ø. Sørensen, and Bertil Tungodden. 2020. "Choice and Personal Responsibility: What is a Morally Relevant Choice?" *Review of Economics and Statistics*, (forthcoming): [10]

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sorensen, and Bertil Tungodden. 2007. "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review* 97 (3): 818–27. [10]

Cappelen, Alexander W., James Konow, Erik Ø. Sørensen, and Bertil Tungodden. 2013. "Just Luck: An Experimental Study of Risk-Taking and Fairness." *American Economic Review* 103 (4): 1398–413. [10]

Cappelen, Alexander W., Karl Ove Moene, Siv-Elisabeth Skjelbred, and Bertil Tungodden. 2020. "The merit primacy effect." *Working Paper*, [10]

Cappelen, Alexander W., Johanna Mollerstrom, Bjørn-Atle Reme, and Bertil Tungodden. 2019. "A Meritocratic Origin of Egalitarian Behavior." *Working Paper*, [26]

Cappelen, Alexander W., Erik Ø. Sørensen, and Bertil Tungodden. 2010. "Responsibility for what? Fairness and individual responsibility." *European Economic Review* 54 (3): 429–41. [10]

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125 (3): 1101–44. [6]

Cassar, Lea, and Arnd H. Klein. 2019. "A Matter of Perspective: How Failure Shapes Distributive Preferences." *Management Science* 65 (11): 4951–5448. [15]

Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Börn Wallace. 2009. "Genetic Variation in Preferences for Giving and Risk Taking." *Quarterly Journal of Economics* 124 (2): 809–42. [13]

Charness, Gary, Uri Gneezy, and Brianna Halladay. 2016. "Experimental methods: Pay one or pay all." *Journal of Economic Behavior & Organization* 131: 141–50. [15]

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77. [26]

Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents." *Research and Politics* 6 (1): 1–14. [19]

Davis, Mark H. 1983. "Measuring individual differences in empathy: Evidence for a multidimensional approach." *Journal of Personality and Social Psychology* 44 (1): 113–26. [22]

Ding, Peng, Avi Feller, and Luke Miratrix. 2016. "Randomization inference for treatment effect variation." *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 78 (3): 655–71. [22]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2012. "The Intergenerational Transmission of Risk and Trust Attitudes." *Review of Economic Studies* 79 (2): 645–77. [13]

Enke, Benjamin, and Florian Zimmermann. 2017. "Correlation Neglect in Belief Formation." *Review of Economic Studies* 86 (1): 313–32. [11]

Eshleman, Andrew. 2016. "Moral Responsibility." In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. [9]

Esponda, Ignacio, and Emanuel Vespa. 2014. "Hypothetical Thinking and Information Extraction in the Laboratory." *American Economic Journal: Microeconomics* 6 (4): 180–202. [11]

Esponda, Ignacio, and Emanuel Vespa. 2019. "Contingent Preferences and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory." *Working Paper*, [11, 28]

Falk, Armin, Fabian Kosse, and Pia Pinger. 2020. "Mentoring and Schooling Decisions: Causal Evidence." *Working Paper*, [6]

Falk, Armin, Thomas Neuber, and Nora Szech. 2020. "Diffusion of Being Pivotal and Immoral Outcomes." *Review of Economic Studies* 87 (5): 2205–29. [10]

Fisman, Raymond, Ilyana Kuziemko, and Silvia Vannutelli. 2020. "Distributional Preferences in Larger Groups: Keeping Up With the Joneses and Keeping Track of the Tails." *Journal of the European Economic Association* jvaa033: [10]

Frank, Robert H. 2016. *Success and Luck: Good Fortune and the Myth of Meritocracy*. Princeton, and Oxford: Princeton University Press. [9]

Gabaix, Xavier. 2019. "Behavioral inattention." In *Handbook of Behavioral Economics: Applications and Foundations*. Edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson. Vol. 2, North-Holland, 261–343. [26]

Giuliano, Paola, and Antonio Spilimbergo. 2013. "Growing up in a Recession." *Review of Economic Studies* 81 (2): 787–817. [10]

Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." *Quarterly Journal of Economics* 132 (3): 1219–60. [6]

Graeber, Thomas. 2021. "Inattentive Inference." *Working Paper*, [11]

Greenfield, Kent. 2011. *The Myth of Choice: Personal Responsibility in a World of Limits*. New Haven, and London: Yale University Press. [9]

Gurdal, Mehmet Y., Joshua B. Miller, and Aldo Rustichini. 2013. "Why Blame?" *Journal of Political Economy* 121 (6): 1205–47. [10]

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. No date. "Designing Information Provision Experiments." *Journal of Economic Literature*, (forthcoming): ().

Han, Yi, Yiming Liu, and George Loewenstein. 2020. "Correspondence Bias." *Working Paper*, [11]

Harrison, David A., David A. Kravitz, David M. Mayer, Lisa M. Leslie, and Dalit Lev-Arey. 2006. "Understanding Attitudes Toward Affirmative Action Programs in Employment: Summary and Meta-analysis of 35 Years of Research." *Journal of Applied Psychology* 91 (5): 1013–36. [40]

Haushofer, Johannes, and Ernst Fehr. 2014. "On the psychology of poverty." *Science* 344 (6186): 862–67. [6]

Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312 (5782): 1900–2. [9, 13, 39]

Henningsen, Arne, and Ott Toomet. 2011. "maxLik: A package for maximum likelihood estimation in R." *Computational Statistics* 26 (3): 443–58. [60]

Kahneman, Daniel, and Dale T. Miller. 1986. "Norm theory: Comparing reality to its alternatives." *Psychological Review* 93 (2): 136–53. [11]

Konow, James. 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review* 90 (4): 1072–91. [6]

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. 2019. "The Formation of Prosociality: Causal Evidence on the Role of Social Environment." *Journal of Political Economy* 128 (2): 434–67. [9, 13, 39]

Kovaleva, Anastassiya. 2012. *The IE-4: Construction and Validation of a Short Scale for the Assessment of Locus of Control*. Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. [22]

Krawczyk, Michał. 2010. "A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution." *Journal of Public Economics* 94 (1-2): 131–41. [10]

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How Elastic Are Preferences for Redistribution? Evidence From Randomized Survey Experiments." *American Economic Review* 105 (4): 1478–508. [10]

Lagnado, David A., and Tobias Gerstenberg. 2017. "Causation in Legal and Moral Reasoning." In *The Oxford Handbook of Causal Reasoning*. Edited by Michael R. Waldmann. New York: Oxford University Press. [11]

Liang, Yucheng. 2021. "Learning from Unknown Information Sources." *Working Paper*, [11]

Markovits, Daniel. 2019. *The Meritocracy Trap*. Penguin Books. [9]

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa. 2019. "Failures in Contingent Reasoning: The Role of Uncertainty." *American Economic Review* 109 (10): 3437–74. [11, 28]

Mollerstrom, Johanna, Bjørn-Atle Reme, and Erik Ø. Sørensen. 2015. "Luck, choice and responsibility — An experimental study of fairness views." *Journal of Public Economics* 131: 33–40. [10, 15]

Nagel, Thomas. 1979. "Moral Luck." In *Mortal Questions*. Cambridge, New York: Cambridge University Press. [10]

Nelkin, Dana K. 2019. "Moral Luck." In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. [9]

Pasek, Josh, Matthew Debell, and Jon A. Krosnick. 2014. "Standardizing and Democratizing Survey Weights: The ANES Weighting System and anesrake." *Working Paper*, [54, 64]

Putnam, Robert D. 2016. *Our Kids: The American Dream in Crisis*. New York: Simon, and Schuster. [9, 39]

Roemer, John E. 1993. "A Pragmatic Theory of Responsibility for the Egalitarian Planner." *Philosophy & Public Affairs* 22 (2): 146–66. [13]

Roese, Neal J. 1997. "Counterfactual thinking." *Psychological Bulletin* 121 (1): 133–48. [11]

Ross, Lee. 1977. "The Intuitive Psychologist and his Shortcomings: Distortions in the Attribution Process." *Advances in Experimental Social Psychology* 10: 173–220. [7, 26]

Sandel, Michael J. 2020. *The Tyranny of Merit: What's Become of the Common Good?* London: Allen Lane. [6, 9]

Schoenberg, Ronald. 1997. "Constrained Maximum Likelihood." *Computational Economics* 10 (3): 251–66. [33, 60]

Sloman, Steven. 2005. *Causal Models: How People Think about the World and Its Alternatives*. New York: Oxford University Press. [11]

Stantcheva, Stefanie. 2021. "Understanding Tax Policy: How Do People Reason?" *Working Paper*, [10]

Taubinsky, Dmitry, and Alex Rees-Jones. 2018. "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment." *Review of Economic Studies* 85 (4): 2462–96. [26]

Young, Michael. 1958. *The Rise of the Meritocracy*. Thames, and Hudson. [9]

Appendix 1.A Samples

Overview. The table provides an overview of all spectator samples used in this study. It lists all samples and describes when and how they were collected.

Sample	When	How	Population	Recruitment	*n*
Main study	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	653
Robustness study "Equal rates"	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	661
Robustness study "Disappointment"	February 2021	Online experiment	US adults	Via survey company Lucid	606
Attention study	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	274
Attention robustness study "Equal rates"	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	267
Counterfactual study	January 2021	Online experiment	US adults (targeted*)	Via survey company Lucid	945
Vignette study	February 2021	Online survey	US adults	Via survey company Lucid	1,222**
				Total n	4,033

*The sampling process targeted a sample that represents the general population in terms of gender, age (3 groups), region (4 groups), income (3 groups), and education (2 groups). The counterfactual study did not target education.
**Wave 1 of the vignette study was attached to the robustness study: disappointment. 595 respondents of the robustness study also participated in the vignette study. The total does not double-count these respondents.

Sample characteristics. Table 1.A.1 summarizes the demographic characteristics of each sample.

Exclusion criteria in online experiments. Exclusion criteria are preregistered (see Appendix 1.D). The samples do not contain the following responses:

1. Respondents who do not complete the first seven redistribution decisions.[30]
2. Respondents who spend less than 30 seconds on the instructions until the first treatment variation is introduced.
3. Duplicate respondents (very rare cases).

30. There is only one redistribution decision in the robustness study. Here, I exclude all respondents who do not complete the study.

Balanced assignment of experimental conditions. Table 1.A.2 and Table 1.A.3 show that the demographic covariates are balanced across experimental conditions in all studies. I test for balanced treatment assignment by regressing the demographic variables on a treatment indicator. Across all studies, the coefficient estimates are mostly small, indicating that the demographic covariates are balanced across treatments. For each study, I also test the joint null hypothesis that *all* treatment differences are zero. None of the highly-powered F-test rejects this hypothesis. For the vignette study, the joint effect is marginally significant ($p = 0.083$), but the effect sizes are relatively minor.

Table 1.A.1. Comparison of all samples to the American Community Survey (ACS)

Variable	ACS (2019)	Main study	Equal rates	Disap-pointment	Atten-tion	Attention equal rates	Counter-factual	Vig-nettes
Gender								
Female	51%	51%	52%	63%	52%	48%	53%	61%
Age								
18-34	30%	30%	28%	11%	32%	33%	23%	15%
35-54	32%	33%	32%	30%	32%	29%	35%	33%
55+	38%	37%	41%	59%	36%	38%	42%	52%
Household net income								
Below 50k	37%	40%	43%	47%	39%	44%	39%	45%
50k-100k	31%	34%	32%	34%	34%	33%	32%	33%
Above 100k	31%	27%	26%	19%	26%	23%	30%	22%
Education								
Bachelor's degree or more	31%	43%	40%	48%	38%	36%	56%	47%
Region								
Northeast	17%	21%	16%	25%	16%	16%	17%	25%
Midwest	21%	21%	22%	25%	18%	21%	21%	23%
South	38%	36%	39%	35%	44%	38%	38%	36%
West	24%	22%	23%	15%	23%	25%	24%	16%
Sample size	2,059,945	653	661	606	274	267	945	1,222

Notes: Column 1 presents data from the American Community Survey (ACS) 2019. The other columns describe the different experimental samples.

Table 1.A.2. Test for balanced treatment assignment – part 1

Main study							
	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	−0.001	0.150	0.754	0.000	−0.012	−0.022	0.031
	(0.039)	(1.339)	(4.488)	(0.039)	(0.032)	(0.038)	(0.032)
Constant	0.511***	47.116***	76.831***	0.426***	0.213***	0.374***	0.204***
	(0.028)	(0.935)	(3.144)	(0.027)	(0.023)	(0.027)	(0.022)
Joint F-test (H_0: all differences between conditions are zero): $p = 0.992$							
Observations	653	653	653	653	653	653	653
R^2	0.000	0.000	0.000	0.000	0.000	0.001	0.001

Robustness study: Equal rates							
	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.022	−1.782	0.715	−0.048	0.022	0.049	−0.063*
	(0.039)	(1.387)	(4.426)	(0.038)	(0.032)	(0.038)	(0.033)
Constant	0.509***	49.357***	74.720***	0.429***	0.208***	0.366***	0.264***
	(0.028)	(1.026)	(3.109)	(0.028)	(0.023)	(0.027)	(0.025)
Joint F-test (H_0: all differences between conditions are zero): $p = 0.306$							
Observations	661	661	661	661	661	661	661
R^2	0.000	0.003	0.000	0.002	0.001	0.003	0.006

Robustness study: Disappointment							
	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	−0.021	0.610	7.267*	0.033	0.008	0.011	−0.064**
	(0.039)	(1.297)	(4.005)	(0.041)	(0.035)	(0.039)	(0.029)
Constant	0.636***	55.844***	62.980***	0.464***	0.245***	0.341***	0.185***
	(0.028)	(0.916)	(2.716)	(0.029)	(0.025)	(0.027)	(0.022)
Joint F-test (H_0: all differences between conditions are zero): $p = 0.214$							
Observations	606	606	606	606	606	606	606
R^2	0.000	0.000	0.005	0.001	0.000	0.000	0.008

Notes: OLS regressions, robust standard errors in parentheses. Each panel represents a study. Within each panel, each column regresses a demographic variable on the treatment dummy to test for imbalanced treatment assignment. In each panel, a joint F-test, estimated in a SUR model, tests the hypothesis that all treatment differences are zero. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 1.A.3. Test for balanced treatment assignment – part 2

Attention study*

	Female	Age	Income (in \$1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Attention	0.011	−1.356	−0.225	−0.042	−0.034	0.064	0.023
	(0.041)	(1.383)	(4.655)	(0.040)	(0.032)	(0.040)	(0.034)
Constant	0.511***	47.116***	76.831***	0.426***	0.213***	0.374***	0.204***
	(0.028)	(0.935)	(3.145)	(0.027)	(0.023)	(0.027)	(0.022)
Joint F-test (H_0: all differences between conditions are zero): $p = 0.400$							
Observations	603	603	603	603	603	603	603
R^2	0.000	0.002	0.000	0.002	0.002	0.004	0.001

Attention "Equal rates" study*

	Female	Age	Income (in \$1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Attention	−0.026	−2.743*	−3.466	−0.069*	0.002	0.012	−0.009
	(0.041)	(1.472)	(4.485)	(0.040)	(0.034)	(0.040)	(0.036)
Constant	0.509***	49.357***	74.720***	0.429***	0.208***	0.366***	0.264***
	(0.028)	(1.026)	(3.109)	(0.028)	(0.023)	(0.027)	(0.025)
Joint F-test (H_0: all differences between conditions are zero): $p = 0.400$							
Observations	589	589	589	589	589	589	589
R^2	0.001	0.006	0.001	0.005	0.000	0.000	0.000

Counterfactual study

	Female	Age	Income (in \$1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Low count.	−0.018	−1.322	3.011	0.017	0.019	−0.019	0.018
	(0.040)	(1.686)	(4.615)	(0.040)	(0.033)	(0.039)	(0.034)
High count.	−0.046	2.631	0.041	0.059	−0.013	−0.014	0.009
	(0.040)	(2.682)	(4.635)	(0.040)	(0.032)	(0.039)	(0.034)
Constant	0.556***	50.869***	79.513***	0.534***	0.211***	0.393***	0.227***
	(0.028)	(1.389)	(3.218)	(0.028)	(0.023)	(0.028)	(0.024)
Joint F-test (H_0: all differences between conditions are zero): $p = 0.717$							
Observations	945	945	945	945	945	945	945
R^2	0.001	0.003	0.001	0.002	0.001	0.000	0.000

*The *Attention* condition of the attention study and the attention "equal rates" study is compared to the *Control* condition of the main study and the robustness "equal rates" study, respectively.

Notes: OLS regressions, robust standard errors in parentheses. Each panel represents a study. Within each panel, each column regresses a demographic variable on the treatment dummy to test for imbalanced treatment assignment. In each panel, a joint F-test, estimated in a SUR model, tests the hypothesis that all treatment differences are zero. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 1.A.4. Test for balanced treatment assignment – part 3

Vignette study							
	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Low count.	0.009	−0.080	3.792	0.054	−0.039	0.084^{**}	0.011
	(0.034)	(1.209)	(3.637)	(0.035)	(0.030)	(0.034)	(0.026)
High count.	−0.016	−0.976	5.773	0.026	−0.020	0.018	−0.031
	(0.034)	(1.161)	(3.590)	(0.035)	(0.030)	(0.033)	(0.025)
Constant	0.612^{***}	53.918^{***}	66.715^{***}	0.448^{***}	0.249^{***}	0.331^{***}	0.163^{***}
	(0.024)	(0.849)	(2.522)	(0.024)	(0.021)	(0.023)	(0.018)
Joint F-test (H_0: all differences between conditions are zero): $p = 0.083$							
Observations	1,222	1,222	1,222	1,222	1,222	1,222	1,222
R^2	0.000	0.001	0.002	0.002	0.001	0.006	0.002

Notes: Results from the vignette study. OLS regressions, robust standard errors in parentheses. Each column regresses a demographic variable on the treatment dummy to test for imbalanced treatment assignment. A joint F-test, estimated in a SUR model, tests the hypothesis that all treatment differences are zero. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Appendix 1.B Supplementary analyses

1.B.1 Treatment effects

Average treatment effects in all experimental studies. Table 1.B.1 and Table 1.B.2 test for differences in merit judgments across the experimental conditions of the main study, the attention study, the "equal rates" robustness study, the "equal rates" attention study, and the counterfactual study.

Histograms for main and counterfactual study. Figure 1.B.1 and Figure 1.B.2 plot the full distribution of reward shares assigned to the disadvantaged worker B in the main study and the counterfactual study, respectively. They show histograms for each experimental condition and each effort scenario.

Heterogeneous treatment effects in main study. Table 1.B.3 tests for heterogeneous treatment effects in the main study.

Robustness study: disappointment. Table 1.B.4 presents the treatment effects in the robustness study.

Table 1.B.1. Average treatment effects on the reward share of the disadvantaged worker

(A) Main study: Treatment – Control

Effort scenario *e*	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	-1.93	-0.33	-1.58	-1.42	0.29	0.20	1.33	-0.49
Standard error	1.46	1.19	1.28	1.40	1.49	1.32	1.39	0.67
CI, 95%	[-4.8, 0.9]	[-2.7, 2]	[-4.1, 0.9]	[-4.2, 1.3]	[-2.6, 3.2]	[-2.4, 2.8]	[-1.4, 4.1]	[-1.8, 0.8]
p-values, t-tests	0.184	0.781	0.218	0.310	0.848	0.879	0.339	0.464
p-value, F-test	0.668							

(B) Robustness study "Equal rates": Treatment – Control

Effort scenario *e*	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	2.36	1.06	0.81	-0.16	-0.52	-1.20	0.41	0.39
Standard error	1.38	1.07	0.63	0.18	0.67	1.17	1.25	0.24
CI, 95%	[-0.3, 5.1]	[-1, 3.2]	[-0.4, 2]	[-0.5, 0.2]	[-1.8, 0.8]	[-3.5, 1.1]	[-2, 2.9]	[-0.1, 0.9]
p-values, t-tests	0.088	0.323	0.200	0.364	0.435	0.307	0.745	0.105
p-value, F-test	0.253							

(C) Attention study: Attention – Control

Effort scenario *e*	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	-1.24	0.88	-0.88	-1.28	-1.38	-0.14	0.04	-0.57
Standard error	1.52	1.31	1.40	1.48	1.52	1.40	1.53	0.72
CI, 95%	[-4.2, 1.7]	[-1.7, 3.4]	[-3.6, 1.9]	[-4.2, 1.6]	[-4.4, 1.6]	[-2.9, 2.6]	[-3, 3]	[-2, 0.8]
p-values, t-tests	0.412	0.504	0.529	0.388	0.366	0.921	0.980	0.423
p-value, F-test	0.583							

(D) Attention robustness study "Equal rates": Attention – Control

Effort scenario *e*	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	-0.88	0.48	0.20	0.14	-0.21	0.04	0.25	0.00
Standard error	1.23	1.13	0.72	0.21	0.76	1.22	1.33	0.23
CI, 95%	[-3.3, 1.5]	[-1.7, 2.7]	[-1.2, 1.6]	[-0.3, 0.5]	[-1.7, 1.3]	[-2.4, 2.4]	[-2.4, 2.9]	[-0.5, 0.5]
p-values, t-tests	0.473	0.672	0.783	0.509	0.778	0.974	0.850	0.998
p-value, F-test	0.897							

Notes: Results from OLS regressions. Each panel presents the results from a different study. Columns "0%" to "100%" present results for each of the seven effort scenarios, and Column "Average" presents results averaged across all scenarios. The outcome variable is the reward share assigned to the disadvantaged worker B. The title of each panel describes which experimental conditions are compared. "Reward diff." denotes the estimated treatment effect. Robust standard errors, 95% confidence intervals, and p-values are reported. The last row, "p-value, F-test", presents the p-value from an F-test that tests the joint null hypothesis that the differences are zero in each effort scenario. It is estimated in a SUR model with standard errors that are clustered on the respondent level.

Table 1.B.2. Counterfactual study: Average treatment effects on the reward share of the disadvantaged worker

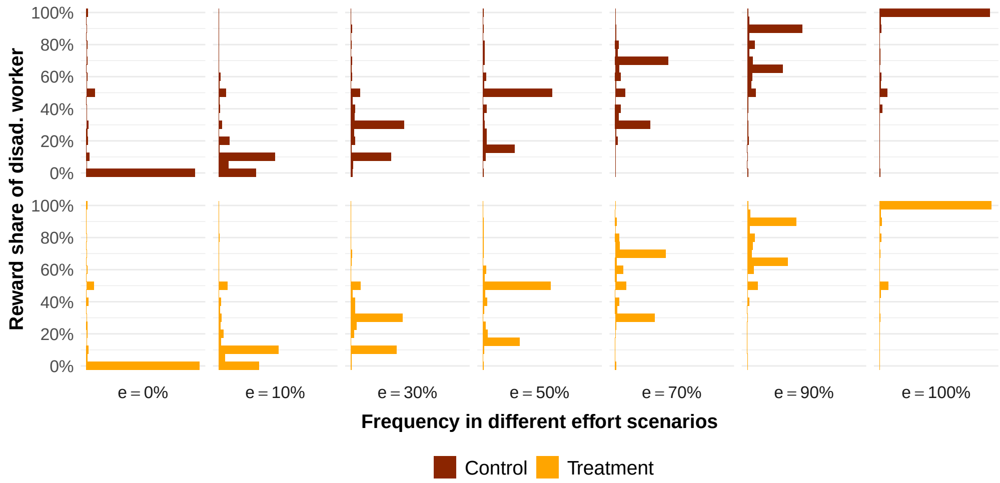
(A) Low counterfactual – No information

Effort scenario *e*	0%	10%	30%	Average
Reward diff.	-0.13	1.58	3.32	1.59
Standard error	1.34	1.31	1.11	1.03
CI, 95%	[-2.8, 2.5]	[-1, 4.1]	[1.1, 5.5]	[-0.4, 3.6]
p-values, t-tests	0.923	0.227	0.003	0.123
p-value, F-test	0.011			

(B) High counterfactual – No information

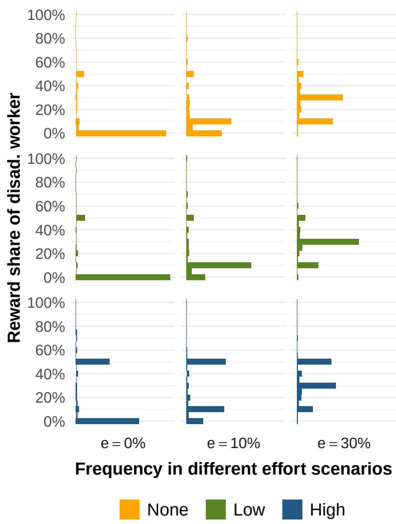
Effort scenario *e*	0%	10%	30%	Average
Reward diff.	12.31	12.75	8.69	11.25
Standard error	1.65	1.49	1.21	1.23
CI, 95%	[9.1, 15.5]	[9.8, 15.7]	[6.3, 11.1]	[8.8, 13.7]
p-values, t-tests	<0.001	<0.001	<0.001	<0.001
p-value, F-test	<0.001			

Notes: Counterfactual study, results from OLS regressions. Panel A compares the *Low counterfactual* with the *No information* condition. Panel B compares the *High counterfactual* with the *No information* condition. Columns “0%” to “30%” present results for each of the three effort scenarios, and Column “Average” presents results averaged across all three scenarios. The outcome variable is the reward share assigned to the disadvantaged worker B. “Reward diff.” denotes the estimated treatment effect. Robust standard errors, 95% confidence intervals, and p-values are reported. The last row, “p-value, F-test”, presents the p-value from an F-test that test the joint null hypothesis that the differences are zero in each effort scenario. It is estimated in a SUR model with standard errors that are clustered on the respondent level.



Notes: Histograms of the reward share assigned to the disadvantaged worker B for each experimental condition and each effort scenario in the main study.

Figure 1.B.1. Main study: Histograms of reward share of disadvantaged worker



Notes: Histograms of the reward share assigned to the disadvantaged worker B for each experimental condition and each effort scenario in the counterfactual study.

Figure 1.B.2. Counterfactual study: Histograms of reward share of disadv. worker

Table 1.B.3. Heterogeneous treatment effects in the main study

	Mean reward share of disadv. worker (in %)
Treatment	9.953
	(8.966)
Female (bin.)	0.024
	(0.993)
College (bin.)	0.570
	(1.092)
Republican (bin.)	−0.852
	(1.002)
Income (log)	0.180
	(0.621)
Empathy (std.)	0.668
	(0.513)
Internal LOC (std.)	0.467
	(0.458)
Treatment × Female (bin.)	0.448
	(1.389)
Treatment × College (bin.)	−0.336
	(1.495)
Treatment × Republican (bin.)	0.764
	(1.394)
Treatment × Income (log)	−0.993
	(0.832)
Treatment × Empathy (std.)	−0.496
	(0.719)
Treatment × Internal LOC (std.)	−1.571
	(0.656)
Constant	42.098
	(6.663)
Observations	634
R^2	0.019

Notes: Results from the main study, OLS regressions, robust standard errors in parentheses. The outcome variable is the reward share assigned to the disadvantaged worker B, averaged across the seven effort scenarios. The independent variables include interaction terms of the treatment dummy with six respondent characteristics: a dummy for female gender, having a Bachelor's degree, and being Republican, logarithmic income, a standardized empathy score, and a standardized internal locus of control score. p-values of the interaction effects (printed in bold) are adjusted for multiple hypotheses testing with the help of the Benjamini-Hochberg procedure. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 1.B.4. Treatment effects in the robustness study: disappointment

	Reward share of disadvantaged worker (in %)	
	(1)	(2)
Treatment	−2.202	−0.763
	(1.422)	(2.122)
Constant	36.695***	35.863***
	(0.973)	(1.387)
Weights	–	✓
Observations	606	606
R^2	0.004	0.000

Notes: Results from the robustness study: disappointment, OLS regressions, robust standard errors in parentheses. The outcome variable is the reward share assigned to worker B (low piece-rate). The independent variable is a treatment indicator. Column 1 reports the unweighted main specification. Column 2 applies post-stratification weights. The weights render the sample representative for the US general population in terms of gender, age, income, education, and census region. I use a raking algorithm (R package `anesrake`) and follow the guidelines of the American National Election Study to calculate the survey weights (Pasek, Debell, and Krosnick, 2014). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

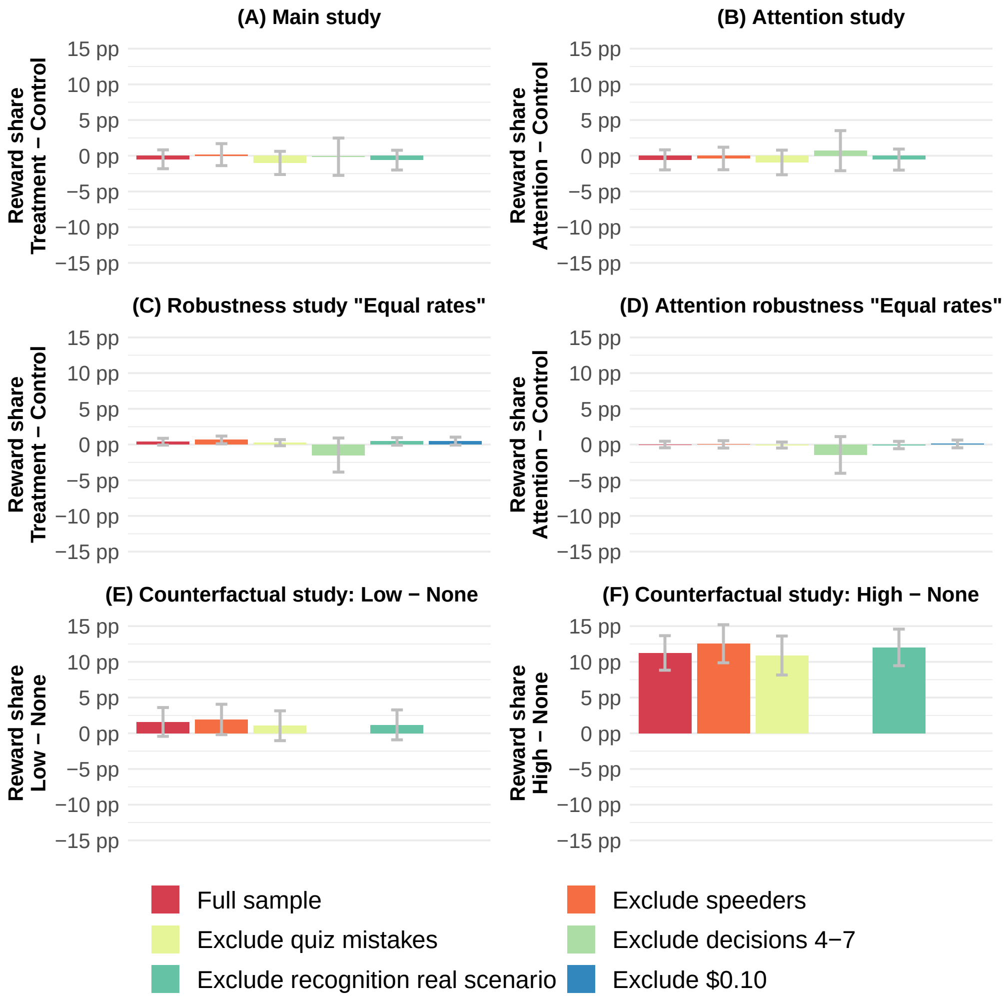
1.B.2 Robustness of treatment effects

Robustness of treatment effects. Figure 1.B.3 explores the robustness of the treatment effects in the main study, the attention study, the "equal rates" robustness study, the "equal rates" attention study and the counterfactual study. The following robustness specifications are estimated.

1. **Full sample**: Full sample, replicates main results.
2. **Exclude speeders**: I exclude the 25% participants with the lowest response duration.
3. **Exclude quiz mistakes** I exclude participants who answer at least 1 question of the quiz wrongly.
4. **Exclude decisions 4-7** I consider only the first three redistribution decisions of each participant. (Note: Not applicable in the counterfactual study, as I always focus on the first three redistribution decisions here.)
5. **Exclude recognition of real scenario** I drop all respondents who are able to distinguish the hypothetical scenarios from the real one, after they saw all scenarios.
6. **Exclude \$0.10** Only applicable to the "equal rates" robustness study and the "equal rates" attention study. The *Control* condition of both studies comes in two variants. Either both workers receive a piece-rate of \$0.10 or both respondents receive a piece-rate of \$0.50. One concern is that only the latter variant can be cleanly compared to the *Treatment* condition in which both workers end up with a piece-rate of \$0.50. This robustness check therefore excludes spectators in the *Control* condition with a piece-rate of \$0.10.

The estimated treatment effects are robust in all studies.

Robustness to the order of workers. In the experiment, I randomize whether worker A or worker B is advantaged or disadvantaged. The main analysis recodes all responses as if A was the advantaged worker to ease analysis and exposition. Here, I test whether a reverse order of workers, that is a worker pair in which worker A is disadvantaged and worker B is advantaged, affects merit judgments. I regress the average reward share respondents assign to the disadvantaged worker on a dummy for reversely ordered worker pairs. Table 1.B.5 shows the results. The random variation in the order of workers does not affect merit judgments.



Notes: Results from the main, attention, "equal rates" robustness, "equal rates" attention, and counterfactual studies. Each panel presents the results from a different study. Each panel plots the treatment effect on the reward share assigned to the disadvantaged worker B (averaged across the effort scenarios) in different robustness specifications. See above for a description. The gray errorbars are 95% confidence intervals.

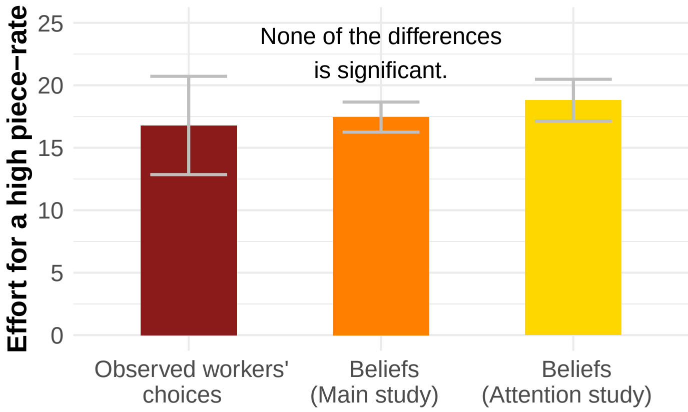
Figure 1.B.3. Robustness of average treatment effects (with 95% CI)

Table 1.B.5. Robustness of merit judgments to the order of workers

	Mean reward share of disadv. worker (in %)				
	Main study	Robustness study "Equal rates"	Attention study	Attention study "Equal rates"	All
	(1)	(2)	(3)	(4)	(5)
Reverse order	−0.327	0.133	−0.058	0.274	−0.037
	(0.674)	(0.243)	(1.064)	(0.344)	(0.302)
Condition FE	✓	✓	✓	✓	✓
Observations	653	661	274	267	1,855
R^2	0.001	0.004	0.000	0.002	0.186

Notes: Results of the main study, the "equal rates" robustness study, the attention study, and the "equal rates" attention study. OLS regressions, robust standard errors in parentheses. The outcome variable is the reward share assigned to the disadvantaged worker, averaged across all seven effort scenarios. The independent variable is a dummy that takes value 1 if worker A is disadvantaged and worker B is advantaged and value 0 for the opposite case. (Note: In the remainder of the paper, I recode all responses as if A was the advantaged worker to ease analysis and exposition.) Columns 1-4 present results from different studies. Column 5 presents a pooled estimate. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

1.B.3 Beliefs about situational influence in the main study



Notes: Results from the main and the attention study. The figure presents the average observed and average perceived effort choices of workers for a high piece-rate of $0.50. The average number of completed tasks for a low piece-rate is 5.04. Red bar: Actual effort decisions of workers. Orange bar: Effort choice that spectators expect in the main study. Yellow bar: Effort choice that spectators expect in attention study. The gray errorbars are 95% confidence intervals. t-tests are used to evaluate the significance of the differences.

Figure 1.B.4. Average beliefs about the piece-rate effect (with 95% CI)

1.B.4 Structural model of merit views

Maximum likelihood estimation

Data. Counterfactual study, decisions 4-7, 945 respondents. In decisions 4-7, respondents face a randomly generated effort scenario.[31] The effort share of worker B and his counterfactual effort share (had he earned a high piece-rate) are drawn as follows.

- Effort of worker A: Uniformly randomly drawn from the set $\{0, 1, ..., 49, 50\}$.
- Effort of worker B: Uniformly randomly drawn from the set $\{0, 1, ..., 24, 25\}$.
- Counterfactual effort of worker B for a high piece-rate: The difference between the counterfactual and observed effort is uniformly randomly drawn from the set $\{0, 1, ..., 24, 25\}$.
- The effort and initial payment shares of both workers follow from the above variables.

In the baseline condition, no information about the counterfactual effort choice of the disadvantaged worker is provided. In the "low counterfactual" and the "high counterfactual" conditions, spectators are informed about what the disadvantaged worker would have made in advantaged circumstances.

Model. Each individual endorses one of the four merit views that are discussed in Section 1.2 of the main text. A respondent i of type t rewards the workers according to her merit view $m_{t(i)}(e,s)$ in scenario s and a normally distributed response error $\varepsilon_{is} \sim_{iid} N(0, \sigma^2)$. That is, $r_{is} = m_{t(i)}(e,s) + \varepsilon_{is}$.

As discussed in Section 1.6, I parametrize the fairness view of comparable choice meritocrats (CCM) as follows:

$$m_{\text{CCM}}(e,s) \;=\; \begin{cases} \rho Ec(e,s) + (1-\rho)e & \text{if counterfactual is uncertain} \\ c(e,s) & \text{if counterfactual is known} \end{cases}$$

This means that comparable choice meritocrats tend to discount the expected counterfactual effort choice if it is uncertain. The discount parameter is ρ.

I also need to estimate spectators' expectation of the counterfactual effort share, $Ec_(e,s)$, when the counterfactual is unknown. In line with the evidence of Section 1.5, I assume that spectators correctly anticipate the average effect of the piece-rate. Moreover, I assume that the real piece-rate effect is constant in line with the discussion in Section 1.2. In the data, I observe that workers are willing to complete about 12.5 tasks more for a high piece-rate (see Table 1.C.1, Column 3). I use

31. The contingent response method allows me to freely vary the effort choices of workers in the hypothetical scenarios without being deceptive.

this estimate to derive spectator's expected counterfactual effort choice of worker B ($EC_B = E_B + 12.5$) for each effort scenario in the baseline condition where the counterfactual effort choice is unknown.[32] This allows me to derive $Ec(e,s) \approx \frac{EC_B}{E_A+EC_B}$. Below, I show that I obtain virtually identical result with an alternative specification of $Ec(e,s)$. The results are insensitive to the calibration $Ec(e,s)$ because the spectators fully discount it anyway.

I estimate six parameters: the population shares $\boldsymbol{\theta}$ of the four merit views ($\sum_t \theta_t = 1$), the discount parameter ρ, and the standard deviation of the response error σ. I impose $0 \leq \theta_t \leq 1 \ \forall t$, $0 \leq \rho \leq 1$, and $\sigma > 0$.

Log-likelihood.

$$\log F(\boldsymbol{r} \mid \boldsymbol{\theta}, \rho, \sigma) = \sum_i \log f_i(\boldsymbol{r}_i \mid \boldsymbol{\theta}, \rho, \sigma) \tag{1.B.1}$$

$$f_i(\boldsymbol{r}_i | \boldsymbol{\theta}, \rho, \sigma) = \sum_t \theta_t \Pr(\boldsymbol{r}_i \mid \theta_t, \rho, \sigma) \tag{1.B.2}$$

$$\Pr(\boldsymbol{r}_i \mid \theta_t, \rho, \sigma) = \prod_s \varphi(r_{is} - m_{t(i)}(s, e, \rho), \sigma^2) \tag{1.B.3}$$

where φ denotes the normal density function.

Estimation. I estimate the model in R with the help of the `maxLik` package (Henningsen and Toomet, 2011). The BFGS algorithm is used to solve the constrained optimization problem. I estimate ρ, σ, and the share of actual choice meritocrats, comparable choice meritocrats, and libertarians. The share of egalitarians follows via $\sum_t \theta_t = 1$.

Computational robustness. I confirm the numerical stability of the maximum likelihood estimator in three steps. First, I replicate the results in 100 estimations with random start parameters. Second, I generate 100 simulated data sets from the model with randomly drawn parameters and confirm that the estimates recover the parameters of the models. Third, I replicate the results with the Nelder-Mead optimization algorithm.

Inference for constrained maximum likelihood. Standard inference in constrained maximum likelihood models can become unreliable if one of the parameters is on or near the boundary (Schoenberg, 1997). Since I estimate a ρ of 0.00 which is on the boundary, caution seems to be warranted. The discussion below indicates, however, that the inference is nevertheless reliable.

First, I obtain virtually identical estimates and standard errors for $\boldsymbol{\theta}$ and σ if I estimate the model without constraints (results available upon request).

Moreover, I assess the coverage of the confidence intervals in an independent simulation experiment. To this end, I generate 1,000 simulated data sets from the

32. Workers can complete at most 50 tasks, so I cap the counterfactual effort choices at 50.

model, assuming that the main estimates in Table 1.6.1 are the true parameter values. In particular, I impose $\rho = 0$. For each simulated data set, I derive the maximum likelihood estimates and their associated 95% confidence intervals. Then, I assess whether the confidence intervals cover the “true” parameters in about 95% of cases. This is indeed the case. The estimated coverage frequency ranges from 93.4% to 97.2%. I obtain similar results if I randomly perturb $\boldsymbol{\theta}$ and σ in each simulation to explore the coverage in the neighborhood of the estimated parameters (here, the coverage ranges from 94.5% to 98.8%).

Robustness of estimates

Table 1.B.6 shows that the results of the maximum likelihood are robust across several different specifications.

- Main: Main specification
- Duration: Excludes respondents with a response duration that is lower than the 25% percentile.
- Quiz: Excludes respondents who answer at least one quiz question wrongly.
- Guess correct: Excludes respondents who are able to distinguish the real scenario from the hypothetical ones.
- Multipl. effort: Here, I calibrate spectators’ expectations of worker B’s counterfactual effort as $EC_B = 3.3 * E_B$, assuming that the effect of the higher piece-rate is multiplicative. In the data, I observe that workers are willing to complete about 3.3 as many tasks for a high piece-rate than for a low piece-rate (see Table 1.C.1, Column 3).
- Bounds adjust: Because the support of normal noise is unbounded, the likelihood function assigns positive probability to reward shares below 0% or above 100% that cannot occur in practice. Here, I limit the support to values that can occur in practice. I rescale each error density by the inverse cumulative density that lies outside the interval [0%-100%].
- Trembling: I explore an alternative error specification. Respondents have a “trembling hand” and their response r_{is} is fully random (uniform over [0%-100%]) with probability α. With probability $1 - \alpha$, their response is very close to their merit view (normal error with a standard deviation of 2 percentage points).

Heterogeneity

The model allows to estimate whether its parameters differ for subgroups of respondents. Consider two groups of respondents A and B. I assume that the population shares of different fairness types and the counterfactual discount parameter are

$(\boldsymbol{\theta},\ \rho)$ in group A. In group B, the population shares are $(\boldsymbol{\theta},\ \rho) + \boldsymbol{\lambda}$. That is, I allow each parameter p to differ by λ_p between both groups.

I estimate this model separately for the following group comparisons: male versus female respondents, respondents with below-median versus above-median income, respondents without versus with college degree, Democrats versus Republicans. Table 1.B.7 displays the resulting estimates of $\boldsymbol{\lambda}$.

Table 1.B.6. Robustness of structural estimation

	(1) **Main**	(2) **Duration**	(3) **Quiz**	(4) **Guess correct**	(5) **Multipl. effort**	(6) **Bounds adjust**	(7) **Trembling**
Population shares							
Actual choice meritocrats	36.7%	35.5%	39.9%	36.8%	36.7%	35.9%	34.7%
	(1.9%)	(2.1%)	(2.3%)	(2.0%)	(1.9%)	(2.1%)	(1.9%)
Comparable choice merit.	26.2%	28.9%	27.3%	26.2%	26.2%	26.2%	29.2%
	(1.7%)	(2.0%)	(2.1%)	(1.9%)	(1.8%)	(2.1%)	(1.8%)
Libertarians	23.0%	23.7%	22.5%	23.4%	23.0%	23.8%	24.8%
	(1.4%)	(1.6%)	(1.7%)	(1.5%)	(1.4%)	(1.4%)	(1.5%)
Egalitarians	14.2%	11.9%	10.4%	13.6%	14.1%	14.1%	11.3%
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
Counterfactual discount parameter							
ρ	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.04)	(0.05)	(0.05)	(0.04)	(0.06)	(0.11)	(0.01)
Error term and sample							
σ noise	9.27	9.16	8.60	9.32	9.27	9.72	
	(0.11)	(0.13)	(0.12)	(0.12)	(0.11)	(0.13)	
α noise							0.23
							(0.01)
Respondents	945	708	656	834	945	945	945
Decisions	3777	2831	2621	3333	3777	3777	3777

Notes: Results from counterfactual study, decisions 4-7. Maximum likelihood estimation of the structural model of merit views. Standard errors in parentheses. The estimates indicate the population shares of different fairness views and the discounting parameter ρ. The columns estimate the model for different specifications. See text above. No standard errors are reported for the share of egalitarians because their share is deduced from the other estimates.

Table 1.B.7. Differences of model parameters ($\boldsymbol{\lambda}$) by group

	(1) **Female** (vs. male)	(2) **Income >median** (vs. ≤median)	(3) **College degree** (vs. none)	(4) **Republican** (vs. Democrats)
Differences in shares				
Actual choice meritocrats	1.7%	-2.2%	0.7%	7.0%*
	(3.7%)	(3.8%)	(3.8%)	(3.8%)
Comparable choice meritocrats	1.6%	0.8%	-1.5%	-0.2%
	(3.5%)	(3.5%)	(3.5%)	(3.6%)
Libertarians	-1.3%	2.4%	1.9%	-4.1%
	(2.9%)	(2.9%)	(2.9%)	(2.9%)
Egalitarians	-2.0%	-1.0%	-1.1%	-2.8%
	(–)	(–)	(–)	(–)
Differences in counterfactual reasoning				
ρ	0.00	0.00	0.00	0.00
	(0.09)	(0.10)	(0.09)	(0.09)
Sample				
Respondents	916	916	916	916
Decisions	3661	3661	3661	3661

Notes: Results from counterfactual study, decisions 4-7. Maximum likelihood estimation of the structural model of merit views which allows for different parameters across two groups of individuals. Standard errors in parentheses. The table reports the estimated differences in parameters ($\boldsymbol{\lambda}$). For the sake of brevity, the baseline estimates ($\boldsymbol{\theta}$ and ρ) as well as the normal error (σ, constant across groups) are not reported. The columns report results from separate estimations. The column labels indicate which two demographic groups are compared. See text above. No standard errors are reported for the share difference of egalitarians because their share is deduced from the other estimates. *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

1.B.5 Vignette study

Robustness of treatment effects. Table 1.B.8 shows that the results of the vignette study are largely insensitive to the exclusion criterion and to survey weights that render the sample representative for the US general population in terms of gender, age, income, education, and census region. I use a raking algorithm (R package `anesrake`) and follow the guidelines of the American National Election Study to calculate the survey weights (Pasek, Debell, and Krosnick, 2014).

- **Main**: Main specification
- **Keep 45s+**: Exclude respondents who complete the vignettes with an average response time of less than 45 seconds (instead of 60s).
- **Keep 75s+**: Exclude respondents who complete the vignettes with an average response time of less than 75 seconds (instead of 60s).
- **Weighted**: Weighted OLS regression.

Results of additional crime vignette. The vignette survey also contained a fourth vignette on criminal behavior (see Appendix 1.F for the full vignette wording).

> **Crime vignette**: In this vignette, the advantaged person grew up in a rich neighborhood with low crime rates. He went to good schools, and his parents made sure he grew up in a loving, nurturing environment. The disadvantaged person grew up in a poor neighborhood with very high crime rates. His parents often neglected him, and both his family and peers committed crimes. While the advantaged person started studying business and works as a salesman, the disadvantaged person started selling drugs and frequently violates the law. Both earn $50,000 a year today.

In contrast to the other vignette, the crime vignette revolves around legal versus illegal behavior instead of hard work or entrepreneurial risk-taking, and both persons earn equal instead of unequal incomes. As a consequence, respondents redistribute money *away* from the disadvantaged, criminal person in the baseline condition, likely because they reject the illegal source of his income. Only 41% accept the initial income equality between both persons (Column 1, Table 1.B.9). This fraction is virtually identical in the low counterfactual treatment, but 12.3 percentage points higher in the high counterfactual treatment, replicating the findings in the other vignettes.

Still, Column 2 suggests that the average reward share of the unlawful person might be slightly lower when respondents know that the person would violate the law even if he had grown up in privileged circumstances. This effect is driven by a slightly larger share of respondents who take all money away from the unlawful person (Column 3). Both effects are however only marginally significant.

Table 1.B.8. Robustness of the results from the vignette study

(A) Share of respondents redistributing towards the disadvantaged worker

	Binary indicator for compensation			
	Main	Keep 45s+	Keep 75s+	Weighted
	(1)	(2)	(3)	(4)
Low counterfactual	−0.004	−0.016	0.002	−0.000
	(0.029)	(0.029)	(0.031)	(0.038)
High counterfactual	0.126***	0.122***	0.122***	0.135***
	(0.029)	(0.029)	(0.031)	(0.037)
Vignette FE	✓	✓	✓	✓
Observations	2,664	2,789	2,390	2,664
R^2	0.028	0.028	0.027	0.024

(B) Mean reward share of disadvantaged person

	Reward share of disadv. person (in %)			
	Main	Keep 45s+	Keep 75s+	Weighted
	(1)	(2)	(3)	(4)
Low counterfactual	−1.539	−1.828*	−0.974	−1.332
	(1.085)	(1.075)	(1.121)	(1.495)
High counterfactual	6.795***	6.921***	6.861***	6.847***
	(1.177)	(1.175)	(1.224)	(1.447)
Vignette FE	✓	✓	✓	✓
Observations	2,664	2,789	2,390	2,664
R^2	0.135	0.133	0.139	0.116

Notes: Results from the vignette study. OLS regressions with standard errors clustered on the respondent level. The dependent variable in Panel A is a binary indicator for whether a respondent compensates the disadvantaged person by redistributing money towards him. The dependent variable in Panel B is the reward share assigned the disadvantaged person. The independent variables are treatment dummies. Columns 1 shows the main specification. Column 2-4 report different robustness checks that are explained above. *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

Table 1.B.9. Vignette study: Results from the crime vignette

	Binary indicator for equal shares	Reward share of disadv. person (in %)	Binary indicator for giving 0% to the disadv. person
	(1)	(2)	(3)
Low counterfactual	−0.031	−3.066*	0.056*
	(0.040)	(1.649)	(0.029)
High counterfactual	0.123***	3.347**	−0.004
	(0.040)	(1.571)	(0.027)
Constant	0.412***	34.111***	0.124***
	(0.028)	(1.114)	(0.019)
Observations	894	894	894
R^2	0.018	0.017	0.006

Notes: Results from the vignette study. OLS regressions with robust standard errors. Column 1 regresses a binary indicator for whether a respondent accepts the reward equality between both persons on treatment dummies. The dependent variable in Column 2 is the reward share assigned the disadvantaged person. In Column 3, the dependent variable is a binary indicator for taking all money away from the unlawful person. *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

Appendix 1.C Endogenous effort choices in the worker setting

This appendix documents that the piece-rates strongly influence how much effort a worker exerts. I study the effort choices of 548 workers who were recruited for the study. 336 workers were recruited for the *main*, *robustness "equal rates"*, *attention*, and *attention "equal rates"* studies (Amazon Mechanical Turk, US, May and June 2020). 212 were recruited for the *counterfactual* study (Amazon Mechanical Turk, US, January 2021).[33]

Table 1.C.1 regresses the number of completed tasks on an indicator for a high piece-rate. Specifically,

- Column 1, Main: "High rate" means a piece-rate of \$0.50 instead of \$0.10.
- Column 2, Robustness "equal rates": "High rate" means (uncertain) piece-rate prospects of \$0.50 or \$0.90 (with equal chance) instead of \$0.10 or \$0.50 (with equal chance).
- Column 3, Counterfactual: "High rate" means a piece-rate of \$0.50 instead of \$0.10. The counterfactual study uses a within-subject design. Each worker decides how much effort he would exert for a high piece-rate and for a low piece-rate.

The higher piece-rate leads to a 333% higher effort in the main condition, a 155% higher effort in the robustness "equal rates" condition, and a 335% higher effort in the counterfactual condition. Thus, the external piece-rate strongly affects how much effort the workers exert.

33. In addition, I recruited 56 workers for the *robustness "disappointment"* study (Amazon Mechanical Turk, US, February 2021). Workers in this condition do not make an effort choice. They have to complete exactly ten tasks.

Table 1.C.1. The effect of a high piece-rate on workers' effort

	Effort (number of completed tasks)		
	Main	Robustness "Equal rates"	Counterfactual
	(1)	(2)	(3)
High rate	11.744***	5.553**	12.547***
	(2.308)	(2.357)	(1.540)
Constant	5.040***	10.044***	5.349***
	(1.135)	(1.226)	(1.043)
Observations	124	212	212
R^2	0.142	0.029	0.149

Notes: OLS regressions, robust standard errors in Columns 1 and 2, standard errors clustered on the worker level in Column 3. The dependent variable is the number of tasks a worker completes. "High rate" is an indicator for high piece-rate (prospects).

Appendix 1.D Research transparency

Preregistration. The main study, the robustness study: equal rates, the robustness study: disappointment, the attention study, the attention "equal rates" study, and the counterfactual study were preregistered as project #AEARCTR-0005811 at the AEA RCT Registry. The preregistration includes details on the experimental design, the full experimental instructions, thus the full list of measured variables, the sampling process and planned sample size, exclusion criteria, hypotheses, and the main analyses. The following notes document where I deviate from the preregistration.

- The preregistration uses a different title and different treatment labels.
- Non-preregistered analyses include the comparison of worker B's reward share, averaged across effort scenarios (a straight-forward summary of the scenario-by-scenario differences), and the structural estimation.
- Wherever I explicitly deviate from the analysis plan, I choose a more conservative approach. For instance, I do not adjust the treatment comparisons in each effort scenario for multiple hypothesis testing. This renders their non-significance even more conservative. The highly significant effects in the counterfactual study survive even conservative adjustments for multiple hypotheses testing.
- The sample size differs slightly from the pre-registered size of about 300 per condition due to the logistics of the sampling process.
- The preregistration defines the difference in payment shares $\Delta p = \frac{P_A}{P_A+P_B} - \frac{P_B}{P_A+P_B}$ as main outcome variable. In contrast, I use worker B's payment share $p = \frac{P_B}{P_A+P_B}$ as main outcome variable. Since both are linearly dependent ($p = \frac{1-\Delta p}{2}$), this difference does not affect the results but eases their interpretation.

The vignette study was not pre-registered.

Ethics approval. The study obtained ethics approval from the German Association for Experimental Economic Research (#HyegJqzx, 12/11/2019).

Data and code availability. All data and code will be made available online.

Competing interests. I declare that I have no competing interests.

Appendix 1.E Extract from the main study's instructions

This appendix shows the central experimental instructions from the main study. The full experimental instructions for all studies are available at https://osf.io/xj7vc/.

Part 1

In what follows, we will ask you to make a series of decisions that might have **consequences for a real-life situation**.

Please read the following pages very carefully. A **quiz** will test your understanding. You can proceed with the study only if you answer all quiz questions correctly.

→

– PAGE BREAK –

The context of your decision

Our institute currently hires adults from the US general public on an online job portal to work on an important task for one of our projects.

Task
These workers search for publicly available email addresses of academic economists. In each task, a worker is given the name of one economist, searches for the economist's personal or university webpage, identifies his or her email address and sends it to us.

The task requires no special qualification or ability, but demands concentration and effort. Typically, it takes about 2 minutes to complete one task.

Workers can freely choose how long they work and how many tasks they want to complete. At most, they can complete 50 tasks.

← →

The context of your decision

Payment
Each worker receives a fixed reward of $1.00 for completing the job as well as a variable payment. The variable payment depends on the number of completed tasks, a piece-rate, and your decisions in this survey. From now on, when we say "payment", we are only referring to this variable payment. It is calculated in two steps:

(1) A worker initially earns a fixed amount for each solved task. We refer to this amount per task as a piece-rate.

variable payment = *number of tasks* *x* *piece-rate*

For example, a worker who has a piece-rate of $0.20 and solves 10 tasks receives a variable payment of $2 (namely $0.20 x 10).

(2) Afterwards, someone else determines the final payments. Workers are informed about this, although they do not know how and why this happens.

This is where you come into play ...

← →

Your decisions

In the last weeks, we hired 200 workers and matched them into 100 pairs. The decisions that you and others make in this study determine their final earnings. We randomly select one study respondent for each pair of workers.

If you are one of the selected respondents, **your decisions determine the final earnings of a pair of workers**. Let us call them *worker A* and *worker B*.

You can redistribute the payments between worker A and worker B. That is, you decide which share of the total payment amount each worker receives.

Example: Worker A receives a payment of $10 and worker B of $5 so that the sum of their payments is $15. You can freely choose how to distribute the total amount of $15 between both workers.

Completely anonymous: Please note that your decisions are completely anonymous. The workers will receive the shares that you choose with no further information. In particular, they will not learn anything about you or the nature of your decisions.

← →

Multiple decisions - each might matter

We ask you to consider **8 different scenarios** corresponding to different possible work outcomes for worker A and worker B. 7 of those scenarios are hypothetical. 1 scenario is real and describes what actually happened when worker A and worker B worked on this task.

You will make **one distribution decision for each scenario**. If you are among the selected respondents, your decision in the real scenario is implemented and determines how much each worker earns. However, you will not be told which scenario really happened, so all of your decisions are important.

Therefore, please take each decision seriously. It might matter a lot to two real workers from the US.

← →

– PAGE BREAK –

The piece-rates

Recall that the piece-rates of the workers determine how much they initially earn for each task. In what follows, we explain how these piece-rates are determined.

← →

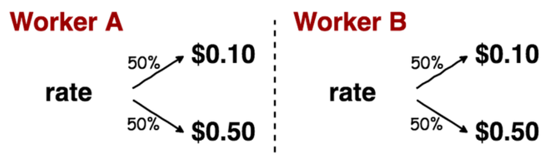
– INFORMATION FOR *CONTROL* GROUP –

The piece-rates

Please read the following information very carefully.

The piece-rate of each worker was determined randomly by a virtual coin flip. Each worker had a 50% chance to get a piece-rate of \$0.10 and a 50% chance to get a piece-rate of \$0.50. One coin flip determined the rate of worker A, and another coin flip determined the rate of worker B.

Thus, the workers had equal prospects to work for the low or the high rate.



Importantly, workers did not know during their work which piece-rate they would get. Only the chances of getting the rates were known. The coin was flipped only after a worker completed and submitted the job. Only then, a worker was informed about his or her definite piece-rate.

In the end, the coin flip determined the following definite rates:

- **Worker A** had a rate of **\$0.50**.
- **Worker B** had a rate of **\$0.10**.

Thus, they worked for a different rate, but they were informed about their rate only after they completed the job.

← →

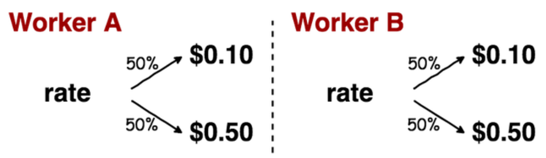
– INFORMATION FOR *TREATMENT* GROUP –

The piece-rates

Please read the following information very carefully.

The piece-rate of each worker was determined randomly by a virtual coin flip. Each worker had a 50% chance to get a piece-rate of \$0.10 and a 50% chance to get a piece-rate of \$0.50. One coin flip determined the rate of worker A, and another coin flip determined the rate of worker B.

Thus, the workers had equal prospects to work for the low or the high rate.



Importantly, workers knew which piece-rate they would get before starting their work. The coin was flipped before the workers started working and workers were informed about the result directly.

The coin flip determined the following definite rates:

- **Worker A** had a rate of **\$0.50**.
- **Worker B** had a rate of **\$0.10**.

Thus, they worked for a different rate.

←

→

– EXAMPLE: REDISTRIBUTION DECISION FOR *CONTROL* GROUP –

Scenario 1

	Rate prospects (known to worker)	Final rate (unknown to worker)	Completed tasks	Initial payment
Worker A	$0.10 or $0.50 50% chance for each	$0.50	**45 tasks** 90% of total work	**$22.50** 98% of total payment
Worker B	$0.10 or $0.50 50% chance for each	$0.10	**5 tasks** 10% of total work	**$0.50** 2% of total payment
			Total payment:	$23.00

Please split the total payment between both workers.
To do so, please specify which share of the total payment each worker gets. The shares need to add up to 100%.

Share of **worker A**	0	%
Share of **worker B**	0	%
Total	0	%

– EXAMPLE: REDISTRIBUTION DECISION FOR *TREATMENT* GROUP –

Scenario 1

	Rate (known to worker)	Completed tasks	Initial payment
Worker A	$0.50	**45 tasks** 90% of total work	**$22.50** 98% of total payment
Worker B	$0.10	**5 tasks** 10% of total work	**$0.50** 2% of total payment
		Total payment:	$23.00

Please split the total payment between both workers.
To do so, please specify which share of the total payment each worker gets. The shares need to add up to 100%.

Share of **worker A**	0	%
Share of **worker B**	0	%
Total	0	%

Appendix 1.F Extract from the vignette study's instructions

This appendix shows the scenario descriptions from the vignette study. The full instructions for the vignette study are available at https://osf.io/xj7vc/.

1.F.1 Scenario "discrimination"

Richard and Oliver work for the same company. In the last months, they competed for a promotion that came with an attractive one-time bonus of $10,000.

However, their boss is notorious for favoring white employees. In fact, he has never promoted a black person before, although he has had plenty of opportunities to do so.

Richard is white. He worked hard to win the promotion.

Oliver is black. He did not work hard to win the promotion.

Who got promoted?

As a consequence of their choices, Richard is promoted and receives the bonus of $10,000. Oliver is not promoted and receives no bonus.

[*Addendum, High counterfactual condition*]

What if the boss did not favor white employees?

Assume that if the boss did not favor white employees, Oliver would have made the same choice as Richard. **Oliver would have worked as hard as Richard did.**

[*Addendum, Low counterfactual condition*]

What if the boss did not favor white employees?

Assume that if his boss did not favor white employees, Oliver would still have made the same choice. **Oliver would not have worked hard.**

1.F.2 Scenario "poverty"

Mike

Mike grew up in a rich family. He was always told, "In this country, you can go as far as your hard work takes you." His family expected him to work hard. Mike went to good, engaging schools that challenged him. He knew he would be popular among his peers if he achieved good grades and worked hard.

Mike has always worked hard in his life.

Paul

Paul grew up in a poor family. He was always told, "In this country, the poor stay poor, and the rich get richer." His family did not expect him to work hard. Paul went to poor-quality schools where he was bored and never challenged. He knew he would be popular among his peers if he was lazy, rebelled against authority, and violated rules.

Paul has never worked hard in his life.

Income today

As a consequence of their choices, Mike earns $125,000 a year, and Paul earns $25,000 a year.

[*Addendum, High counterfactual condition*]

What if Paul had grown up in Mike's environment?

Assume that if Paul had grown up in the same environment as Mike, he would have made the same choices as Mike. **Paul would always have worked as hard as Mike did.**

[*Addendum, Low counterfactual condition*]

What if Paul had grown up in Mike's environment?

Assume that if Paul had grown up in the same environment as Mike, he would still have made the same choices. **Paul would never have worked hard in his life.**

1.F.3 Scenario “start-up”

Frank

Frank always dreamed of founding his own software start-up. He knew that he would **inherit a considerable fortune**. Therefore, he knew that he had enough money to launch his start-up, and that even if his first attempts failed, he would have enough money left to try again and pursue a new business idea.

Frank decided to take the risk and founded his own software start-up.

Ray

Ray always dreamed of founding his own software start-up, too. However, Ray’s parents were poor and he had **very little money**. Therefore, he knew that it would be difficult to find enough money to launch a start-up, and he knew that if his first attempt failed, he would be broke.

Ray decided not to take the risk. Instead, he works as a software developer for a local company.

Income today

As a consequence of their choices, Frank earns $200,000 a year, and Ray earns $50,000 a year.

[*Addendum, High counterfactual condition*]

What if Ray had had as much money as Frank?

Assume that if Ray had had as much money as Frank, he would have made the same choices as Frank. **Ray would have taken the risk and founded his own software start-up.**

[*Addendum, Low counterfactual condition*]

What if Ray had had as much money as Frank?

Assume that if Ray had had as much money as Frank, he would still have made the same choices. **Ray would have decided not to take the risk. Instead, he would work as a software developer for a local company.**

1.F.4 Scenario "crime"

Robert

Robert grew up in a rich neighborhood with very low crime rates. His parents made sure he grew up in a loving, nurturing environment. Robert has always been told, "In this country, you can rise as far as you want if you play by the rules." Robert went to good, engaging schools that challenged him. Many of his peers planned to study at a university.

Robert started studying business at the age of 20. Today, he works as salesman. He never does anything illegal.

John

John grew up in a poor neighborhood with very high crime rates. His parents often neglected him. Once his father was caught selling drugs and had to spend several years in jail. John has always been told, "Playing by the rules means nothing when the rules are stacked against you." He went to poor-quality schools where he was bored and never challenged. Many of his peers had already committed crimes by the time they reached their teenage years.

John committed his first crime at the age of 20. Today, he sells drugs. He frequently violates the law.

Income today

As a consequence of their choices, Robert earns $50,000 a year, and John earns $50,000 a year.

[*Addendum, High counterfactual condition*]

What if John had grown up in Robert's environment?

Assume that if John had grown up in the same environment as Robert, he would have made the same choices as Robert. **John would never do anything illegal.**

[*Addendum, Low counterfactual condition*]

What if John had grown up in Robert's environment?

Assume that if John had grown up in the same environment as Robert, he would still have made the same choices. **John would sell drugs and frequently violate the law.**

Chapter 2

Subjective Models of the Macroeconomy

Joint with Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart

Abstract: We study people's subjective models of the macroeconomy and shed light on their attentional foundations. To do so, we measure beliefs about the effects of macroeconomic shocks on unemployment and inflation, providing respondents with identical information about the parameters of the shocks and previous realizations of macroeconomic variables. Within samples of both 6,500 US households and 1,500 experts, beliefs are widely dispersed, even about the directional effects of shocks, and there are large differences in average beliefs between households and experts. Part of this disagreement seems to arise from selective retrieval of different propagation channels of macroeconomic shocks. We confirm this mechanism causally by exogenously shifting households' attention to either supply-side or demand-side channels. Moreover, households with different personal experiences recall different propagation channels of the shocks, while experts tend to recall textbook models. Our findings offer a new perspective on the widely documented disagreement in macroeconomic expectations.

Acknowledgments: We thank the editor (Nicola Gennaioli) and three anonymous referees for extremely useful suggestions. We also thank Rudi Bachmann, Carola Binder, Pedro Bordalo, Benjamin Born, Felix Chopra, Olivier Coibion, Francesco D'Acunto, Stefano DellaVigna, Thomas Dohmen, Armin Falk, Andreas Fuster, Xavier Gabaix, Dimitris Georgarakos, Yuriy Gorodnichenko, Thomas Graeber, Michael Haliassos, Lukas Hensel, Chi Hyun Kim, Gizem Kosar, Michael Kosfeld, Fabian Krüger, Matt Lowe, Michael McMahon, Valerie Ramey, Aakaash Rao, Sonja Settele, Andrei Shleifer, Uwe Sunde, Johannes Stroebel, Giorgio Topa, Egon Tripodi, Michael Weber, Mirko Wiederholt, Basit Zafar, Florian Zimmermann, as well as participants at various conferences and seminars. We thank Dorine Boumans, Johanna Garnitz, Andreas Peichl, and the ifo Institute for including our module in the World Economic Survey and Valentin Reich for help with the analysis. We are grateful to the data services of the IDSC of IZA. Hrishikesh Iyengar, Apoorv Kanoongo, Melisa Kurtis, Anna Lane and David Zeimentz provided excellent research assistance. **Funding**: We thank the Joachim Herz Foundation as well as the Fritz-Thyssen Foundation for financial support. Funding by the Deutsche Forschungsgemeinschaft (DFG) through CRC TR 224 (Project A01) is gratefully acknowledged. The activities of the Center for Economic Behavior and Inequality (CEBI) are financed by the Danish National Research Foundation, Grant DNRF134. Roth: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866. **Ethics approval**: We received ethics approval from Goethe University Frankfurt and the University of Warwick. **Instructions**: The experimental instructions can be found at the following link: https://osf.io/6mxaz/. **Disclaimer**: The views expressed in this paper are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

2.1 Introduction

Individuals usually exhibit substantial disagreement in their expectations about macroeconomic outcomes. This holds true for consumers, firm managers, retail investors, and even professional forecasters (Mankiw, Reis, and Wolfers, 2003; Dovern, Fritsche, and Slacalek, 2012; Coibion and Gorodnichenko, 2015a; Link, Peichl, Roth, and Wohlfart, 2020; Giglio, Maggiori, Stroebel, and Utkus, 2021). Disagreement in turn has major implications for the transmission of shocks and of fiscal and monetary policy (Ball, Mankiw, and Reis, 2005; Paciello and Wiederholt, 2014; Angeletos and Lian, 2018). There are two broad views on what is driving disagreement in expectations. Disagreement is most commonly attributed to differences in information about the current state of the economy (Mankiw and Reis, 2002; Reis, 2006; Coibion and Gorodnichenko, 2012). According to such explanations, conditional on the same information set, economic agents make homogeneous predictions about the reaction of the economy to shocks. Alternatively, disagreement could be due to heterogeneity in subjective models, that is, the way agents think about the functioning of the economy (Bray and Savin, 1986; Marcet and Sargent, 1989; Molavi, 2019; Angeletos, Huo, and Sastry, 2020). Such heterogeneity generates disagreement in expectations even when all agents observe the same shock and have the same information about previous realizations of macroeconomic variables.

In this paper, we provide the first direct empirical evidence on people's subjective models of the macroeconomy and their origins. We propose that heterogeneity in subjective models is a consequence of selective recall of specific economic mechanisms, which differ across individuals and contexts. We use a new approach to measure people's subjective models, which we apply to samples of about 6,500 respondents representative of the US population and about 1,500 academic and non-academic experts. Our approach relies on vignettes in which respondents predict future unemployment and inflation under different hypothetical macroeconomic shocks. We focus on four different shocks that are among the most commonly studied in macroeconomics: an oil supply shock, a monetary policy shock, a government spending shock, and an income tax shock. The vignettes make sure that all respondents observe the shock, and provide information about the source of the shock and previous realizations of unemployment and inflation. This ensures comparable information sets across respondents and enables us to characterize heterogeneity in forecasts to the extent it arises from differences in subjective models.

For each vignette, we elicit the respondents' expectations about the unemployment rate and the inflation rate twice: first, under a hypothetical baseline scenario in which no shock occurs; second, under a hypothetical shock scenario in which the shock variable unexpectedly changes. In the oil price vignette, we tell our respondents that the oil price will be on average $30 higher over the following 12 months. In the monetary policy vignette, the federal funds rate increases by 0.5 p.p. In the government spending vignette, the government announces a major new spending program on defense, while in the income tax vignette, the government increases income taxes by 1 p.p. for every US household for one year. To establish the exogeneity of the shocks, we tell respondents that the change in the oil price is due to problems with the local production technology in the Middle East, that the federal funds rate is increased even though the Fed does not change its assessment of economic conditions, and that government spending or taxes are increased without any changes in the government's assessment of national security or economic conditions. By taking the difference in the forecasts of unemployment and inflation between the shock scenario and the baseline scenario, we identify each respondent's beliefs about the effects of the shock, while taking out differences in baseline expectations across individuals.

We document five key results. Our first main result is that there is substantial heterogeneity in forecasts about the effects of macroeconomic shocks, among experts, among households, and between the two groups. For example, in the monetary policy vignette, 72% of experts predict an increase in unemployment in response to the rise in the federal funds rate, 12% expect no change, and the remaining 15% expect a decrease. Among households, 51% predict an increase in unemployment, 16% expect no change, and 33% expect a decrease. Similarly, there is strong heterogeneity in beliefs about the inflation response to interest rate hikes, with both increases and decreases being predicted by substantial fractions of households (57% vs. 30%) and experts (19% vs 72%). Across all vignettes, there is more disagreement among households than among experts. Average predictions of households and experts are often similar but differ substantially in three cases: Experts predict inflation to decrease in response to a hike in the federal funds rate, while households forecast an increase in inflation. Similarly, households predict inflation to increase in response to the income tax hike, while experts predict it to decrease. Finally, households predict a muted unemployment response to a government spending program, while experts predict a decrease. The high levels of disagreement in a setting where individuals have comparable information about past realizations of macroeconomic variables indicates an important role for heterogeneity in subjective models in expectation formation.

In a second step, we explore the origins of this heterogeneity. Specifically, we examine the possibility that individuals selectively retrieve specific propagation mechanisms of the shocks, while neglecting others. Selective memory has been shown to be important in shaping people's thoughts and behavior in various contexts (Tversky and Kahneman, 1973; Kahana, 2012; Bordalo, Coffman, Gennaioli, Schwerter, and Shleifer, 2020). In our setting, differences in associations across individuals and contexts could be a key driver of heterogeneity in forecasts. Based on an additional tailored survey, we provide direct evidence on this conjecture. We directly measure what comes to respondents' minds when they think about the shocks using a combination of unstructured textual responses as well as responses to more structured questions. Our second main finding is that the propagation channels that are on respondents' minds vary systematically within and between our samples of households and experts. Across vignettes, experts tend to recall channels that are central in textbook models, while households in many cases neglect these channels and think of channels that are conventionally seen as less important. For example, households are relatively more likely than experts to think of a "cost channel" in the context of the monetary policy shock, according to which firms pass on higher costs of borrowing to consumers in the form of higher prices. By contrast, experts are more likely to think of demand-side mechanisms, such as intertemporal substitution or an investment channel.

In a third step, we ask whether the propagation channels that are on top of respondents' minds are related to their predictions. Our third finding is that thoughts of the different propagation channels are significantly correlated with respondents' unemployment and inflation forecasts, in expected directions. Thoughts of different propagation channels also reconcile part of the differences in forecasts between experts and households.

In a fourth step, we provide proof-of-concept evidence that selective retrieval of specific propagation channels is a causal driver of households' forecasts of the effects of macroeconomic shocks. We conduct an additional experiment with a representative sample in which we use a priming intervention to exogenously shift households' attention to either supply-side or demand-side channels in the context of the monetary policy shock. Our fourth main result is that being primed on demand-side factors significantly increases respondents' retrieval of negative demand-side implications of an increase in the federal funds rate, and has a negative effect on respondents' predicted inflation response to the shock. The finding that drawing households' attention to a specific aspect of the shock changes their forecasts suggests that households' subjective models are not fixed. Instead, these models may be formed "on the fly", depending on the associations triggered by the context. This suggests that news or actual events in the economy may systematically affect which models people entertain. Rather than sticking to one particular model, economic agents retrieve specific memories when cued by events, which in turn shape the economic mechanisms they think of.

Finally, in a fifth step, we test the prediction of selective recall that differences in personal experiences in the memory database should be a key driver of differences in associations and forecasts. Our fifth main result confirms this prediction: households' personal experiences are correlated with selective recall of specific propagation mechanisms, which in turn is reflected in individuals' forecasts about the effects of macroeconomic shocks. For instance, under the government spending shock, which focuses on an increase in defense spending, previous employment by suppliers of the military is associated with a greater tendency to think of mechanisms related to increases in product demand and labor demand. This experience is also associated with a stronger predicted unemployment decrease. Furthermore, in the oil price vignette, having experienced the OPEC crisis in the 1970s is associated with significantly stronger retrieval of cost-push mechanisms, which is reflected in higher predicted unemployment and inflation responses.

Our findings offer a new perspective on the strong heterogeneity in macroeconomic expectations – one of the most well-documented empirical facts in the literature (Mankiw, Reis, and Wolfers, 2003; Coibion and Gorodnichenko, 2012). Our results imply that, even if agents hold comparable information about previous realizations of macroeconomic variables, associative recall of different economic mechanisms generates heterogeneity in expectations. In this view, the subjective models individuals rely on are not fully stable, but depend on what is cued by the context and on individuals' past experiences. Incorporating associative recall into a macroeconomic model could thus be a fruitful avenue for future research.

The main contribution of our paper is to provide the first direct evidence on heterogeneity in subjective models of the macroeconomy and their origins. Our paper builds on previous work studying the relationships between beliefs about different macroeconomic variables. Carvalho and Nechio (2014), Dräger, Lamla, and Pfajfar (2016), and Kuchler and Zafar (2019) use observational data to examine how households' beliefs about unemployment, inflation and interest rates are correlated with each other. A series of papers have used information experiments to study households' beliefs about the autocorrelation of macroeconomic variables (Armantier, Nelson, Topa, van der Klaauw, and Zafar, 2016; Cavallo, Cruces, and Perez-Truglia, 2017; Armona, Fuster, and Zafar, 2018; Fuster, Perez-Truglia, Wiederholt, and Zafar, 2020). Other information experiments have studied how respondents update their expectations about one macroeconomic variable in response to information about a different macroeconomic variable (Coibion, Gorodnichenko, and Kumar, 2018; Coibion, Georgarakos, Gorodnichenko, and van Rooij, 2019; Coibion, Gorodnichenko, and Ropele, 2020; Roth and Wohlfart, 2020). While the randomized provision of information in these experiments allows for causal identification, the interpretation is complicated by the fact that respondents' beliefs about the sources of changes in inflation or GDP growth are unrestricted. In contrast to previous literature, our approach directly measures households' beliefs about the causal effects of macroeconomic shocks on unemployment and inflation, controlling for information about previous realizations of macroeconomic variables and about the sources of the shocks.[1]

1. More generally, we contribute to a growing literature studying the formation of macroeconomic expectations of experts, households and firms, and the role of these expectations in economic and financial decisions (Fuster, Laibson, and Mendel, 2010; Malmendier and Nagel, 2011; Coibion and Gorodnichenko, 2012; Fuster, Hebert, and Laibson, 2012; Armantier, Bruine de Bruin, Topa, Klaauw, and Zafar, 2015; Bachmann, Berg, and Sims, 2015; Coibion and Gorodnichenko, 2015a; Coibion and Gorodnichenko, 2015b; Malmendier and Nagel, 2016; Binder and Rodrigue, 2018; Bordalo, Gennaioli, and Shleifer, 2018; Acosta and Afrouzi, 2019; Afrouzi, 2019; Kamdar, 2019; Vellekoop and Wiederholt, 2019; Binder and Makridis, 2020; Bordalo, Gennaioli, Ma, and Shleifer, 2020; Goldfayn-Frank

Our work relates to research on attention and memory in people's belief formation and decision-making (Gennaioli and Shleifer, 2010; Lacetera, Pope, and Sydnor, 2012; Bordalo, Coffman, Gennaioli, and Shleifer, 2016; Gabaix, 2019; Bordalo, Coffman, Gennaioli, Schwerter, et al., 2020; Enke, Schwerter, and Zimmermann, 2020; Graeber, 2021). Bordalo, Coffman, Gennaioli, Schwerter, et al. (2020) propose a model of choice in which a choice option cues recall of similar past experiences. We contribute to this literature by documenting what comes to people's mind when they think about a set of canonical macroeconomic shocks and by providing causal evidence on the role of associations in shaping the predictions that individuals make. This relates to work by Stantcheva (2020), who provides descriptive evidence on what people think about when they evaluate economic policies, such as estate taxation or health insurance. Our combination of unstructured text responses with priming interventions allows us to characterize how associations causally affect expectation formation.

We also contribute to the literature on the role of personal experiences in macroeconomic expectation formation (Malmendier and Nagel, 2011, 2016; Kuchler and Zafar, 2019; Malmendier, Nagel, and Yan, 2021). While the existing literature has focused on the reduced-form effects of experiences on unconditional expectations of macroeconomic variables, we study how experiences shape forecasts of these variables conditional on the occurrence of shocks. Moreover, our paper provides novel evidence on the link between personal experiences and selective recall of propagation channels, highlighting a potential attentional mechanism underlying experience effects.

Finally, the paper contributes to a small literature that investigates the views and beliefs of academic economists (e.g., Gordon and Dahl, 2013; Sapienza and Zingales, 2013; DellaVigna and Pope, 2018; Andre and Falk, 2021). We document how economists assess and think about four commonly studied macroeconomic shocks.

The rest of this paper is structured as follows. Section 2.2 provides an overview of the samples of households and experts, and the survey design. Section 2.3 presents our evidence on experts' and households' predictions in the different vignettes. Section 2.4 provides evidence on selective recall as a driver of heterogeneity in forecasts. Section 2.5 discusses the implications of our findings for understanding heterogeneity in survey data and for modeling the formation of macroeconomic expectations. Section 2.6 concludes.

and Wohlfart, 2020; Bachmann, Carstensen, Lautenbacher, and Schneider, 2021; Roth, Settele, and Wohlfart, 2021).

2.2 Data and design

2.2.1 Samples

Household survey. For our main online survey, we collect a sample of about 2,200 respondents that is representative of the US population in terms of gender, age, region, total household income, and education. We collect the data in two waves. The first wave was launched in February and March 2019 in collaboration with the market research company Dynata, and the second wave was conducted in July 2019 with the survey company Lucid. Both online panel providers are commonly used in economics and social science research (Haaland, Roth, and Wohlfart, 2021). The pooled sample from Waves 1 and 2 closely matches the characteristics of the general population. For instance, 55% of our respondents are female, compared to 51% in the 2019 American Community Survey (ACS, see Appendix Table 2.B.1). 32% of the respondents in our sample have at least a bachelor's degree, compared to 31% in the ACS. The median income in our sample is $62,500 compared to $65,712 in the ACS.

Expert survey. In parallel to both household survey waves, we recruit two samples of approximately 1,100 experts in total. For the first wave, we invited economists who were authors or discussants at leading macroeconomic conferences.[2] In total, 180 experts completed the first wave of the survey. 83% of these experts are from academic institutions, while 16% work at policy institutions, such as the IMF and central banks (for more details, see Appendix Table 2.B.2). For the second wave, we included our module in the World Economic Survey (WES) – a global survey of economic experts, run by the ifo Institute (Boumans and Garnitz, 2017). 908 experts participated in our module. 56% of these experts are from academia, 16% from policy institutions, 16% work at a bank or a private company, while the remaining 12% have another type of employer. 65% of the experts have a Ph.D., and they predominantly come from North America or Western Europe (50%) (for more details, see Appendix Table 2.B.2). Table 2.B.3 provides an overview of the different data sets used in the paper.

2.2.2 Structure of the survey

Respondents to the household survey start by completing a series of demographic questions. Then, they receive brief non-technical definitions of the unemployment rate and the inflation rate to establish a common-ground definition of the two terms at the start of the survey, and are informed about the current values of these rates. In

2. For details on the conferences considered, see Appendix 2.J. We also invited a few Ph.D. students, experts from several policy institutions, as well as several experts working in the broader areas of expectation formation and macroeconomic forecasting.

the subsequent main part of the survey, participants make predictions about unemployment and inflation under two hypothetical vignettes.[3] Finally, we collect data on some additional respondent characteristics. The expert survey consists of a subset of the household survey. After being introduced to the question format, experts directly proceed to the prediction task in two randomly selected vignettes. We do not include the definitions of inflation and unemployment, but still provide the experts with the most recent values of both variables.[4]

2.2.3 Hypothetical vignettes

To measure our respondents' beliefs about the effects of different macroeconomic shocks, we use hypothetical vignettes in which we introduce our respondents to different scenarios and ask them to predict future unemployment and inflation. This approach allows us to provide individuals with identical information about the source and the parameters of the shock. The vignettes focus on four different exogenous shocks, which are among the most commonly studied in macroeconomics: an oil supply shock, a government spending shock, a monetary policy shock, and a tax shock. This enables us to compare respondents' predictions with estimates from a rich macroeconomic literature. At the same time, these shocks have the advantage that they can be explained to individuals without an economics degree. Our participants are randomly assigned to make predictions for two out of four hypothetical vignettes, which are presented in random order.[5]

Each vignette follows the same structure (summarized in Appendix Figure 2.A.1). All start with a short introduction that familiarizes respondents with the setting of the vignette. For example, in the income tax vignette, they are informed about the average US income tax rate and the amount that the median household currently pays in taxes on labor income. Then, respondents are presented with a *baseline scenario* in which they are asked to imagine that the variable of interest (e.g., income tax rates) does not change. We elicit people's expectations about the unemployment rate in 12 months and the inflation rate over the next 12 months under this scenario. Thereafter, respondents are asked to predict unemployment and

3. A series of papers uses hypothetical vignettes to study belief formation in contexts such as human capital (Wiswall and Zafar, 2017; Delavande and Zafar, 2019) or consumption behavior (Christelis, Georgarakos, Jappelli, Pistaferri, and Van Rooij, 2019; Fuster, Kaplan, and Zafar, 2020).

4. The median household respondent spends about 14 minutes to complete the survey (10th percentile: 7-8 min, 90th percentile: 27-33 min, depending on the wave). The median expert in wave 1 needs 5 minutes to complete the shorter expert survey (10th percentile: 3 min, 90th percentile: 14 min). The survey completion rates are close to 80%. See Table 2.B.4 for further details. Appendix Figure 2.A.1 summarizes the structure of both surveys. The full set of experimental instructions for Wave 1 and Wave 2 of the surveys can be found under the following link: https://osf.io/6mxaz/.

5. In Wave 2 of the expert survey, it was not feasible to randomize the order of vignettes. Instead, the vignettes were ordered as follows: 1. income tax shock, 2. federal funds rate, 3. government spending shock, 4. oil supply shock. Respondents received two randomly selected vignettes.

inflation in a *shock scenario* in which an exogenous shock to the economy is introduced. Specifically, we randomize respondents into a rise-scenario with an increase in the shock variable (e.g., all income tax rates rise by 0.5 p.p.) and a fall-scenario with a decrease in the shock variable (e.g., all income tax rates fall by 0.5 p.p.). To simplify the exposition, we reverse all predictions for the fall-scenarios and analyze them together with predictions for the rise-scenario.[6] Our main outcome variable is respondents' beliefs about the effect of a shock, i.e., the difference in predictions between the shock and the baseline scenario. Eliciting beliefs under both a baseline and a shock scenario has two important methodological advantages: first, it decomposes and simplifies the prediction problem for households; second, divergent beliefs about baseline trends of the US economy that are present in both scenarios cancel out.

Respondents indicate the expected unemployment and inflation rates on two sliders that range from 0% to 10% for unemployment and from -2% to 8% for inflation. The default position of each slider is the value of the respective rate at the time of each survey. The sliders ease the task for our respondents and reduce noise and cognitive strain.[7] In what follows, we provide details on each of the four vignettes.

Oil supply shock. In the introduction to the oil vignette, respondents learn about the current average price of one barrel of crude oil. Then, in the baseline scenario, our respondents are told to imagine that the average price of crude oil stays constant over the next 12 months. Thereafter, they are randomly assigned to either an "oil price rise scenario" or an "oil price fall scenario". Specifically, respondents in the "oil price rise scenario" receive the following instructions:

> *Imagine the average price of crude oil unexpectedly rises due to problems with the local production technology in the Middle East. On average, the price will be $30 higher for the next 12 months than the current price. That is, the price will be on average $84 for the next 12 months.*[8]

As is the case for all other vignettes, instructions for the fall-scenario are analogous to the rise-scenario.

Government spending shock. This vignette first provides respondents with information on the size of yearly government spending in the US and its usual growth rate. In the baseline scenario, our respondents are told to imagine that federal government spending grows as usual over the next 12 months. In the rise-scenario, our respondents receive the following instructions:

6. In appendix Section 2.D.1, we compare predictions across the rise and fall scenarios. Asymmetries occur more often for households than for experts, but are mostly minor.

7. Finally, to account for potential order effects, we cross-randomize whether respondents first receive the question on the inflation rate or the question on the unemployment rate. For each participant, the order of the inflation and unemployment questions is identical across all scenarios.

8. The last sentence of the vignette was not included in Wave 2.

> *Imagine federal government spending unexpectedly grows to a larger extent than usual over the next 12 months due to a newly announced spending program on defense. In particular, total government spending grows by 2.4 p.p. more than the usual growth that took place in the previous years.*
>
> *The government announces: The change is temporary and occurs despite no changes in the government's assessment of national security or economic conditions. Moreover, federal taxes do not change in response to the spending program.*

Monetary policy shock. We familiarize respondents with the federal funds target rate and its current value. The baseline scenario asks our respondents to imagine that the Federal Open Market Committee announces that it will keep the federal funds target rate constant. In the subsequent rise-scenario, our respondents receive the following instructions:

> *Imagine the federal funds target rate is unexpectedly 0.5 percentage points higher. That is, in its next meeting, the Federal Open Market Committee announces that it is raising the rate from 2.5% to 3%.*
>
> *Imagine the committee announces it does so with no changes in their assessment of the economic conditions.*

Tax shock. After a brief explanation of federal income taxes in the US, the baseline scenario tells our respondents to imagine that income tax rates stay constant for all US citizens over the next 12 months. The subsequent rise-scenario is described as follows:

> *Imagine that income tax rates are unexpectedly 1 percentage point higher for all households in the US over the next 12 months. This means that the typical US household would pay about $400 more in taxes.*
>
> *The government announces: The tax change is temporary and occurs despite no changes in the government's assessment of the economic conditions. Moreover, government spending does not change in response to the tax increase.*

Discussion of the design. Our design allows us to interpret belief disagreement as arising from heterogeneity in respondents' subjective models of the economy. We measure a respondent's belief about the effects of a shock as the difference in the respondent's forecasts between the rise/fall and the baseline scenario. By focusing on the difference in forecasts across scenarios, we already control for differences in the baseline level of expected inflation or unemployment across respondents. This aspect of our design shuts down information frictions – the key alternative explanation for belief disagreement – to a large extent. Of course, holding different information about the state of the economy could still affect forecasts of the *effect* of a shock, even under the same subjective model. However, our design choice to provide individuals with identical information about past unemployment, inflation and the realization and parameters of the shock strongly mitigates this remaining concern. As a result, heterogeneity in forecasts across respondents should be due to heterogeneity in the way individuals think about the functioning of the economy – the subjective models they rely on.[9]

Since we work with a general population sample, we face a trade-off between the precision of the vignettes and the ease of understanding them. To avoid cognitive overload among respondents from the general population sample, we make the vignettes as simple to understand as possible. At the same time, we are careful to make clear that the shocks are exogenous to the US economy, which makes our estimates comparable to theoretical models and empirical evidence. For instance, we attribute the oil supply shock to changes in the local production technology in the Middle East. Similarly, in the interest rate scenario, we explicitly state that the change in interest rates occurs with no changes in the Fed's assessment of economic conditions. Moreover, we also fix people's beliefs about the duration of the shocks by clarifying that the changes in taxation and government spending only last for one year.[10] For the government spending and taxation shocks, we clarify that the temporary nature of the shock is common knowledge by using the wording "the government announces".

Furthermore, many of our design choices are motivated by common modeling assumptions in DSGE models and by empirical evidence from VARs in order to ensure comparability of our survey responses to these external benchmarks. For example, empirical evidence on government spending shocks often focuses on defense spending (e.g., Nakamura and Steinsson, 2014; Basso and Rachedi, 2019; Auerbach, Gorodnichenko, and Murphy, 2020) as this type of spending does not affect the economy's productivity and does not directly redistribute resources across the income distribution.

Theoretical and empirical benchmarks. We draw from seminal studies in the theoretical and empirical literature to obtain benchmark estimates for the inflation and unemployment responses to each shock.[11] These values broadly illustrate the view on the effects of shocks established in the literature and put respondents' estimates into context. For example, for the oil price shock, our empirical benchmark is derived from the VAR estimate of Blanchard and Galí (2010) for the Great Moderation period, while the theoretical benchmark is based on Bodenstein, Erceg, and

9. Part of the heterogeneity in forecasts in our vignettes could reflect measurement error. However, much of our descriptive analysis in Section 3 focuses on directional predictions, for which measurement error should be strongly mitigated. In addition, in our analysis of the role of thoughts of different propagation channels in Section 4, forecasts are used as dependent variables, so (classical) measurement error should not bias coefficient estimates.

10. We do not fix beliefs about the duration of the change in interest rates under the monetary policy shock, since the interest rate should react endogenously to changes in inflation and unemployment in response to the shock through the Taylor rule.

11. We found no established benchmark estimate for the inflation response to the income tax shock.

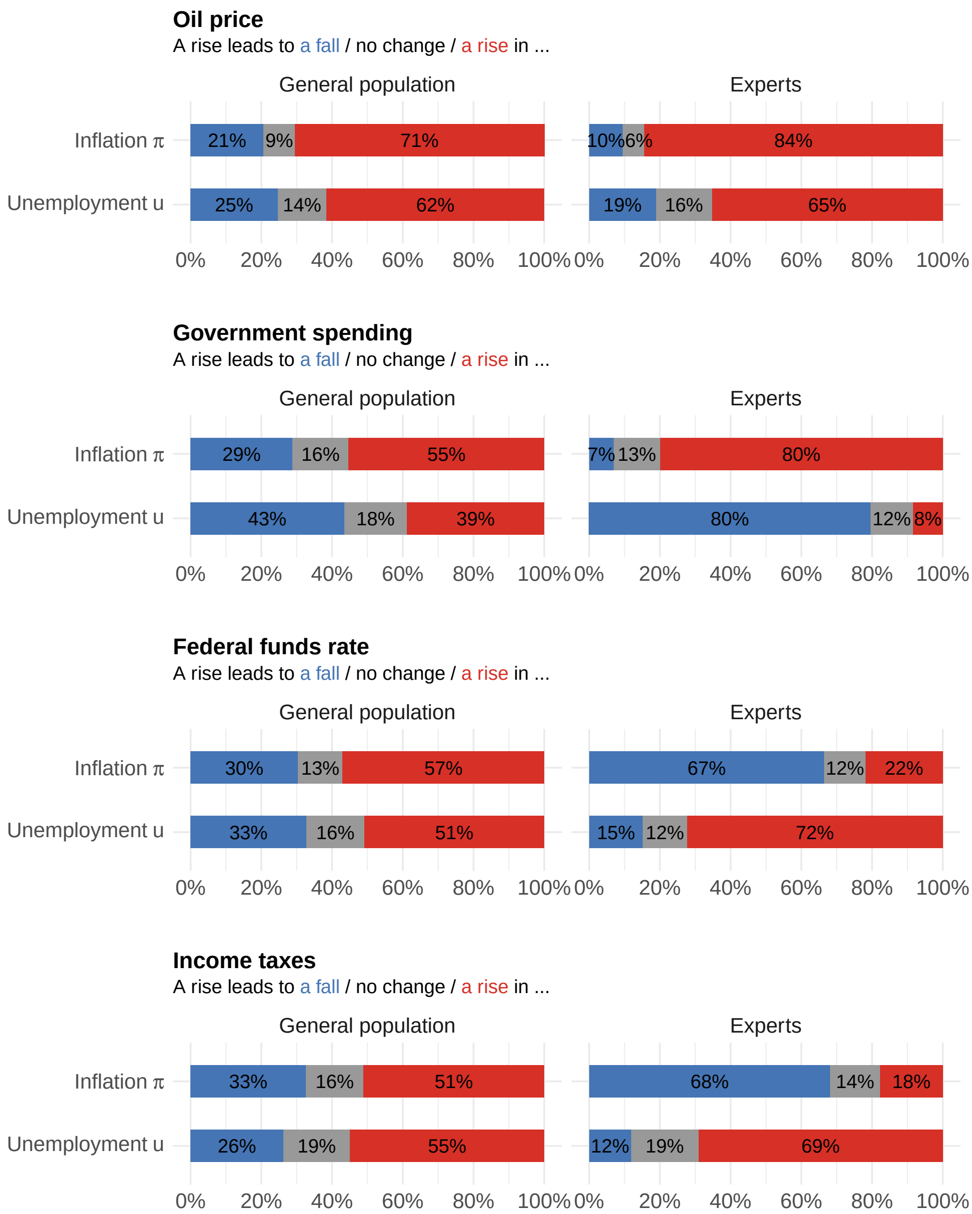
Guerrieri (2011) and Balke and Brown (2018). The former paper models the US as a purely oil-importing country and the latter treats the US as both oil-producing and oil-importing. Naturally, given the always ongoing debates in the respective areas, these benchmarks neither represent “correct” values nor do they fully capture the degree of estimates across the entire literature on each topic. Appendix 2.C provides details on the derivations of the benchmarks and lists the main studies that we consulted.

Differences between Waves 1 and 2. We introduce a couple of minor wording changes to the instructions of Wave 2 to confirm that the results are robust to these modifications. First, our main object of interest are individuals’ beliefs about the effects of the shocks accounting for potential endogenous responses by policymakers. We, therefore, explicitly tell respondents in Wave 2 of both the household and the expert survey to account for potential responses of the government and the central bank when making the predictions. Second, to ensure that the respondents do not just interpret our questions as a test of their knowledge of economics, we tell them that we are interested in their own subjective views on what would actually happen under the different scenarios. Despite these differences in instructions across Waves 1 and 2, there are barely any differences in responses, neither in the household nor in the expert survey. We therefore focus on the pooled sample in our main analysis.

2.3 Predicted unemployment and inflation responses to shocks

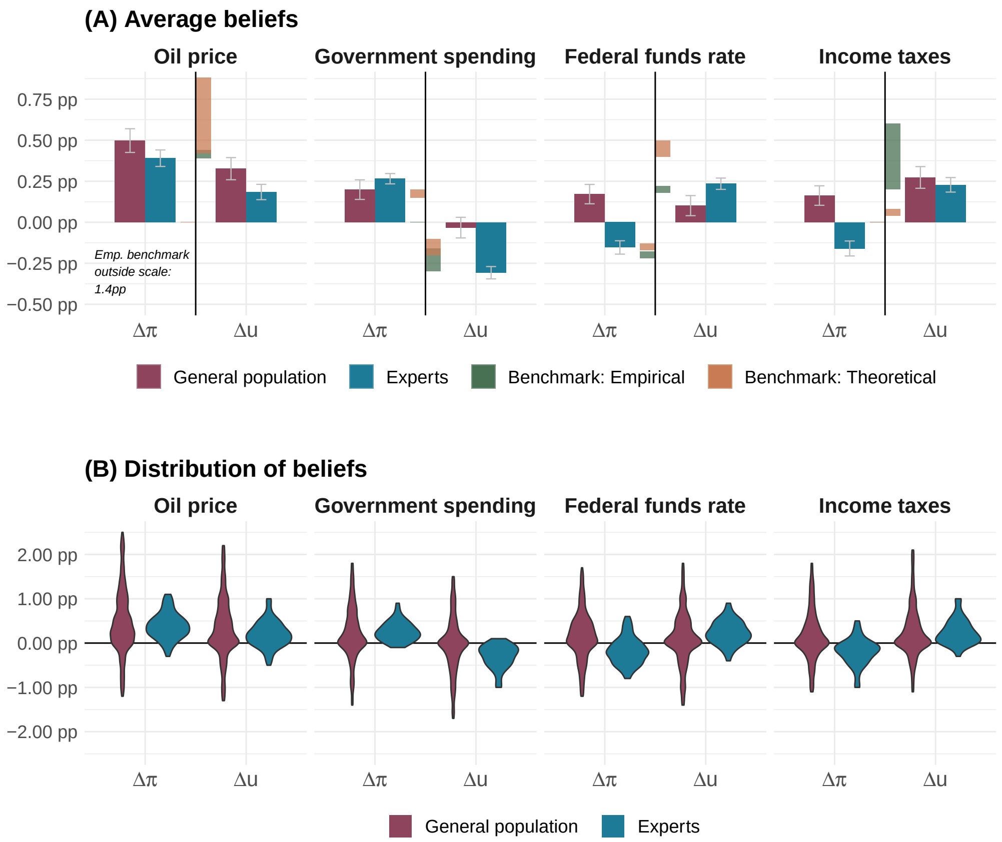
In this section, we present our results on experts’ and households’ forecasts of the effects of macroeconomic shocks. For each shock, we discuss the heterogeneity in predictions within the expert sample, within the household sample, and between both groups. Figure 2.3.1 presents the fractions of experts and households who predict a fall, no change, or rise of inflation and unemployment for each shock, respectively. We focus mostly on the qualitative directions of forecasts as those are less susceptible to extreme predictions.[12] Panel A of Figure 2.3.2 then presents the average quantitative predictions as well as the benchmark estimates from the empirical and theoretical literature. Panel B of Figure 2.3.2 displays the full distribution of the quantitative predictions in separate violin plots.

12. Given the large sample size, even minor differences in households’ and experts’ directional predictions are statistically different ($p < 0.01$, χ^2-tests). Moreover, disagreement is always significantly larger among households than among experts (see Appendix Table 2.B.5). We also confirm the robustness of our results in several checks. Appendix 2.D.3 discusses order effects and the effect of incentives on predictions of households. Figure 2.A.2 showcases the stability of the expert results in different subsamples of experts.



Notes: This figure presents the forecasts of the directional effects of macroeconomic shocks on the inflation rate and the unemployment rate, using Wave 1 and Wave 2 data. It compares the forecasts of the general population (left column) to those of experts (right column). Predictions in the fall scenarios are reversed to render them comparable to rise predictions.

Figure 2.3.1. Forecasts of the directional effects of macroeconomic shocks



Notes: Panel A displays the average forecasts of the effects of macroeconomic shocks on the inflation rate ($\Delta\pi$) and the unemployment rate (Δu), using Wave 1 and Wave 2 data. It compares responses in the representative sample (red bars) with those of experts (blue bars). Error bars present 95% confidence intervals, using robust standard errors. The green and yellow rectangles depict the range of benchmark estimates that we compile from the empirical and theoretical macroeconomic literature. Panel B plots the distribution of responses (with trimmed 5% tails), using kernel density estimators. Both panels pool forecasts for the "rise" and "fall" scenarios. Predictions in the fall scenarios are reversed to make them comparable to rise predictions.

Figure 2.3.2. Forecasts of the quantitative effects of macroeconomic shocks

Oil price shock. Experts mostly agree on the directional response of inflation to an exogenous increase in the oil price, with 84% of experts predicting an increase, 6% expecting no change, and 10% predicting a decrease. There is more disagreement about the unemployment response, with 65% predicting an increase, 16% forecasting no change, and 19% predicting a decrease. Disagreement among households is higher than among experts. Only 71% of households predict an increase in inflation, and only 62% expect an increase in unemployment.

Thus, our data suggest that both experts and households primarily hold the conventional view that an oil shock increases both inflation and unemployment, although this view is more pronounced among experts. In terms of quantitative predictions, both households and experts on average predict positive responses of inflation and unemployment to the oil price shock. The quantitative magnitudes of the average predicted responses are higher among households, but below the benchmarks from the empirical and theoretical literature.[13]

Government spending shock. For the government spending shock, Figure 2.3.1 displays similar levels of disagreement as in the oil vignette among experts, and much higher levels of disagreement among households. The majority of experts predict an increase in inflation (80%) and a decrease in unemployment (80%) in response to a government spending program. Among households, only 55% predict an increase in inflation, while 29% predict a decrease. For the unemployment rate, disagreement among households is even larger: Only 43% expect a decrease in unemployment in response to an increase in government spending, while 39% forecast higher unemployment.

The high level of disagreement about the unemployment response among households is reflected in a muted average predicted response close to zero (-0.03 p.p., see Figure 2.3.2), while experts on average predict a decrease in unemployment by 0.31 p.p. For inflation, households predict an average response of 0.20 p.p., while experts predict a response of 0.26 p.p. The average expert predictions are close to the benchmarks from the empirical and theoretical literature.

Interest rate shock. We uncover substantial disagreement about the effect of an unexpected hike in the federal funds target rate – both within and between the samples of experts and households. 67% of experts predict a decrease in inflation in response to an unexpected interest rate hike and 22% predict an increase. 15% of experts think that the unemployment rate would decrease, whereas 72% predict an increase. Households' beliefs are more dispersed than those of experts. A majority of respondents believe that the inflation rate will increase in response to the interest rate hike (57%), while only 30% expect a decrease. 51% of households predict an increase in unemployment and 33% a decrease.

The differences in qualitative inflation predictions between households and experts are also reflected in their quantitative forecasts: While households on average predict an increase in inflation by 0.17 p.p., experts predict a decrease in inflation by 0.15 p.p.[14] Average predictions about unemployment have the same direction in the two samples but are more muted among households than among experts. Experts' average predictions are close to the empirical benchmarks for both unemployment and inflation.[15]

13. Bordalo, Gennaioli, Ma, et al. (2020) propose a framework to study over- and underreaction of individual and consensus forecasts to news.

14. In Section 2.4.7 we show that only a very small fraction of households seem to misperceive the interest rate hike as the Fed's endogenous reaction to a higher inflation outlook.

15. These patterns also become apparent if we study the predictions of the joint response of inflation and unemployment (see appendix 2.D.2.1). For instance, 55% of experts express the conventional

Tax shock. For the tax shock, we find very similar patterns as for the monetary policy shock. While the view that tax hikes are inflationary is prevalent among households (51%), experts overwhelmingly predict a negative response of inflation (68%). The majority of both households (55%) and experts (69%) expect an increase in unemployment. Again, experts are on average close to the empirical and theoretical benchmarks.

Summary. Taken together, our first main result can be summarized as follows:

Result 1. *There is substantial heterogeneity in forecasts of the effects of macroeconomic shocks, among experts and among households. Average predictions of households and experts are similar in many cases but differ substantially for the inflation response to monetary policy and income tax shocks as well as for the unemployment response to government spending shocks. Disagreement in forecasts in a setting where respondents have comparable information about past realizations of macroeconomic variables indicates an important role for heterogeneity in subjective models in expectation formation.*

2.4 The role of selective recall

What drives the heterogeneity in unemployment and inflation forecasts within and between the household and expert samples? One possibility is that individuals selectively retrieve different propagation mechanisms of the shocks. Selective recall has been shown to be important in shaping people's thoughts and behavior in various contexts (Tversky and Kahneman, 1973; Kahana, 2012; Bordalo, Coffman, Gennaioli, Schwerter, et al., 2020). In our setting, experts may tend to think of textbook models, which account for the full general equilibrium effects of a shock. Households may selectively retrieve specific partial equilibrium effects and propagation channels, for instance driven by their personal experiences. Associations of propagation channels may be strongly context-dependent, as the same individual may recall different memories when confronted with different economic shocks. Moreover, the propagation channels that immediately come to households' minds may not necessarily coincide with the mechanisms that are most central to the transmission of a shock.

To shed light on the role of associations, we conduct additional surveys in which we directly measure respondents' thoughts while they make their predictions. We also implement an experiment that exogenously shifts households' attention to two different propagation mechanisms and allows for a causal analysis of the effect of selective recall of particular propagation channels. Finally, we shed light on the role of personal experiences as a source of households' associations.

---

view that the interest rate shock increases unemployment and decreases inflation, compared to 11% of households.

2.4.1 Samples

Household sample (Wave 3). We recruit a sample of 2,126 respondents in February 2021 in collaboration with the survey company Lucid. Our sample is again broadly representative of the US population in terms of a set of basic demographic variables (see Table 2.B.1).

Expert sample (Wave 3). We identify the email addresses of all economists who published in the top 20 economics journals on JEL code "E: Macroeconomics and Monetary Economics" in the years 2015-2019. We also invite experts from our Wave 1 expert survey and Ph.D. students from 22 leading research institutions (see Appendix 2.J.2 for more details). The expert survey was run in March 2021, shortly after the household survey. In total, 375 experts completed our survey, of which 40% are Ph.D. students (see Appendix Table 2.B.2).

2.4.2 Design

Our design closely follows the main experiment, with some important modifications tailored to measure the thoughts that underlie respondents' predictions. The baseline vignettes are identical to the main survey. However, instead of predicting the level of each rate twice, once in the baseline and once in the shock scenario, respondents directly predict *differences* in each rate between the shock and baseline scenario. This approach allows us to elicit what comes to respondents' minds when they think about the *effect* of a shock. To reduce the cognitive strain of respondents, they indicate their predictions on discrete scales, proceeding in steps of 0.25 p.p. from "1 (or more) p.p. lower" to "1 (or more) p.p. higher". We only collect data on rise-scenarios and each respondent completes only one vignette to keep the collection parsimonious.[16]

Our main object of interest is measuring what people think about while making the prediction. We collect two complementary measures of respondents' associations. First, we ask respondents to tell us about their "main considerations in making the prediction" and about how they "come up with [their] prediction" in an open-text box. This open-response question is placed on the same page as the shock scenario, just below the inflation and unemployment predictions. Second, on the subsequent survey page, we present respondents with a structured list of seven to eight shock-specific propagation channels and ask them to indicate which of these channels – if any – they were thinking about when they made their predictions. For each vignette, we select propagation channels that play a key role in canonical models and chan-

16. We replicate our main results for both the directional and the quantitative predictions (see Appendix Figures 2.A.3 and 2.A.4). This highlights the robustness of our findings across time and to changes in the design, such as the prediction scales or the simultaneous measurement of thoughts.

nels that were frequently mentioned in open-text responses from pilot studies.[17] Because many propagation channels are only meaningful for a specific shock and to avoid mental overload among respondents, the structured questions focus on a different subset of propagation channels in each vignette. For instance, in the oil price vignette, these channels include a reduction in firms' labor demand due to higher production costs and a reduction in households' spending due to lower purchasing power, among others. In the case of the monetary policy vignette, the survey question includes a cost channel, an intertemporal substitution channel, a channel capturing changes in household spending due to changes in income, as well as several other channels. In several parts of our analysis, we focus on groups of those channels, such as negative supply-side mechanisms (e.g. higher production costs for firms) or negative demand-side mechanisms (e.g. reduced household spending due to lower purchasing power). Appendix 2.E provides an overview of the full instructions used in the structured questions on propagation channels.

For ease of exposition, we focus mostly on the structured questions in our main analysis. These structured questions also offer several advantages compared to the open-text questions. First, the responses to the structured questions are straightforward to compare across respondents, while there is likely large variation in the way individuals respond to the open-text questions. Second, the structured questions allow us to measure thoughts of full, clearly defined propagation channels, while this is more difficult with the open-text responses, which are often not sufficiently nuanced. Third, the structured questions require less effort by the respondents, which may result in lower measurement error. Finally, responses to the structured questions do not need to be categorized and interpreted before the analysis, which avoids judgment calls on the part of researchers.

One potential concern is that responses to the structured questions may be prone to ex-post rationalization of forecasts. To address this concern, we also make use of the open-text responses as an additional data source. These responses offer a unique lens into respondents' associations without priming them on any particular propagation channel that could be at play, and should therefore be more immune to ex-post rationalization. We use the open-text responses i) to validate responses to the structured questions, ii) to demonstrate the robustness of our findings, and iii) to capture additional features of thinking not covered by the structured questions (e.g., general equilibrium thinking, mentioning models, and perceived endogeneity of the shock).

COVID-19 pandemic. At the time of the data collection, the coronavirus pandemic was still affecting the US economy. To avoid respondents' thoughts being captured by the COVID-19 pandemic, we ask them to assume that "it is the 1st of January

17. The order of response options is randomized across individuals to address potential order effects.

2025. The COVID-19 pandemic is over. The US economy has fully recovered and is back to 'business as usual'." In particular, we ask our respondents to assume that the inflation rate is at 1.8% and that the unemployment rate is at 3.6% on the 1st of January 2025, similar to our main data collection from February and July 2019.

2.4.3 Results: Propagation mechanisms that come to Mind

Figure 2.4.1 summarizes respondents' thoughts of propagation channels based on the structured questions. We first describe variation of thoughts within the household and within the expert sample, and then discuss differences between the two groups.

Heterogeneity within the household sample. For each of the vignettes, there is a lot of heterogeneity in the thoughts that come to households' minds. Very few of the propagation channels are selected by more than half of the respondents.

How do households' thoughts vary across the different shocks? Supply-side mechanisms related to price increases or layoffs due to higher costs are most frequently mentioned under the oil vignette (about 50% for each). For the interest rate and the income tax shock, which are conventionally seen as demand-side shocks, smaller but still sizable fractions (between 30% and 40%) think of the different negative supply-side channels.

Moreover, many households indicate reductions in product demand due to lower purchasing power or job loss in the oil vignette (about 40% for each channel). By contrast, only 25% of households indicate increases in demand due to higher incomes in the government spending vignette, and only 31% and 27% indicate lower spending due to lower incomes or due to intertemporal substitution in the interest rate vignette, even though these shocks are commonly considered to be classical demand-side shocks.

These patterns are in line with households selectively retrieving specific mechanisms, where the types of mechanisms that are recalled depend on the context. Our evidence also suggests that in many cases households neglect mechanisms that may plausibly play a major role in reality, and that may be useful in forecasting responses of unemployment and inflation.

Heterogeneity within the expert sample. We also observe substantial heterogeneity in the propagation channels experts think of within each of the vignettes. However, the within-vignette variation is smaller than among households, and experts' thoughts tend to be more concentrated in specific channels. This suggests that there is more agreement among experts about which propagation channels are important under each shock.

The variation in experts' thoughts across vignettes largely reflects differences in how the shocks are typically viewed in textbooks. For instance, thoughts of negative supply-side channels associated with increases in production costs are most frequently stated in the oil price vignette (79% and 57% for price increases and reductions in labor demand due to higher costs, respectively). Experts think much less frequently of supply-side channels under the three demand-side shocks (ranging from 5% to 26% for different channels across the three vignettes).

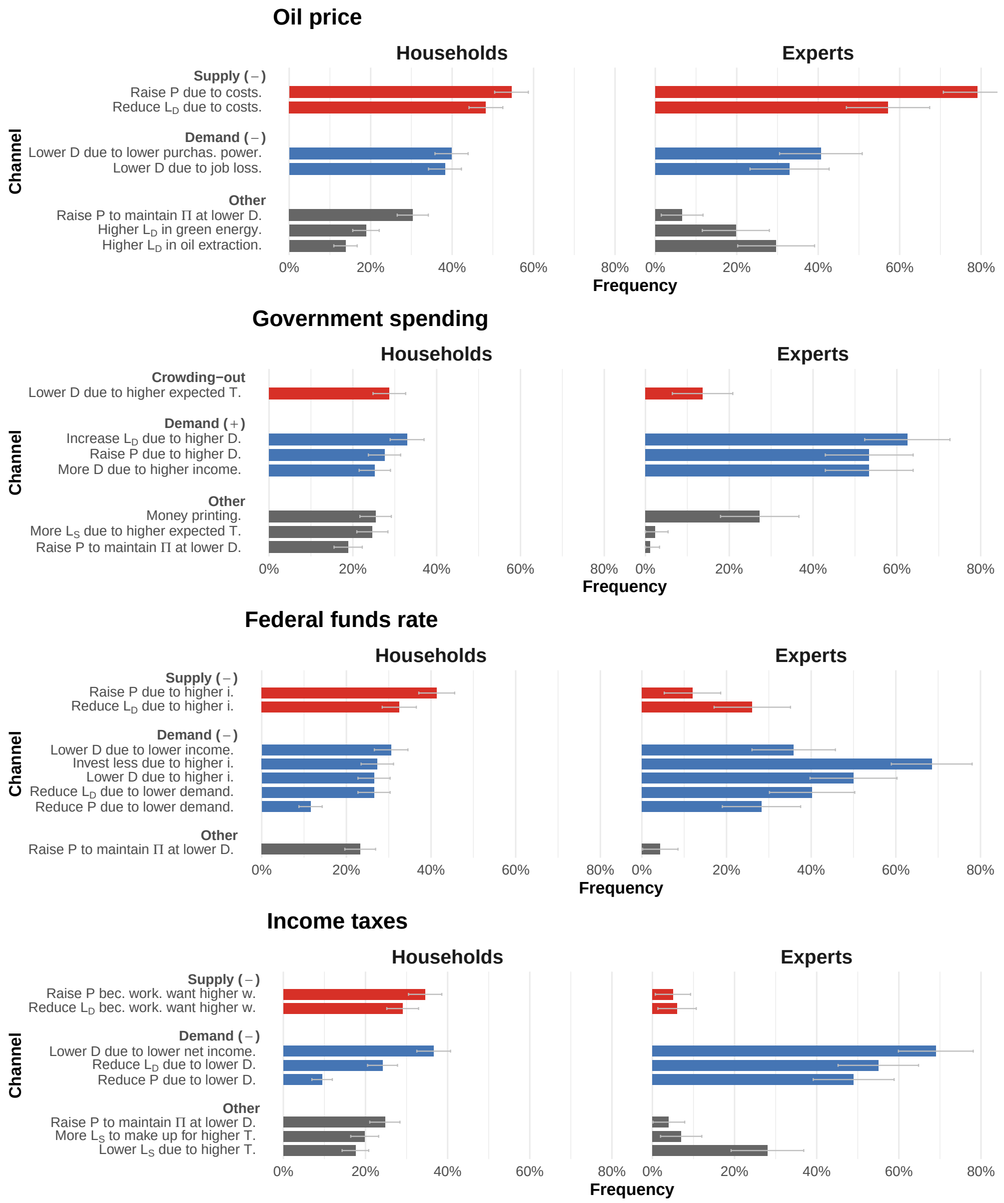
Sizable fractions of experts indicate thoughts of demand reductions under the oil vignette due to lower purchasing power (41%) or job loss (33%), consistent with second-round effects in standard models. Under the three shocks conventionally seen as demand-side ones, even higher fractions select demand-side channels that are prominent in textbook models. For instance, 68% of experts think of a reduction in firms' investment expenditure in response to an interest rate hike, while 50% think of a reduction in household spending due to intertemporal substitution. 53% and 69% of experts indicate changes in household spending due to changes in incomes under the government spending vignette and the tax vignette, respectively.

Overall, the variation in experts' thoughts across vignettes suggests that many experts retrieve textbook models when they are confronted with the different macroeconomic shocks.[18]

Similarities and differences between households and experts. We next compare households' and experts' associations under each of the shocks.

Households and experts think about similar propagation mechanisms in the context of the oil price vignette. In the other three vignettes, however, there are marked differences between households and experts in the propagation mechanisms respondents think about. Most importantly, compared to experts, households tend to attach lower relative importance to demand-side channels and higher relative importance to supply-side mechanisms in the interest rate and income tax vignettes. For instance, in the interest rate vignette, households choose the two supply-side mechanisms – higher costs leading firms to increase prices and to reduce labor demand – more often than any of the channels related to negative demand-side effects. The patterns are reversed among experts. Thus, many households seem to attribute an important role to a cost-channel in the transmission of monetary policy, where firms pass on higher borrowing costs to consumers in the form of higher prices (Barth and Ramey, 2002). Experts' views are much more closely aligned with the common textbook view that interest rate shocks primarily operate through reductions in product demand. To illustrate households' thoughts in the interest rate vignette, Table 2.4.1 provides example responses for households mentioning a cost channel or a demand channel in the open-text response. Similarly, under the income tax vignette, 35% of households indicate propagation channels according to which firms need to raise wages to compensate employees for the higher tax rate and pass the higher cost on

18. Figure 2.A.6 shows that thoughts of the different propagation channels are very similar across different subgroups in the expert sample. For instance, experts that are PhD students think of very similar channels as non-PhD student experts.



Notes: This figure shows which propagation channels are on respondents' minds when they make their predictions, using Wave 3 data. Respondents can select the channels from a list. The results are displayed separately for each vignette and for households (left panel) and experts (right panel). Error bars display 95% confidence intervals. P abbreviates "firm prices", L_D "labor demand", D "product demand", Π "firm profits", T "taxes", i "interest rates", w "wages", and L_S "labor supply". The full wording of the channels is available in Appendix 2.E.

Figure 2.4.1. Thoughts of propagation channels

Table 2.4.1. Associations in the federal funds rate vignette: Examples of households' open-text responses

Thoughts about a cost channel	Thoughts about demand-side channels
"If the cost to borrow funds goes up, then a business will have to pay more to pay back a loan. Thus, businesses will have to raise prices. This will result in inflation. A business may not be able to pay employees and have to let them go or a business will not be able to pay back the load and the business will fail. The employees will lose their jobs and raises unemployment"	"with change in fed funds rate upward, unemployment is likely to rise (as cost to business to borrow increases and invest less in expansion) and inflation should in theory be kept in check and even fall."
"I believe if the fed rate increases, the inflation rate will as well because companies will be paying more on their credit and they will pass that on to consumers. Do not think it will affect unemployment."	"Interest rates rising will increase the cost of investment. This will make companies lay people off. However, with higher interest rates, less money will be invested and it will cause inflation to fall."
"If the Fed rate is increased, the following usually happens – the cost of borrowing money for businesses increases –the business has to raise prices –there is usually a corresponding effect on the unemployment rate as employers find they have to cut staff to remain competitive "	"when the interest rate goes up I believe the unemployment rate goes up as well. Inflation will also hurt the job market. If people are not buying the jobs decrease."
"The higher federal funds rate causes the cost of borrowing to rise. As a result, prices are raised. And employment is lowered to cover cost of borrowing."	"the demand will decrease and the investment will be less then usual also saving will be increased"
"When the interest rate rises that would mean that it would cost more for companies to borrow money and so they would charge more for their products (inflation would go up) and they would not have money to expand and hire more people (unemployment would go up). I really don't know if the exact amounts of the inflation and unemployment rises would be the same as the % that the inflation rate rose but I thought maybe it would."	"With the target rate going up, money will become more expensive to borrow, consumer credit rates will rise. This will cause consumer demand to drop and possibly put people out of work"
"The cost of business goes up so business will try to raise prices to make a profit. Business will try to cut costs by employing fewer workers."	"when interest rates increase there is less spending no new jobs"
	"Interest rate hike will cause less overall spending slightly more unemployment and greater inflation as prices adjust to this rate hike."

Notes: This table displays examples for households' responses to the open-text question, focusing on the monetary policy vignette. The left-hand side focuses on responses explicitly referring to a cost channel and neglecting demand-side mechanisms. The right-hand side focuses on responses pointing to demand-side channels.

to consumers in the structured question, while only 5% of experts think of such a channel.

Moreover, across all vignettes, sizable fractions of households (about 20% to 30%) indicate thoughts that firms react to reductions in demand by *increasing* prices to maintain profit levels – a channel that has no role in standard models, and which is selected by almost none of the experts. Households' positive predicted inflation response to interest rate or income tax hikes – the most striking deviation from experts' forecasts – could thus be partially driven by i) relatively higher attention to supply-side factors, and ii) a different view on how firms adjust their prices in response to changes in product demand.

In the government spending shock, households select channels working through increases in product demand much less frequently than experts (between 25% and 33% among households compared to between 53% and 63% among experts). By contrast, households are almost twice as likely as experts to indicate *reductions* in household spending due to an increase in expected future taxes (29% vs 14%). Together, these patterns could explain households' more muted average prediction about the unemployment response to higher government spending.

Finally, we use the open-ended data to document that experts are more likely to account for general equilibrium effects in their forecasts than households based on two facts. First, 10% and 6% of experts refer to endogenous reactions of the central bank to the oil shock and to the government spending shock, respectively, in the open-text question (see Figure 2.A.7). Virtually none of the households refer to reactions by the Fed to these shocks. Second, 22% of the experts explicitly refer to an economic model (such as the New-Keynesian model), compared to none of the households, suggesting that experts are more likely to think about the shocks through the lens of economic theories. These theories in turn account for general equilibrium effects of the shocks.[19]

Discussion. Taken together, we find strong heterogeneity in the propagation mechanisms respondents think about, both within and between our samples of households and experts. The responses by experts suggest that many experts retrieve textbook models when making their forecasts. These models in turn account for general equilibrium effects of the shocks. Heterogeneity within the expert sample could, for instance, be driven by differences in academic backgrounds or fields of expertise.[20]

19. These findings are in line with participants' responses to a question about the approach they pursued in their forecasts. Figure 2.A.5 shows that 88% of experts report that they drew on their knowledge of economics compared to only 29% of households. This is consistent with the notion that experts are more likely to think about the shocks through the lens of textbook models. In contrast to experts, households are relatively more likely to rely on their memories of past economic events and their gut feeling when making their predictions.

20. Our surveys are not tailored to study the drivers of heterogeneity in associations within the expert sample due to space constraints.

Households frequently choose channels that are less important in textbook models, and often neglect mechanisms that are commonly considered to be central. Their forecasting seems to be based on a patchwork of partial equilibrium responses that strongly differs across contexts and individuals. Households often do not account for second-round effects, such as policy responses, or disagree on their direction, such as for the pricing response of firms to changes in product demand. We explore the role of heterogeneous personal experiences as one driver of differences in associations within the household sample in Section 2.4.6 below.

Taking together the evidence presented above, our second main result is the following:

Result 2. *The propagation channels that are on top of respondents' minds vary systematically within and across our samples of households and experts. Experts tend to recall channels that are central in textbook models, while households in many cases neglect these channels and think of channels that are conventionally seen as less important.*

Robustness: Open-ended responses. We also leverage responses to the open-text question eliciting participants' thoughts on the prediction screen to demonstrate the robustness of our findings to a different measurement technique. First, Appendix Figure 2.A.7 highlights how frequently different word groups are mentioned in the open-ended question across vignettes and samples. While naturally the levels are not comparable between structured and unstructured data of thoughts, we replicate differences between households and experts in terms of the relative importance of different mechanisms. Second, in Online Appendix 2.F, we develop a coding scheme to manually categorize open-ended responses into thoughts of different mechanisms. Each response is independently coded by two coders, with high inter-rater reliability. The hand-coded measures of thoughts are strongly correlated with our main measures based on the structured question (see Tables 2.F.5 and 2.F.6), and are similarly distributed across vignettes (see Figure 2.F.2). These findings validate our measures based on the structured questions and mitigate concerns related to ex-post rationalization of forecasts in the structured questions.

2.4.4 Correlations between associations and predictions

Is heterogeneity in thoughts about propagation channels driving heterogeneity in inflation and unemployment forecasts? Table 2.4.2 shows that the propagation mechanisms selected in the structured question are strongly associated with inflation and unemployment forecasts in both the expert and the household sample across all four vignettes. For presentational convenience, we use dummies indicating whether a respondent selects at least one (positive/negative) demand-side or supply-side channel, respectively.[21]

21. In Appendix 2.F we demonstrate robustness of these correlations to using the hand-coded measures of thoughts based on the open-text data.

Table 2.4.2. Thoughts of propagation channels correlate with predictions

Oil price

	Households		Experts	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)
Supply (–)	0.343***	0.145***	−0.010	0.329***
	(0.047)	(0.046)	(0.081)	(0.063)
Demand (–)	0.069*	0.230***	0.174**	0.165**
	(0.040)	(0.041)	(0.077)	(0.076)
Constant	0.173***	0.090**	0.296***	−0.076*
	(0.044)	(0.042)	(0.065)	(0.046)
Observations	557	557	91	91
R^2	0.113	0.078	0.058	0.150
R^2 (all 7 channel indicators)	0.168	0.214	0.095	0.440

Government spending

	Households		Experts	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)
Crowding-out	0.140***	0.236***	−0.036	0.057
	(0.050)	(0.055)	(0.071)	(0.046)
Demand (+)	−0.067	−0.249***	0.076	−0.299***
	(0.045)	(0.047)	(0.076)	(0.057)
Constant	0.329***	0.080**	0.195***	0.009
	(0.038)	(0.037)	(0.067)	(0.051)
Observations	519	519	88	88
R^2	0.023	0.102	0.014	0.266
R^2 (all 7 channel indicators)	0.062	0.180	0.178	0.438

Table continued on next page.

Table 2.4.2 (continued): Thoughts of propagation channels correlate with predictions

Federal funds target rate

	Households		Experts	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)
Supply (–)	0.188***	0.142***	0.090	−0.094
	(0.041)	(0.044)	(0.075)	(0.063)
Demand (–)	−0.053	0.088**	−0.324***	0.340***
	(0.040)	(0.044)	(0.096)	(0.068)
Constant	0.229***	0.068*	0.068	−0.012
	(0.034)	(0.039)	(0.084)	(0.063)
Observations	520	520	92	92
R^2	0.041	0.032	0.175	0.199
R^2 (all 8 channel indicators)	0.088	0.068	0.167	0.206

Income taxes

	Households		Experts	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)
Supply (–)	0.217***	0.188***	0.018	0.004
	(0.041)	(0.044)	(0.074)	(0.074)
Demand (–)	0.024	0.054	−0.150***	0.212***
	(0.041)	(0.043)	(0.046)	(0.038)
Constant	0.254***	0.130***	−0.035	0.041
	(0.032)	(0.034)	(0.041)	(0.030)
Observations	530	530	100	100
R^2	0.053	0.039	0.095	0.169
R^2 (all 8 channel indicators)	0.128	0.129	0.375	0.277

Notes: This table shows data from Wave 3. It regresses the predicted inflation ($\Delta\pi$) and unemployment (Δu) changes on the propagation channels that were on respondents' minds while they made their predictions (see Figure 2.4.1). Each panel presents results for a different vignette. In each panel, Columns (1) and (2) present results for households, Columns (3) and (4) present results for experts. "Supply (–)" takes value 1 for respondents who choose a negative supply-side propagation channel. "Demand (–)" and "Demand (+)" take value 1 for respondents choosing a negative or positive demand-side propagation channel, respectively. In the government spending vignette, "Crowding-out" takes value 1 for respondents who select the channel that demand falls due to higher expected future taxes (see Figure 2.4.1 for more details). Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Most of the correlational patterns uncovered in Table 2.4.2 go into the expected direction. For example, households thinking of negative supply-side propagation channels expect higher increases of inflation ($p < 0.01$) and unemployment ($p < 0.01$) in response to an oil price shock. Experts choosing supply-side propagation channels also expect higher increases in unemployment ($p < 0.01$) in response to oil price hikes, but do not expect higher levels of inflation. In the context of the government spending shock, we uncover robust negative correlations between choosing propagation channels related to positive demand-side shocks and expected changes in unemployment rates ($p < 0.01$). Among households, we also find a strong positive association between choosing channels related to crowding-out and predicted increases in inflation ($p < 0.01$) in response to a government spending increase, while for experts this association is more muted. For households, we document strong positive associations between choosing supply-related propagation mechanisms and predicted increases in inflation ($p < 0.01$) and unemployment ($p < 0.01$) in response to both an interest rate hike and an increase in income taxes, while for experts these patterns are less pronounced. For experts, on the other hand, we find that choosing demand-related mechanisms is associated with lower inflation ($p < 0.01$) and higher unemployment ($p < 0.01$) predictions in response to both an interest rate and an income tax hike.

Table 2.4.2 illustrates that, across shocks, dummies for thoughts about different propagation channels have significant explanatory power for forecasts. Regressing forecasts on dummies for all vignette-specific channels gives an R-squared between 6% and 21% for households, and between 10% and 44% for experts. These values are sizeable given the low R-squared often documented in studies of the determinants of survey expectations, such as individual characteristics or experiences (Malmendier and Nagel, 2011; Kuchler and Zafar, 2019; Das, Kuhnen, and Nagel, 2020; Giglio et al., 2021). The actual explanatory power of associations is likely even larger than measured in our survey given i) the potential measurement error in associations, ii) the fact that we do not measure the perceived strength of the different channels, and iii) the possibility that we do not capture all relevant channels that respondents have on their minds.

Can differences in associations account for differences in average predictions between households and experts? Table 2.4.3 examines the extent to which the gap in predictions between experts and households can be explained by differences in responses to the structured question on propagation mechanisms. Our analysis zooms in on the three predictions for which the average gap between households and experts is most pronounced. Columns 1 and 2 show that the average differences in unemployment predictions in the government spending vignette are fully explained by differences in the selected propagation mechanisms. Columns 3 and 4 show that the propagation channels explain approximately one third of the gap in inflation predictions in the interest rate vignette. Finally, they explain about one third of the

Table 2.4.3. Thoughts of propagation channels account for differences between experts' and households' predictions

	Government spending Unemployment Δu		**Federal funds rate** Inflation $\Delta\pi$		**Income taxes** Inflation $\Delta\pi$	
	(1)	(2)	(3)	(4)	(5)	(6)
Expert	-0.215^{***}	-0.003	-0.462^{***}	-0.323^{***}	-0.517^{***}	-0.347^{***}
	(0.036)	(0.035)	(0.037)	(0.048)	(0.030)	(0.041)
Constant	0.013	0.040	0.297^{***}	0.207^{***}	0.368^{***}	0.248^{***}
	(0.025)	(0.035)	(0.020)	(0.030)	(0.021)	(0.030)
p_F: Expert coeff. equal		**<0.001**		**<0.001**		**<0.001**
Channels	–	✓	–	✓	–	✓
Observations	608	607	614	612	631	630
R^2	0.020	0.203	0.127	0.199	0.152	0.258

Notes: This table uses data from Wave 3 of the expert and household surveys. It tests whether thoughts of different propagation channels (see Figure 2.4.1) can account for the differences in experts' and households' predicted inflation ($\Delta\pi$) and unemployment (Δu) changes. We consider the three cases for which large differences in experts' and households' predictions can be found: Unemployment in the government spending vignette (columns 1-2), inflation in the federal funds rate vignette (columns 3-4), and inflation in the income tax rate vignette (columns 5-6). "Expert" takes value 1 for respondents from the expert sample. Results in columns (2), (4), and (6) control for the selected propagation channels (7-8 indicators, depending on the vignette, see Figure 2.4.1 for all propagation channels). p-values result from an F-test of equality of the "Expert" coefficient with and without channel controls (estimated using seemingly unrelated regressions). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

prediction gap in the tax vignette (Columns 5 and 6). Taking together the evidence presented above, our third main result is the following:

Result 3. *Thoughts of specific propagation channels are correlated with forecasts of the effects of macroeconomic shocks on inflation and unemployment in the expected directions, and account for part of the differences in forecasts between households and experts.*

One important caveat about our descriptive evidence is that omitted variables could be driving both thoughts of propagation channels and forecasts about unemployment and inflation. To provide evidence of a causal effect of thoughts and selective recall of propagation mechanisms, we conduct an additional experiment, which we discuss in the next subsection.

2.4.5 The causal effect of associations

To shed light on the causal effects of selective retrieval of particular propagation mechanisms on households' inflation and unemployment forecasts, we conduct a simple experiment. We focus on beliefs about the effect of a federal funds rate hike on the inflation rate as this is one of the cases where predictions differ the most between households and experts.[22] Moreover, monetary policy innovations are the most studied type of shock in the theoretical and empirical literature. The experiment aims to provide a proof of concept that an exogenous shift in people's selective retrieval of propagation mechanisms can causally affect their beliefs about the effects of macroeconomic shocks. If an exogenous change in attention to specific aspects of the prediction problem changes respondents' forecasts, this would suggest that individuals do not hold a "fixed" subjective model, but instead form their models "on the fly", depending on the associations triggered by the context.

Sample. We conduct this experiment with a sample of 1,521 respondents provided by Lucid in February 2021 (Wave 4 of the household survey). Our sample is again broadly representative of the US population in terms of a set of basic demographic variables (see Table 2.B.1).

Design. Our design closely follows the descriptive survey on associations, except that it only focuses on inflation expectations and the interest rate vignette (see Figure 2.A.1 for a visual summary). In the experiment, we randomize respondents into one of three treatments: Respondents in the "cost treatment" are asked two additional questions on firms' costs of doing business before making their inflation prediction. First, they are asked whether US firms face higher or lower costs of doing business when the federal funds rate rises. Second, they are asked to describe their main considerations in making their prediction about costs in an open-text box. In the "demand treatment", respondents are asked about the demand for firms' products before they forecast effects on inflation. First, they are asked whether firms face higher or lower demand for their goods and services when the federal funds rate rises. Second, as in the cost treatment, they describe their main considerations in making the prediction about demand in an open-ended question. Respondents in the "control treatment" do not receive any additional prompt before they make their inflation prediction. Respondents in all three groups report in an open-text box what considerations are on their mind while they make their inflation prediction.[23]

At the end of the survey, respondents in the control treatment are asked either the same two additional questions on costs ("cost control group") or the same two

22. Given the nature of attention, focusing on one macroeconomic variable (inflation) gives us more control over the respondents' thoughts while they make their predictions.

23. Appendix Section 2.G provides an overview of the prediction screens across all three treatment arms.

Table 2.4.4. Results of the priming study (households only)

	Word usage (open-text data)		**Inflation prediction**
	Cost-related words	Demand-related words	$\Delta\pi$
	(1)	(2)	(3)
Costs prime	0.086***	0.007	0.021
	(0.023)	(0.020)	(0.031)
Demand prime	−0.021	0.077***	−0.057**
	(0.017)	(0.023)	(0.029)
Constant	0.093***	0.106***	0.366***
	(0.010)	(0.011)	(0.017)
p: Costs = Demand	<0.001	0.007	0.028
Observations	1,521	1,521	1,521
R^2	0.017	0.010	0.004

Notes: This table presents results from the priming study which focuses on the interest rate vignette (Wave 4 of the household survey). “Costs prime” takes value 1 for respondents randomly assigned to be primed on the costs of production. “Demand prime” takes value 1 for respondents randomly assigned to be primed on product demand. Columns (1) and (2) show effects on word usage in the open-text responses, and Column (3) presents the effects on the inflation forecast. The variable “Cost-related words” takes value 1 for responses which include the word (stem) “cost”. “Demand-related words” takes value 1 for responses which use the words or word stems “demand”, “buy”, “purchas”, “invest”, “spend”, and “consum”. $\Delta\pi$ denotes the perceived reaction of the inflation rate. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

additional questions on demand (“demand control group”). This allows us to characterize heterogeneity in beliefs and to study whether the effects of our attention treatments depend on participants’ beliefs about the direction of the effect of the federal funds rate hike on costs or demand.[24]

The purpose of asking respondents to forecast the response of costs or demand to the shock before they make their inflation forecast is to exogenously draw their attention to different propagation channels of the interest rate shock. For instance, if households’ forecasts of a positive inflation response to interest rate hikes are partially driven by relative inattention to demand-side compared to supply-side mechanisms, then our demand treatment should reduce respondents’ inflation forecasts by

24. For this analysis to be valid, beliefs about the directions in which costs and demand change need to be balanced between the treatment and control groups, which we confirm empirically.

increasing their retrieval of demand-side mechanisms. We believe that drawing respondents' attention to a particular mechanism by asking a question on the decision screen is a relatively subtle way of manipulating associations, which mitigates concerns about experimenter demand effects (de Quidt, Haushofer, and Roth, 2018).

Results. We leverage the text data in which respondents describe what is on their mind while making the inflation prediction to shed light on the "first-stage" effects of our treatments on selective retrieval of propagation mechanisms. Columns 1 and 2 of Table 2.4.4 present the effects of the treatments on the words that respondents use to describe their thoughts.[25] Respondents in the "cost treatment" arm are 8.6 p.p. ($p < 0.01$) more likely to use words related to firms' costs (control mean: 9.3%). The demand treatment increases the use of words related to demand by 7.7 p.p. ($p < 0.01$) – a 75% increase compared to the control group mean of 10.6%. There are no spillovers of the cost treatment on the use of demand-related words, or vice versa. The overall small fractions mentioning such words should be viewed in light of the unstructured nature of the open-text data. Taken together, our treatments seem to be successful in drawing respondents' attention to supply-side or demand-side mechanisms, respectively.

We next turn to the effects on respondents' inflation forecasts. Column 3 of Table 2.4.4 shows that while the cost prime increases inflation predictions insignificantly by 0.021 p.p. ($p = 0.50$), the demand prime significantly decreases inflation predictions by 0.057 p.p. ($p < 0.05$). The stronger response of inflation forecasts to the demand treatment could be due to the fact that many households already predict a positive inflation response by default, potentially due to higher attention to supply-side mechanisms. This could limit the scope for further increases in inflation forecasts.

Despite the relatively large first-stage effects on word usage, the effects on inflation forecasts we uncover are relatively small in magnitude. There are at least three potential explanations. First, the effect of attention to changes in costs or product demand on inflation forecasts should depend on respondents' beliefs about the direction of changes in costs or product demand in response to the rate hike. If there is disagreement on the directions of these changes, this will attenuate the average effects of attention to costs or demand on inflation forecasts. Consistent with this conjecture, Table 2.B.6 in the Online Appendix shows that the demand treatment decreases inflation forecasts by 0.10 p.p. ($p < 0.05$) among respondents expecting a decrease in demand, while it has no significant effect among those expecting an increase. Similarly, the cost treatment increases inflation predictions by 0.05 p.p. among respondents who expect an increase in costs ($p = 0.20$), while it de-

25. In Online Appendix 2.F, we show similar patterns using measures of thoughts based on hand-coding of the open-text data. We do not use structured measures as those were not included in this data collection.

creases predictions by 0.15 p.p. among respondents who expect a decrease in costs ($p = 0.14$). Second, even among respondents with beliefs about changes in costs or changes in demand in the same direction, there could be disagreement about the direction of firms' pricing response to a given change in costs or demand. Indeed, as documented in Section 2.4.3, households seem to disagree about the direction in which firms adjust their prices in response to decreases in demand. Such disagreement implies that higher attention to demand or costs shifts different households' inflation forecasts in different directions, which may further attenuate the average effects on inflation forecasts. Third, inattention to the demand- or supply-side may only be part of the story, i.e. people could hold differential beliefs about the importance of demand- and supply-side channels in the transmission. Hence, even if respondents are made attentive to these channels, only part of them might think this is important for inflation.

Taking together the evidence presented above, our fourth main result is the following:

Result 4. *An exogenous shift in attention to demand-side factors has a negative causal effect on households' predicted inflation response to interest rate hikes. The fact that an exogenous change in retrieval of propagation mechanisms of shocks changes households' forecasts suggests that households may not form their expectations based on a fixed subjective model. Instead, individuals may form their subjective models "on the fly", in line with the associations that come to their minds depending on the context.*

This suggests that news or actual events in the economy may systematically affect which models people entertain. Rather than sticking to one particular model, economic agents retrieve specific memories when cued by events, which in turn shape the economic mechanisms they think of.

2.4.6 The role of experiences

A key open question is what determines households' recall of specific propagation channels when they think about macroeconomic shocks. Human memory is known to be associative, selective, and to draw on personal experiences (Kahana, 2012; Bordalo, Gennaioli, and Shleifer, 2017; Enke, Schwerter, and Zimmermann, 2020). Different personal experiences in the memory database should therefore be reflected in differences in associations and forecasts. In this subsection, we use an additional data collection on the government spending vignette among households (Wave 5) and data on the oil price vignette from Wave 3 of the household survey to shed light on this conjecture.

Experiences with the propagation channels of military spending. In an additional data collection (Wave 5 of the household survey, n=486), we collect data on the government spending vignette using identical baseline instructions as in Wave

3.[26] In addition, we include two main sets of variables to gauge the role of personal experiences.

First, we ask respondents to assess their overall experience with the mechanisms that we listed in our structured question on propagation channels, such as an increase in household spending due to higher incomes (see Figure 2.4.1). Respondents rate the extent to which they themselves or their family and friends have been part of each mechanism on a five-point scale ranging from "no experiences" to "a lot of experiences". For the analysis, we compute two summary indices, namely the standardized sum of experiences with positive demand-side channels and a standardized version of experience with "crowding-out" channels. The two indices provide measures of respondents' cumulative first-hand and second-hand experiences with propagation channels.

Second, we also zoom in on a more specific experience by eliciting whether the respondent or anyone among their friends and family members has ever been employed by a company receiving contracts from the US military. This, in turn, allows us to capture one specific way in which a respondent could have direct personal experience with the demand-side mechanisms and, in particular, the potential labor market effects of military spending increases.

Panel A of Table 2.4.5 shows that respondents who indicate to have more experiences with positive demand-side mechanisms are more likely to choose demand channels ($p < 0.01$) and somewhat less likely to choose channels related to crowd-out ($p < 0.10$) in the structured question, and are more likely to mention words related to product demand ($p < 0.10$) and labor demand ($p < 0.10$) in the open-text question. Conversely, respondents who have more experiences with crowd-out channels are more likely to choose propagation channels related to crowd-out ($p < 0.01$) and less likely to choose channels related to demand ($p < 0.01$) in the structured question, and somewhat more likely to mention words related to costs ($p < 0.10$) and less likely to mention words related to labor demand ($p < 0.05$) in the open-text question. These differences in the propagation channels respondents think of are reflected in a more negative predicted unemployment response to the spending program among those with positive demand-side experiences ($p < 0.01$) and a more positive predicted unemployment response among those with crowd-out experiences ($p < 0.01$).

Panel B of Table 2.4.5 shows that respondents who were either personally employed by a company receiving contracts from the US military or have someone among their friends and family members who was employed by such a company are somewhat more likely to choose propagation channels related to demand in the structured question ($p < 0.10$), and are more likely to use words related to labor demand in the open-ended question ($p < 0.01$) when they make their forecasts.

26. Our respondents in this sample are on average somewhat older and more educated compared to our other data collections (see Table 2.B.1).

They also predict a stronger decrease in the unemployment rate in response to the increase in government spending ($p < 0.01$).[27]

Experiences with oil supply shocks. To provide further evidence on the role of personal experiences, we leverage variation in whether respondents lived through the OPEC crisis in the 1970s – a singular and particularly memorable event. Building on prior work by Binder and Makridis (2020), we proxy personal experiences of the 1970s oil crisis with an indicator for whether the respondent was born before 1962 (teenagers by the late 1970s). Given that the oil price shocks of the 1970s are conventionally seen as supply-side shocks, we would expect respondents with personal experiences of the OPEC crisis to be more likely to recall channels related to production cost increases.

Panel C of Table 2.4.5 shows that individuals born before 1962 are indeed more likely to choose propagation channels related to the supply-side ($p < 0.01$) and more likely to use words related to costs ($p < 0.05$) when making predictions in the oil vignette. Consistent with the associations on top of their mind, respondents who experienced the OPEC crisis predict stronger increases in unemployment and inflation ($p < 0.01$) (Panel C of Table 2.4.5).[28]

Our fifth and final result can be summarized as follows:

Result 5. *Personal experiences are correlated with selective recall of specific propagation mechanisms, which is reflected in individuals' beliefs about the effects of macroeconomic shocks.*

Personal experiences typically vary widely across individuals and are hence likely to be a key driver of heterogeneity in associations regarding macroeconomic shocks. At the same time, personal experiences are likely not the only source of households' associations. For instance, individuals could retrieve things they have recently heard in the news, recall things about economics they learned in college or school, or think of the immediate consequences of a shock for themselves.

Table 2.B.8 uses responses to a question on which approaches households followed in making their forecasts (see Figure 2.A.5) to examine how thoughts of different channels vary across different sources of associations. Households that use knowledge of economics in their predictions are more likely to have associations of channels that are important in textbook models, moving their thoughts closer to those of experts. Respondents whose predictions are shaped by their personal situations are more likely to think of demand-side channels, such as changes in household spending, across the different vignettes. Finally, retrieving macroeconomic ex-

27. Table 2.B.7 shows that we obtain similar results using alternative measures of personal employment experience with government suppliers.

28. In Online Appendix 2.F, we show similar patterns for the effect of experiences on associations using measures of thoughts based on hand-coding of the open-text data.

periences or things heard in the news is significantly associated with having more thoughts of both supply-side and demand-side channels in the different vignettes.

Future research could provide more systematic evidence on how personal experiences or media exposure trigger different associations across contexts. Such an exercise could be guided by a model of memory of own experiences that makes predictions on how experiences affect associations across contexts.

2.4.7 Other drivers of forecasts

In the previous subsections, we provided descriptive and causal evidence highlighting that selective recall of different propagation channels is driving heterogeneity in forecasts both between and within our samples of households and experts. In this subsection, we study a range of other factors that could be important for households' unemployment and inflation forecasts, and compare their quantitative importance to the role of thoughts about propagation channels.

Data. In our Wave 3 data collection, we also collect rich data capturing (i) respondents' knowledge of different aspects of the economy, (ii) their beliefs about historical correlations of different macroeconomic variables, (iii) the extent to which they consider knowledge of how the economy works useful for making good economic decisions, (iv) their numeracy, and (v) a range of other background characteristics – all of which are described in further detail below and in Appendix 2.H.

Specifications. To ease presentation, we examine correlates of whether a prediction is benchmark-consistent, that is whether it is directionally aligned with the literature benchmarks, using data from Wave 3 of the household survey.[29] We pool unemployment and inflation forecasts for this exercise. Column 1 of Table 2.4.6 depicts bivariate regression coefficients for different potential determinants (coded as dummy variables, see table notes), while Column 2 shows multivariate regression coefficients. Each coefficient can be interpreted as the increase in probability that a forecast is benchmark-consistent. In the description of the results, we focus on the bivariate regressions, but the patterns are very similar for the multivariate ones.

Thoughts of propagation channels. We start by assessing the role of associations. Table 2.4.6 corroborates our main finding that respondents' selective retrieval of propagation mechanisms affects predictions. When respondents report thinking about a propagation channel that is in line with the benchmark, they are 17 p.p.

29. We have at least one theoretical or empirical benchmark in all cases except for the effects of income tax shocks on inflation. In this case, we rely on the conventional view of income tax shocks as demand-side shocks.

Table 2.4.5. Households' experiences correlate with mechanism associations and forecasts

(A) Government spending: Experience with propagation channels (std. indices)

	Propagation channels		**Word usage** (open-text data)			**Predictions**	
	Crowding-out	Demand (+)	Costs	Demand	Labor	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Exp. crowd.-out	0.093^{***}	-0.081^{***}	0.023^{*}	-0.034	-0.050^{**}	0.004	0.106^{***}
	(0.026)	(0.026)	(0.012)	(0.025)	(0.025)	(0.026)	(0.029)
Exp. demand +	-0.046^{*}	0.107^{***}	0.003	0.044^{*}	0.048^{*}	0.038	-0.109^{***}
	(0.026)	(0.026)	(0.014)	(0.025)	(0.026)	(0.025)	(0.030)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	483	483	483	483	483	483	483
R^2	0.113	0.076	0.038	0.191	0.092	0.142	0.180

(B) Government spending: Ever worked for military supplier (self/friend, binary indicator)

	Propagation channels		**Word usage** (open-text data)			**Predictions**	
	Crowding-out	Demand (+)	Costs	Demand	Labor	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Yes	-0.010	0.081^{*}	-0.005	0.036	0.121^{***}	-0.024	-0.101^{**}
	(0.043)	(0.046)	(0.020)	(0.043)	(0.042)	(0.045)	(0.049)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	483	483	483	483	483	483	483
R^2	0.088	0.050	0.025	0.187	0.098	0.137	0.155

(C) Oil price: Experienced OPEC crisis (born before 1962, binary indicator)

	Propagation channels		**Word usage** (open-text data)			**Predictions**	
	Supply (–)	Demand (–)	Costs	Demand	Labor	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Yes	0.114^{***}	0.036	0.100^{**}	0.039	0.011	0.208^{***}	0.202^{***}
	(0.040)	(0.045)	(0.040)	(0.031)	(0.041)	(0.044)	(0.043)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	521	521	521	521	521	521	521
R^2	0.064	0.040	0.053	0.020	0.026	0.080	0.074

Notes: This table presents results from Wave 3 (Panel C) and Wave 5 (Panel A and B) of the household survey. In Columns (1) and (2), it asks whether respondents who made experiences related to the vignettes think about different propagation mechanisms (binary indicators; see Figure 2.4.1). In Columns (3) to (5), it tests whether respondents with vignette-related experiences use different word (stems) in their open-text responses (binary indicators; "Costs": cost; "Demand": demand, buy, purchas, invest, spend, consum; "Labor": layoff, fire, hire, labor, work, job). In Columns (6) and (7), it tests whether they make different forecasts (inflation: $\Delta\pi$, unemployment: Δu). The right-hand-side experience variable varies across panels. In Panel A, "Experienced crowding-out" and "Experienced demand (+)" are standardized indices of self-rated experiences (familiarity) with crowding-out and positive demand-side channels, respectively. In Panel B, "Yes" is a binary dummy taking value 1 if respondents themselves or friends/family of them ever worked for a company that sells to the US military. In Panel C, "Yes" is a binary dummy taking value 1 if respondents were born before 1962, a proxy that they experienced the OPEC crisis. Control variables comprise age (except for Panel C), log income, inflation and unemployment forecasts in the baseline scenario, as well as binary indicators for gender, college education, being a Republican, having taken an economics course at the college level, and census regions. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

($p < 0.01$) more likely to make a benchmark-consistent prediction.[30] This effect is sizable, given an overall fraction of benchmark-consistent predictions of 48%.

Perceived past correlations. In Section 2.4.6, we showed that personal experiences are correlated with the associations respondents have on their mind when thinking about macroeconomic shocks. Here, we examine how a respondent's perception of the historical correlation between the shock variable (e.g., the oil price) and the prediction variable (e.g., inflation) is related to their forecasts (for details, see Appendix 2.H.1). Table 2.4.6 highlights that respondents who perceive a correlation between the variables that is consistent with the benchmark are 18 p.p. ($p < 0.01$) more likely to make a benchmark-consistent prediction.[31] Thus, we document that perceived experienced joint movements of macroeconomic variables are related to households' forecasts of the effects of macroeconomic shocks – similar to the reduced-form relationship between average experienced realizations of macroeconomic variables and unconditional expectations of these variables documented by previous literature (Malmendier and Nagel, 2011, 2016; Kuchler and Zafar, 2019). This relationship could reflect a direct effect or could partially be driven by associative recall of specific propagation channels, in line with our evidence presented in Section 2.4.6.

Rational inattention. We also examine whether individuals who consider it necessary to be knowledgeable about macroeconomic relationships to make good economic decisions are more likely to make benchmark-consistent forecasts, in line with a premise of rational inattention models (Sims, 2003; Maćkowiak and Wiederholt, 2015; for details see Appendix 2.H.2). Table 2.4.6 shows that there is a small but statistically significant positive association of respondents' perceived importance of understanding the working of the economy with giving benchmark-consistent responses. The table also highlights that an objective measure of respondents' knowledge about the economy is significantly positively correlated with benchmark-consistent forecasts, but the coefficient estimate is comparatively small (see Appendix 2.H.3 for details).

Other correlates of forecasts. Moreover, we find a small but significant effect of cognitive ability as proxied by numeracy skills, which has been shown to be an important driver of inflation expectations (D'Acunto, Hoang, Paloviita, and Weber, 2021). A wide set of basic demographics, such as education, age, income, and respondents' political affiliation, are at most weakly correlated with the tendency to

30. The explanatory variable is based on the structured question on propagation channels. Supply-side propagation channels are in line with the benchmarks in the oil price vignette, while demand-side propagation channels are in line with the benchmarks in the other three vignettes.

31. One concern is that respondents may derive their estimates for historical correlations from their causal understanding of the economy as measured in the vignette forecasts. This could give rise to reverse causality, which would upward-bias the estimated coefficient.

Table 2.4.6. Correlates of benchmark-consistent forecasts (households only)

	Indicator for benchmark-consistent prediction	
	Separate bivariate models	Multivariate model
	(1)	(2)
Consistent channel association	0.170***	0.143***
	(0.015)	(0.015)
Consistent perceived correlation	0.181***	0.157***
	(0.017)	(0.016)
Importance of model (1 if >median)	0.040***	0.015
	(0.015)	(0.015)
Knowledge (1 if >median)	0.077***	0.034**
	(0.015)	(0.015)
Numeracy (1 if >median)	0.062***	0.031**
	(0.015)	(0.014)
Female	−0.026*	−0.008
	(0.015)	(0.014)
Age (1 if >median)	0.057***	0.032**
	(0.015)	(0.014)
College degree	0.022	−0.001
	(0.015)	(0.015)
Income (1 if >median)	0.012	−0.001
	(0.016)	(0.016)
Republican	0.022	0.021
	(0.016)	(0.015)
Mean share of benchmark-consistent pred.	0.480	0.480
Fixed effects	Vignette ⊗ rate	Vignette ⊗ rate
Observations	3,860	3,860
R^2	–	0.237

Notes: This table presents results from Wave 3 of the household survey. It presents the effect of various binary covariates on the likelihood of making inflation or unemployment predictions (pooled) that are consistent with the benchmarks, i.e. directionally aligned with the empirical and theoretical literature benchmarks. Each coefficient can be interpreted as the increase in probability that a forecast is benchmark-consistent. Column (1) shows the results from separate bivariate regressions, while Column (2) shows the results from a multivariate model. "Consistent channel association" takes value 1 if the respondent chooses a channel that suggests a benchmark-consistent prediction (e.g. a negative demand-side channel for the federal funds rate vignette). Likewise, "Consistent perceived correlation" takes value 1 if respondents believe in a past correlation between the shock variable (e.g. oil price) and the target variable (e.g. inflation) that is in line with a benchmark-consistent prediction. "Importance of model" measures respondents' assessment of how important knowledge of the functioning of the economy is to them for making good economic decisions. "Knowledge" measures information about the current state of the economy. "Numeracy" is respondents' score on a numeracy test. "1 if >median" indicates that a variable is binarized and takes value 1 for respondents with an above-median value. We include fixed effects for each vignette-rate combination (e.g. oil-inflation). Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

make benchmark-consistent forecasts of unemployment and inflation.[32,33] Overall, thoughts of propagation channels seem to play a more important role than several other plausible drivers of respondents' forecasts.

Misperceived endogeneity. All vignettes are carefully worded to make clear that the shocks are exogenous. However, there may still be a concern that respondents in the representative sample believe that the shock is endogenous and the result of changing economic conditions. This concern is particularly relevant for the monetary policy vignette since exogenous changes in the federal funds rate are atypical and might be particularly difficult to imagine for respondents. For instance, participants might believe that the higher interest rate indicates that the Fed is reacting to an expected rise in inflation and therefore predict higher inflation. However, the open-text data of Wave 3 suggest that only a small fraction of respondents misperceive the shocks as endogenous. We hand-code whether respondents erroneously mistake a shock as a signal for another change to the US economy. Across vignettes, only 1.5% of respondents falsely interpret a shock as endogenous. Even in the monetary policy vignette, the fraction is negligible (2.9%). Moreover, our main results are robust to excluding the relevant respondents.[34]

2.5 Implications

In this section, we discuss the broader implications of our findings for understanding macroeconomic expectation formation and for modeling choices.

Understanding disagreement in expectations. One of the most well-documented empirical facts on the macroeconomic expectations of households, firms, and experts is that there is a substantial amount of disagreement about the future development of the economy (Mankiw, Reis, and Wolfers, 2003; Coibion and Gorodnichenko, 2012; Dovern, Fritsche, and Slacalek, 2012; Link et al., 2020; Giglio et al., 2021). This evidence is at odds with traditional models of full information and rational expectations. There are two broad views on the origins of disagreement in macroeconomic expectations. The most prominent explanation for belief disagreement brought forward by the theoretical literature is that agents have different information on the current state of the economy, which may be driven by infrequent updating of information sets (Mankiw and Reis, 2002; Reis, 2006) or by noise in

32. We also find no significant political heterogeneity in quantitative forecasts – not even in the government spending vignette (see Table 2.B.9).

33. In Appendix 2.H.4, we discuss the potential of a simple affective heuristic to explain variation in making benchmark-consistent predictions across respondents.

34. If respondents misperceive the interest rate change as endogenous, their predictions should be shaped by their beliefs about how the Fed endogenously responds to changes in inflation or unemployment. We measure these beliefs and show that they cannot explain the patterns in our data (Appendix 2.H.5).

private signals about the economy (Sims, 2003; Woodford, 2003). According to such explanations, if agents have the same information sets, they fully agree on how the economy responds to shocks. In contrast to this view, we document strong heterogeneity in unemployment and inflation forecasts even in a setting where all individuals observe the same shock and hold similar information about current realizations of macroeconomic variables. This finding is more in line with the alternative view that dispersion in expectations is (partially) due to individuals relying on different subjective models of the economy (Bray and Savin, 1986; Marcet and Sargent, 1989; Andrade, Crump, Eusepi, and Moench, 2016; Molavi, 2019; Angeletos, Huo, and Sastry, 2020). Accordingly, economic agents evaluate the same news about the economy through the lens of their own model. Since there is strong heterogeneity in these models, disagreement about the future arises even when agents have comparable information about current realizations of macroeconomic variables and shocks.

Relation to existing theories featuring disagreement about the model. Can existing theories featuring disagreement about the model of the economy explain our findings? For instance, in theories of learning and model misspecification, agents may disagree about structural parameters of the economy, such as the persistence of inflation (Bray and Savin, 1986; Marcet and Sargent, 1989; Orphanides and Williams, 2005; Milani, 2007; Evans and Honkapohja, 2012; Bhandari, Borovicka, and Ho, 2019; Molavi, 2019; Angeletos, Huo, and Sastry, 2020). In models of learning from experience (Malmendier and Nagel, 2016), individuals only use realizations of macroeconomic variables observed during their lifetimes to estimate the data-generating process, leading to disagreement in inflation expectations across cohorts even if everyone observes the same current realization. While heterogeneous beliefs about structural parameters from this literature find support in our results, these models cannot quantitatively account for the large heterogeneity in beliefs about the impact of the shocks we document, including disagreement even about the directional responses to shocks. More importantly, our priming evidence that changes in attention to different aspects of the problem affect forecasts is at odds with these models.[35]

Associative recall and subjective models. Instead, our evidence is consistent with the idea that heterogeneity in macroeconomic expectations is partially due to associative recall of different propagation mechanisms of shocks (Gennaioli and Shleifer,

35. A literature in behavioral macroeconomics has proposed k-level thinking (Farhi and Werning, 2019), a lack of common knowledge (Angeletos and Lian, 2017), or myopia (Gabaix, 2020) in macroeconomic expectation formation to explain muted responses of output and consumption to shocks. These models mostly do not directly speak to disagreement in expectations. Moreover, in models of diagnostic expectations, disagreement arises from economic agents' use of the representativeness heuristic to learn from noisy private signals about the economy (Bordalo, Gennaioli, and Shleifer, 2018; Bordalo, Gennaioli, Ma, et al., 2020). In our survey, we provide agents with identical information.

2010; Bordalo, Coffman, Gennaioli, Schwerter, et al., 2020). In this view, heterogeneity in the models individuals rely on is not fully stable, but depends on what is cued by the context and on individuals' past experiences. Given our evidence, we believe that incorporating associative recall could be a fruitful avenue for macroeconomic modeling.

While formulating a model of associative recall as a driver of heterogeneity in subjective models is beyond the scope of our paper, in Appendix 2.I we compare the predictions of a canonical sticky-information model (Mankiw and Reis, 2002) with those of a basic framework that features heterogeneity in beliefs about the effects of macroeconomic shocks, being agnostic about the sources of heterogeneity in these beliefs. We calibrate both models to the empirical results from our vignettes and show that subjective models can produce either an under- or over-reaction of expectations relative to the true inflation response to shocks, and a rise in disagreement of comparable magnitude to that of the sticky-information model. However, unless some frictions in observation of shocks (i.e., sticky information) are assumed, the subjective models framework cannot explain empirical evidence on the persistence in forecast errors by Coibion and Gorodnichenko (2012). Hence, a subjective models approach does not fully substitute but rather complements information frictions.

Disagreement in more natural settings. How do our findings speak to disagreement in more natural settings? Our vignettes describe different hypothetical shocks with a small number of parameters, including previous realizations of the shock variable, unemployment and inflation, as well as the duration of the shock and the information structure. Real-world macroeconomic shocks likely feature a higher number of relevant parameters. Moreover, the simplifying common knowledge assumption about the duration of changes in government spending or taxes in our vignettes will rarely be fulfilled in the real world. These points suggest that disagreement about the effects of shocks in more natural settings, both among households and among experts, may be even larger than measured in our surveys.

2.6 Conclusion

Using samples of about 6,500 households representative of the US population and samples of about 1,500 experts, we use a new vignette-based approach to measure individuals' subjective models of the economy and investigate their attentional foundations. We document substantial disagreement, even about the directional effects of macroeconomic shocks, both within and between samples of households and experts, in a setting where individuals have similar information about previous realizations of macroeconomic variables. Part of this disagreement seems to be due to selective recall of different propagation mechanisms of the shocks. While experts tend to retrieve textbook models, households often neglect channels that are commonly viewed as central to the transmission of a shock. We confirm a causal role for selective retrieval of specific propagation channels by exogenously shifting households' attention to either supply-side or demand-side channels. Finally, we show that personal experiences are correlated with households' associations about specific propagation channels when they are confronted with the shocks. Our findings highlight selective recall as a new explanation for disagreement in macroeconomic expectations.

We believe that our approach of measuring beliefs about the effects of shocks can be applied to many other questions in macroeconomics. For example, it could be fruitful to apply our approach to other structural shocks that are commonly found to be quantitatively important, such as total factor productivity or sentiment shocks. In addition, we believe that our approach of measuring what is on top of people's mind while they make their predictions is a widely applicable tool that could help to better understand how associations drive belief formation.

Our findings also have several implications for policymakers.[36] First, in recent years policy institutions have made efforts to reach broader groups with their communication to increase the effectiveness of fiscal and monetary policy (Haldane and McMahon, 2018). Such efforts could be less fruitful if households disagree about the direction in which policy shocks affect macroeconomic outcomes. Second, our evidence suggests that the way a policy is communicated – for example, whether demand-side implications rather than supply-side implications are emphasized – could substantially alter its effect on individuals' expectations. Finally, our finding of substantial heterogeneity in households' beliefs about macroeconomic relationships implies a large degree of variation in the effectiveness of monetary policy and fiscal policy in shifting expectations and behavior for different subpopulations of interest.

36. The role of macroeconomic expectations in households' spending decisions is still being debated in the literature. Some studies find a positive association of inflation expectations with consumption (D'Acunto, Hoang, and Weber, 2021), while others document a muted (Bachmann, Berg, and Sims, 2015; Galashin, Kanz, and Perez-Truglia, 2021) or negative relationship (Coibion, Georgarakos, Gorodnichenko, and van Rooij, 2019). The evidence on the role of expectations about aggregate unemployment and growth is more limited, but there is some evidence suggesting a role in households' spending decisions (Roth and Wohlfart, 2020; Coibion, Georgarakos, Gorodnichenko, Kenny, and Weber, 2021).

References

Acosta, Miguel, and Hassan Afrouzi. 2019. "Rationally Confused: On the Aggregate Implications of Information Provision Policies." *Working Paper*, [88]

Afrouzi, Hassan. 2019. "Strategic Inattention, Inflation Dynamics and the Non-Neutrality of Money." *Working Paper*, [88]

Andrade, Philippe, Richard K Crump, Stefano Eusepi, and Emanuel Moench. 2016. "Fundamental Disagreement." *Journal of Monetary Economics* 83: 106–28. [123]

Andre, Peter, and Armin Falk. 2021. "What's Worth Knowing? Economists' Opinions about Economics." *Working Paper*, [89]

Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry. 2020. "Imperfect Macroeconomic Expectations: Evidence and Theory." In *NBER Macroeconomics Annual 2020, volume 35*. NBER Chapters. National Bureau of Economic Research, Inc. [85, 123]

Angeletos, George-Marios, and Chen Lian. 2017. "Dampening General Equilibrium: From Micro to Macro." *Working Paper*, [123]

Angeletos, George-Marios, and Chen Lian. 2018. "Forward Guidance without Common Knowledge." *American Economic Review* 108 (9): 2477–512. [85]

Arias, Jonas E., Dario Caldara, and Juan F. Rubio-Ramírez. 2019. "The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure." *Journal of Monetary Economics* 101: 1–13. [151]

Armantier, Olivier, Wändi Bruine de Bruin, Giorgio Topa, Wilbert Klaauw, and Basit Zafar. 2015. "Inflation Expectations and Behavior: Do Survey Respondents Act on their Beliefs?" *International Economic Review* 56 (2): 505–36. [88]

Armantier, Olivier, Scott Nelson, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar. 2016. "The Price Is Right: Updating Inflation Expectations in a Randomized Price Information Experiment." *Review of Economics and Statistics* 98 (3): 503–23. [88]

Armona, Luis, Andreas Fuster, and Basit Zafar. 2018. "Home Price Expectations and Behaviour: Evidence from a Randomized Information Experiment." *Review of Economic Studies* 86 (4): 1371–410. [88]

Auerbach, Alan J, Yuriy Gorodnichenko, and Daniel Murphy. 2020. "Local Fiscal Multipliers And Fiscal Spillovers in the United States." *IMF Economic Review* 68: 195–229. [94, 150]

Auerbach, Alan J., and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4 (2): 1–27. [150, 151]

Bachmann, Rüdiger, Tim O Berg, and Eric R Sims. 2015. "Inflation Expectations and Readiness to Spend: Cross-Sectional Evidence." *American Economic Journal: Economic Policy* 7: 1–35. [88, 125]

Bachmann, Rüdiger, Kai Carstensen, Stefan Lautenbacher, and Martin Schneider. 2021. "Uncertainty and Change: Survey Evidence of Firms' Subjective Beliefs." *Working Paper*, [89]

Balke, Nathan S., and Stephen P.A. Brown. 2018. "Oil Supply Shocks and the U.S. Economy: An Estimated DSGE Model." *Energy Policy* 116 (February): 357–72. [95, 150]

Ball, L, D Leigh, and P Loungani. 2017. "Okun's Law: Fit at 50?" *Journal of Money, Credit and Banking* 49 (7): 1413–41. [149]

Ball, Laurence, N Gregory Mankiw, and Ricardo Reis. 2005. "Monetary Policy for Inattentive Economies." *Journal of Monetary Economics* 52 (4): 703–25. [85]

Barth, Marvin J III, and Valerie Ramey. 2002. "The Cost Channel of Monetary Transmission." In *NBER Macroeconomics Annual 2001*. Edited by Ben Bernanke and Kenneth Rogoff. Cambridge, MA: MIT Press, 199–240. [103]

Basso, Henrique S., and Omar Rachedi. 2019. "The Young, the Old, and the Government: Demographics and Fiscal Multipliers." *Working Paper*, [94, 150]

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz. 2005. "Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics* 120 (1): 387–422. [151]

Bernanke, Ben S., and Ilian Mihov. 1998. "Measuring Monetary Policy Expectations." *Quarterly Journal of Economics* 113 (3): 869–902. [151]

Bhandari, A., J. Borovicka, and P. Ho. 2019. "Survey Data and Subjective Beliefs in Business Cycle Models." *Federal Reserve Bank of Richmon Working Paper Series*, (19-14): [123]

Binder, Carola, and Christos Makridis. 2020. "Stuck in the Seventies: Gas Prices and Macroeconomic Expectations." *Review of Economics and Statistics*, [88, 117]

Binder, Carola, and Alex Rodrigue. 2018. "Household Informedness and Long-Run Inflation Expectations: Experimental Evidence." *Southern Economic Journal* 85 (2): 580–98. [88]

Blanchard, Olivier, and Jordi Galí. 2010. "The Macroeconomic Effects of Oil Price Shocks: Why Are the 2000s so Different from the 1970s? Chapter." In *International Dimensions of Monetary Policy*. [94, 149, 150]

Blanchard, Olivier, and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117 (4): 1329–68. [150–152]

Bodenstein, Martin, Christopher J. Erceg, and Luca Guerrieri. 2011. "Oil Shocks and External Adjustment." *Journal of International Economics* 83 (2): 168–84. [94, 150]

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, Frederik Schwerter, and Andrei Shleifer. 2020. "Memory and Representativeness." *Psychological Review* 128 (1): 71. [86, 89, 99, 124]

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *Quarterly Journal Economics* 131 (4): 1753–94. [89]

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *American Economic Review* 110 (9): 2748–82. [88, 98, 123]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2017. "Memory, Attention, and Choice." *Quarterly Journal of Economics*, [115]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2018. "Diagnostic Expectations and Credit Cycles." *Journal of Finance* 73 (1): 199–227. [88, 123]

Boumans, Dorine, and Johanna Garnitz. 2017. "Ifo World Economic Survey Database–An International Economic Expert Survey." *Jahrbücher für Nationalökonomie und Statistik* 237 (1): 71–80. [90]

Bray, Margaret M, and Nathan E Savin. 1986. "Rational Expectations Equilibria, Learning, and Model Specification." *Econometrica* 54: 1129–60. [85, 123]

Carroll, Christopher D. 2003. "Macroeconomic Expectations of Households and Professional Forecasters." *Quarterly Journal of Economics* 118 (1): 269–98. [211, 213]

Carvalho, Carlos, and Fernanda Nechio. 2014. "Do People Understand Monetary Policy?" *Journal of Monetary Economics* 66: 108–23. [88]

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia. 2017. "Inflation Expectations, Learning and Supermarket Prices: Evidence from Field Experiments." *American Economic Journal: Macroeconomics* 9 (3): 1–35. [88]

Christelis, Dimitris, Dimitris Georgarakos, Tullio Jappelli, Luigi Pistaferri, and Maarten Van Rooij. 2019. "Asymmetric Consumption Effects of Transitory Income Shocks." *Economic Journal* 129 (622): 2322–41. [91]

Christiano, L J, M Eichenbaum, and C Evans. 1999. "Monetary Policy Shocks: What Have We Learned and to What End? Handbook of Macroeconomics." *Handbook of Macroeconomics* A: [151]

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, Geoff Kenny, and Michael Weber. 2021. "The Effect of Macroeconomic Uncertainty on Household Spending." *Working Paper*, [125]

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Maarten van Rooij. 2019. "How Does Consumption Respond to News about Inflation? Field Evidence from a Randomized Control Trial." *National Bureau of Economic Research Working Paper*, [88, 125]

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What Can Survey Forecasts Tell us about Information Rigidities?" *Journal of Political Economy* 120 (1): 116–59. [85, 88, 122, 124, 210, 213]

Coibion, Olivier, and Yuriy Gorodnichenko. 2015a. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review* 105 (8): 2644–78. [85, 88]

Coibion, Olivier, and Yuriy Gorodnichenko. 2015b. "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation." *American Economic Journal: Macroeconomics* 7 (1): 197–232. [88]

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar. 2018. "How do firms form their expectations? new survey evidence." *American Economic Review* 108 (9): 2671–713. [88]

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele. 2020. "Inflation Expectations and Firm Decisions: New Causal Evidence." *Quarterly Journal of Economics* 135 (1): 165–219. [88]

D'Acunto, Francesco, Daniel Hoang, Maritta Paloviita, and Michael Weber. 2021. "IQ, Expectations, and Choice." *Review of Economic Studies*, [120]

D'Acunto, Francesco, Daniel Hoang, and Michael Weber. 2021. "Managing Households' Expectations with Simple Economic Policies." *Review of Financial Studies*, [125]

Das, Sreyoshi, Camelia M Kuhnen, and Stefan Nagel. 2020. "Socioeconomic Status and Macroeconomic Expectations." *Review of Financial Studies* 33 (1): 395–432. [110]

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." *American Economic Review* 108 (11): 3266–302. [114]

Delavande, Adeline, and Basit Zafar. 2019. "University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints." *Journal of Political Economy* 127 (5): 2343–93. [91]

DellaVigna, Stefano, and Devin Pope. 2018. "What Motivates Effort? Evidence and Expert Forecasts." *Review of Economic Studies* 85 (2): 1029–69. [89]

Dovern, Jonas, Ulrich Fritsche, and Jiri Slacalek. 2012. "Disagreement Among Forecasters in G7 Countries." *Review of Economics and Statistics* 94 (4): 1081–96. [85, 122]

Dräger, Lena, Michael J Lamla, and Damjan Pfajfar. 2016. "Are Survey Expectations Theory-consistent? The Role of Central Bank Communication and News." *European Economic Review* 85: 84–111. [88]

Enke, Benjamin, Frederik Schwerter, and Florian Zimmermann. 2020. "Associative Memory and Belief Formation." *Working Paper*, [89, 115]

Evans, George W, and Seppo Honkapohja. 2012. *Learning and Expectations in Macroeconomics*. Princeton University Press. [123]

Farhi, Emmanuel, and Iván Werning. 2019. "Monetary policy, Bounded Rationality, and Incomplete Markets." *American Economic Review* 109 (11): 3887–928. [123]

Favero, Carlo, and Francesco Giavazzi. 2012. "Measuring Tax Multipliers: The Narrative Method in Fiscal VARs." *American Economic Journal: Economic Policy* 4 (2): 69–94. [152]

Fuster, Andreas, Benjamin Hebert, and David Laibson. 2012. "Natural Expectations, Macroeconomic Dynamics, and Asset Pricing." *NBER Macroeconomics Annual* 26 (1): 1–48. [88]

Fuster, Andreas, Greg Kaplan, and Basit Zafar. 2020. "What Would You Do With $500? Spending Responses to Gains, Losses, News and Loans." *Review of Economic Studies*, rdaa076. [91]

Fuster, Andreas, David Laibson, and Brock Mendel. 2010. "Natural Expectations and Macroeconomic Fluctuations." *Journal of Economic Perspectives* 24 (4): 67–84. [88]

Fuster, Andreas, Ricardo Perez-Truglia, Mirko Wiederholt, and Basit Zafar. 2020. "Expectations with Endogenous Information Acquisition: An Experimental Investigation." *Review of Economics and Statistics, forthcoming*, 1–54. [88]

Gabaix, Xavier. 2019. "Behavioral inattention." In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2, Elsevier, 261–343. [89]

Gabaix, Xavier. 2020. "A Behavioral New Keynesian Model." *American Economic Review* 110 (8): 2271–327. [123]

Galashin, Misha, Martin Kanz, and Ricardo Perez-Truglia. 2021. "Macroeconomic Expectations and Credit Card Spending." *Working Paper*, [125]

Galí, Jordi, Frank Smets, and Rafael Wouters. 2011. "Unemployment in an Estimated New Keynesian Model." *NBER Macroeconomics Annual* 26 (1): 329360. [149–151]

Gennaioli, Nicola, and Andrei Shleifer. 2010. "What Comes to Mind." *Quarterly Journal of Economics* 125 (4): 1399–433. [89, 123]

Gigerenzer, Gerd, and Peter M Todd. 1999. *Simple Heuristics that Make Us Smart*. Oxford University Press, USA. [203]

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus. 2021. "Five Facts about Beliefs and Portfolios." *American Economic Review* 111 (5): 1481–522. [85, 110, 122]

Goldfayn-Frank, Olga, and Johannes Wohlfart. 2020. "Expectation Formation in a New Environment: Evidence from the German Reunification." *Journal of Monetary Economics* 115: 301–20. [88]

Gordon, Roger, and Gordon B. Dahl. 2013. "Views among Economists: Professional Consensus or Point-Counterpoint?" *American Economic Review* 103 (3): 629–35. [89]

Graeber, Thomas. 2021. "Inattentive Inference." *Working Paper*, [89]

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2021. "Designing Information Provision Experiments." *Journal of Economic Literature*, [90]

Haldane, Andrew, and Michael McMahon. 2018. "Central Bank Communications and the General Public." In *AEA Papers and Proceedings*. Vol. 108, 578–83. [125]

Kahana, Michael Jacob. 2012. *Foundations of Human Memory*. Oxford University Press. [86, 99, 115]

Kamdar, Rupal. 2019. "The Inattentive Consumer: Sentiment and Expectations." [88, 203]

Kuchler, Theresa, and Basit Zafar. 2019. "Personal Experiences and Expectations about Aggregate Outcomes." *Journal of Finance* 74 (5): 2491–542. [88, 89, 110, 120]

Lacetera, Nicola, Devin G Pope, and Justin R Sydnor. 2012. "Heuristic thinking and limited attention in the car market." *American Economic Review* 102 (5): 2206–36. [89]

Leiser, David, and Ronen Aroch. 2009. "Lay Understanding of Macroeconomic Causation: The Good-begets-good Heuristic." *Applied Psychology* 58 (3): 370–84. [203]

Leiser, David, and Zeev Krill. 2017. "How Laypeople Understand the Economy." *Economic Psychology*, 139–54. [203]

Link, Sebastian, Andreas Peichl, Christopher Roth, and Johannes Wohlfart. 2020. "Information Frictions Among Firms and Households." *Working Paper*, [85, 122]

Maćkowiak, Bartosz, and Mirko Wiederholt. 2015. "Business Cycle Dynamics under Rational Inattention." *Review of Economic Studies* 82 (4): 1502–32. [120, 203]

Malmendier, Ulrike, and Stefan Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk-taking?" *Quarterly Journal of Economics* 126 (1): 373–416. [88, 89, 110, 120]

Malmendier, Ulrike, and Stefan Nagel. 2016. "Learning from Inflation Experiences." *Quarterly Journal of Economics* 131 (1): 53–87. [88, 89, 120, 123]

Malmendier, Ulrike, Stefan Nagel, and Zhen Yan. 2021. "The Making of Hawks and Doves." *Journal of Monetary Economics* 117: 19–42. [89]

Mankiw, N Gregory, and Ricardo Reis. 2002. "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics* 117 (4): 1295–328. [85, 122, 124, 209, 210]

Mankiw, N Gregory, Ricardo Reis, and Justin Wolfers. 2003. "Disagreement About Inflation Expectations." *NBER Macroeconomics Annual* 18: 209–48. [85, 88, 122]

Marcet, Albert, and Thomas J Sargent. 1989. "Convergence of Least Squares Learning Mechanisms in Self-referential Linear Stochastic Models." *Journal of Economic Theory* 48 (2): 337–68. [85, 123]

Mertens, Karel, and Morten O Ravn. 2012. "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks." *American Economic Journal: Economic Policy* 4 (2): 145–81. [152]

Mertens, Karel, and Morten O. Ravn. 2014. "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers." *Journal of Monetary Economics* 68 (dec): S1–S19. [152]

Milani, Fabio. 2007. "Expectations, Learning and Macroeconomic Persistence." *Journal of Monetary Economics* 54 (7): 2065–82. [123]

Molavi, Pooya. 2019. "Macroeconomics with Learning and Misspecification: A General Theory and Applications." *Working Paper*, [85, 123]

Nakamura, Emi, and Jòn Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104 (3): 753–92. [94, 150]

Orphanides, Athanasios, and John C Williams. 2005. "The Decline of Activist Stabilization Policy: Natural Rate Misperceptions, Learning, and Expectations." *Journal of Economic Dynamics and Control* 29 (11): 1927–50. [123]

Paciello, Luigi, and Mirko Wiederholt. 2014. "Exogenous Information, Endogenous Information, and Optimal Monetary Policy." *Review of Economic Studies* 81 (1): 356–88. [85]

Perotti, Roberto. 2012. "The Effects of Tax Shocks on Output: Not so Large, but Not Small Either." *American Economic Journal: Economic Policy* 4 (2): 214–37. [152]

Primiceri, Giorgio E. 2005. "Time Varying Structural Vector Autoregressions and Monetary Policy." *Review of Economic Studies* 72: 821–52. [151]

Ramey, Valerie A. 2011. "Can Government Spending Stimulate the Economy?" *Journal of Economic Literature* 49 (3): 673–85. [150, 151]

Reis, Ricardo. 2006. "Inattentive Consumers." *Journal of Monetary Economics* 53 (8): 1761–800. [85, 122]

Romer, Christina D, and David H Romer. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 94 (4): 1055–84. [151]

Romer, Christina D, and David H Romer. 2010. "The Macroeconomic Effects of Tax Changes : Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100 (3): 763–801. [152]

Roth, Christopher, Sonja Settele, and Johannes Wohlfart. 2021. "Risk Exposure and Acquisition of Macroeconomic Information." *American Economic Review: Insights, forthcoming*, [89]

Roth, Christopher, and Johannes Wohlfart. 2020. "How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior?" *Review of Economics and Statistics* 102 (4): 731–48. [88, 125]

Sapienza, Paola, and Luigi Zingales. 2013. "Economic Experts versus Average Americans." *American Economic Review* 103 (3): 636–42. [89]

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90. [120, 123, 203]

Smets, Frank, and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97 (3): 586–606. [149–151]

Stantcheva, Stefanie. 2020. "Understanding tax policy: How do people reason?" Working paper. National Bureau of Economic Research. [89]

Stock, James H., and Mark W Watson. 2001. "Vector Autoregressions." *Journal of Economic Perspectives* 15: 101–16. [151]

Tversky, Amos, and Daniel Kahneman. 1973. "Availability: A heuristic for judging frequency and probability." *Cognitive psychology* 5 (2): 207–32. [86, 99]

Uhlig, Harald. 2005. "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure." *Journal of Monetary Economics* 52 (2): 381–419. [151]

Vellekoop, Nathanael, and Mirko Wiederholt. 2019. "Inflation Expectations and Choices of Households." *Working Paper*, [88]

Wiswall, Matthew, and Basit Zafar. 2017. "Preference for the Workplace, Investment in Human Capital, and Gender." *Quarterly Journal of Economics* 133 (1): 457–507. [91]

Woodford, Michael. 2003. "Imperfect Common Knowledge and The Effects of Monetary Policy." In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*. Edited by Philippe Aghion, Roman Frydman, Joseph E. Stiglitz, and Michael Woodford. Princeton, NJ: Princeton Univ. Press. [123]

Zubairy, Sarah. 2014. "On Fiscal Multipliers: Estimates from a Medium Scale DSGE Model." *International Economic Review* 55 (1): 169–95. [150–152]

Appendix 2.A Additional figures

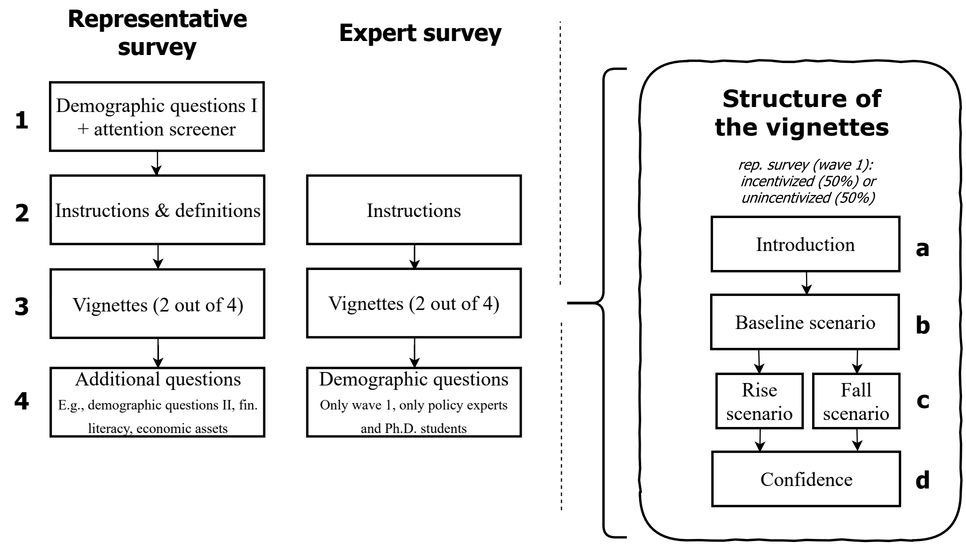


Figure 2.A.1. Main survey (Waves 1 and 2). Overview of the survey structure and the structure of the vignettes

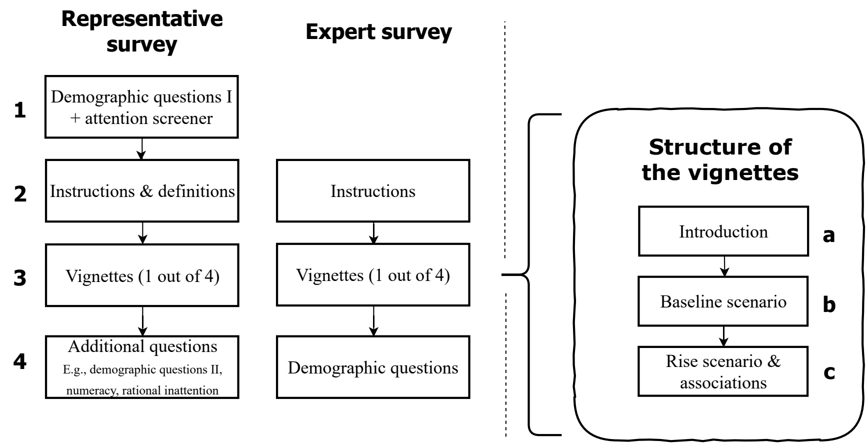


Figure 2.A.1 (continued): **Associations survey** (Wave 3). Overview of the survey structure and the structure of the vignettes

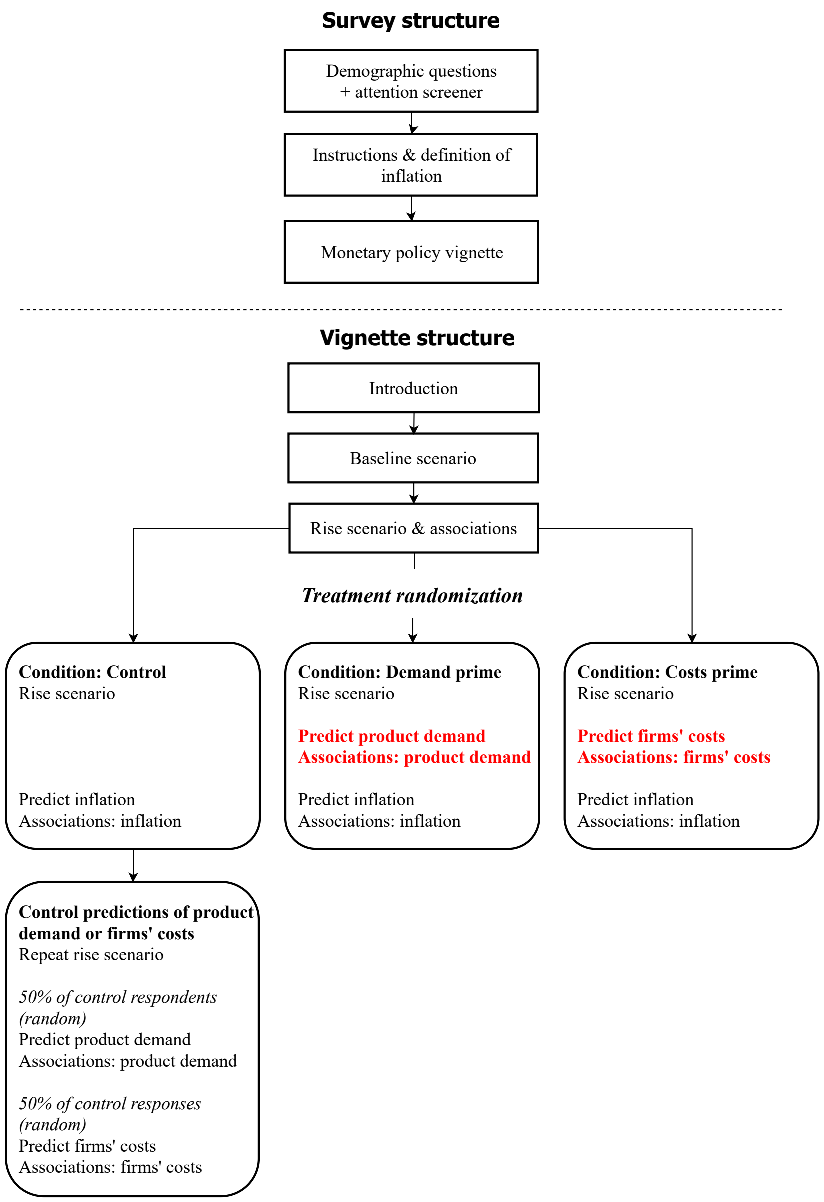


Figure 2.A.1 (continued): **Priming study** (households only, Wave 4). Overview of the survey structure and the structure of the vignettes and treatments

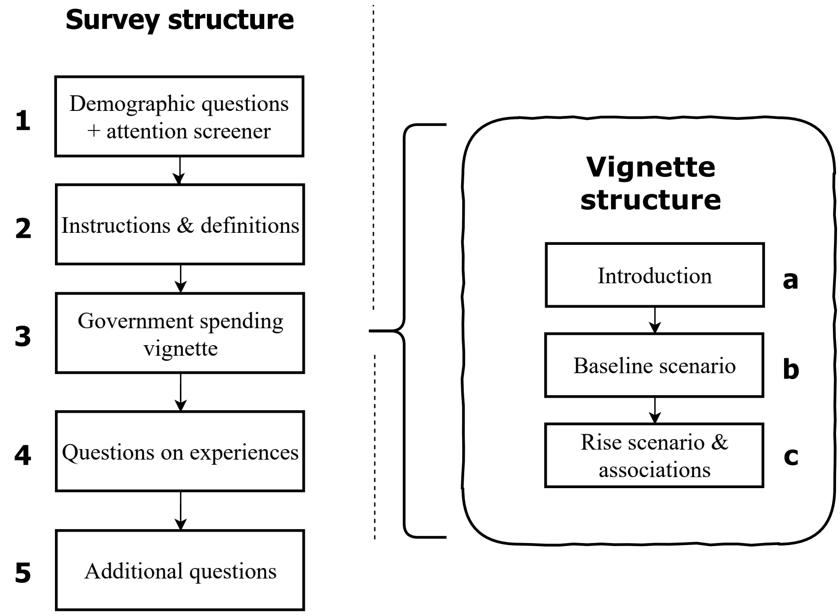
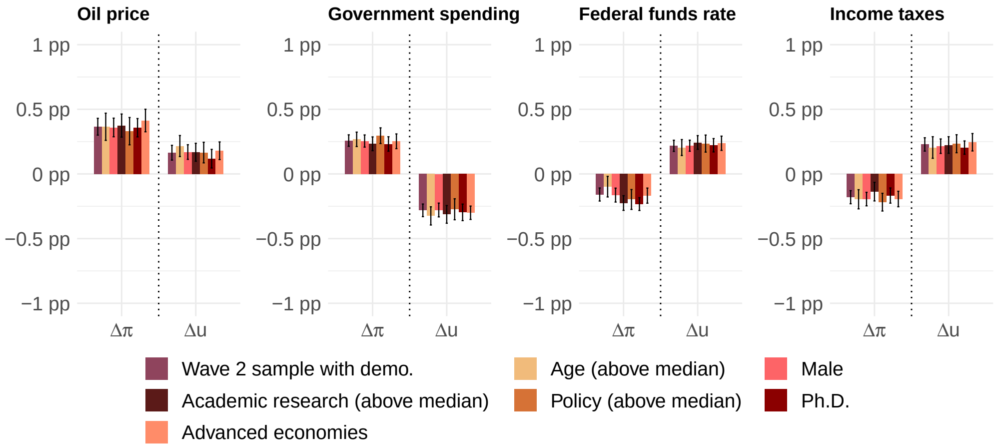
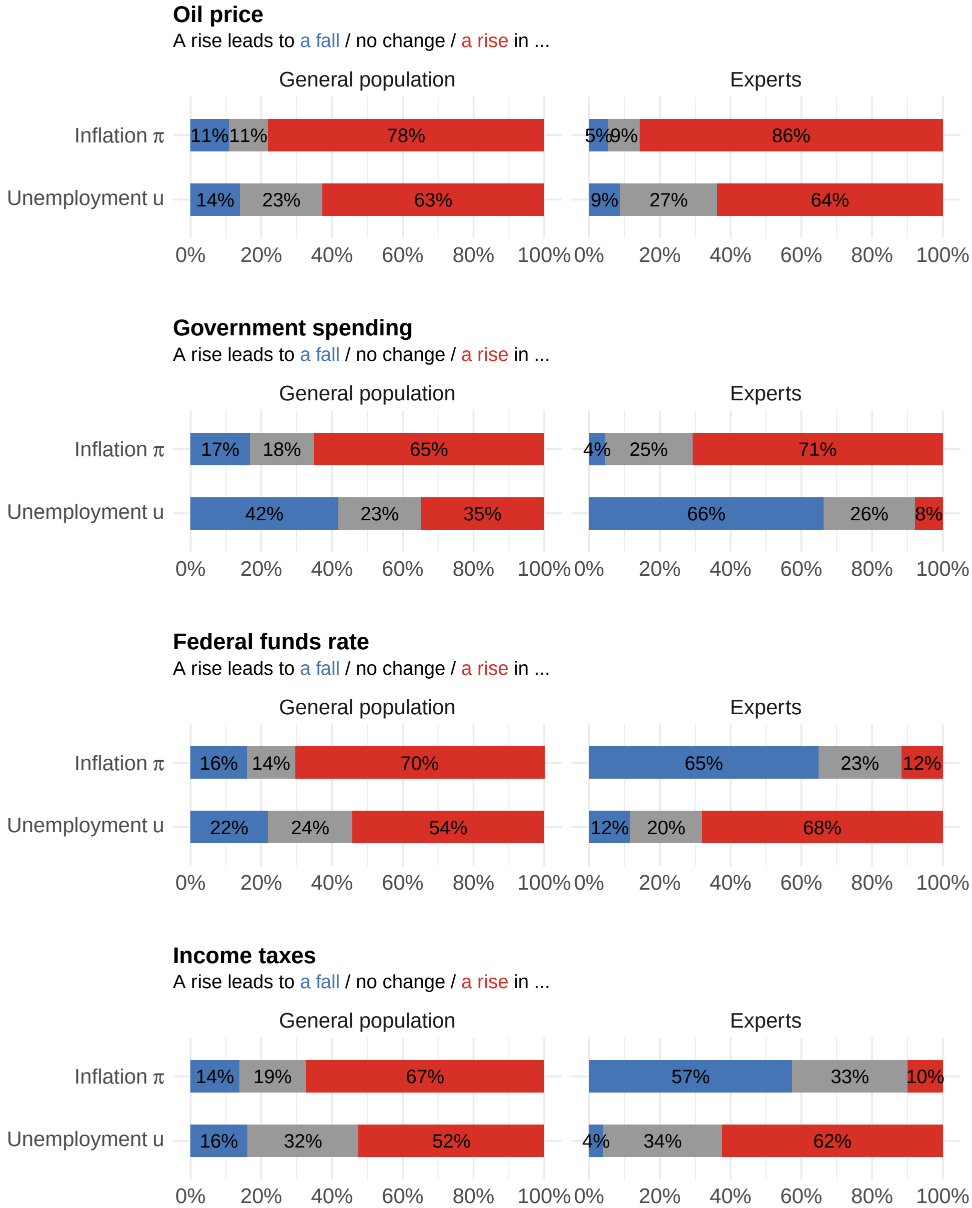


Figure 2.A.1 (continued): **Experience survey** (households only, Wave 5). Overview of the survey structure and the structure of the vignettes



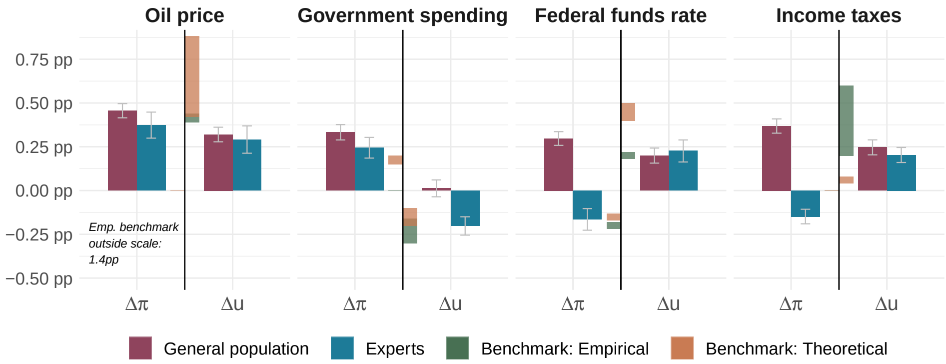
Notes: This figure shows the stability of expert forecasts across various subsamples of the expert Wave 2 sample. It repeats the main analysis for different subsamples and plots expected changes in the unemployment rate (Δu) and the inflation rate ($\Delta\pi$) for each of the different vignettes separately. Predictions in the fall scenarios are reversed to render them comparable to rise predictions. Error bars show the 95% confidence intervals. "Wave 2 sample with demo." denotes the full sample for which background data are available ($n = 596$). "Age (above-median)" contains only respondents with above-median age. "Male" contains only male respondents. "Academic research (ab.-median)" focuses on respondents that spend an above-median percentage of their working time on academic research, while "Policy (ab.-median)" restricts the sample to those who do an above-median amount of policy work. "Ph.D." contains only respondents with a Ph.D., and "Advanced economies" contains only respondents that are registered at the WES to make forecasts about an advanced economy (as classified by the IMF).

Figure 2.A.2. Robustness of experts' forecasts across different subsamples



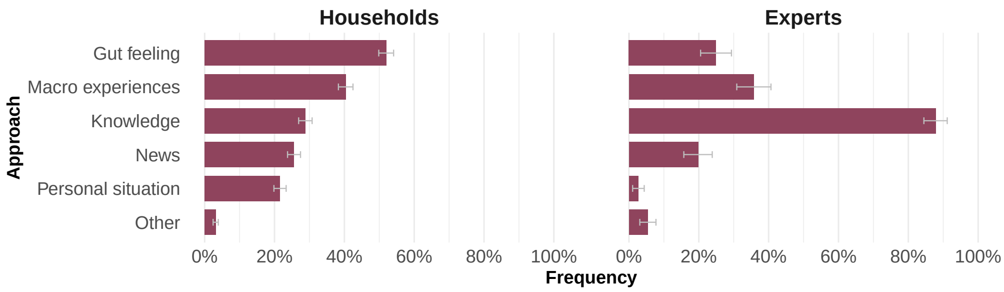
Notes: This figure presents the forecasts of the directional effects of macroeconomic shocks on the inflation rate and the unemployment rate, using Wave 3 data. It compares the forecasts of the general population (left column) to those of experts (right column). Predictions in the fall scenarios are reversed to render them comparable to rise predictions.

Figure 2.A.3. Wave 3: Forecasts of the directional effects of macroeconomic shocks



Notes: This figure displays the average forecasts of the effects of macroeconomic shocks on the inflation rate ($\Delta\pi$) and the unemployment rate (Δu), using Wave 3 data. It compares responses in the representative sample (red bars) with those of experts (blue bars). Error bars present 95% confidence intervals, using robust standard errors. The green and yellow rectangles depict the range of benchmark estimates that we compile from the empirical and theoretical macroeconomic literature. The figure pools forecasts for the "rise" and "fall" scenarios. Predictions in the fall scenarios are reversed to render them comparable to rise predictions.

Figure 2.A.4. Wave 3: Forecasts of the quantitative effects of macroeconomic shocks



Notes: This figure presents the prediction approaches adopted by households and experts in Wave 3, averaged across all four vignettes. In a multiple response question, respondents report which factors they thought most about when making their predictions. Error bars display 95% confidence intervals. "Gut feeling" denotes respondents choosing "I simply responded based on my gut feeling." "Macro experiences": "My memories of economic events in the past." "Knowledge": "My knowledge of economics." "News": "Things I read or heard in the news." "Personal situation": "My personal economic situation today."

Figure 2.A.5. Prediction approaches by households and experts

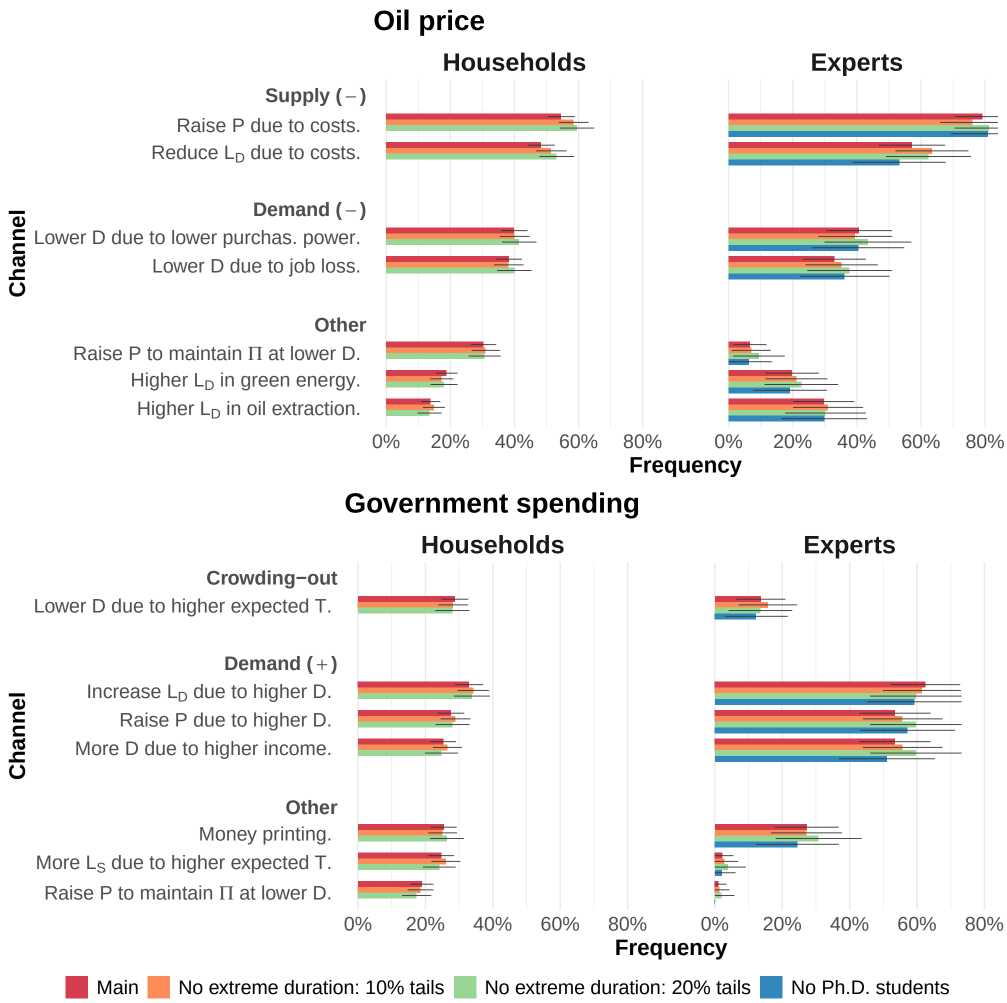
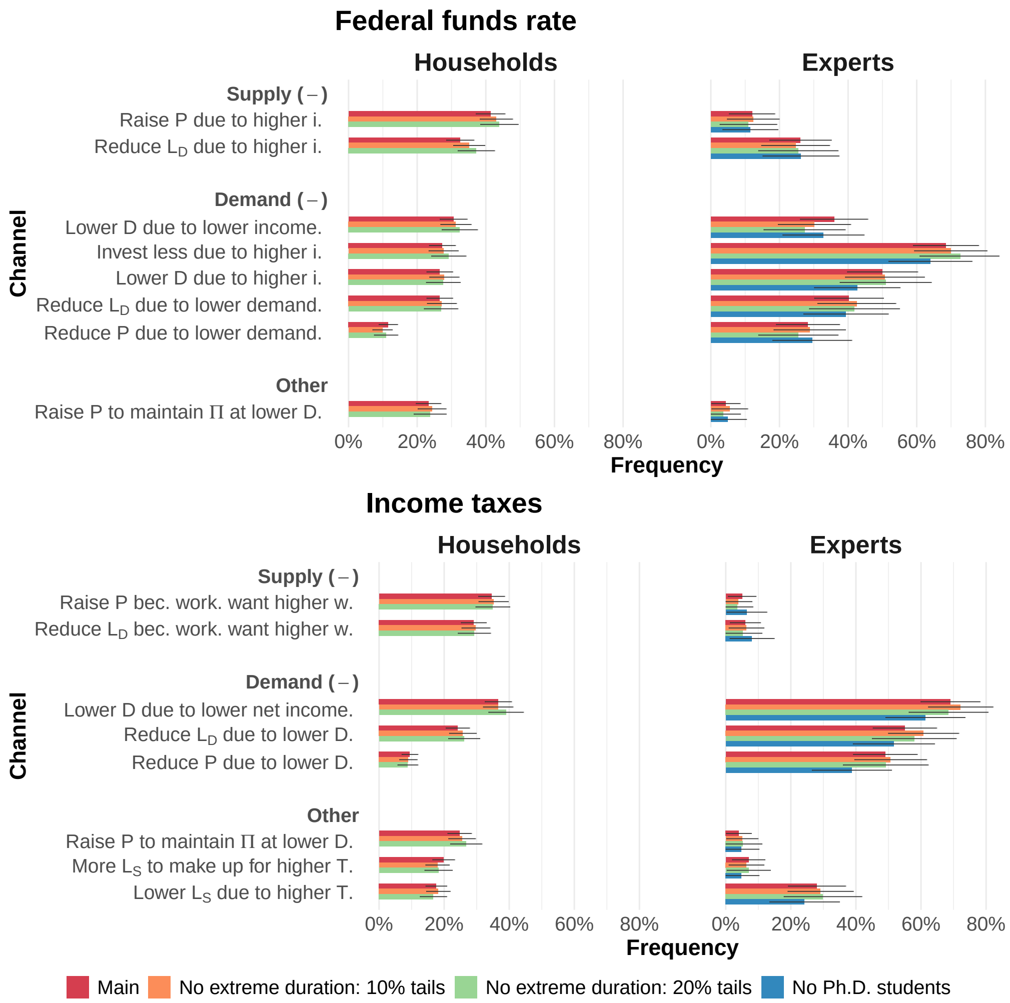


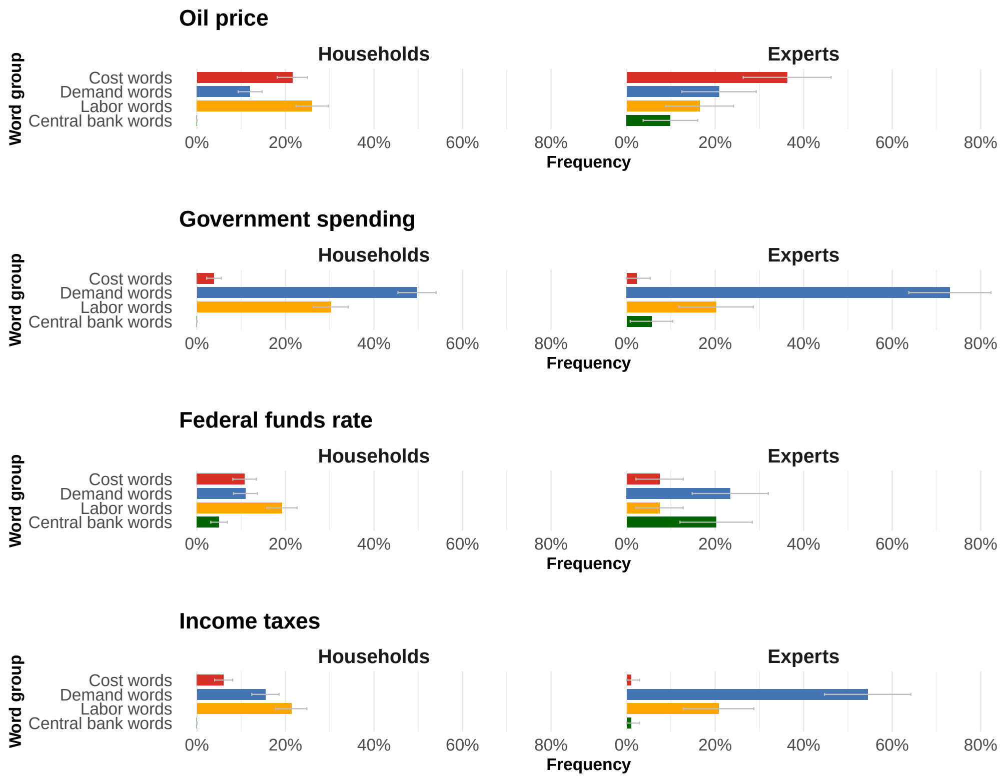
Figure continued on next page.

Figure 2.A.6. Robustness of thoughts of propagation channels



Notes: This figure shows the robustness of the relevant propagation channels that respondents' have on their minds when they make their predictions, using Wave 3 data. Specifically, the figure presents the selected propagation channels for different subsamples. "Main" is the full sample. "No extreme duration: 10% tails" drops the 10% of respondents with the lowest and the 10% of respondents with the highest response duration. Analogously, "No extreme duration: 20% tails" drops both 20% tails of the response duration distribution. "No Ph.D. students" drops all graduate students from the expert sample. The results are displayed separately for each vignette and for households (left panel) and experts (right panel). Error bars display 95% confidence intervals. *P* abbreviates "firm prices", L_D "labor demand", *D* "product demand", Π "firm profits", *T* "taxes", *i* "interest rates", *w* "wages", and L_S "labor supply". The full wording of the channels is available in Appendix 2.E.

Figure 2.A.6 (continued): Robustness of thoughts of propagation channels



Notes: This figure shows the shares of households (left panel) and experts (right panel) mentioning words from three groups in their open-text response in Wave 3. Cost words include the word (stem) "cost". Demand words include the words or word stems "demand", "buy", "purchas", "invest", "spend", and "consum". Labor words include the words or word stems "layoff", "fire", "hire", "labor", "work", and "job". Central bank words include "monetary policy" and "fed* funds (target) rate". The error bars indicate 95% confidence intervals.

Figure 2.A.7. Word usage across vignettes (open-text data)

Appendix 2.B Additional tables

Table 2.B.1. Summary statistics: Covariates in the general population samples

Variable	**ACS (2019)**	**Waves 1-2**	**Wave 3**	**Wave 4**	**Wave 5**
Gender					
Female	51%	55%	54%	56%	54%
Age					
18-34	30%	28%	27%	27%	18%
35-54	32%	39%	27%	33%	17%
55+	38%	33%	46%	40%	65%
Household net income					
Median income, in USD	65,712	62,500	62,500	62,500	62,500
Education					
Bachelor’s degree or more	31%	32%	38%	46%	51%
Region					
Northeast	17%	21%	22%	22%	24%
Midwest	21%	22%	24%	22%	21%
South	38%	41%	36%	40%	36%
West	24%	16%	18%	16%	19%
Sample size	2,059,945	2,214	2,126	1,521	486

Notes: This table compares the distributions of individual characteristics in our different household waves with those in the American Community Survey (ACS) 2019.

Table 2.B.2. Summary statistics: Covariates in the expert samples

Variable	Wave 1	Wave 2	Wave 3
Gender, age			
Female	26%	14%	17%
Age (median)		52	34
Institution			
Policy institution	16%	16%	11%
Academia	83%	56%	88%
Private sector	0%	16%	1%
Academic position (wave 1 and wave 3)			
Full professor	21%		19%
PhD student	18%		40%
Field of study (WES only)			
Economics		84%	
Business		7%	
Completed Ph.D.		65%	
Region of expertise (WES only)			
Western Europe		42%	
Eastern Europe		12%	
CIS		7%	
North America		8%	
Latin America		10%	
Africa		7%	
Middle East		2%	
Asia		10%	
Oceania		2%	
Sample size	180	908	375

Notes: This table provides an overview of the covariates in the expert sample. Different covariates were collected in the three waves. Demographic data are not available for all respondents.

Table 2.B.3. Overview of data collections

Data collection	Sample	Treatments arms	Mechanism questions
Households Wave 1 (February/March 2019) (N=1,063)	Online panel in collaboration with Research Now	None	Beliefs about propagation mechanisms, financial literacy
Households Wave 2 (July 2019) (N=1,151)	Online panel in collaboration with Lucid	None	Good-bad heuristic, rational inattention, numeracy, beliefs about supply-side mechanisms, subjective interest rate rule
Households Wave 3 (February 2021) (N=2,126)	Online panel in collaboration with Lucid	None	Open-text mechanism associations, structured propagation channels, structured prediction approaches, good-bad heuristic, rational inattention, numeracy, perceived past correlations, knowledge
Households Wave 4 (February 2021) (N=1,521)	Online panel in collaboration with Lucid	Demand prime, cost prime and pure control group	None
Households Wave 5 (June 2021) (N=486)	Online panel in collaboration with Lucid	None	Open-text mechanism associations, structured propagation channels, structured prediction approaches, experiences
Experts Wave 1 (February/March 2019) (N=180)	Experts recruited via email invitation (for details see Section 2.J)	None	None
Experts Wave 2 (July 2019) (N=908)	Experts recruited via the ifo World Economic Survey	None	None
Experts Wave 3 (February 2021) (N=375)	Experts recruited via email invitation (for details see Section 2.J)	None	Open-text mechanism associations, structured propagation channels, structured prediction approaches

Table 2.B.4. Response times

Survey	**Wave**	10%	25%	50%	75%	90%	Completion rate
Households	1	7m 3s	9m 56s	13m 46s	19m 28s	26m 57s	78%
	2	7m 38s	10m 27s	14m 36s	21m 51s	32m 48s	79%
	3	9m 4s	12m 35s	17m 37s	25m 47s	36m 22s	74%
	4	4m 0s	5m 52s	8m 52s	13m 38s	19m 42s	70%*
	5	7m 4s	10m 22s	15m 8s	23m 6s	31m 36s	68%
Experts	1	2m 44s	3m 50s	5m 21s	8m 11s	14m 2s	78%
	3	5m 12s	6m 55s	9m 14s	14m 18s	22m 53s	62%

*The completion rates vary only negligibly and insignificantly across treatments (control: 70%, costs prime: 67%, demand prime: 72%).

Notes: This table summarizes quantiles of the distribution of the response duration for all survey waves, except Wave 2 of the expert survey. In Wave 2 of the expert survey, collected via the World Economic Survey, we were not able to collect data on response duration. The last column additionally displays the survey completion rate. Households: the fraction of respondents passing the attention check who complete the survey. Experts: the fraction of respondents passing the general instructions who complete the survey.

Table 2.B.5. Disagreement in perceived effects on inflation and unemployment

		Inflation $\Delta\pi$			**Unemployment Δu**		
Vignette	**Case**	$\sigma_{experts}$	$\sigma_{gen.\ pop.}$	p	$\sigma_{experts}$	$\sigma_{gen.\ pop.}$	p
Oil price	rise	0.28	0.74	<0.01	0.27	0.64	<0.01
	fall	0.32	0.71	<0.01	0.28	0.69	<0.01
Gov. spend.	rise	0.22	0.54	<0.01	0.27	0.61	<0.01
	fall	0.20	0.61	<0.01	0.24	0.63	<0.01
Fed. funds rate	rise	0.31	0.52	<0.01	0.27	0.55	<0.01
	fall	0.28	0.59	<0.01	0.25	0.63	<0.01
Inc. taxes	rise	0.29	0.52	<0.01	0.26	0.55	<0.01
	fall	0.25	0.58	<0.01	0.27	0.56	<0.01
Weighted mean		0.27	0.60		0.26	0.61	

Notes: This table reports data from Waves 1 and 2 of the household and expert surveys. It reports the standard deviations of predicted changes in inflation and unemployment in response to shocks for experts and for the general population, respectively, as well as p-values from a Levene's test of equality of variance (trimmed, median-based, bootstrapped) for each rise or fall scenario. For both the household and the expert sample, we exclude extreme predictions, namely both 5% tails of the distribution, to reduce the influence of outliers. The last row presents the average within-scenario standard deviation, weighted by the differential number of respondents across scenarios.

Table 2.B.6. Heterogeneity of priming effects (households only)

Costs prime: Heterogeneous effects

	Word usage (open-text data)		**Inflation prediction**
	Cost-related words	Demand-related words	$\Delta\pi$
	(1)	(2)	(3)
Costs prime	0.065*	0.003	−0.147
	(0.037)	(0.041)	(0.101)
Costs prime × Costs rise	0.022	−0.009	0.193*
	(0.046)	(0.048)	(0.107)
Costs rise	0.109***	0.089***	0.138*
	(0.017)	(0.033)	(0.072)
Constant	0.000	0.040	0.245***
		(0.028)	(0.069)
Observations	761	761	761
R^2	0.029	0.008	0.032

Demand prime: Heterogeneous effects

	Word usage (open-text data)		**Inflation prediction**
	Cost-related words	Demand-related words	$\Delta\pi$
	(1)	(2)	(3)
Demand prime	−0.043	0.132***	−0.102**
	(0.033)	(0.038)	(0.049)
Demand prime × Demand rises	0.056	−0.084*	0.094
	(0.038)	(0.049)	(0.069)
Demand rises	−0.113***	−0.054*	−0.027
	(0.028)	(0.030)	(0.049)
Constant	0.143***	0.118***	0.378***
	(0.025)	(0.023)	(0.035)
Observations	760	760	760
R^2	0.028	0.040	0.007

Notes: This table presents results from the priming study which focuses on the interest rate vignette (Wave 4 of the household survey). "Costs prime" takes value 1 for respondents randomly assigned to be primed on the costs of production. "Costs rise" takes value 1 for respondents who think that firms' costs increase in response to an increase in the federal funds rate, and zero otherwise. "Demand prime" takes value 1 for respondents randomly assigned to be primed on product demand. "Demand rises" takes value 1 for respondents who think that the demand for firms' products increases in response to an increase in the federal funds rate, and zero otherwise. Columns (1) and (2) show effects on word usage in the open-text responses, and Column (3) presents the effects on the inflation forecast. The variable "Cost-related words" takes value 1 for responses which include the word (stem) "cost". "Demand-related words" takes value 1 for responses which use the words or word stems "demand", "buy", "purchas", "invest", "spend", and "consum". $\Delta\pi$ denotes the perceived reaction of the inflation rate. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.B.7. Robustness of experience analysis for government spending vignette (households only)

(A) *Respondent* worked for military supplier (binary indicator)

	Crowding-out	Demand (+)	Costs	Demand	Labor	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Yes	0.029	0.114**	−0.031	−0.022	0.063	−0.072	−0.155**
	(0.056)	(0.057)	(0.021)	(0.055)	(0.054)	(0.055)	(0.066)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	483	483	483	483	483	483	483
R^2	0.088	0.051	0.028	0.186	0.084	0.139	0.158

(B) *Friend/family* of respondent worked for military supplier (binary indicator)

	Crowding-out	Demand (+)	Costs	Demand	Labor	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Yes	−0.009	0.063	0.003	0.047	0.127***	−0.040	−0.131***
	(0.044)	(0.047)	(0.020)	(0.044)	(0.043)	(0.046)	(0.049)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	483	483	483	483	483	483	483
R^2	0.088	0.047	0.025	0.188	0.099	0.138	0.160

(C) Ever worked for *government* supplier (self/friend, binary indicator)

	Crowding-out	Demand (+)	Costs	Demand	Labor	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Yes	−0.005	0.077	0.002	0.044	0.141***	−0.057	−0.132***
	(0.045)	(0.048)	(0.022)	(0.045)	(0.044)	(0.045)	(0.051)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	483	483	483	483	483	483	483
R^2	0.088	0.049	0.025	0.188	0.103	0.139	0.159

Notes: This table presents results from Wave 5 of the household survey. In Columns (1) and (2), it asks whether respondents who made experiences related to the vignettes think about different propagation mechanisms (binary indicators; see Figure 2.4.1). In Columns (3) to (5), it tests whether respondents with vignette-related experiences use different word (stems) in their open-text responses (binary indicators; "Costs": cost; "Demand": demand, buy, purchas, invest, spend, consum; "Labor": layoff, fire, hire, labor, work, job). In Columns (6) and (7), it tests whether they make different forecasts (inflation: $\Delta\pi$, unemployment: Δu). The right-hand-side experience variable varies across panels. In Panel A, "Yes" is a binary dummy taking value 1 if respondents *themselves* ever worked for a company that sells to the US military. In Panel B, "Yes" is a binary dummy taking value 1 if respondents' *friends/family* ever worked for a company that sells to the US military. In Panel C, "Yes" is a binary dummy taking value 1 if respondents themselves or friends/family of them ever worker for a company that sells to the US *government*. Control variables comprise age, log income, inflation and unemployment forecasts in the baseline scenario, as well as binary indicators for gender, college education, being a Republican, having taken an economics course at the college level, and census regions. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.B.8. Relationship between prediction approaches and thoughts of propagation channels for households

	Propagation channels in different vignettes							
	Oil price		**Government spending**		**Federal funds rate**		**Income taxes**	
	Supply (–)	Demand (–)	Crowd.-out	Demand (+)	Supply (–)	Demand (–)	Supply (–)	Demand (–)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
News	0.007	0.121**	0.081*	0.102**	0.135***	0.113**	0.189***	0.069
	(0.044)	(0.048)	(0.048)	(0.050)	(0.052)	(0.047)	(0.052)	(0.052)
Knowledge	0.119***	0.026	−0.031	0.098*	0.030	0.160***	0.090	0.199***
	(0.045)	(0.050)	(0.048)	(0.051)	(0.053)	(0.049)	(0.055)	(0.054)
Personal sit.	0.054	0.236***	0.138***	0.077	0.057	0.161***	0.070	0.199***
	(0.047)	(0.050)	(0.052)	(0.053)	(0.060)	(0.051)	(0.052)	(0.052)
Macro exp.	0.192***	0.164***	0.055	0.012	0.209***	0.177***	0.219***	0.152***
	(0.041)	(0.046)	(0.044)	(0.048)	(0.048)	(0.045)	(0.047)	(0.048)
Gut feeling	0.074*	0.062	0.109**	0.010	0.068	0.101**	−0.005	0.051
	(0.043)	(0.047)	(0.048)	(0.051)	(0.051)	(0.050)	(0.048)	(0.048)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Observations	521	521	492	492	477	477	490	490
R^2	0.111	0.119	0.077	0.110	0.089	0.134	0.147	0.120

Notes: This table shows data from the household survey of Wave 3. It regresses the propagation channels that are on respondents' minds while they make their predictions (see Figure 2.4.1) on respondents' prediction approaches (see Figure 2.A.5). Each column presents results for a different propagation channel: "Supply (–)" takes value 1 for respondents who choose a negative supply-side propagation channel. "Demand (–)" and "Demand (+)" takes value 1 for respondents choosing a negative or positive demand-side propagation channel, respectively. In the government spending vignette, "Crowding-out" takes value 1 for respondents who select the channel that demand falls due to higher expected future taxes. The explanatory variables are binary and defined as follows: "News" denotes respondents choosing "Things I read or heard in the news." "Knowledge": "My knowledge of economics." "Personal situation": "My personal economic situation today." "Macro experiences": "My memories of economic events in the past." "Gut feeling": "I simply responded based on my gut feeling." Control variables comprise age, log income, inflation and unemployment forecasts in the baseline scenario, as well as binary indicators for gender, college education, being a Republican, having taken an economics course at the college level, and census regions. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.B.9. Households: Political heterogeneity in forecasts

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Democrat	−0.014	0.018	−0.086	0.026	0.072	0.020	−0.018	−0.019
	(0.076)	(0.069)	(0.065)	(0.067)	(0.063)	(0.064)	(0.063)	(0.071)
Constant	0.503***	0.319***	0.232***	−0.043	0.144***	0.094**	0.170***	0.280***
	(0.046)	(0.046)	(0.036)	(0.040)	(0.038)	(0.040)	(0.038)	(0.041)
Joint F-test does not detect a significant effect of Democrat.								
p = 0.891								
Observations	1,099	1,099	1,085	1,085	1,123	1,123	1,121	1,121
R^2	0.000	0.000	0.002	0.000	0.001	0.000	0.000	0.000

Notes: This table reports data from Waves 1 and 2 of the representative general population sample. It provides an overview of political heterogeneity in the predicted changes in inflation ($\Delta\pi$) and unemployment (Δu) for each of the different vignettes separately. The joint F-test results from Seemingly Unrelated Regressions (SUR) with respondent-level clustered standard errors and tests for an overall zero effect of *Democrat*. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Appendix 2.C Theoretical and empirical benchmarks

We compile a set of quantitative benchmarks for each shock from the theoretical and empirical macroeconomic literature. This enables us to compare the forecasts of experts and the general population with how the macroeconomic literature conventionally assesses the effect of each shock on inflation and unemployment.

The references for the empirical benchmarks are chosen from frequently cited papers reporting results that are generally referred to as conventional and/or seminal. The majority of these studies apply Vector Auto-Regression (VAR) models. For the theoretical benchmarks, when possible, we consider as an immediate benchmark the most comparable shock in a model that is widely accepted as a standard medium-size New Keynesian DSGE model such as Smets and Wouters (2007) and Galí, Smets, and Wouters (2011).

To ensure comparability between the size of shocks in the literature and the vignettes, we first calculate the relative size of the shock in each paper relative to the corresponding shock in the vignettes and rescale appropriately. Since most papers focus on output as the main variable of real activity, we translate the responses into changes in the unemployment rate using Okun's Law using a coefficient of -0.4, based on Ball, Leigh, and Loungani (2017), which implies a 0.4 p.p. rise in unemployment associated to a 1% fall in output over the course of a year.

In each case, the following five key steps are involved: 1) identifying the size of the shock in the source paper(s), 2) identifying the size of the response of the variables of interest in the source paper(s), 3) determining the size of the shock in the vignettes, 4) rescaling the shocks from the source papers to be of the same size as those from the vignettes, 5) translating output changes into unemployment changes when needed.

Below, we describe the derivation of benchmarks for each vignette. Δy indicates a percent fall in output over four quarters, and $\Delta\pi$ and Δu are the respective four-quarter changes of inflation and the unemployment rate in p.p.[37] All calculations contain a small degree of approximation.

Oil price. Blanchard and Galí (2010) show that since 1984, a date conventionally considered as the beginning of the Great Moderation, the response of the US economy to oil price fluctuations has become milder. We thus derive our benchmark from the authors' post-1984 VAR results. As shown in Table 2.C.1, the benchmark unemployment rate change for an oil price rise of $30 is 0.4 to 0.45 p.p. For inflation, we derive an empirical benchmark rise of 1.25 to 1.5 p.p.

37. In the case of government and tax shocks in the model of Galí, Smets, and Wouters (2011), the responses of output and unemployment exhibited very low persistence, likely due to the specification of the shock process itself. We, therefore, opted for using the average change over four quarters rather than the change in the fourth quarter only.

We choose two papers as theoretical references: Bodenstein, Erceg, and Guerrieri (2011) and Balke and Brown (2018). Both papers model the effect of shocks to oil supply outside the US. While the former paper models the US as a purely oil-importing country, the latter treats the US as both oil-producing and oil-importing, providing us with a theoretical benchmark effect ranging from 0.35 to 0.8 p.p. (see Table 2.C.1). Neither of these papers studies the impact of oil supply shocks on domestic inflation.

Oil price - Empirical. Source: Blanchard and Galí (2010), Figure 1, Panel B (i.e. post-84). 1) Shock is 10% change in price. 2) $\Delta y = -0.2$, $\Delta \pi = 0.25$. 3) Size of shock in vignette 55% (Wave 2) or 56% (Wave 1) so we approximately multiply the original shocks by 5.5. 4) $\Delta y = -1.1$, $\Delta \pi = 1.4$. 5) Okun's Law: $\Delta u = 0.425$.

Oil price - Theory. Source: Bodenstein, Erceg, and Guerrieri (2011), Figure 2. 1) Shock is an 8% change in price 2) $\Delta y = -0.15$. 3) Size of shock in vignette 55% (Wave 2) 56% (Wave 1) so we approximately multiply the original shocks by 7. 4) $\Delta y = -1.05$ 5) Okun's Law: $\Delta u = 0.42$.

Source: Balke and Brown (2018), Figure 3. 1) Shock is 2.5% change in price 2) $\Delta y = -0.1$. 3) Size of shock in vignette 55% (Wave 2) 56% (Wave 1) so we approximately multiply the original shocks by 22. 4) $\Delta y = -2.2$ 5) Okun's Law: $\Delta u = 0.88$.

Government spending. Regarding government spending, the growing body of studies focusing narrowly on defense spending shocks (Nakamura and Steinsson, 2014; Basso and Rachedi, 2019; Auerbach, Gorodnichenko, and Murphy, 2020) would in theory constitute the optimal comparison for the vignette. However, these studies compute fiscal multipliers at the local level (e.g., metro area or state), which are not necessarily applicable to the national level. We therefore refer to studies that examine the impact of spending at the national level and that utilize the same methodologies (i.e., VAR models) as the papers we consider for the other shocks. For the effect of government spending increases on unemployment, we compute an empirical reference range of −0.1 to −0.2 p.p. (Blanchard and Perotti, 2002; Ramey, 2011; Auerbach and Gorodnichenko, 2012). No results are available for the effect on inflation. On the theoretical side, we interpret the exogenous spending shock in Smets and Wouters (2007) and Galí, Smets, and Wouters (2011) as a government spending shock. A third source is the government spending shock in Zubairy (2014).[38] The theoretical reference range of values for the change in unemployment after a rise in spending of 0.5% of GDP, reported in Table 2.C.1, is between −0.1 to −0.2 p.p., while the benchmark rise in inflation is 0.15 to 0.2 p.p.

38. Note that we do not use this paper as a benchmark for the response of inflation. Although inflation dynamics resulting from fiscal policy are embedded in the model, they are not discussed in detail by the author.

Government spending - Empirical. Source: Blanchard and Perotti (2002), Ramey (2011) and sources therein, Auerbach and Gorodnichenko (2012). 1) Shock is 1% of GDP 2) $\Delta y = 0.8$ to 1.5. 3) Size of shock in vignette is 2.4% of 4.2 trillion of government spending. US 2018 GDP is 20.89 trillion according to the Bureau of Economic Analysis, so the shock is about 2.4% of 20% of GDP, which is 0.5% of GDP. So we divide the original shock by 2. 4) $\Delta y = 0.4$ to 0.75. 5) Okun's Law: $\Delta u = -0.16$ to -0.3.

Government spending - Theory. Source: Galí, Smets, and Wouters (2011), Figure 3. 1) Size of shock is 0.47, with exogenous spending formulated in percent of output, so it can be interpreted as 0.5% of GDP. 2) $\Delta u = -0.1$, $\Delta\pi = 0.2$. 3) The shock in the vignette is very similar in size, so there is no need to scale it. 4) $\Delta u = -0.1$, $\Delta\pi = 0.2$.

Source: Smets and Wouters (2007), Figure 2. 1) Size of shock is 0.5, with exogenous spending formulated in percent of output, so it can be interpreted as 0.5% of GDP. 2) $\Delta y = 0.3$, $\Delta\pi = 0.15$. 3) The shock in the vignette is very similar in size, so there is no need to scale it. 4) $\Delta y = 0.3$, $\Delta\pi = 0.15$. 5) Okun's Law: $\Delta u = -0.12$.

Source: Zubairy (2014), Table 2. 1) Size of shock is 1% of GDP. 2) $\Delta y = 1$. 3) Divide by 2 to make it comparable to the vignette. 4) $\Delta y = 0.5$. 5) Okun's Law: $\Delta u = -0.2$.

Monetary policy. Arias, Caldara, and Rubio-Ramírez (2019) gives an empirical benchmark effect of 0.2 p.p. on unemployment and 0.2 p.p. on inflation for our federal funds rate rise by 50 basis points. This is largely in line with a large and consistent body of VAR evidence since the late 1990's (Bernanke and Mihov, 1998; Christiano, Eichenbaum, and Evans, 1999; Stock and Watson, 2001; Romer and Romer, 2004; Bernanke, Boivin, and Eliasz, 2005; Primiceri, 2005; Uhlig, 2005). As a theoretical reference, we again use Smets and Wouters (2007) and Galí, Smets, and Wouters (2011) and arrive at a benchmark of 0.4 to 0.5 p.p. for unemployment and a benchmark of -0.15 p.p. for inflation.

Monetary policy - Empirical. Source: Arias, Caldara, and Rubio-Ramírez (2019) Figure 5 (i.e. estimation on full post-WWII sample, imposing a zero restriction on the systematic response of monetary policy to commodity prices). 1) Shock size is 0.25 p.p. 2) $\Delta y = -0.25$, $\Delta\pi = -0.1$. 3) To make the shock comparable to the vignette, we multiply by 2. 4) $\Delta y = -0.5$, $\Delta\pi = -0.2$. 5) Okun's Law: $\Delta u = 0.2$.

Monetary policy - Theory. Source: Galí, Smets, and Wouters (2011), Figure 3. 1) Size of shock is 0.15 p.p. 2) $\Delta u = -0.15$, $\Delta\pi = -0.05$. 3) We approximately multiply by 3.3 to make it comparable to the vignette. 4) $\Delta u = 0.5$, $\Delta\pi = -0.16$.

Source: Smets and Wouters (2007), Figure 2. 1) Size of shock is 0.175. 2) $\Delta y = -0.35$, $\Delta\pi = -0.05$. 3) We approximately multiply by 3 to make it comparable to the vignette. 4) $\Delta y = -1$, $\Delta\pi = -0.15$. 5) Okun's Law: $\Delta u = 0.4$.

Income tax rate. The empirical benchmark for the unemployment change in response to the increase in the income tax rate by 1 p.p. on average ranges between 0.2 and 0.6 p.p. (Blanchard and Perotti, 2002; Romer and Romer, 2010; Favero and Giavazzi, 2012; Mertens and Ravn, 2012; Perotti, 2012; Mertens and Ravn, 2014). To our knowledge, the only paper modeling the impact of labor income tax rate fluctuations in a New Keynesian model is Zubairy (2014). For the theoretical benchmark of the effect on unemployment, we derive a value of 0.06.[39]

Tax rate change - Empirical. Source: Blanchard and Perotti (2002), Romer and Romer (2010), Favero and Giavazzi (2012), Mertens and Ravn (2012, 2014), and Perotti (2012). 1) Shock size is a 1% of GDP increase in tax revenue. 2) Range of empirical output multipliers at 4 to 6 quarters is 1 to 3% of GDP. 3) The shock size in the vignette is approximately 0.5% of GDP. So we divide by 2 to make the shock comparable to the vignette. 4) $\Delta y = 0.5$ to 1.5. 5) Okun's Law: 0.2 to 0.6.

Tax rate change - Theory. Source: Zubairy (2014), Table 2. 1) Size of shock is 1% of GDP. 2) $\Delta y = 0.32$. 3) Divide by 2 to make it comparable to the vignette. 4) $\Delta y = 0.15$. 5) Okun's Law: $\Delta u = -0.06$.

Table 2.C.1. Benchmarks for the sign and size of the effects of different shocks

Shock		Unemployment Response		Inflation Response	
		Sign	Value (p.p.)	Sign	Value (p.p.)
Oil price rise	Theory	+	0.42 to 0.88		
(55% higher price)	Empirical	+	0.42	+	1.4
Government spending rise (2.4%	Theory	–	−0.1 to −0.2	+	0.15 to 0.2
higher growth rate)	Empirical	–	−0.16 to −0.3		
Interest rate rise	Theory	+	0.4 to 0.5	–	−0.15
(0.5 b.p. higher rate)	Empirical	+	0.2	–	−0.2
Tax rate rise	Theory	+	0.06		
(1 p.p. higher rates)	Empirical	+	0.2 to 0.6		

Notes: The table reports the benchmarks for changes in the unemployment rate and the inflation rate four quarters after the respective shock from the theoretical and empirical literature. The values are adjusted to be comparable to the size of the shocks in our survey. Empty fields indicate that – to the best of our knowledge – there is no robust and rigorous evidence on the effect of a given shock on the respective outcome variable of interest.

39. Once again, we do not use this paper as a benchmark for the response of inflation. Although inflation dynamics resulting from fiscal policy are embedded in the model, they are not discussed in detail by the author.

Appendix 2.D Additional results on forecasts

In this Appendix, we provide more detailed results on the household and expert predictions, mostly based on data from Waves 1 and 2.

2.D.1 Predictions by direction of the shock

While we pooled the responses from the rise and (reverse-coded) fall scenarios in the main text, in this section, we present the expert and household predictions separately by the rise and fall scenarios. For brevity, we focus on differences in average (absolute) forecasts about inflation and unemployment between rise and fall scenarios. Both rise scenarios and fall scenarios feature considerable amounts of heterogeneity. As in our baseline analysis, we characterize beliefs about the effects of macroeconomic shocks pooling responses from Waves 1 and 2 as we do not find any qualitative differences in predictions across waves.

2.D.1.1 Expert predictions

Oil supply shock. Experts predict an increase in unemployment of 0.24 p.p. and a rise in inflation of 0.45 p.p. in the scenario where the oil price increases by $30 (Columns 1 and 2 in Table 2.D.1, Panel A). In the scenario in which the oil price decreases by $30, they predict that the unemployment rate would be lower by 0.13 p.p. and that the inflation rate would be lower by 0.33 p.p. The table reveals that the absolute values of the predictions for the rise and fall scenarios are not statistically distinguishable.

Government spending shock. Experts predict a 0.31 p.p. lower unemployment rate and a 0.30 p.p. higher inflation rate in the rise-scenario (Columns 3 and 4 in Table 2.D.1, Panel A). In the fall scenario, they predict that the unemployment rate would be higher by 0.30 p.p. and that the inflation rate would be lower by 0.22 p.p. The absolute value of the unemployment and inflation predictions are not statistically distinguishable between the rise and fall scenarios.

Interest rate shock. Our experts predict that unemployment would be higher by 0.29 p.p., while inflation would be lower by 0.15 p.p. in response to an unexpected increase in the interest rate. In the fall scenario, experts predict that unemployment would be lower by 0.19 p.p., and that inflation would be higher by 0.16 p.p. While the absolute value of inflation forecasts is almost identical between rise and fall scenarios, the magnitude of unemployment forecasts is significantly different ($p < 0.05$).

Tax shock. On average, experts predict a 0.22 p.p. higher unemployment rate and a 0.11 p.p. lower inflation rate under the rise-scenario (Columns 7 and 8 in Table

2.D.1, Panel A). For the fall-scenario, experts predict a 0.24 p.p. lower unemployment rate and a 0.21 p.p. higher inflation rate. The absolute values of predictions from the rise and fall scenarios are fairly similar and not statistically distinguishable from each other.

Summary. Taken together, experts on average perceive no strong asymmetry between the effects of increases or decreases of the different shock variables.

2.D.1.2 Household predictions

We continue with the forecasts from the general population, which are displayed in Panel B of Table 2.D.1.

Oil supply shock. Households on average predict the unemployment rate to be 0.45 p.p. higher and the inflation rate to be 0.67 p.p. higher in the scenario where the oil price rises by $30. In the oil price fall-scenario, they expect the unemployment rate to be 0.21 p.p. lower and the inflation rate to be 0.33 p.p. lower. Households predict a significantly larger response of inflation and unemployment (in absolute values) in the rise scenario compared to the fall scenario ($p < 0.05$).

Government spending shock. Households believe that inflation will be 0.26 p.p. lower in response to an exogenous reduction in government spending, and that it would be higher by 0.13 p.p. in response to an increase in government spending. Households predict a significantly larger response of inflation (in absolute values) in the fall scenario compared to the rise scenario ($p < 0.05$), but similar magnitudes for unemployment. Households on average think that unemployment neither responds to increases nor decreases in government spending.

Interest rate shock. Respondents think that unemployment would be 0.17 p.p. higher in response to a rise in interest rates. However, they expect unemployment to remain unchanged in response to a decrease in interest rates. Respondents expect a 0.15 p.p. *lower* inflation rate in response to a fall in the federal funds target rate and a 0.19 p.p. *higher* inflation rate in response to a rise. While the responses to the rise and fall scenario are fairly symmetric for inflation rate predictions, they are statistically different for unemployment ($p < 0.05$).

Tax shock. Respondents think higher taxes would lead to a 0.30 p.p. higher unemployment rate, and that lower taxes would result in a 0.25 p.p. lower unemployment rate. Moreover, they predict that a tax hike would result in a 0.21 p.p. higher inflation rate, while they forecast a 0.12 p.p. lower inflation rate in response to a tax cut. The absolute values of predictions from the rise and fall scenarios are not statistically distinguishable from each other.

Summary. Taken together, different from the expert forecasts, we find some evidence of asymmetry in households' predicted responses of inflation and unemployment between the rise and the fall scenarios.

Table 2.D.1. Inflation and unemployment forecasts by direction of the shocks

(A) Experts

	Oil price		Government spending		Federal funds rate		Income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fall	−0.327***	−0.130***	−0.224***	0.303***	0.155***	−0.188***	0.209***	−0.235***
	(0.042)	(0.037)	(0.024)	(0.026)	(0.027)	(0.025)	(0.031)	(0.028)
Rise	0.449***	0.235***	0.299***	−0.311***	−0.152***	0.289***	−0.107***	0.221***
	(0.030)	(0.030)	(0.021)	(0.028)	(0.033)	(0.025)	(0.035)	(0.036)
p-values from additional tests								
Fall≠Rise	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
|Fall|≠|Rise|	0.018	0.028	0.019	0.837	0.953	0.004	0.028	0.772
Observations	482	481	474	475	517	513	515	521
R^2	0.333	0.120	0.373	0.352	0.096	0.270	0.093	0.164

(B) General Population

	Oil price		Government spending		Federal funds rate		Income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fall	−0.331***	−0.210***	−0.261***	0.042	−0.150***	−0.028	−0.122***	−0.250***
	(0.050)	(0.049)	(0.045)	(0.045)	(0.048)	(0.045)	(0.043)	(0.051)
Rise	0.667***	0.445***	0.135***	−0.023	0.193***	0.174***	0.206***	0.298***
	(0.053)	(0.047)	(0.041)	(0.046)	(0.037)	(0.043)	(0.043)	(0.044)
p-values from additional tests								
Fall≠Rise	<0.001	<0.001	<0.001	0.310	<0.001	0.001	<0.001	<0.001
|Fall|≠|Rise|	<0.001	0.001	0.038	0.762	0.481	0.019	0.171	0.471
Fall: (A)≠(B)	0.954	0.192	0.467	<0.001	<0.001	0.002	<0.001	0.794
Rise: (A)≠(B)	<0.001	<0.001	<0.001	<0.001	<0.001	0.020	<0.001	0.180
Observations	1,099	1,099	1,085	1,085	1,123	1,123	1,121	1,121
R^2	0.159	0.085	0.042	0.001	0.029	0.014	0.027	0.056

Notes: This table compares beliefs about the quantitative effects of the different macroeconomic shocks across the fall and rise scenarios for experts (Panel A) and for households (Panel B). "Fall" takes value 1 for the predictions in the fall scenario, and "Rise" takes value 1 for the rise scenario. $\Delta\pi$ denotes the predicted change in the inflation rate compared to the baseline scenario. Δu denotes the predicted change in the unemployment rate compared to the baseline scenario. Additionally, p-values from the following tests are reported: tests whether there is a difference between rise and fall predictions (Fall≠Rise), tests whether there is a difference in the absolute size of rise and fall predictions (|Fall|≠|Rise|), tests whether there is a difference in fall predictions between experts and the general population (Fall: (A)≠(B)), and tests whether the rise predictions differ between experts and households (Rise: (A)≠(B)). Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

2.D.2 Co-movement predictions of inflation and unemployment

In the main text, we focus on forecasts about inflation and unemployment separately. In this subsection, we examine households' and experts' forecasts of the joint response of inflation and unemployment to macroeconomic shocks.

2.D.2.1 Predicted co-movement (Waves 1 and 2)

To describe households' and experts' predicted co-movement, we focus on our main data collection in Waves 1 and 2, and pool predictions under the rise scenario with (reversed) predictions under the fall scenario. Throughout this section, we divide forecasts into five categories: i) inflation and unemployment both decrease; ii) inflation falls, while unemployment increases; iii) inflation rises, while unemployment falls; iv) both increase; v) other, which includes all cases where one variable is predicted to stay constant. In our discussion we focus on categories i)-iv), and treat category v) as residual. The fractions predicting joint movement in different directions are displayed in Figure 2.D.1.

Oil supply shock. Among households, a view that both inflation and unemployment increase in response to oil supply shocks is most prevalent (49% of responses). Fractions between 9% and 13% predict both to fall or movement in opposite directions.

Among experts, a majority (59%) predict both unemployment and inflation to increase, while 15% predict inflation to increase and unemployment to fall. Only very small fractions predict co-movement featuring a decrease in inflation.

Government spending shock. We find strong heterogeneity in households' beliefs about the co-movement of inflation and unemployment in response to government spending shocks. Fractions between 17% and 26% predict both variables to fall, inflation to rise and unemployment to fall, or both variables to rise, while 8% forecast a fall in inflation and a rise in unemployment.

Strikingly, experts' views on the joint response of unemployment and inflation are much more homogeneous. 68% of experts view the government spending shock as a demand-side shock featuring an increase in inflation and a fall in unemployment, while only 6% predict both variables to rise, and almost no expert predicts both variables to fall or unemployment to increase and inflation to decrease.

Interest rate shock. Households disagree strongly on the co-movement of unemployment and inflation in response to monetary policy shocks. A view that both variables will increase is most prevalent (36%), while fractions between 11% and 17% predict both variables to fall or movement in opposite directions.

Experts agree much more on the joint response, with a majority of 55% taking the common view that an interest rate hike increases unemployment and decreases inflation. Fractions of 7% and 12% predict inflation to rise and unemployment to fall or both to rise, respectively, while only 7% predict both to fall.

Tax shock. Households' beliefs about co-movement under the tax shock are very similar as under the interest rate shock, with a view that both variables will increase being most prevalent (32%).

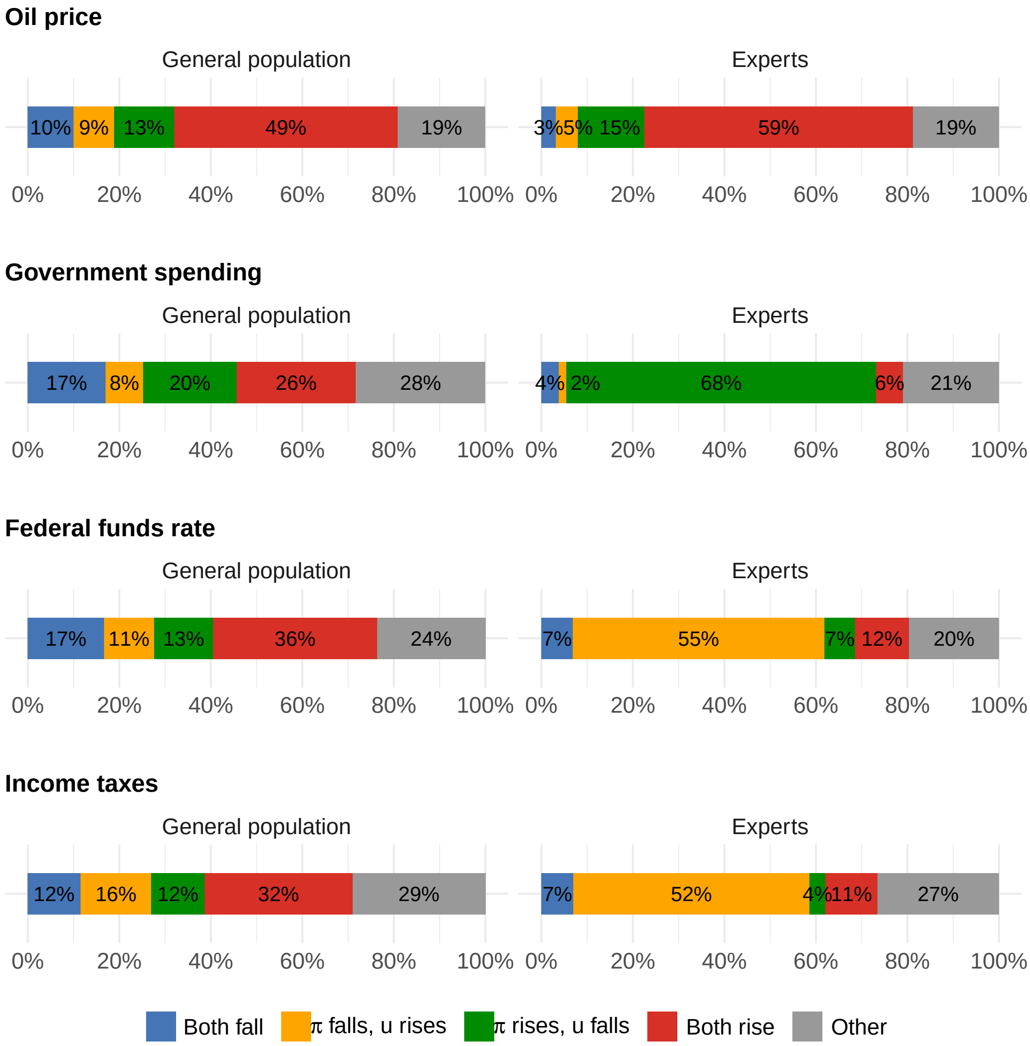
Similarly, a majority (52%) of experts view a tax hike as increasing unemployment and decreasing inflation, while only fractions between 4% and 11% predict both variables to move but in different ways than predicted by the most common view in the literature.

Summary. Taken together, these results highlight that there is strong disagreement among households about the joint response of inflation and unemployment to aggregate shocks, with majorities of households predicting co-movement different to literature benchmarks under all shocks. In contrast, heterogeneity in expert predictions is almost exclusively driven by making a forecast in line with theory benchmarks vs falling into the "other" category, where one variable is predicted to stay constant.

2.D.2.2 Associations and predicted co-movement (Wave 3)

We also use our data from Wave 3 of the household survey to study the correlation of thoughts about propagation mechanisms with forecasts of the co-movement of unemployment and inflation. We focus on correlations of thoughts with a dummy indicating a predicted co-movement in line with literature benchmarks. For the household survey, Table 2.D.2 shows correlations of forecasts with thoughts about propagation channels elicited under the structured survey questions. In the context of the oil price vignette, thoughts about propagation channels featuring reductions in demand or supply are strongly positively associated with predicting a benchmark-consistent co-movement. In the context of the government spending vignette, thoughts about channels indicating reductions in demand are negatively related to forecasting benchmark-consistent co-movement, while thoughts about channels featuring demand increases are positively related. In the context of the interest rate vignette, supply-side propagation channels are negatively related to predicting a benchmark-consistent co-movement, while demand-side propagation mechanisms are positively related. In the context of the income tax vignette, structured propagation channels related to a negative demand shock are positively associated with predicting a benchmark-consistent co-movement.

Table 2.D.3 repeats the analysis for the expert survey of Wave 3. The results show that experts who think of propagation channels that are conventionally featured in macroeconomic models are more likely to forecast a co-movement of inflation and unemployment that is in line with the benchmarks from the literature.



Notes: This figure presents the forecasts of the joint movement of inflation and unemployment in response to macroeconomic shocks measured in Waves 1 and 2. Directional predictions in the fall scenarios are reversed to render them comparable to rise predictions. It compares the forecasts of the general population (left column) to those of experts (right column).

Figure 2.D.1. Forecasts of the joint movement of inflation and unemployment in response to macroeconomic shocks

Table 2.D.2. Households: Thoughts of propagation channels correlate with benchmark-consistent co-movement of unemployment and inflation predictions

	Indicator: Both π and u prediction in line with benchmarks			
	Oil price	Government spending	Federal funds target rate	Income taxes
	(1)	(2)	(3)	(4)
Oil: Supply (–)	0.270***			
	(0.045)			
Oil: Demand (–)	0.201***			
	(0.041)			
Gov.: Crowding-out		−0.094**		
		(0.037)		
Gov.: Demand (+)		0.188***		
		(0.036)		
Fed.: Supply (–)			−0.065***	
			(0.024)	
Fed.: Demand (–)			0.080***	
			(0.022)	
Tax: Supply (–)				−0.013
				(0.019)
Tax: Demand (–)				0.059***
				(0.019)
Constant	0.259***	0.167***	0.058***	0.029**
	(0.039)	(0.028)	(0.018)	(0.013)
Observations	557	519	520	530
R^2	0.116	0.066	0.030	0.018

Notes: This table presents results from wave 3 of the household survey. The outcome variable is a dummy which takes value 1 if the directions of both the inflation and the unemployment forecast are consistent with the literature benchmarks. "Supply (–)" takes value 1 for respondents who choose a negative supply-side propagation channel in the structured question. "Demand (–)" and "Demand (+)" take value 1 for respondents choosing a negative or positive demand-side propagation mechanism in the structured survey question, respectively. "Crowding-out" takes value 1 for respondents who select the channel that demand falls due to higher expected future taxes.

Table 2.D.3. Experts: Thoughts of propagation channels correlate with benchmark-consistent co-movement of unemployment and inflation predictions

	Indicator: Both π and u prediction in line with benchmarks			
	Oil price	Government spending	Federal funds target rate	Income taxes
	(1)	(2)	(3)	(4)
Oil: Supply (–)	0.460***			
	(0.098)			
Oil: Demand (–)	0.280***			
	(0.099)			
Gov.: Crowding-out		−0.063		
		(0.146)		
Gov.: Demand (+)		0.553***		
		(0.095)		
Fed.: Supply (–)			−0.116	
			(0.106)	
Fed.: Demand (–)			0.478***	
			(0.104)	
Tax: Supply (–)				0.137
				(0.154)
Tax: Demand (–)				0.381***
				(0.098)
Constant	−0.010	0.139*	0.186**	0.138*
	(0.072)	(0.075)	(0.092)	(0.082)
Observations	91	88	92	100
R^2	0.198	0.227	0.142	0.102

Notes: This table presents results from wave 3 of the expert survey. The outcome variable is a dummy which takes value 1 if the directions of both the inflation and the unemployment forecast are consistent with the literature benchmarks. "Supply (–)" takes value 1 for respondents who choose a negative supply-side propagation channel in the structured question. "Demand (–)" and "Demand (+)" take value 1 for respondents choosing a negative or positive demand-side propagation mechanism in the structured survey question, respectively. "Crowding-out" takes value 1 for respondents who select the channel that demand falls due to higher expected future taxes.

2.D.3 Robustness

Robustness to incentives. To examine the role of effort and attention in responses to the hypothetical vignettes, we provide a random subset of respondents with monetary incentives in Wave 1 of the household survey. We inform these respondents that we asked experts the same questions and that for one randomly selected case they can earn an additional $0.50 if their response is at most 0.2 p.p. away from the average expert response. This amount corresponds to approximately one third of the show-up fee. Of course, this incentivizes households to state their second-order beliefs about experts' beliefs, which might differ from the households' own beliefs. To address this concern, we also measure the perceived objectivity and accuracy of experts.

Incentives moderately increase the fraction of benchmark-consistent predictions of inflation by 4 p.p. (Table 2.D.4 Column 1), while the predictions regarding unemployment are completely unaffected (Column 2). In a joint test, no effect of incentives on consistency of predictions with the benchmarks can be detected (Column 4), even though incentivized respondents spend roughly 40 seconds longer in the vignettes – a 25% increase in response time (Column 6). The effect of incentives does not significantly vary with a measure of trust in experts (Panel B of Table 2.D.4).

Order effects. To account for potential order effects, we randomize both the order of vignettes as well as the order in which unemployment and inflation forecasts are elicited. Figure 2.D.2 shows average quantitative forecasts in the vignettes for households pooling across Waves 1 and 2, separately for (i) all forecasts, (ii) forecasts under the first vignette faced by each respondent, (iii) forecasts for the first variable (either unemployment or inflation) in both vignettes faced by a respondent. The figure highlights that the responses are very similar, indicating a limited relevance of order effects.

Attention to the survey. Figure 2.D.2 also displays forecasts separately (iv) for a restricted sample excluding respondents in the upper and lower 10% tails of the survey time distribution, and (v) for a restricted sample excluding the 20% of respondents with the largest absolute difference in predictions in the baseline scenarios across the two vignettes to which they responded.[40] Our figure highlights very similar patterns for those two different samples, suggesting that a lack of attention to the survey does not account for the patterns observed in the household sample.

40. Given that the baseline scenarios ask respondents to assume no change in the shock variable of interest, large differences in predictions between the two baseline scenarios each respondent faced could indicate inattention or random response behavior.

Table 2.D.4. Households: Robustness: Incentive effects

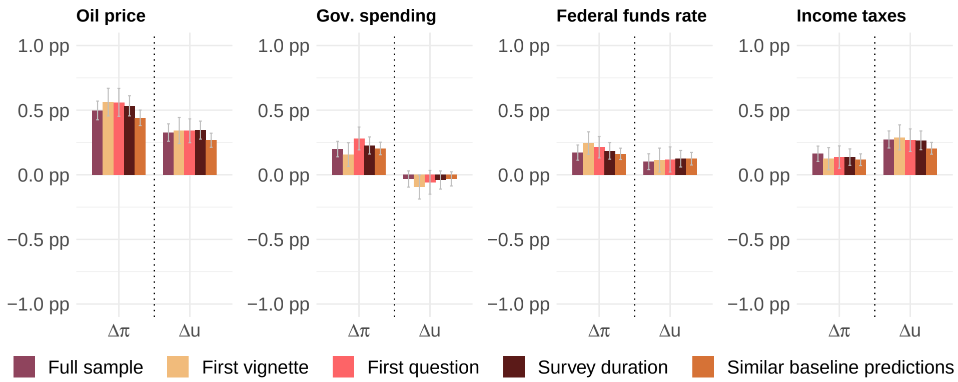
Panel A: Incentives

	$\Delta\pi\checkmark$	$\Delta u\checkmark$	both✓	all✓	time instructions	time vignettes
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	0.044**	−0.000	0.038**	0.022	−0.537	38.589***
	(0.022)	(0.023)	(0.019)	(0.016)	(10.361)	(13.236)
Constant	0.447***	0.508***	0.216***	0.477***	112.689***	165.001***
	(0.015)	(0.017)	(0.013)	(0.011)	(9.261)	(6.490)
Observations	1,063	1,063	1,063	1,063	1,063	1,063
R^2	0.004	0.000	0.004	0.002	0.000	0.008

Panel B: Incentives crossed with subjective perception of expert accuracy

	$\Delta\pi\checkmark$	$\Delta u\checkmark$	both✓	all ✓	time instructions	time vignettes
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	0.040*	0.001	0.038**	0.020	−1.128	38.692***
	(0.022)	(0.023)	(0.019)	(0.016)	(10.555)	(13.029)
Experts acc.	0.006	−0.015	0.012	−0.005	7.261	5.663
	(0.015)	(0.017)	(0.013)	(0.012)	(7.068)	(4.976)
Incent. × Exp. acc.	−0.017	−0.007	−0.029	−0.012	0.246	7.697
	(0.022)	(0.024)	(0.018)	(0.017)	(9.717)	(17.501)
Constant	0.449***	0.506***	0.217***	0.477***	113.222***	165.118***
	(0.015)	(0.017)	(0.013)	(0.011)	(9.502)	(6.577)
Observations	1,049	1,049	1,049	1,049	1,049	1,049
R^2	0.004	0.003	0.006	0.003	0.002	0.010

Notes: This table provides an overview of the effect of monetary incentives on the response behavior of the general population in Wave 1. A forecast is classified as benchmark-consistent if it follows the same qualitative direction as literature benchmark. Panel A displays the effect on the benchmark-consistency of forecasts and response times. *Incentives* constitutes a binary variable that takes value 1 for incentivized respondents. For each individual, $\Delta\pi\checkmark$ measures the fraction of benchmark-consistent inflation forecasts (out of two), $\Delta u\checkmark$ the fraction of benchmark-consistent unemployment forecasts (out of two), *both*✓ the fraction of vignettes in which both forecasts are benchmark-consistent (out of two), and *all*✓ the overall fraction of benchmark-consistent forecasts (out of four). Thus, the coefficients can be interpreted as the effect of incentives on the probability of a benchmark-consistent forecast. Columns 5 and 6 show effects on the time spent reading the instructions and the total time spent on the vignettes. Panel B examines heterogeneity according to the respondents' perceived accuracy of experts (*Experts acc.*, standardized) to rule out that incentives might be ineffective merely because expert forecasts are perceived as inaccurate. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.



Notes: This figure provides an overview of procedural robustness checks that repeat the main analysis for different sub-samples of households from Waves 1 and 2. It pools beliefs for the "rise" and "fall" scenario. Predictions in the fall scenarios are reversed to render them comparable to rise predictions. Error bars show 95% confidence intervals using robust standard errors. *Δu* denotes the expected change in the unemployment rate compared to the baseline scenario. *Δπ* denotes the expected change in the inflation rate compared to the baseline scenario. "Full sample" denotes the full sample and, thus, replicates the results of the main Figure 2.3.2. "First vignette" contains only the responses to the first vignette, while "First question" focuses only on responses to the first forecast question (in both vignettes). "Survey duration" excludes both 10% tails in the survey duration distribution, and "Similar baseline prediction" excludes the 20% respondents with the largest absolute difference in baseline predictions across the two vignettes they responded to.

Figure 2.D.2. Households: Procedural robustness of quantitative beliefs

Appendix 2.E Structured question on propagation channels

The order of items is randomized across participants, except for *None of the above* which is always the last response option.

Oil vignette

How did you come up with your predictions?

The following statements describe different thoughts you might have had on your mind while making your predictions for the alternative scenario. **Did you have any of these thoughts on your mind?** Please tick all that you had on your mind.

- ☐ Due to lower incomes or job loss, households cut back on their spending.
- ☐ Businesses face lower demand for their products, so they increase their product prices to keep profits at the same level.
- ☐ The higher cost of oil makes it more attractive to use alternative energy sources and energy-saving technologies, which leads to job creation.
- ☐ To make up for the higher cost of production, businesses reduce their workforce.
- ☐ Because higher product prices lower their purchasing power, households cut back on their spending.
- ☐ To make up for the higher cost of production, businesses increase product prices.
- ☐ The US oil extraction industry profits from the higher oil price, which leads to job creation.
- ☐ *None of the above.*

Government spending vignette

How did you come up with your predictions?

The following statements describe different thoughts you might have had on your mind while making your predictions for the alternative scenario. **Did you have any of these thoughts on your mind?** Please tick all that you had on your mind.

- ☐ Because of higher incomes, households increase their spending.
- ☐ Because there is more demand for their products, businesses increase their product prices.
- ☐ Businesses face lower demand for their products, so they increase their product prices to keep profits at the same level.
- ☐ Households expect to pay higher taxes in the future, which may be needed to pay back the new government debt. Therefore, households work more.
- ☐ Households expect to pay higher taxes in the future, which may be needed to pay back the new government debt. Therefore, households cut back on their spending.
- ☐ Because there is more demand for their products, businesses increase their workforce.
- ☐ To help the government finance the additional spending, the central bank prints money.
- ☐ *None of the above.*

Interest rate vignette

How did you come up with your predictions?

The following statements describe different thoughts you might have had on your mind while making your predictions for the alternative scenario. **Did you have any of these thoughts on your mind?** Please tick all that you had on your mind.

- ☐ Because there is less demand for their products, businesses reduce their workforce.
- ☐ Businesses face lower demand for their products, so they increase their product prices to keep profits at the same level.
- ☐ To make up for the higher cost of borrowing, businesses reduce their workforce.
- ☐ Because there is less demand for their products, businesses reduce their product prices.
- ☐ Because higher interest rates make it more attractive to save and less attractive to borrow, households cut back on their spending.
- ☐ Due to the higher cost of borrowing, businesses pursue fewer investment projects.
- ☐ To make up for the higher cost of borrowing, businesses increase product prices.
- ☐ Because of lower incomes or job loss, households cut back on their spending.
- ☐ *None of the above.*

Taxation vignette

How did you come up with your predictions?

The following statements describe different thoughts you might have had on your mind while making your predictions for the alternative scenario. **Did you have any of these thoughts on your mind?** Please tick all that you had on your mind.

- ☐ Because workers demand higher wages to make up for the higher income taxes, businesses reduce their workforce.
- ☐ Because workers demand higher wages to make up for the higher income taxes, businesses increase their product prices.
- ☐ Because there is less demand for their products, businesses reduce their product prices.
- ☐ Because of lower disposable incomes, households cut back on their spending.
- ☐ Because higher taxes make it less attractive to work, households work less.
- ☐ Because there is less demand for their products, businesses reduce their workforce.
- ☐ Businesses face lower demand for their products, so they increase their product prices to keep profits at the same level.
- ☐ To make up for their reduced disposable incomes, households work more.
- ☐ *None of the above.*

Appendix 2.F Hand-coded measures of thoughts (open-text data)

In Wave 3 of our household and expert surveys, we elicit respondents' thoughts while they are forecasting the changes in unemployment and inflation in response to hypothetical shocks using an open-text question on the forecast survey screen. Specifically, we ask respondents to "tell us how [they] come up with their prediction" and about "[their] main considerations in making the prediction". In this appendix section, we demonstrate robustness of our findings on thoughts about different propagation channels presented in Section 2.4 to using alternative measures of thoughts based on hand-coding of the open-text responses.

2.F.1 Response types

Coding scheme. We first classify the open-ended text responses into broad response type categories. We use the following categories: i) "Mechanism", including all responses mentioning thoughts related to economic propagation mechanisms of the shocks; ii) "Model", including all responses mentioning a specific economic theory or model (only done by experts); iii) "Guess", indicating responses that express uncertainty or explicitly mention that the forecast is a guess; iv) "Politics", which includes general political or normative statements; v) "Historical", including reference to how things "typically" evolve or how things have evolved in the past; vi) "Misunderstanding", denoting whether respondents misunderstood features of the vignette; vii) "Restates prediction", indicating whether the open-text response repeats or summarizes the provided forecasts of unemployment and inflation; viii) "Endogenous shock", indicating whether respondents mention that e.g. changes in interest rates are the Fed's response to other developments in the economy; ix) "Other", which is a residual category for responses not falling into any of the other categories (e.g. including statements about general trends in the economy not specifically related to the shocks). We allow each response to fall into more than one category. Table 2.F.1 provides an overview of the different categories, including example responses falling into each category.

Inter-rater reliability. Each open-text response is independently coded by two reviewers.[41,42] This allows us to estimate the inter-rater reliability, i.e. the degree of agreement among independent raters. We observe a high inter-rater reliability for the response type classification: For 77% of the assigned codes, both reviewers agree.

41. For training purposes, about 50 responses of each vignette were coded independently by four coders. The coding of these responses was subsequently discussed. We do not use these responses to estimate the inter-rater reliability.

42. In Wave 4 (priming study), each response is coded by only one reviewer.

Table 2.F.1. Response type categories

Category	Explanation	Examples
Mechanism	Mentions (part of the) propagation mechanisms of the shock.	"I think people will cutback on expenses." "Banks will be more reluctant to borrow more money, which leads to charging people more interest, which will decrease their economic activity. When businesses have less ability to take out loans, they may hire less people." "An increase in oil prices will cause increases in costs of delivery, driving, heating and production. It would also increase the price of goods derived from oil such as plastics and fertilizers."
Model	Makes reference to a specific economic theory or model. *Occurs only among experts.*	"I am thinking of a textbook NK model of an economy at a steady state experiencing a nominal interest rate shock. Quantitatively, I expect relatively small effects but I do not have much confidence in the actual magnitudes." "transmission via Phillips curve" "The classic AD/AS model is still my reference point for these questions, and in that model this is an aggregate supply shock, raising prices and lowering output through a rise in the costs of production. The increase of 30% is substantial, but the US economy is relatively insulated from oil price fluctuations in the current economy, so I chose relatively small movements in both inflation and unemployment."
Guess	States that response was a guess or indicates uncertainty or low confidence in response.	"I just took a guess." "I don't know anything about any of this stuff. I'm just making guesses. It's not at all easy to understand or speculate about." "I'm not an economist nor do I follow economic trends so these really are wild guesses. It depends on who is in office and how the rest of the world is doing."
Politics	Makes a political statement or talks about what policymakers should do.	"This pandemic has turned this world upside down, due to the Non responce from the former President, it will take forever to get back to even a new normal" "The government should take measures to reduce unemployment" "The effects of a Biden presidency will be disastrous."

Notes: See next page.

Table 2.F.1 (continued): Response type categories

Category	Explanation	Examples
Historical	Mentions how things have evolved in the past or how things typically evolve.	"Any time government spending goes up so does every thing else. Unemployment also tends to go up." "i based on the situation presented to me. i also thought about how it has happened in the past" "what i have seen in the past"
Misunder-standing	Gives a response that indicates misunderstanding of or unwillingness to accept key features of the vignette.	"In the alternative scenario the Federal Target Rate rises and there is no change in economic condition." "The above states that everything would remain the same so I don't believe anything would change with the tax rate" "Even when the government announces that a rate increase is temporary, it never is. The government is so far in debt right now that the only way they can raise more money to service that debt is to inflate the taxes paid by workers."
Re-states prediction	Repeats (part of) the forecasts about unemployment and inflation.	"The unemployment rate wouldn't rise but the inflation rate may rise slightly" "I believe that if the interest rate goes up, so will the inflation rate."
Endogenous shock	Misunderstood the exogenous shocks as happening in response to other events, i.e. as being endogenous.	"The main consideration in my predictions is due to the economy must be doing well in order to have the income tax rate increase. In order to achieve this unemployment must be low. People's income must have jump too." "The federal reserve doesn't change rates just to do so. There has to be a reason. Given that I expect unemployment and or the inflation rate to be a little higher than expected."

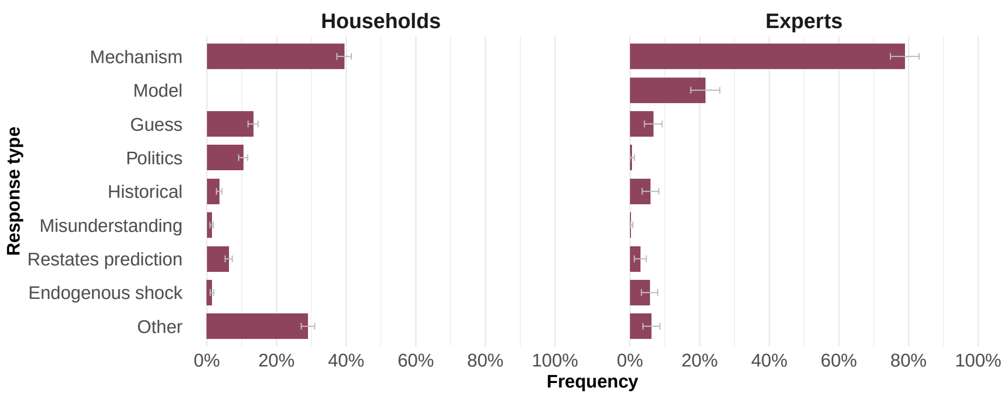
Notes: See next page.

Table 2.F.1 (continued): Response type categories

Category	Explanation	Examples
Other	Residual category.	"the unemployment rate will be high" "I would assume that supply and demand will start to return once the pandemic is largely controlled" "Unless they're going to make more jobs, its going to raise "

Notes: This table provides details and examples on the response types into which we classify responses to the open-text question. Each response is allowed to fall into more than one category. The examples (except for the "model" category) are based on the household sample, given that there is much more heterogeneity in relevant response categories among households than among experts.

Results. The results are shown in Figure 2.F.1. By far the largest fraction of open-text responses among both households and experts are classified as "mechanism associations", i.e. thoughts related to economic propagation mechanisms, such as changes in consumer spending or the hiring decisions of firms. 39% of households express thoughts about propagation mechanisms, while 79% of experts do so. In addition, 22% of experts explicitly refer to an economic model (such as the New-Keynesian model), compared to none of the households. These patterns suggest that experts are more likely than households to think about the shocks through the lens of economic theories. In contrast, households are more likely to comment on general political issues (10%), such as which political party is in power. Almost none of the experts do so. Likewise, 13% of households indicate that their prediction is based, at least in part, on a guess, compared to only 7% of experts. References to economic events from the past are equally frequent among households and experts (4% and 6%, respectively), and only few responses reveal that participants misunderstand features of the vignette or perceive the macroeconomic shocks as endogenous. Finally, 29% of household responses fall into a residual category, which e.g. captures general statements about the economy, compared to 6% among experts. Thus, we find overall similar patterns as based on our structured question of what approach respondents used in their predictions (see Figure 2.A.5).



Notes: This figure presents the manually-coded "response type" classification (open-text data) of households' and experts' responses in Wave 3, averaged across all four vignettes. Error bars display 95% confidence intervals. "Mechanism": Respondent mentions an economic mechanism through which the shock could affect the economy. "Model": Response explicitly mentions an economic model. "Guess": Respondent indicates that they made a guess. "Politics": Political issues and statements. "Historical": Reference to historical events or data. "Misunderstanding": Misunderstanding of vignette instructions. "Restates prediction": Restatement of prediction in the open-ended text. "Endogenous shock": Respondents who misperceive the exogenous shock as endogenous. "Other": Residual category.

Figure 2.F.1. Manually coded "response types" in the open-text data

2.F.2 Mechanism associations

Next, we zoom in on those 39% of responses describing elements of a propagation mechanism (i.e. all responses falling into the "Mechanism" category).

2.F.2.1 Coding scheme

Variable codes. For each response in the "Mechanism" category, we hand-code which specific elements of a propagation mechanism are mentioned. For example, a respondent might mention that firms' costs of borrowing increase, or that households cut back on their spending. We use the following coding scheme: i) We identify variables that are mentioned by the respondent, e.g. "costs borrowing firms" or "labor demand". For many variables, we have a general form (e.g. "demand") and a more specific form (e.g. "demand households") to accommodate different levels of detail given in the open-text responses. ii) When applicable, we code the direction of the change the respondent mentions for this variable ("+" for an increase, "–" for a decrease, "o" for no change; no extension if no direction is mentioned). In addition, we include a range of codes not referring to any specific variable but indicating why a respondent thinks that a shock has no major effect (e.g. "minor" if respondents argue that the shock is too small in size to have any effect). Table 2.F.2 presents all codes we use to classify elements of propagation channels. Table 2.F.4 displays examples of assigned variable codes for selected mechanism responses.

Aggregation into broader mechanisms. Finally, we classify the most commonly mentioned elements of propagation channels into broader classes of mechanism associations. For each of our vignettes, we identify the most prevalent mechanism associations. For example, in the case of the interest rate vignette the most common mechanism associations concern 1) increases in firms' costs, 2) reductions in product demand, and 3) a reduction in labor demand. For instance, if a respondent mentions that firms face higher costs of borrowing or "pass on" higher costs to the consumer, this falls under 1) increases in firms' costs, while 2) reductions in product demand subsume decreases in households' spending or lower investment by firms. If a respondent mentions that firms fire workers or that there are fewer job opportunities, we code this as 3) a reduction in firms' labor demand. Other elements of propagation mechanisms are mentioned less frequently and remain unassigned. For experts, we also subsume some responses coded as "model", since they clearly refer to one of the specific broader mechanisms: "supply shock", "demand shock", and "multiplier". Table 2.F.3 describes how the different elements of mechanism associations are aggregated into broader vignette-specific mechanisms. Table 2.F.4 provides examples for the coding of mechanism responses, including both the specific codes and the aggregate categories that are assigned to the example responses.

Table 2.F.2. Mechanism associations: Variable codes

Code	Explanation
Economic variables:	
borrowing borrowing firms borrowing household borrowing government	Amount borrowed (debt) by different groups, or amount lent by banks to these groups
costs costs firms costs household	costs: Costs faced by different groups other than borrowing costs. costs firms: Production costs, including costs of input goods, wages paid; "Firms need to cover"; "firms need to make up for it", … costs household: Costs of subsistence goods, e.g. heating, gasoline, …
costs borrowing costs borrowing firms costs borrowing household costs borrowing banks costs borrowing government	Borrowing rates and/or access to credit faced by different groups.
demand demand firms demand household demand government	Demand for goods, spending, consumption, …
expected inflation expected unemployment	Expectations of future realizations of macroeconomic variables as propagation mechanisms.
firm prices	Firms' decisions about pricing.
government taxes government finances	government taxes: Tax revenue collected by the government. government finances: Residual category referring to unspecified improvements or deterioration in the government's budget.
growth	GDP growth, overall growth of the economy.
housing housing demand housing supply	Quantity of housing demanded, supplied, or unspecified whether demand or supply is meant.
income	Household income, wages received, purchasing power.
interest interest household	General interest rate category if agent not specified or if not specified whether households' rates on borrowing vs saving are meant.
investment	Investment (expenditure) of firms.
labor demand labor supply labor	labor demand: "Job creation", firm's/government's demand for employees, "Job opportunities". labor supply: Changes in households' desired work hours. labor: Residual category for cases where it is unclear whether the respondent is thinking about labor demand or supply, e.g. "more people work".

Notes: See next page.

Table 2.F.2 (continued): Mechanism associations: Variable codes (continued)

Code	Explanation
Economic variables (continued):	
money	Overall amount of money in circulation, money printing by the central bank.
policy rate	Target rate, central bank rate, policy rate, federal funds rate.
prices stock prices house	Asset prices, "the stock market".
production	Firms' production / supply of goods and services
profit	Firms' profits or profit margin, including firms facing pressure to take actions to keep the profit margin at a certain level.
saving	Amount saved by households.
saving rate	Interest rate earned on savings.
Other codes:	
crowd-out	Crowd-out of other types of demand due to increased government spending e.g. on defense, including crowd-out of other types of government spending.
domestic	Only relevant for the oil price shock. Highlights responses indicating that the domestic oil industry will benefit from the shock/buffer the shock.
green	Only relevant for the oil price shock. Highlights responses indicating that the economy is less dependent on oil than historically due to growing importance of new energy sources.
lag	Shock is perceived to only affect the economy with a lag.
minor	Shock is perceived as too small to have a pronounced effect on the economy.
sector-specific	Shock is perceived to have a small effect because it only affects a certain sector.
state-dependence	Impact of the shock is perceived to depend on initial conditions of the economy.
temporary	Shock is perceived to have no major effect because it is only temporary.

Notes: This table provides details on the definition of variables for the hand-coding of the open-text responses classified as “mechanism” responses. There is no limit on the number of variables that can be used for the coding of each individual response.

Table 2.F.3. Mechanism associations: Aggregated

Aggregate category	Variable codes
	Oil price shock
Firms' costs (+)	costs firms +, firm prices +, supply shock –*
Product demand (–)	costs household +, demand –, demand household –, income –
Labor demand (–)	labor demand –
Oil dependency (–)	domestic, green
	Government spending shock
Crowding out	borrowing government +, costs firms +, crowd-out, government finances –, government taxes +, labor demand –
Product demand (+)	demand +, demand household +, demand shock +*, income +, money +, multiplier*
Labor demand (+)	labor demand +
	Interest rate shock
Firms' costs (+)	costs borrowing firms +, costs firms +, firm prices +
Product demand (–)	costs borrowing household +, costs household +, demand –, demand household –, demand shock –*, income –, investment –, money –
Labor demand (–)	labor demand –
	Income tax shock
Firms' costs (+)	costs firms +, firm prices +
Product demand (–)	costs household +, demand –, demand household –, income –
Labor demand (–)	labor demand –

Notes: This table shows how specific variable codes were aggregated into broader mechanism categories across the four vignettes. *References to economic models occur only in expert responses.

Table 2.F.4. Mechanism associations: Coding examples

Text response	Mechanism code(s)	Aggregated code(s)
Households:		
(Oil price shock) “Unemployment will go up, because business will not be able to afford the increase in oil costs that go with it. Inflation will rise, since so many products are oil based, or use oil in some way, transportation etc.”	costs firms +	*Firms’ costs (+)*
(Government spending shock) “The government will be spending more money, which means the public has less money, I believe. I’m not exactly sure how it works, but that’s my prediction as an uninformed citizen.”	crowd-out, income -	*Crowding out*
(Government spending shock) “Defense spending leads to higher inflation. Higher government spending that’s not really driving income-generation. The impact of unemployment depends on what they’re spending on - are we making more weapons/planes that require more workers. Not sure”	income o, labor demand, sector-specific	None
(Interest rate shock) “with change in fed funds rate upward, unemployment is likely to rise (as cost to business to borrow increases and invest less in expansion) and inflation should in theory be kept in check and even fall.”	costs borrowing firms +, investment -	*Firms’ costs (+), Product demand (–)*
(Interest rate shock) “The rate of federal funds affects banks which in turn affects spending/inflation”	costs borrowing banks, demand	None
(Interest rate shock) “If the money that corporations borrow costs them more, they will charge more accordingly. If the merchandise costs businesses more, they will compensate by raising prices and by eliminating business expenses which means job termination raising the unemployment rate.”	costs borrowing firms +, firm prices +, labor demand -	*Firms’ costs (+), Labor demand (–)*
(Income tax shock) “The income tax increase would decrease disposable income by $400 per household, thus decreasing spending. Plus, an increase in the income tax rate can lower inflation. However, in this scenario the 1% increase does not seem high enough to affect unemployment or inflation significantly in the short-term.”	demand household -, income -, lag, minor	*Product demand (–)*
(Income tax shock) “I believe that the unemployment rate will increase as people will be looking for higher paying jobs. They will need higher paying jobs to off set the increase of taxes.”	costs firms +, costs household +, labor supply -	*Firms’ costs (+), Product demand (–)*

Notes: See next page.

Table 2.F.4 (continued): Mechanism associations: Coding examples

Text response	**Mechanism code(s)**	**Aggregated code(s)**
Experts:		
(Oil price shock) "I see it as a (near)-stagflation scenario: - production costs rise pushing inflation up - unemployment is constant or marginally higher - the Fed does not intervene because it has to balance two objectives (control inflation and support employment) and it knows that monetary expansions are ineffective against aggregate supply shocks"	costs firms +, policy rate o, supply shock -	*Firms' costs (+)*
(Oil price shock) "Negative supply shocks are contractionary and inflationary, since they raise firms' marginal costs. I don't know exactly what the pass-through of oil prices is, but it seems reasonable to expect a moderate pass-through over a short period of time, also given the muted effects of oil price movements of the last decade."	costs firms +, firm prices +, supply shock -	*Firms' costs (+)*
(Government spending shock) "In the short term, the government spending will create more jobs, which decreases u. It can't go down by much though because it's already at the natural rate/ very low. Inflation increases because the government consumption has to come from somewhere and it is likely that it's financed with bonds/seigniorage/ some form of money creation that will raise prices."	demand +, government finances -, labor demand +, money +, state-dependence	*Crowding out, Product demand (+), Labor demand (+)*
(Government spending shock) "Inflation will increase because people have more income and will consume more. Unemployment will decrease because more government spending can finance more employment in the economy."	demand household +, income +, labor demand +	*Product demand (+), Labor demand (+)*
(Interest rate shock) "With higher interest rates, consumption and investment will be typically lower and moved to the future. "	demand household -, investment -	*Product demand (–)*
(Income tax shock) "The tax change is temporary, so should affect non-liquidity constrained households too much. However, some households at the constraint will spend less, which underlies my projection for slightly higher unemployment (and slightly lower inflation)."	demand household -, temporary	*Product demand (–)*
(Income tax shock) "households will decrease working hours but if no other change takes place they should borrow to smooth out the negative shock so that the shock should not affect inflation as it is temporary. For unemployment, I would expect a fall in labor hours/employment because of the temporal shift in labor supply, so a slight fall in unemployment should occur, for me here the substitution effect is higher than the income effect, as the shock is pretty temporary. As workers will not want to break the employment relationship temporarily I expect the fall to be small"	borrowing household +, labor supply -, temporary	None

Notes: This table provides selected examples for how responses to the open-text question included in Wave 3 falling under the mechanism category are coded and which broader, more aggregated mechanisms are assigned. Wave 3 only contained rise scenarios, so all examples refer to scenarios where the shock variable of interest increases.

Inter-rater reliability. We observe a high inter-rater reliability for the aggregated mechanism associations among households. In 95% of cases, the two coders independently agree whether or not a specific mechanism association should be assigned to the response. The inter-rater reliability is only slightly lower for expert responses (92%). If we restrict attention to cases where at least one coder detects a "mechanism" response, the inter-rater-reliability of mechanism associations is still very high (88% and 91% for households and experts respectively).

2.F.2.2 Validation of structured measures of thoughts

Table 2.F.5 shows that the structured measures of thoughts and the hand-coded mechanism associations based on the open-ended data are mostly strongly and statistically significantly correlated in the expected directions under the different vignettes. For instance, indicating a negative supply-side channel under the structured question in the oil vignette increases the likelihood of mentioning an increase in firms' cost in the open-text data by 26 p.p. Selecting negative demand-side channels in the structured question under the income tax vignette increases the likelihood of mentioning decreases in product demand by 17.9 p.p. The coefficients are naturally smaller than one because i) there is no one-to-one correspondence between the hand-coded mechanism associations and the more nuanced channels in the structured questions, and ii) both structured and open-ended data likely contain measurement error.

Table 2.F.6 shows similar results for experts.

Table 2.F.5. Households: Structured propagation channels predict manually-coded open-text data

Panel A

	Oil price			Government spending		
	Firms' costs (+)	Product D (–)	Labor D (–)	Crowding-out	Product D (+)	Labor D (+)
	(1)	(2)	(3)	(4)	(5)	(6)
Supply (–)	0.260***	0.144***	0.219***			
	(0.028)	(0.031)	(0.025)			
Demand (–)	0.018	0.145***	0.038			
	(0.034)	(0.032)	(0.032)			
Crowding-out				0.111***	−0.073***	−0.132***
				(0.035)	(0.023)	(0.029)
Demand (+)				−0.139***	0.119***	0.237***
				(0.029)	(0.025)	(0.031)
Constant	0.024	0.019	0.002	0.165***	0.066***	0.105***
	(0.018)	(0.023)	(0.017)	(0.026)	(0.017)	(0.020)
Observations	557	557	557	519	519	519
R^2	0.083	0.067	0.072	0.081	0.054	0.127

Panel B

	Federal funds rate			Income tax rates		
	Firms' costs (+)	Product D (–)	Labor D (–)	Firms' costs (+)	Product D (–)	Labor D (–)
	(1)	(2)	(3)	(4)	(5)	(6)
Supply (–)	0.163***	0.057**	0.157***			
	(0.026)	(0.026)	(0.028)			
Demand (–)	−0.029	0.044*	0.045			
	(0.028)	(0.026)	(0.028)			
Supply (–)				0.067***	−0.019	0.055**
				(0.022)	(0.032)	(0.027)
Demand (–)				0.015	0.179***	0.086***
				(0.021)	(0.032)	(0.027)
Constant	0.036**	0.044**	0.018	0.025*	0.090***	0.038**
	(0.016)	(0.020)	(0.018)	(0.013)	(0.021)	(0.015)
Observations	520	520	520	530	530	530
R^2	0.067	0.016	0.064	0.020	0.056	0.029

Notes: This table presents data from Wave 3 of the household survey. It regresses the manually-coded mechanism associations that respondents mention in their open-text response on the selected propagation channels in the structured question (see Figure 2.4.1). *Explanatory variable, propagation channels, structured question*: "Supply (–)" takes value 1 for respondents who choose a negative supply-side propagation channel in the structured question. "Demand (–)" and "Demand (+)" take value 1 for respondents choosing the negative or positive demand-side propagation mechanism in the structured question respectively. "Crowding-out" takes value 1 for respondents who select the channel that demand falls due to higher expected future taxes. *Outcome variable, manually-coded mechanism associations, open-text data*: "Firm's costs (+)" takes value 1 for respondents who mention an increase in firms' costs. "Product D (–)" takes value 1 for respondents who mention a decrease in product demand. Likewise, "Labor D (+)" represents an increase in labor demand, "Labor D (–)" a decrease in labor demand, "Crowding-out" the negative effects of increases in government spending, and "Product D (+)" an increase in product demand. See appendix Section 2.F for further details on the coding of the open-text data which varies across vignettes. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.F.6. Experts: Structured propagation channels predict manually-coded open-text data

Panel A

	Oil price			Government spending		
	Firms' costs (+)	Product D (−)	Labor D (−)	Crowding-out	Product D (+)	Labor D (+)
	(1)	(2)	(3)	(4)	(5)	(6)
Supply (−)	0.539***	0.022	0.188***			
	(0.060)	(0.088)	(0.047)			
Demand (−)	0.031	0.044	0.010			
	(0.100)	(0.065)	(0.079)			
Crowding-out				0.295*	0.022	0.127
				(0.151)	(0.155)	(0.144)
Demand (+)				−0.051	0.225*	0.182**
				(0.096)	(0.116)	(0.071)
Constant	−0.010	0.069	−0.003	0.168*	0.272***	0.040
	(0.034)	(0.089)	(0.026)	(0.086)	(0.097)	(0.046)
Observations	91	91	91	88	88	88
R^2	0.137	0.006	0.030	0.071	0.039	0.059

Panel B

	Federal funds rate			Income tax rates		
	Firms' costs (+)	Product D (−)	Labor D (−)	Firms' costs (+)	Product D (−)	Labor D (−)
	(1)	(2)	(3)	(4)	(5)	(6)
Supply (−)	0.033	0.077	0.036			
	(0.063)	(0.103)	(0.073)			
Demand (−)	0.019	0.336***	0.114***			
	(0.062)	(0.059)	(0.039)			
Supply (−)				0.191	−0.033	−0.095**
				(0.131)	(0.119)	(0.040)
Demand (−)				−0.027	0.622***	−0.019
				(0.038)	(0.083)	(0.069)
Constant	0.050	−0.013	−0.006	0.044	0.090	0.103
	(0.057)	(0.018)	(0.012)	(0.028)	(0.068)	(0.064)
Observations	92	92	92	100	100	100
R^2	0.005	0.102	0.030	0.089	0.292	0.009

Notes: This table presents data from Wave 3 of the expert survey. It regresses the manually-coded mechanism associations that respondents mention in their open-text response on the selected propagation channels in the structured question (see Figure 2.4.1). *Explanatory variable, propagation channels, structured question*: "Supply (−)" takes value 1 for respondents who choose a negative supply-side propagation channel in the structured question. "Demand (−)" and "Demand (+)" take value 1 for respondents choosing the negative or positive demand-side propagation mechanism in the structured question respectively. "Crowding-out" takes value 1 for respondents who select the channel that demand falls due to higher expected future taxes. *Outcome variable, manually-coded mechanism associations, open-text data*: "Firm's costs (+)" takes value 1 for respondents who mention an increase in firms' costs. "Product D (−)" takes value 1 for respondents who mention a decrease in product demand. Likewise, "Labor D (+)" represents an increase in labor demand, "Labor D (−)" a decrease in labor demand, "Crowding-out" the negative effects of increases in government spending, and "Product D (+)" an increase in product demand. See appendix Section 2.F for further details on the coding of the open-text data which varies across vignettes. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

2.F.2.3 Hand-coded mechanism associations across vignettes

We first demonstrate robustness of our descriptive evidence on variation in thoughts of propagation channels across vignettes and samples presented in Section 2.4.3. Figure 2.F.2 shows the fractions of respondents mentioning different aggregated mechanism associations, while Figure 2.F.3 displays the results for the more detailed mechanism codes. We focus our discussion on the aggregated mechanism codes.

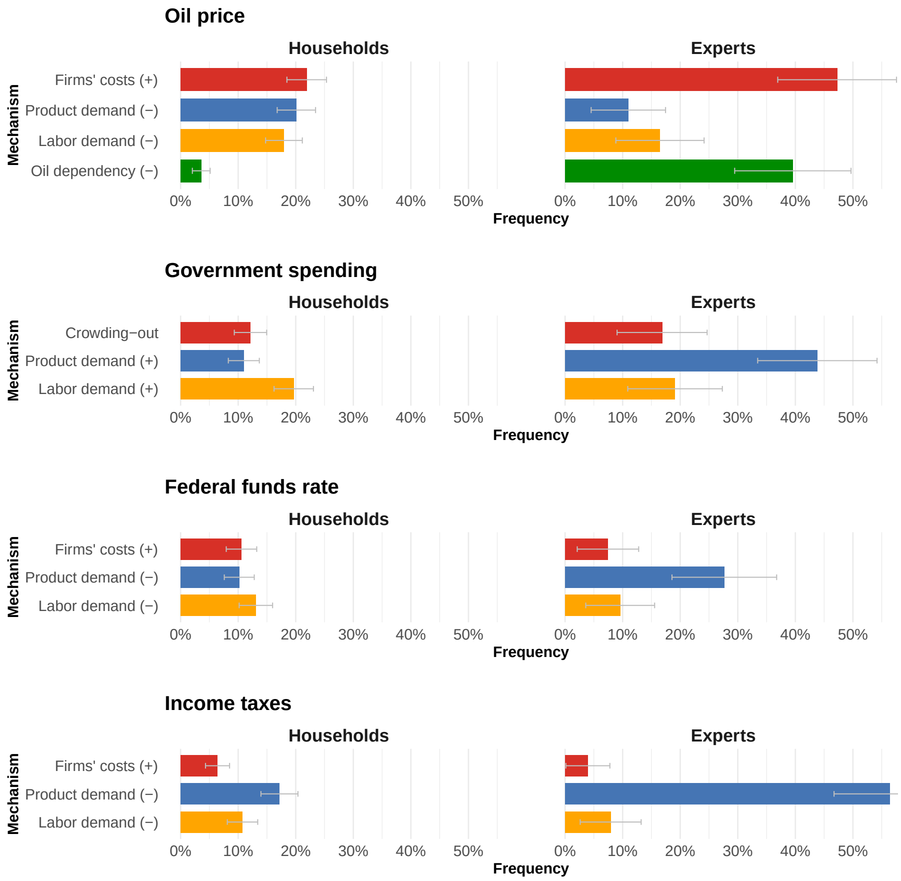
Heterogeneity within the household sample. The left column of Figure 2.F.2 presents the fractions of households mentioning different mechanisms across the four vignettes. Overall, smaller fractions of households mention the different mechanisms than in the structured data. This should be seen in light of the fact that only 39% of households provide a response that we classify as "mechanism response".

How do households' thoughts vary across the different shocks? Similarly as in our results based on the structured data, increases in firms' costs are most frequently mentioned under the oil price vignette (22%), but are still quite common in both the income tax vignette (6%) and the monetary policy vignette (11%). Similarly, the product demand channel is most commonly mentioned in the oil price vignette (20%), and less frequently in the income tax vignette, the government spending and interest rate vignette, even though these shocks are conventionally viewed as demand-side shocks. The labor demand channel is most commonly mentioned in the government spending vignette (20%), but also frequently mentioned in the oil price vignette (18%) as well as the interest rate and income tax vignettes (13% and 11%, respectively). Taken together, the hand-coded mechanism associations provide a similar picture as the structured measures of thoughts: households think of different propagation channels depending on the context, but the variation across contexts is not fully aligned with textbook models.

Heterogeneity within the expert sample. The right column of Figure 2.F.2 presents the fractions of experts mentioning different mechanisms across the four vignettes. The figure highlights that there is a lot of variation in the mechanisms that come to experts' minds across vignettes. Firms' costs are very frequently mentioned in the oil price vignette (47%), and much less frequently in the interest rate vignette (7%) and the income tax vignette (4%). Similarly, there is substantial variation in how frequently the product demand channel is mentioned across vignettes. While a very large fraction of experts mention the product demand channel in the income tax and government spending vignettes (56% and 44%, respectively), smaller fractions mention this mechanism in the federal funds rate and oil price vignettes (28% and 11%, respectively). Finally, there is somewhat less variation across vignettes in the labor demand associations. While labor demand is somewhat more frequently mentioned in the oil price and government spending vignettes (16% and 19%, respectively), it is less frequently mentioned in the interest rate and income tax vignettes (10% and 8%, respectively). Taken together, the data on hand-coded mechanism associations confirm the findings from the structured measures of thoughts: the variation in experts' associations across contexts suggests that experts retrieve textbook models when thinking about macroeconomic shocks.

Similarities and differences between households and experts. The hand-coded data on mechanism associations reveal striking differences between households and experts in terms of mechanism associations that come to their minds.[43] While in the context of the oil price vignette the differences in the relative importance of mechanisms between the household and expert samples are more muted, differences are quite striking in all other three vignettes. Experts are relatively more likely to think of mechanism associations related to product demand compared to households. Conversely, households are relatively more likely to think of mechanism associations related to either crowd-out or increases in firms' costs compared to experts. The hand-coded data on mechanism associations thus paint a very similar picture as the structured data on mechanism associations.

43. Since the fraction of respondents mentioning any mechanism is much larger among experts compared to households, we focus our description of results not on the importance of levels but instead on differences in the relative importance that households and experts attach to different mechanisms.



Coding

Oil price

Firms' costs (+): costs firms +, firm prices +, supply shock –*

Product demand (–): costs household +, demand –, demand household –, income –

Labor demand (–): labor demand –

Oil dependency (–): domestic, green

Government spending

Crowding out: borrowing government +, costs firms +, crowd-out, government finances –, government taxes +, labor demand –

Product demand (+): demand +, demand household +, demand shock +*, income +, money +, multiplier*

Labor demand (+): labor demand +

Federal funds rate

Firms' costs (+): costs borrowing firms +, costs firms +, firm prices +

Product demand (–): costs borrowing household +, costs household +, demand –, demand household –, demand shock –*, income –, investment –, money –

Labor demand (–): labor demand –

Income taxes

Firms' costs (+): costs firms +, firm prices +

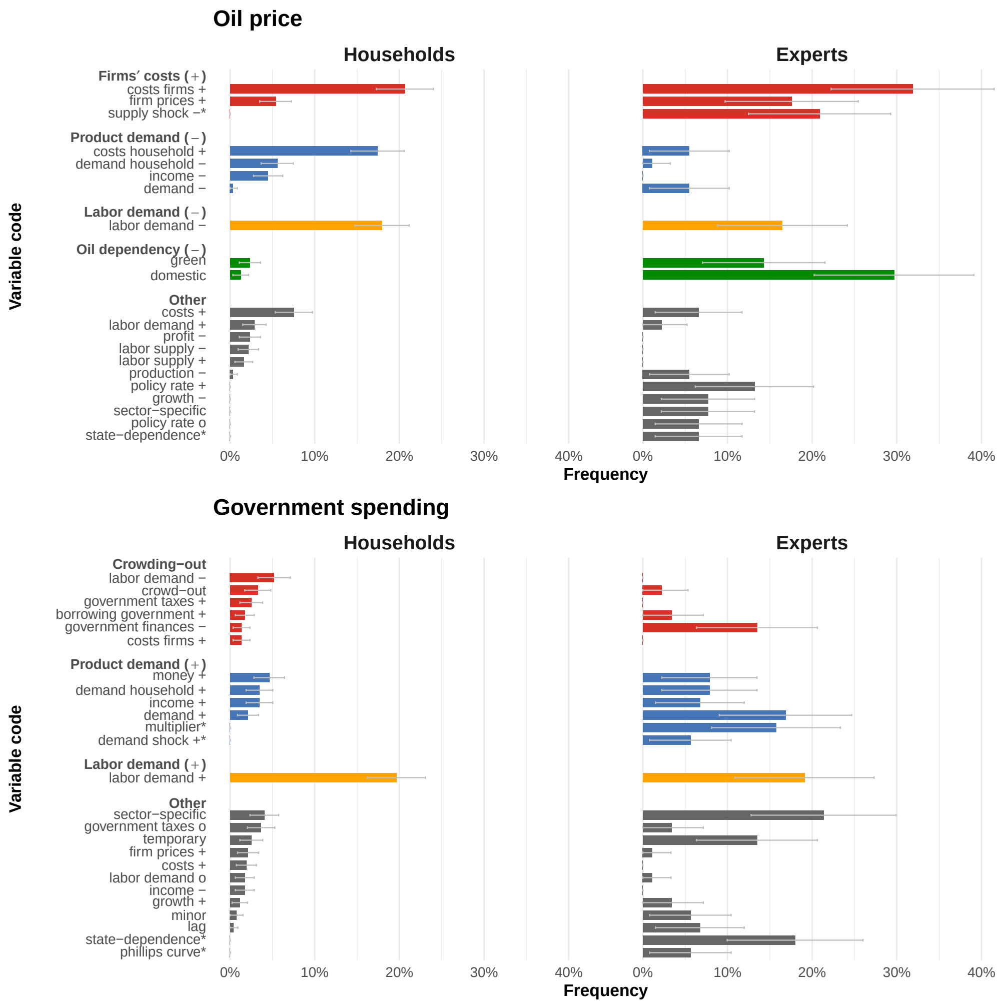
Product demand (–): costs household +, demand –, demand household –, income –

Labor demand (–): labor demand –

*References to economic models occur only in expert responses.

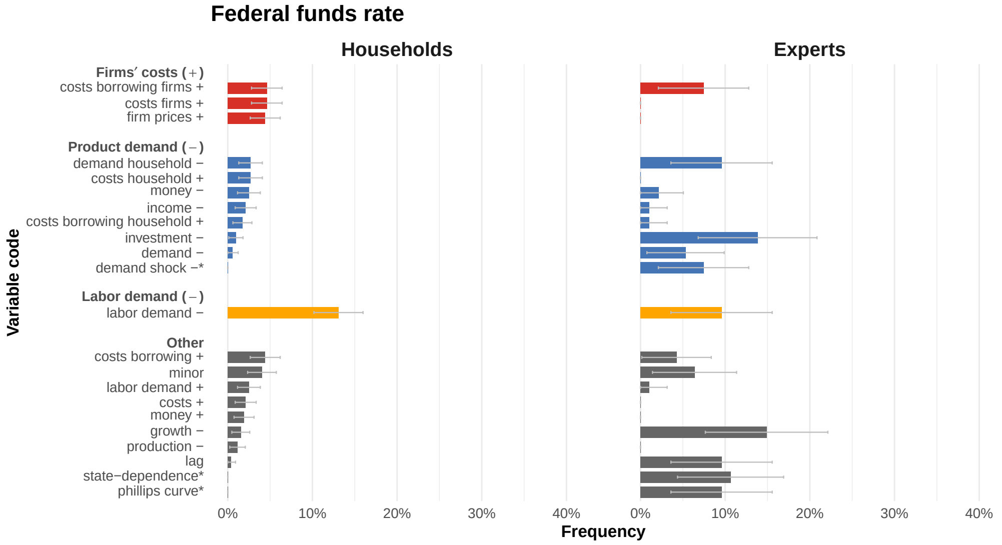
Notes: This figure shows the shares of households (left panel) and of experts (right panel) mentioning various mechanisms in their open responses in Wave 3. The responses are manually reviewed and assigned to various variable codes, which are then grouped into different mechanisms (see "Coding" panel). The results are displayed for each vignette. Error bars display 95% confidence intervals.

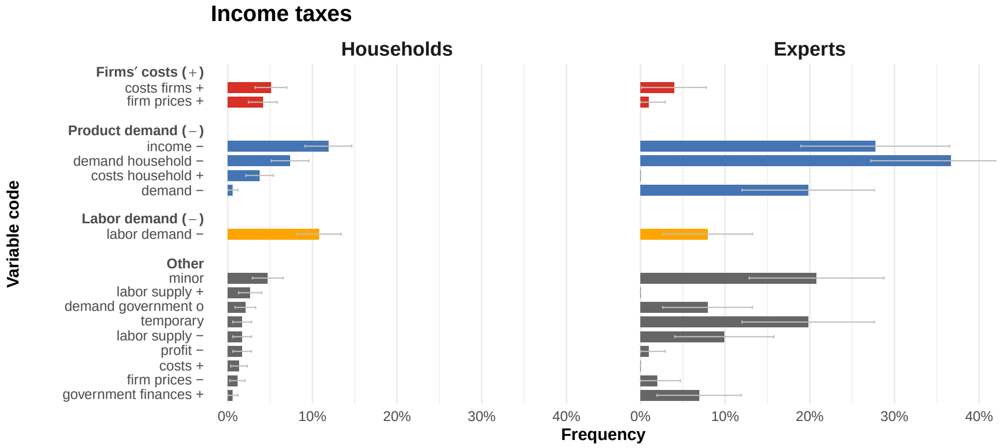
Figure 2.F.2. Mechanism associations across vignettes (open-text question)



Notes: See next page.

Figure 2.F.3. Mechanism associations across vignettes (open-text data)





Notes: This figure shows the shares of households (left panel) and experts (right panel) mentioning various mechanisms in their open-text response in Wave 3. The responses are manually reviewed and assigned to various variable codes which are then grouped into different mechanisms (bold labels). Only variable codes that are mentioned by at least 1% of households or 5% of experts are displayed. The error bars indicate 95% confidence intervals.

Figure 2.F.3 (continued): Mechanism associations across vignettes (open-text data)

Correlations between associations and predictions. Table 2.F.7 documents correlations between the hand-coded mechanism associations and forecasts in the household sample. Households mentioning decreases in product demand and decreases in labor demand expect a larger increase in unemployment in response to the oil price shock. Moreover, mentioning a decreasing oil dependency of the US economy is associated with lower inflation and unemployment increases in response to an increase in oil prices. Crowd-out associations are robustly associated with larger increases in inflation and unemployment in response to an increase in government spending. Moreover, mentioning increases in labor demand is robustly associated with predicting larger decreases in the unemployment rate in response to government spending increases. While mechanism associations do not robustly correlate with predictions about the inflation response in the interest rate vignette, they do strongly correlate with unemployment predictions. Households mentioning decreases in product demand, and decreases in labor demand predict a larger increase in unemployment in response to an interest rate hike. Households that mention increases in firms' costs and decreases in labor demand predict higher increases in inflation in response to income tax hikes. Households that mention decreases in labor demand also predict higher increases in unemployment in response to income tax hikes. Taken together, similarly as based on the structured data of thoughts (shown in Table 2.4.2), thoughts of the different propagation channels are mostly significantly associated with households' forecasts of inflation and unemployment responses to the different shocks in the expected directions.

Table 2.F.8 reports an analogous analysis for Wave 3 of the expert survey. The thoughts of experts correlate with their forecasts in directions suggested by conventional macroeconomic models. For instance, experts who think of a rise of product demand in the government vignette predict higher inflation but lower unemployment. Likewise, experts who think of a fall in product demand in the income tax vignette predict lower inflation and higher unemployment.

Table 2.F.7. Households: Relationship between manually-coded mechanism associations (open-text data) and predictions

Oil price

	Inflation $\Delta\pi$		Unemployment Δu	
	(1)	(2)	(3)	(4)
Firms' costs (+)	0.126***	0.043	0.002	0.033
	(0.048)	(0.055)	(0.045)	(0.052)
Product dem. (–)	0.158***	0.056	0.119**	0.156***
	(0.048)	(0.054)	(0.047)	(0.054)
Labor demand (–)	0.148***	0.084	0.292***	0.315***
	(0.053)	(0.055)	(0.049)	(0.052)
Oil depend. (–)	−0.222*	−0.386***	−0.304**	−0.245*
	(0.117)	(0.126)	(0.120)	(0.128)
Any mech.		0.207***		−0.075
		(0.060)		(0.062)
Constant	0.378***	0.339***	0.254***	0.268***
	(0.026)	(0.029)	(0.027)	(0.030)
Observations	557	557	557	557
R^2	0.086	0.104	0.091	0.094

Government spending

	Inflation $\Delta\pi$		Unemployment Δu	
	(1)	(2)	(3)	(4)
Crowding-out	0.245***	0.225***	0.260***	0.364***
	(0.059)	(0.073)	(0.069)	(0.080)
Product dem. (+)	0.138**	0.128*	−0.100*	−0.051
	(0.064)	(0.071)	(0.051)	(0.054)
Labor dem. (+)	−0.042	−0.060	−0.479***	−0.387***
	(0.054)	(0.066)	(0.045)	(0.060)
Any mech.		0.026		−0.132**
		(0.066)		(0.064)
Constant	0.296***	0.292***	0.086***	0.105***
	(0.028)	(0.031)	(0.030)	(0.033)
Observations	519	519	519	519
R^2	0.031	0.031	0.170	0.174

Notes: See next page.

Table 2.F.7 (continued): Households: Relationship between manually-coded mechanism associations (open-text data) and predictions

Federal funds target rate

	Inflation $\Delta\pi$		Unemployment Δu	
	(1)	(2)	(3)	(4)
Firms' costs (+)	0.081	0.074	0.083	0.126*
	(0.067)	(0.072)	(0.065)	(0.071)
Product dem. (–)	−0.019	−0.029	0.149**	0.213***
	(0.072)	(0.077)	(0.067)	(0.073)
Labor demand (–)	0.042	0.034	0.247***	0.298***
	(0.065)	(0.069)	(0.064)	(0.071)
Any mech.		0.018		−0.115*
		(0.056)		(0.061)
Constant	0.285***	0.282***	0.143***	0.164***
	(0.023)	(0.025)	(0.024)	(0.027)
Observations	520	520	520	520
R^2	0.006	0.006	0.057	0.064

Income taxes

	Inflation $\Delta\pi$		Unemployment Δu	
	(1)	(2)	(3)	(4)
Firms' costs (+)	0.273***	0.309***	−0.025	0.037
	(0.070)	(0.073)	(0.096)	(0.096)
Product dem. (–)	−0.088	−0.019	−0.027	0.091
	(0.055)	(0.062)	(0.052)	(0.060)
Labor demand (–)	0.140**	0.183***	0.236***	0.311***
	(0.063)	(0.068)	(0.073)	(0.075)
Any mech.		−0.107**		−0.183***
		(0.054)		(0.051)
Constant	0.351***	0.371***	0.228***	0.261***
	(0.024)	(0.027)	(0.025)	(0.029)
Observations	530	530	530	530
R^2	0.035	0.041	0.019	0.035

Notes: This table shows data from Wave 3 of the household survey. It regresses the predicted inflation changes ($\Delta\pi$) and unemployment changes (Δu) on the manually-coded mechanism associations that respondents mention in their open-text response. "Firm's costs (+)" takes value 1 for respondents who mention an increase in firms' costs. "Product dem. (–)" takes value 1 for respondents who mention a decrease in product demand. Likewise, "Labor demand (–)" represents a decrease in labor demand. "Oil depend." a decrease in the US economy's dependency on oil, "Crowding-out" the negative effects of increases in government spending, and "Product dem. (+)" an increase in product demand. "Any mech." takes value 1 if the response mentions at least one economic mechanism through which the shock could affect the economy. See appendix Section 2.F for further details on the coding of the open-text data which varies across vignettes. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.F.8. Experts: Relationship between manually-coded mechanism associations (open-text data) and predictions

Oil price

	Inflation $\Delta\pi$		Unemployment Δu	
	(1)	(2)	(3)	(4)
Firms' costs (+)	0.157**	0.125*	0.032	0.019
	(0.076)	(0.067)	(0.075)	(0.076)
Product dem. (–)	−0.080	−0.089	−0.134	−0.138
	(0.111)	(0.109)	(0.170)	(0.172)
Labor demand (–)	0.093	0.083	0.323***	0.318***
	(0.092)	(0.092)	(0.105)	(0.107)
Oil depend. (–)	−0.056	−0.094	−0.286***	−0.301***
	(0.074)	(0.067)	(0.072)	(0.076)
Any mech.		0.148		0.061
		(0.193)		(0.163)
Constant	0.315***	0.219	0.351***	0.311**
	(0.084)	(0.186)	(0.078)	(0.147)
Observations	91	91	91	91
R^2	0.072	0.087	0.239	0.241

Government spending

	Inflation $\Delta\pi$		Unemployment Δu	
	(1)	(2)	(3)	(4)
Crowding-out	−0.023	−0.041	0.043	0.054
	(0.094)	(0.092)	(0.071)	(0.071)
Product dem. (+)	0.150**	0.134**	−0.137**	−0.128**
	(0.060)	(0.057)	(0.057)	(0.057)
Labor dem. (+)	−0.004	−0.023	−0.085	−0.074
	(0.047)	(0.048)	(0.056)	(0.056)
Any mech.		0.108		−0.062
		(0.109)		(0.091)
Constant	0.183***	0.106	−0.133***	−0.089
	(0.045)	(0.105)	(0.036)	(0.084)
Observations	89	89	89	89
R^2	0.067	0.084	0.102	0.109

Notes: See next page.

Table 2.F.8 (continued): Experts: Relationship between manually-coded mechanism associations (open-text data) and predictions

Federal funds target rate

	Inflation $\Delta\pi$		Unemployment Δu	
	(1)	(2)	(3)	(4)
Firms' costs (+)	−0.021	0.009	−0.040	−0.077
	(0.138)	(0.140)	(0.160)	(0.159)
Product dem. (−)	−0.133**	−0.097	0.112*	0.067
	(0.053)	(0.059)	(0.058)	(0.063)
Labor demand (−)	0.019	0.040	0.135	0.109
	(0.126)	(0.124)	(0.098)	(0.102)
Any mech.		−0.103		0.126*
		(0.071)		(0.071)
Constant	−0.128***	−0.088*	0.185***	0.136***
	(0.041)	(0.051)	(0.041)	(0.050)
Observations	94	94	94	94
R^2	0.037	0.060	0.053	0.085

Income taxes

	Inflation $\Delta\pi$		Unemployment Δu	
	(1)	(2)	(3)	(4)
Firms' costs (+)	0.210*	0.178	0.255**	0.277**
	(0.115)	(0.117)	(0.115)	(0.117)
Product dem. (−)	−0.149***	−0.180***	0.136***	0.158***
	(0.040)	(0.044)	(0.043)	(0.046)
Labor demand (−)	0.153**	0.148**	−0.052	−0.049
	(0.063)	(0.065)	(0.075)	(0.075)
Any mech.		0.141**		−0.096
		(0.065)		(0.076)
Constant	−0.085***	−0.194***	0.120***	0.194***
	(0.032)	(0.054)	(0.033)	(0.067)
Observations	101	101	101	101
R^2	0.198	0.229	0.113	0.126

Notes: This table shows data from Wave 3 of the expert survey. It regresses the predicted inflation changes ($\Delta\pi$) and unemployment changes (Δu) on the manually-coded mechanism associations that respondents mention in their open-text response. "Firm's costs (+)" takes value 1 for respondents who mention an increase in firms' costs. "Product dem. (−)" takes value 1 for respondents who mention a decrease in product demand. Likewise, "Labor demand (−)" represents a decrease in labor demand. "Oil depend." a decrease in the US economy's dependency on oil, "Crowding-out" the negative effects of increases in government spending, and "Product dem. (+)" an increase in product demand. "Any mech." takes value 1 if the response mentions at least one economic mechanism through which the shock could affect the economy. See appendix Section 2.F for further details on the coding of the open-text data which varies across vignettes. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

2.F.2.4 Effects of priming exercise on hand-coded mechanism associations

In this section, we discuss the effects of the priming treatment on hand-coded mechanism associations. Table 2.F.9 shows that respondents whose attention is directed towards costs are 6.3 p.p. more likely to mention associations related to increases in firms' costs ($p < 0.01$), compared to a control mean of 7.3 percent. They are also 3 percentage points less likely to mention associations related to decreases in product demand ($p < 0.05$), compared to a control mean of 8.3 percent. Respondents in the "demand prime condition" are 4.1 p.p. more likely to mention associations related to decreases in product demand. The differences in the likelihood of mentioning increases in firms' costs and decreases in product demand are significantly different across the costs prime and demand prime conditions ($p = 0.01$ and $p < 0.01$, respectively). Thus, our results based on the hand-coded mechanism associations are consistent with the findings based on the word-counting exercise presented in Table 2.4.4. This gives us further reassurance that our treatments successfully shifted attention to the cost or the demand side of the propagation of the shock.

Table 2.F.9. Effects of priming study on manually-coded mechanism associations (open-text data)

	Mechanism associations (open-text data)		**Inflation prediction**
	Firms' costs (+)	Product demand (–)	$\Delta\pi$
	(1)	(2)	(3)
Costs prime	0.063***	−0.030**	0.021
	(0.020)	(0.015)	(0.031)
Demand prime	−0.008	0.041**	−0.057**
	(0.016)	(0.019)	(0.029)
Constant	0.073***	0.083***	0.366***
	(0.009)	(0.010)	(0.017)
***p*: Costs = Demand**	**0.001**	**<0.001**	**0.028**
Observations	1,521	1,521	1,521
R^2	0.010	0.008	0.004

Notes: This table presents results from the priming study which focuses on the interest rate vignette (Wave 4 of the household survey). “Costs prime” takes value 1 for respondents randomly assigned to be primed on the costs of production. “Demand prime” takes value 1 for respondents randomly assigned to be primed on product demand. As before (see Figure 2.F.2), “Firms' costs (+)” takes value 1 for respondents who mention increases in firms' borrowing costs, firms' costs, or firms' prices. “Product demand (–)” takes value 1 for respondents who mention increases in costs of borrowing for households, increases in the costs of households more generally, decreases in the demand of households or their income, or a decrease in firms' investments. $\Delta\pi$ denotes the perceived reaction of the inflation rate. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

2.F.2.5 Effects of experiences on hand-coded mechanism associations

In this section, we show correlations between experiences and thoughts as measured with hand-coded mechanism associations. Panel A of Table 2.F.10 shows that respondents who indicate to have more experiences with positive demand side-mechanisms are more likely to mention increases in labor demand in the open-text question ($p < 0.05$) under the government spending shock. Conversely, respondents who have more experiences with crowd-out channels are more likely to mention mechanism associations related to crowding-out effects ($p < 0.05$), and are less likely to indicate increases in product demand ($p < 0.01$) or labor demand ($p < 0.01$).

Panel B of Table 2.F.10 shows that respondents who were either personally employed by a company receiving contracts from the US military or have someone among their friends and family members who was employed by such a company think more about mechanisms related to increases in labor demand when they make their forecasts of the effects of government spending shocks ($p < 0.05$).

Panel C of Table 2.F.10 shows that individuals born before 1962 are more likely to think of increases in production costs ($p < 0.10$) and decreases in product demand ($p < 0.05$) when making predictions under the oil vignette.

These results based on the hand-coded mechanism associations are consistent with the results from the word-counting exercise and the structured questions on thoughts presented in Table 2.4.5. This gives us further reassurance that experiences are significantly associated with the thoughts that come to respondents' minds.

Table 2.F.10. Households' experiences correlate with manually-coded mechanism associations (open-text data)

(A) Government spending: Experience with propagation channels (std. indices)					
	Mechanisms associations (open-text data)			**Predictions**	
	Crowding-out	Prod. demand (+)	Labor demand (+)	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)
Exp. crowding-out	0.043**	−0.045***	−0.056***	0.004	0.106***
	(0.018)	(0.017)	(0.021)	(0.026)	(0.029)
Exp. prod. demand (+)	0.018	0.021	0.050**	0.038	−0.109***
	(0.016)	(0.017)	(0.021)	(0.025)	(0.030)
Controls	✓	✓	✓	✓	✓
Observations	483	483	483	483	483
R^2	0.093	0.074	0.164	0.142	0.180
(B) Government spending: Ever worked for military supplier (self/friend, binary indicator)					
	Crowding-out	Prod. demand (+)	Labor demand (+)	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)
Yes	−0.011	0.017	0.081**	−0.024	−0.101**
	(0.031)	(0.027)	(0.035)	(0.045)	(0.049)
Controls	✓	✓	✓	✓	✓
Observations	483	483	483	483	483
R^2	0.069	0.059	0.159	0.137	0.155
(C) Oil price: Experienced OPEC crisis (born before 1962, binary indicator)					
	Firms' costs (+)	Prod. demand (–)	Labor demand (–)	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)
Yes	0.074*	0.075**	0.079**	0.208***	0.202***
	(0.039)	(0.037)	(0.038)	(0.044)	(0.043)
Controls	✓	✓	✓	✓	✓
Observations	521	521	521	521	521
R^2	0.066	0.072	0.041	0.080	0.074

Notes: This table presents results from Wave 3 (Panel C) and Wave 5 (Panel A and B) of the household survey. It asks whether respondents who made experiences related to the vignettes have different manually-coded mechanism associations (columns 1-3, open-text data) and make different forecasts (inflation: $\Delta\pi$, unemployment: Δu; columns 4-5). The right-hand-side experience variable varies across panels. In Panel A, "Experienced crowding-out" and "Experienced product demand (+)" are standardized indices of self-reported experiences with crowding-out and positive demand-side channels, respectively. In Panel B, "Yes" is a binary dummy taking value 1 if respondents themselves or friends/family of them ever worked for a company that sells to the US military. In Panel C, "Yes" is a binary dummy taking value 1 if respondents were born before 1962, a proxy that they experienced the OPEC crisis. Control variables comprise age (except for Panel C), log income, inflation and unemployment forecasts in the baseline scenario, as well as binary indicators for gender, college education, being a Republican, having taken an economics course at the college level, and census regions. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Appendix 2.G Key screenshots for priming experiment

Demand prime

Step 1: Prediction of demand for firms' goods and services

Think about the demand for products and services that US firms face. Please complete the following sentence.

In the alternative scenario (federal funds rate rises), US firms face _______ than in the baseline scenario (federal funds rate constant).

Your thoughts in step 1

What are your main considerations in making the above prediction?

Please respond in 2-3 sentences.

Step 2: Prediction of inflation

Think about the US inflation rate over the 12 months from January to December 2025. Please complete the following sentence.

In the alternative scenario (federal funds rate rises), the inflation rate will be approximately _______ than in the baseline scenario (federal funds rate constant).

Your thoughts in step 2

Above, you predict how the change in the alternative scenario affects the US inflation.

Please tell us how you come up with your predictions.

What are your main considerations in making the predictions in step 2?

Please respond in 2-3 sentences.

Cost prime

Step 1: Prediction of firms' costs

Think about the costs of doing business that US firms face. Please complete the following sentence.

In the alternative scenario (federal funds rate rises), US firms face ________ than in the baseline scenario (federal funds rate constant).

Your thoughts in step 1

What are your main considerations in making the above prediction?

Please respond in 2-3 sentences.

Step 2: Prediction of inflation

Think about the US inflation rate over the 12 months from January to December 2025. Please complete the following sentence.

In the alternative scenario (federal funds rate rises), the inflation rate will be approximately ________ than in the baseline scenario (federal funds rate constant).

Your thoughts in step 2

Above, you predict how the change in the alternative scenario affects the US inflation. Please tell us how you come up with your predictions.

What are your main considerations in making the predictions in step 2?

Please respond in 2-3 sentences.

Control

Prediction of inflation

Think about the US inflation rate over the 12 months from January to December 2025.
Please complete the following sentence.

In the alternative scenario (federal funds rate rises), the inflation rate will be approximately ________ than in the baseline scenario (federal funds rate constant).

Your thoughts

Above, you predict how the change in the alternative scenario affects the US inflation rate.
Please tell us how you come up with your prediction.
What are your main considerations in making the prediction?
Please respond in 2-3 sentences.

Only for control group: Demand and cost question

Important!

On the next page, your task is to predict how the change between the baseline and the alternative scenario affects another aspect of the US economy and to write down **what considerations you have on your mind** while you make your prediction.

Therefore, while you read the scenario description and think about its consequences for the US economy, please **pay special attention to what comes to your mind**. Of course, there are no right or wrong answers. Just write down your thoughts. Your response is very valuable for this research project.

Please take your time to respond carefully.

Federal funds target rate

Please think about the following two hypothetical scenarios once more.

Compare the alternative scenario to the baseline scenario where the federal funds rate stays constant.

Reminder: Please assume it is the 1st of January 2025. The COVID-19 pandemic is over. The US economy has fully recovered and is back to "business as usual".

Baseline scenario: Federal funds target rate stays constant

Imagine the **federal funds target rate** stays **constant**. That is, in its first meeting in 2025, the Federal Open Market Committee announces that it will keep the interest rate constant at 2.5%.

The committee announces it does so with no changes in their assessment of the economic conditions.

Alternative scenario: Federal funds target rate rises

Imagine the **federal funds target rate** is unexpectedly **0.5 percentage points higher**. That is, in its first meeting in 2025, the Federal Open Market Committee announces that it is raising the interest rate from 2.5% to 3%.

The committee announces it does so with no changes in their assessment of the economic conditions.

Demand: control (same page)

Prediction of demand for firms' goods and services

Think about the demand for products and services that US firms face. Please complete the following sentence.

In the alternative scenario (federal funds rate rises), US firms face ________ than in the baseline scenario (federal funds rate constant).

Your thoughts

What are your main considerations in making the above prediction?
Please respond in 2-3 sentences.

Cost: control (same page)

Prediction of firms' costs

Think about the costs of doing business that US firms face. Please complete the following sentence.

In the alternative scenario (federal funds rate rises), US firms face ________ than in the baseline scenario (federal funds rate constant).

Your thoughts

What are your main considerations in making the above prediction?

Please respond in 2-3 sentences.

Appendix 2.H Alternative explanations

2.H.1 Perceived past correlations

In this subsection, we provide additional details on our measurement of households' perceived past correlations of macroeconomic variables (used in Section 2.4.7), and the role of these beliefs in inflation and unemployment forecasts under the different shocks.

Measurement. We elicit respondents' perceived past correlations in two ways. A random half of our respondents report their perceived correlations of past *changes* in macroeconomic variables. We first tell these participants that "we have analyzed data about the development of the US economy in the last 50 years (1969-2019). We studied which economic outcomes tend to move in the same direction, which tend to move in opposite directions, and which move independently of each other. Consider for example the average unemployment rate and the inflation rate. We calculated the share of years in which these outcomes (i) moved in the same direction, i.e. both rise or both fall, and (ii) moved in opposite directions, i.e. one rises, but the other one falls. The two variables moved independently of each other if they moved in the same direction 50% of the time and moved in opposite directions 50% of the time." We then ask our respondents to consider two variables (e.g. the oil price and the unemployment rate) and ask them what percent of the time these two variables moved (i) in the same direction and (ii) in opposite directions over the last 50 years.

The other half of our respondents report their beliefs about the correlation of levels of macroeconomic variables in the past, using a survey question with qualitative response categories.

Role in forecasts by shock. In Section 2.4.7, we show that perceived past correlations of shock variable and outcome variable of interest are strongly correlated with households' forecasts in the vignettes, where we pool forecasts i) across inflation and unemployment, ii) across shocks, and iii) across the two ways of eliciting past correlations. We now study the role of perceived past correlations separately for forecasts of each outcome under each shock and for each elicitation method.

As shown in Table 2.H.1, the perceived past correlations between inflation and the shock variables are strongly associated with respondents' forecasts. These patterns hold across vignettes and both for perceived correlation of levels and changes. Our findings hold for forecasts of both the unemployment rate and the inflation rate. This evidence is purely correlational and should be interpreted cautiously, as it could be confounded by omitted variables or reverse causality, as explained in the main text.

Table 2.H.1. Households: Perceived past correlations

Inflation $\Delta\pi$

	Pooled		Oil price		Gov. spend.		Fed. funds rate		Inc. taxes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Past corr.	0.139***	0.147***	0.184***	0.241***	0.071	0.131**	0.130*	0.077**	0.148**	0.177***
	(0.035)	(0.024)	(0.069)	(0.061)	(0.070)	(0.051)	(0.074)	(0.038)	(0.062)	(0.049)
Type	Changes	Levels	Changes	Levels	Changes	Levels	Changes	Levels	Changes	Levels
Vignette	All	All	Oil	Oil	Gov.	Gov.	Fed.	Fed.	Tax	Tax
Vig. FE	✓	✓	–	–	–	–	–	–	–	–
Obs.	1,026	1,059	285	265	245	264	229	280	267	250
R^2	0.027	0.060	0.031	0.083	0.004	0.025	0.016	0.016	0.022	0.060

Unemployment Δu

	Pooled		Oil price		Gov. spend.		Fed. funds rate		Income taxes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Past corr.	0.066*	0.209***	0.139*	0.219***	0.061	0.314***	0.018	0.152***	0.032	0.133***
	(0.038)	(0.022)	(0.073)	(0.050)	(0.096)	(0.043)	(0.080)	(0.037)	(0.062)	(0.048)
Type	Changes	Levels	Changes	Levels	Changes	Levels	Changes	Levels	Changes	Levels
Vignette	All	All	Oil	Oil	Gov.	Gov.	Fed.	Fed.	Tax	Tax
Vig. FE	✓	✓	–	–	–	–	–	–	–	–
Obs.	1,026	1,059	285	265	245	264	229	280	267	250
R^2	0.062	0.117	0.015	0.077	0.002	0.181	0.000	0.057	0.001	0.030

Notes: This table presents results from Wave 3 of the household survey. Two measures of past perceived correlations ("Past corr.") are used. Changes: "Past corr." ranges from -1 (the two variables always move in opposite directions) to 1 (the two variables always move in the same direction). Levels: "Past corr." takes either value -1, 0 or 1, where 1 means that respondents think that the two variables are positively correlated, -1 means that respondents think that the two variables are negatively correlated and 0 means that respondents think that the two variables are uncorrelated. $\Delta\pi$ denotes the expected difference in inflation between the rise scenario and the baseline scenario. Δu denotes the expected difference in unemployment between the rise scenario and the baseline scenario. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

2.H.2 Perceived importance of knowledge about economy

In Section 2.4.7, we also examine whether individuals who consider it necessary to be knowledgeable about macroeconomic relationships to make good economic decisions are more likely to make benchmark-consistent forecasts, in line with a premise of rational inattention models (Sims, 2003; Maćkowiak and Wiederholt, 2015). In Wave 3 of the household survey, we ask how useful respondents consider it to have knowledge about different issues for making good economic decisions. Among others, we ask this question for "knowledge about how the US economy works" and e.g. "knowledge about how income tax rates affect the US economy". In the second case, the question varies depending on the vignette a respondent is assigned to. We construct our index of the perceived usefulness of understanding the functioning of the economy as the average of a participant's responses to the questions on these two issues.

2.H.3 Objective measure of knowledge about economy

We measure households' knowledge about the economy through questions on beliefs about the current and the 2019 unemployment rate, and beliefs about inflation over the previous 12 months, as well as questions on self-reported acquisition of information about unemployment and inflation over the last three months. For the quantitative beliefs we calculate deviations from the true values. We z-score deviations from the true values for the beliefs and the responses to the qualitative questions on information acquisition using the means and standard deviations of the variables. We construct our index of economic knowledge as the average over the resulting variables.

2.H.4 Good-bad heuristic

Research from psychology suggests that individuals revert to simple heuristics in complex decision environments in which there is a lot of uncertainty (Gigerenzer and Todd, 1999). In light of this evidence, we consider whether a simple heuristic, namely that good things only lead to good things and bad things only lead to bad things, can explain the heterogeneity in predictions in the representative sample. We refer to this as the good-bad heuristic (GBH). It postulates that households perceiving two variables as both good or both bad (symmetric affective evaluation) are more likely to predict a positive co-movement between them, while predicting a movement in opposing directions if they perceive one variable as good and the other one as bad (asymmetric affective evaluation). Our evidence is related to evidence from psychology studying understanding of the macroeconomy using student samples where a similar idea has been discussed under the label "good-begets-good heuristic" (Leiser and Aroch, 2009; Leiser and Krill, 2017). It also relates to theoretical work in economics (Kamdar, 2019).

To test this hypothesis, we measure whether respondents consider higher values of the four shock variables, unemployment, and inflation as good or bad for the US economy and for their own household. Households can respond on two 7-point scales, ranging from very bad (-3) to very good (3). For each variable, we average the evaluations for the US economy and their own households. Then, we derive the directional prediction that follows from the GBH for each forecast. If a respondent evaluates the two variables underlying a forecast (e.g. government spending and inflation) symmetrically (asymmetrically), the GBH implies a predicted change of the outcome variable in the same (opposite) direction as the change in the shock variable. For example, if a respondent perceives both higher government spending and higher inflation as bad, the GBH predicts that she expects that inflation will increase in response to an exogenous increase in government spending. If a respondent perceives higher government spending as good but higher inflation as bad, the GBH predicts that she expects that inflation will decrease in response to an exogenous increase in government spending. If at least one variable is evaluated neutrally (neither good nor bad), no change is predicted. Finally, we construct a dummy that takes value 1 whenever the predicted change suggested by the GBH is in line with the literature benchmarks, that is, whenever following the GBH would result in making a benchmark-consistent forecast. This dummy is used in our analyses.

We uncover a fairly large explanatory power of the good-bad heuristic. On average, forecast consistency with benchmarks increases by 14 p.p. when the GBH makes a benchmark-consistent prediction (Table 2.H.2).

One potential concern with the GBH evidence, however, is that respondents' forecasts in the vignettes might be driving their affective encoding of the different variables. We somewhat mitigate this concern in Wave 3 of the data collection by moving the questions on affective evaluations of different variables to the very end of the survey. Nonetheless, future work could randomize the order of the questions on affective evaluations and the vignettes to deal with this concern. To provide causal evidence on the GBH, future work could try to manipulate the affective encoding of different variables, for example, by providing individuals with personal payoffs associated with the rise or fall of macroeconomic variables.

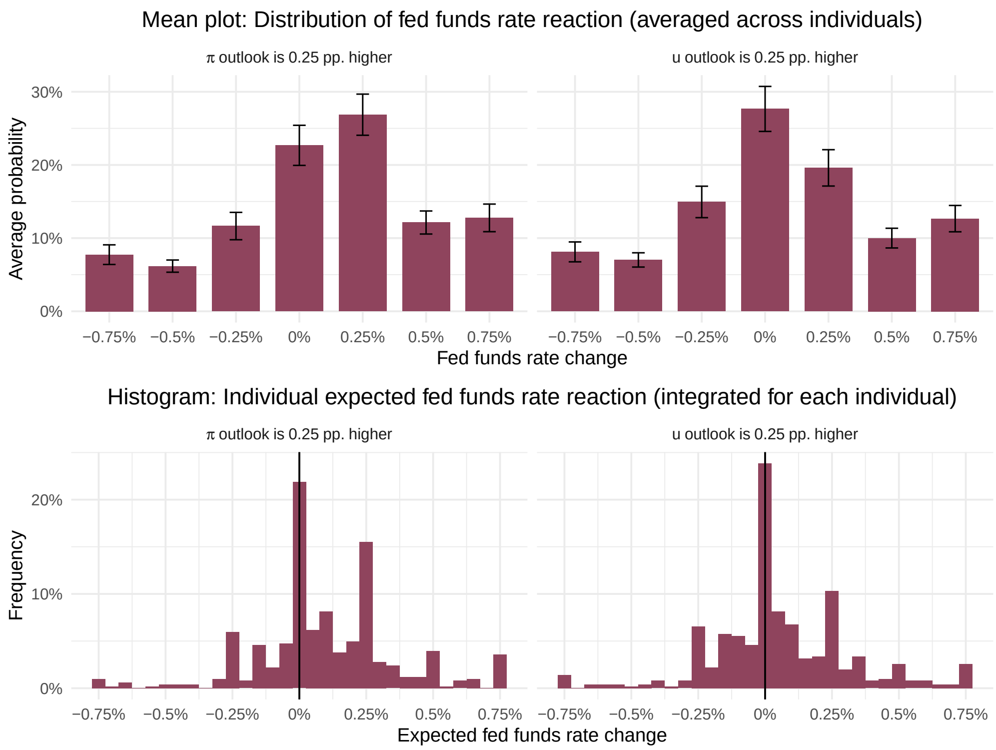
Table 2.H.2. Households: Good-bad-heuristic: Predictors of benchmark-consistent forecasts

	Indicator for benchmark-consistent prediction	
	Separate bivariate models	Multivariate model
	(1)	(2)
Consistent Good-Bad-Heur.	0.138***	0.102***
	(0.019)	(0.019)
Consistent channel association	0.172***	0.141***
	(0.015)	(0.015)
Consistent perceived correlation	0.181***	0.149***
	(0.017)	(0.016)
Importance of model (1 if >median)	0.042***	0.014
	(0.015)	(0.015)
Knowledge (1 if >median)	0.077***	0.031**
	(0.015)	(0.015)
Numeracy (1 if >median)	0.062***	0.027*
	(0.015)	(0.014)
Female	−0.026*	−0.007
	(0.015)	(0.014)
Age (1 if >median)	0.058***	0.028*
	(0.015)	(0.015)
College degree	0.022	−0.004
	(0.015)	(0.015)
Income (1 if >median)	0.012	0.000
	(0.016)	(0.016)
Republican	0.021	0.018
	(0.016)	(0.015)
Mean share of benchmark-consistent pred.	0.480	0.480
Fixed effects	Vignette ⊗ rate	Vignette ⊗ rate
Observations	3,844	3,844
R^2	–	0.245

Notes: This table presents results from Wave 3 of the household survey. It presents the effect of various binary covariates on the likelihood of making inflation or unemployment predictions (pooled) that are consistent with the benchmarks, i.e. directionally aligned with the literature benchmark. Each coefficient can be interpreted as the increase in probability that a forecast is benchmark-consistent. Column (1) shows the results from separate bivariate regressions, while Column (2) shows the results from a multivariate model. "Consistent Good-Bad-Heur." takes value 1 if the good-bad-heuristic is directionally aligned with a benchmark-consistent prediction. "Consistent channel association" takes value 1 if the respondent chooses a channel (structured question) that suggests a benchmark-consistent prediction (e.g. a negative demand-side channel for the federal funds rate vignette). Likewise, "Consistent perceived correlation" takes value 1 if respondents believe in a past correlation between the shock variable (e.g. oil price) and the target variable (e.g. inflation) that is in line with a benchmark-consistent prediction. "Importance of model" measures respondents' assessment of how important knowledge of the functioning of the economy is to them for making good economic decisions. "Knowledge" measures information about the current state of the economy. "Numeracy" is respondents' score on a numeracy test. "1 if >median" indicates that a variable is binarized and takes value 1 for respondents with an above-median value. We include fixed effects for each vignette-rate combination (e.g. oil-inflation). Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

2.H.5 Misperceived endogeneity in the interest rate vignette

In Wave 2 of the household survey we conduct an additional quantitative robustness check against the possibility that respondents misperceive the monetary policy shock as the Fed's endogenous response to a change in its outlook for inflation. We elicit subjective beliefs about how the Fed usually adjusts interest rates to an unexpected increase in the outlook for (i) inflation and (ii) unemployment. For inflation, we ask our respondents to "imagine that the FOMC changes their outlook for inflation over the next 12 months due to data revisions, while there is no change in the outlook for unemployment. Specifically, the Fed believes that the inflation rate will be 0.25 p.p. higher than their initial estimate." We provide similar instructions for the change in the outlook for unemployment. Thereafter, we measure respondents' beliefs about how the Fed would adjust the federal funds rate. Figure 2.H.1 shows that there is substantial heterogeneity in beliefs on how the Fed would adjust interest rates. If our results were driven by respondents attributing a higher fed funds rate to a change in the Fed's outlook for inflation, we would expect stronger predicted increases in inflation among those respondents who believe that the Fed reacts more strongly to a higher outlook for inflation. However, there is no significant heterogeneity along this dimension and, if anything, the patterns go in the opposite direction of what would be predicted by this potential confound (see Table 2.H.3). Likewise, we can rule out that respondents interpret the interest rate change as a signal that the FOMC changed its unemployment outlook.



Notes: This figure analyzes the distribution of responses to the subjective interest rate rule questions in Wave 2 of the household survey. Respondents are asked to estimate the likelihood of different federal funds target rate changes in response to a 0.25 pp. increase in the Fed's outlook for the inflation rate or the unemployment rate. For each possible federal funds target rate reaction, the "Mean plot" summarizes the average probability assigned to this event (averaged across individuals), with 95% confidence intervals. The histogram plots the distribution of individual-level expected changes in the federal funds target rate in response to increases in the Fed's outlook for inflation or unemployment (integrated for each individual).

Figure 2.H.1. Households: Descriptive statistics for the subjective interest rate rules

Table 2.H.3. Households: Misperceived endogeneity of interest rate shock

Panel A: Binary monetary policy reaction

	fed. funds rate			
	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)
$1(\alpha > 0)$	−0.104		−0.088	0.055
	(0.099)		(0.099)	(0.097)
$1(\beta > 0)$		−0.098	−0.059	−0.113
		(0.092)	(0.097)	(0.095)
Constant	0.259***	0.152**	0.279***	0.126
	(0.076)	(0.064)	(0.087)	(0.079)
Observations	503	503	503	503
R^2	0.002	0.002	0.003	0.003

Panel B: Expected monetary policy reaction

	fed. funds rate			
	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)
$\alpha/4$	−0.332		−0.304	−0.143
	(0.211)		(0.201)	(0.188)
$\beta/4$		−0.068	−0.079	−0.015
		(0.188)	(0.204)	(0.188)
Constant	0.232***	0.107**	0.234***	0.119**
	(0.053)	(0.047)	(0.054)	(0.050)
Observations	503	503	503	503
R^2	0.006	0.000	0.007	0.001

Notes: This table reports regressions that test for a misperception of the interest rate shock as an endogenous reaction of the Fed to a changed outlook in inflation or unemployment, using Wave 2 household data. α denotes the perceived coefficient on π^e in the Fed's linear forward-looking interest rate rule, and β denotes the perceived coefficient on u^e. Δu denotes the predicted change in the unemployment rate compared to the baseline scenario. $\Delta\pi$ denotes the predicted change in the inflation rate compared to the baseline scenario. Panel A regresses both variables on $1(\alpha > 0)$ – a dummy taking value 1 if the respondent believes that the Fed would increase the federal funds target rate in response to an unexpected increase in the outlook for future inflation – and $1(\beta > 0)$ – a dummy taking value 1 if the respondent believes that the Fed would increase the federal funds target rate in response to an unexpected increase in the outlook for future unemployment. Panel B uses α and β which are the respondents' estimates of the coefficients in the forward-looking interest rate rule. They are divided by 4 because the inflation and unemployment outlook change by 0.25 p.p. (rather than 1 pp.) in the survey questions. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Appendix 2.I A simple formal framework for inflation expectations

In this section, we present a simple framework to include subjective models of the propagation of shocks into the formation of inflation expectations. Without loss of generality, the same example is applicable to other macroeconomic variables. We consider how the resulting dynamics compare to those of the sticky-information model of Mankiw and Reis (2002). In particular, we discuss whether the framework can produce three empirical features of inflation expectations:

1. The cross-sectional average of inflation expectations across individuals under-reacts to a shock compared to actual inflation;
2. The sluggish response to a shock of the average expectation implies that forecast errors for inflation of period $T > t_0$ are correlated between periods t_0 and $T-1$;
3. Disagreement (i.e., the cross-sectional dispersion among individuals in expectations) increases in response to a shock.

Model. Assume that inflation follows an AR(1) process:

$$\pi_t = \rho\pi_{t-1} + \omega_t \tag{2.I.1}$$

End-of-period inflation π_t is not observed at the beginning of period t, when agents fully observe past end-of-period inflation π_{t-1} and the beginning-of-period shock ω_t. However, while the true inflation process remains (2.I.1), agents hold different beliefs on the impact of the shock, such that they derive different expectations of π_t. Specifically an individual i receives a draw of the coefficient α such that:

$$\mathbb{E}_t^\alpha \pi_t = \pi_{t|t}^\alpha = \rho\pi_{t-1} + \alpha\omega_t, \tag{2.I.2}$$

$$\alpha \sim N(\mu_\alpha, \sigma_\alpha^2) \tag{2.I.3}$$

Note that α can be below 0, implying that an agent thinks that the shock affects the economy with opposite sign.

The individual expectation of inflation h periods ahead at the beginning of period t is

$$\begin{aligned}\mathbb{E}_t^\alpha \pi_{t+h} &= \mathbb{E}_t\Big(\rho^{h+1}\pi_{t-1} + \alpha\rho^h\omega_t + \sum_{k=0}^{h-1}\rho^{h-k}\omega_{t+h-k}\Big) \\ &= \rho^{h+1}\pi_{t-1} + \alpha\rho^h\omega_t\end{aligned}$$

The cross-sectional average of expectations for inflation h periods ahead, in period t is

$$\begin{aligned}\overline{\pi_{t+h|t}} = \overline{\mathbb{E}}(\pi_{t+h|t}(\alpha)) &= \int \mathbb{E}_t^\alpha \pi_{t+h} \quad \mathrm{d}F(\alpha) \\ &= \int \left(\rho^{h+1}\pi_{t-1} + \alpha\rho^h\omega_t\right) \ \mathrm{d}F(\alpha) \\ &= \rho^{h+1}\pi_{t-1} + \mu_\alpha\rho^h\omega_t\end{aligned}$$

If $\mu_\alpha = 1$, the cross-sectional average across individuals is unbiased. If $\mu_\alpha < 1$, the cross-sectional average of inflation expectations under-reacts to a shock, which would be consistent with one of the empirical facts established by the literature mentioned above.

Disagreement is defined as the cross-sectional variance in the expectations:

$$\begin{aligned}\mathbb{V}_t\pi_{t+h} &= \int \left(\mathbb{E}_t^\alpha \pi_{t+h} - \overline{\pi}_{t+h|t}\right)^2 \mathrm{d}F(\alpha) \\ &= \int \left(\rho^{h+1}\pi_{t-1} + \alpha\rho^h\omega_t - \rho^{h+1}\pi_{t-1} - \mu_\alpha\rho^h\omega_t\right)^2 \mathrm{d}F(\alpha) \\ &= \int \left(\alpha - \mu_\alpha\right)^2\left(\rho^h\omega_t\right)^2 \mathrm{d}F(\alpha) \\ &= \left(\rho^h\omega_t\right)^2\sigma_\alpha^2\end{aligned}$$

Since disagreement is a function of the squared value of ω_t, it follows that disagreement rises with the absolute size of shocks, independently of their sign.

In the context of our empirical results, the model could be generalized to include multiple observed shocks ω_t^j, where j represents the nature of the shock, e.g., an oil, monetary policy, tax, or government spending shock. Moreover, each agent would have a draw of α^j corresponding to the shock j and capturing beliefs on the propagation of the specific shock. In the exercise we carry out below, we use the survey results to calibrate the parameters μ_α and σ_α for the oil shock and the interest rate shock. We then discuss how the computed response of inflation expectations and disagreement compares to that ensuing under the assumption of sticky information.

Comparison with sticky-information assumption. As a benchmark for comparison we consider the sticky-information inflation expectations model of Mankiw and Reis (2002), as described in Coibion and Gorodnichenko (2012). The key assumption of the model is that, although agents form expectations according to (2.I.1), in each period t they observe the true inflation π_t with probability $(1 - \lambda)$, where $0 < \lambda < 1$. Hence, a shock ω_t is gradually incorporated into expectations as more and more agents observe the new level of inflation over time. We refer the readers to Coibion and Gorodnichenko (2012) for a more detailed discussion of this model.

Responses of expectations and disagreement to shocks. To compare the two models quantitatively, we compute the response of inflation expectations and disagreement to a shock ω_t starting from a steady state of $\pi_s = 0$ for $s < t_0$.

We parametrize the process of inflation by estimating an AR(1) equation on US inflation between 1960 and 2019 at quarterly frequency, which yields an estimate for ρ of 0.9.[44] For the size of the true shock ω_t we use the theoretical benchmarks that we derived for the discussion of our empirical results (see Appendix 2.C for details).

For each shock, the values of μ_α and σ_α are derived from the survey results. μ_α is computed as the ratio of the average predicted change among household to the benchmark for the impact of the shock, while σ_α is the ratio of the standard deviation of the change predicted by households to the benchmark.[45]

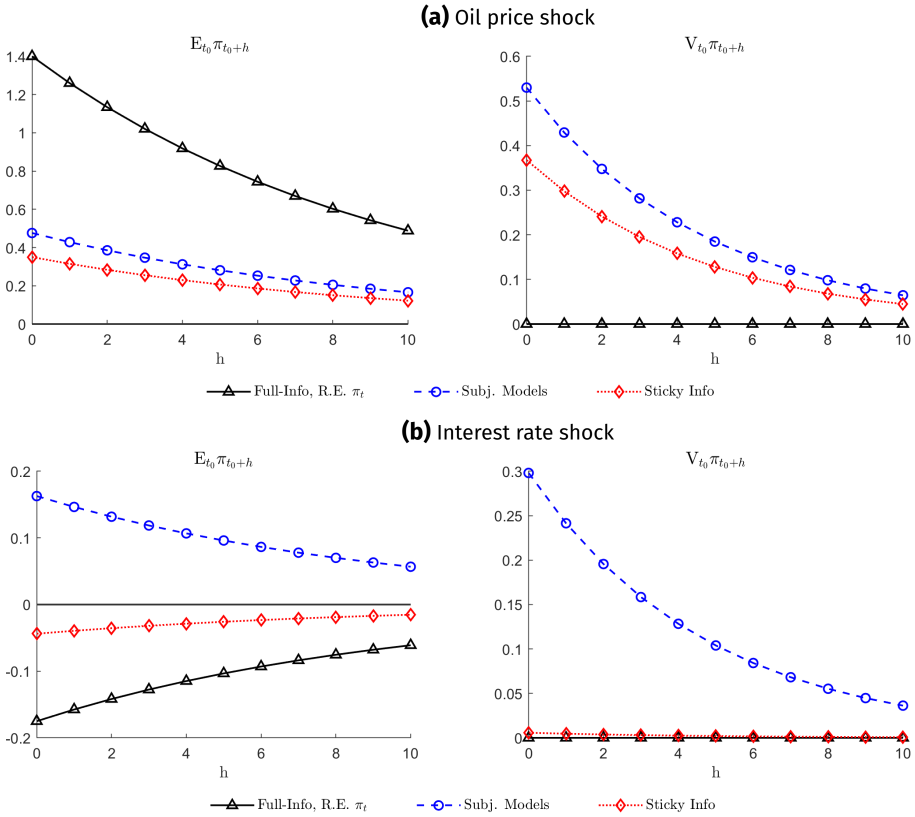
For the sticky-information expectations, we set $\lambda = 0.75$, consistent with the findings by Carroll (2003) for non-expert households. This value implies that, on average, agents update their expectations for inflation once a year (i.e., every four quarters).

For the oil price (top panels) and interest rate (bottom panel) shocks, Figure 2.I.1 reports the response of (i) the true inflation process, as it would also be elicited by full-information rational expectations (black triangles), (ii) the subjective-models expectations (blue circles), and (iii) the sticky-information expectations (red diamonds). The left panels report the change in the expectations at time t_0 of inflation for times $t_0 + h$. The right panels report the change in disagreement at time t_0 with respect to expectations for π_{t_0+h}. In the oil price shock, the subjective models specification produces an average expected inflation and disagreement that are approximately 25 to 35 percent larger than the sticky-information one. For the interest rate shock, the average expectation under subjective models is close in magnitude to the true shock but has the opposite sign compared to both the true inflation response and the sticky-information model. This is consistent with our baseline results showing that on average households expect inflation to rise after a tightening of monetary policy. Moreover, the large level of disagreement across households on the effect of monetary policy, as elicited in our survey, entails a rise in disagreement over expected inflation under the subjective models specification that is many times larger than for the sticky-information model (bottom right panel).

Persistence of expectations. The previous figures focused on the contemporaneous response of average expected inflation and disagreement over a forecast horizon. However, a key empirical regularity of average inflation expectations, as discussed

44. For inflation we use the annualized growth rate in the urban CPI downloaded from the FRED Database of the Federal Reserve of St. Louis.

45. For the oil price shock, the values are $\mu_\alpha = 0.34$ and $\sigma_\alpha = 0.52$. For the interest rate shock, $\mu_\alpha = -0.93$ and $\sigma_\alpha = 3.12$



Notes: The panels report the response of the cross-sectional average and variance of inflation expectations at time t_0 over the forecasting horizon $t + h$, to a one-standard deviation shock to the true inflation process, where $h = 0, ..., 10$. The black triangles plot the response for the true inflation process, also consistent with a full-information rational-expectations model. The blue circles plot the response under the subjective-models specification. The red diamonds report the response under the sticky-information specification. See the main text in Appendix 2.I for more details. The size of the shock is based on the empirical benchmarks used in the main text for the analysis of the results (see Appendix 2.C for more details).

Figure 2.I.1. Responses of inflation expectations and disagreement to an exogenous shock under alternative specifications of expectations formation

by Coibion and Gorodnichenko (2012), is the persistence in forecast errors for inflation at a given time T. In other words, $\pi_T - \overline{\pi}_{T|t}$ and $\pi_T - \overline{\pi}_{T|t+1}$ are positively correlated. The sticky-information model captures this feature because information on new shocks is only acquired by a fraction of agents in each period. In contrast, the subjective-models specification by itself does not produce any persistence in forecast errors under the assumption that past inflation is perfectly observed at time t, implying a fully updated information set up to the beginning of the period.

Combining sticky information and subjective models. While our survey does not speak to the persistence of forecast errors, for illustrative purposes, below we discuss the behavior of forecast errors and disagreement over time in response to a shock – rather than just in the concurrent period. As an example we use the parametrization of the oil vignette. Moreover, to ease the comparison, we consider a third specification of expectations that combines sticky information and subjective models. In this set-up, at time t a fraction $1-\lambda$ of agents fully observes past inflation π_{t-1} and the current shock ω_t. This extra specification provides an indication of how some degree of information friction can be added to the subjective models assumption to replicate the empirical persistence in forecast errors.

The cross-sectional average of expectations of this specification is:

$$\overline{\pi_{t+h|t}}^{SI+SM} = (1-\lambda)\sum_{k=0}^{\infty}\lambda^k(\rho^{k+h+1}\pi_{t-k-1} + \mu_\alpha \rho^{k+h}\omega_{t-k}) = (1-\lambda)\sum_{k=0}^{\infty}\lambda^k \overline{\pi_{t+h|t}}^{SM},$$

where the superscript SM represents the subjective-models expectation and the superscript $SI+SM$ represents the joint sticky-information and subjective models specification. Disagreement is as follows:

$$\mathbb{V}_t^{SI+SM}\pi_{t+h} = (1-\lambda)\sum_{k=0}^{\infty}\lambda^k\Big(\mathbb{V}_t^{SM}\pi_{t+h} + (\overline{\pi_{t+h|t-k}}^{SM} - \overline{\pi_{t+h|t}}^{SM+SI})^2\Big).$$

For illustrative purposes, we examine the dynamics of the $SI+SM$ model using two alternative values of λ. The first is $\lambda^{SI+SM} = 0.75$, equal to the value used in the pure sticky-information model from Carroll (2003). The second is a calibrated value of λ^{SI+SM} such that the initial reaction of inflation expectations to the shock is equal to that of the sticky-information model. Based on the other parameter values used for the previous exercise, the resulting value is $\lambda^{SI+SM} = 0.265$.

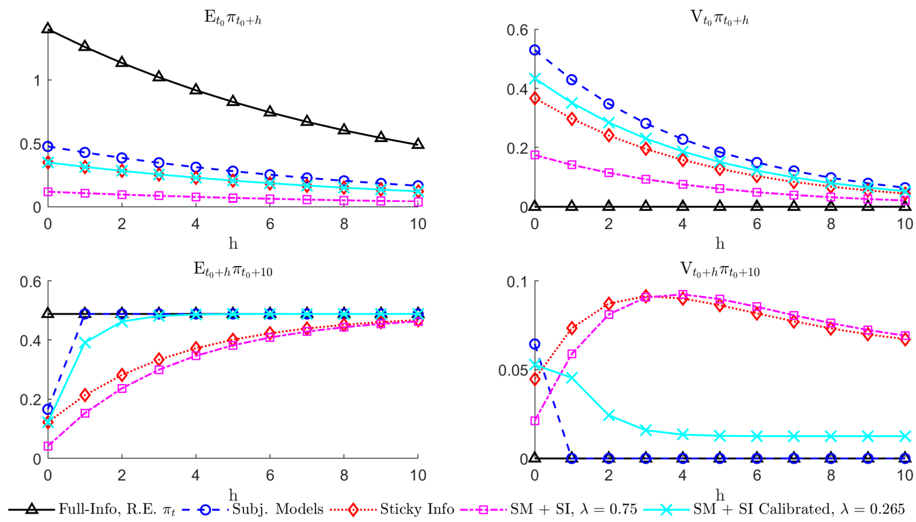
Figure 2.I.3 reports the results of the exercise. The top panels are the same as in Figure 2.I.1 except for the addition of the $SI+SM$ specifications (magenta squares for the baseline specification with $\lambda^{SI+SM} = \lambda = 0.75$ and light blue X's for the alternative calibrated $\lambda^{SI+SM} = 0.265$). In the top left panel, the response of average expectations under the baseline $SI+SM$ specification is more muted than both the subjective models and the sticky-information specifications by virtue of the parameter choice. Intuitively, in the first period the only response of expectations comes from the fraction $1-\lambda < 1$ of agents observing the shock, whose average

update of expectation is a multiple $\mu_\alpha < 1$ of the true shock. Hence, the two behavioral features of expectations compound each other in a multiplicative way. The average expectations in the alternative, calibrated $SI + SM$ model entirely overlap those of the sticky-information model. By construction, matching the response in period t_0 implies the same level of persistence in forecast across the two specifications also at longer time horizons. Intuitively, the addition of subjective models to sticky-information implies that the same level of under-reaction to a shock can be obtained with a lower level of information friction (0.265 instead of 0.75 in this case).

The response of disagreement in period t_0 under the baseline $SI + SM$ specification is also more muted than those of the two constituent models. While this result is dependent on parameter choices, and the fact that $\mu_\alpha < 1$ for the oil vignette, it intuitively derives from the fact that both specifications mute the true response of inflation: Disagreement is limited when fewer agents observe any new information that can lead to heterogeneous expectations or when they under-react to it. When re-calibrating the λ^{SI+SM} parameter, however, disagreement rises closer to the level of the subjective models specification since a larger fraction of agents noisily observe the shock.

The bottom panels of Figure 2.I.3 report the response of expectations and disagreement at time $t_0 + h$ with respect to a fixed time in the future, in this case $t_0 + 10$. The subjective models specification assumes that after period t_0 all agents perfectly observe past inflation. Hence, average expectations converge to the true expected inflation and disagreement disappears after one period (left panel). Meanwhile, under sticky information, average expectations converge slowly, leading to the correlation in forecast errors, and disagreement has a hump-shaped response. In this case, $\mu_\alpha < 1$ implies that the baseline $SI + SM$ specification increases the lag in the convergence of average expectations compared to the sticky-information model. Meanwhile, the variance of expectations in the combined model maintains a path similar to the sticky-information model (right panel), as the inability to perfectly observe past realizations of inflation is the main driver of the persistence in disagreement. Having a lower degree of information frictions, the alternative, calibrated $SI + SM$ specification features a persistence in expectations and disagreement between that of the sticky-information model and the absence of persistence of the subjective models specification.

Take-aways from the combined model. This exercise shows that, in the case when $\mu_\alpha < 1$, the combination of subjective models and sticky information generates dynamics of expectations that compound the effect of the two models with respect to the immediate reaction of expectations to a shock but that more closely resemble those of the sticky information model in terms of persistence in expectations and disagreement over time. Moreover, adding heterogeneity in beliefs about the model of the economy implies that the sluggish response of inflation expectations



Notes: The panels report the response of the cross-sectional average and the variance of inflation expectations to a one-standard deviation shock to the true inflation process. The top panels focus on the path of expectations in period t_0, when the shock realizes, for the time horizon $t + h$, $h = 0, ..., 10$. The bottom panels focus on the expectations in period $t + h$, $h = 0, ..., 10$ for inflation at time $T = 10$. The black triangles plot the response for the true inflation process, also consistent with a full-information rational-expectations model. The blue circles plot the response under the subjective-models specification. The red diamonds report the response under the sticky-information specification. The magenta squares plot the specification combining sticky information and subjective models. The light blue X's report the specification combining sticky information and subjective models with the parameter λ calibrated to match the inflation expectation response of the sticky-information model. See the main text in Appendix 2.I for more details. The size of the shock is based on the empirical benchmarks used in the main text for the analysis of the results (see Appendix 2.C for more details).

Figure 2.I.3. Responses of inflation expectations and disagreement to the oil price shock under alternative expectations formation models at time t_0 and a time $t_0 + h$

to a shock need not be driven entirely by information stickiness. It is hence possible that empirical estimates of the frequency at which agents update their information, if solely recovered from the time-series properties of average expectations, may over-estimate or under-estimate the level of information stickiness if subjective models are not accounted for. In the case of the oil shock, we find that information stickiness may be over-estimated. However, given that the parameters of the subjective models specification are dependent on the specific vignette, the result may differ for other shocks. In cases where the average expectation among respondents is larger than the true reaction of inflation (i.e., $\mu_\alpha > 1$), information stickiness would be under-estimated. Finally, in cases where agents update their beliefs in the wrong direction, such as in the interest rate vignette, the interaction of subjective models and information frictions is likely more complicated.

Appendix 2.J Details on expert surveys

2.J.1 Wave 1

We compiled a list of participants of the following conferences:

- SITE Macroeconomics of Uncertainty and Volatility (2018, 2017, 2016)
- SITE Macroeconomics and Inequality (2018)
- Cowles Macro Conference (2018, 2017, 2016)
- NBER Annual Conference on Macroeconomics (2018, 2017, 2016)
- ifo Conference on "Macroeconomics and Survey Data" (2018, 2017, 2016)
- Venice Summer Institute on Expectation Formation (2018)
- Workshop on Subjective Expectations NY Fed (2016)

We also recruited a sample of graduate students in macroeconomics from the following institutions:

- University of Bonn
- Goethe University Frankfurt
- University of Oxford

Finally, we also recruited a sample of economists from the following policy institutions:

- The Federal Reserve Board, Washington D.C.
- The International Monetary Fund, Washington D.C.
- Bank for International Settlements, Basel
- Deutsche Bundesbank, Frankfurt
- European Central Bank, Frankfurt
- ifo centre, Munich

Below is a list of the institutions that our experts (from Wave 1) have as one of their main institutions: Kellogg School of Management, Northwestern University, University of Cologne, Haverford College, University of Minnesota, Ross School of Business, University of Michigan, Federal Reserve Bank of Boston, University of Amsterdam, Boston University, Questrom School of Business, Federal Reserve Bank of St. Louis, Goethe University Frankfurt, LMU Munich, University of Notre Dame, University of California San Diego, University of Oxford, Temple University, International Monetary Fund, University of Toronto, Carleton University, Yale University, Federal Reserve Board, University of Copenhagen, University of Bologna, Georgia Institute of Technology Atlanta, Statistics Norway, Deutsche Bundesbank, Frankfurt School of Finance & Management, Johns Hopkins University, Baltimore, Brandeis University, Federal Reserve Bank of Cleveland, Bank of England, MIT Sloan

School of Management, Rand Corporation, University of Copenhagen, International Monetary Fund, Swiss National Bank, Boston College, University of Reading, UNC Kenan-Flagler Business School, Bonn Graduate School of Economics, Institute for Employment Research Friedrich-Alexander University (FAU) Erlangen-Nuremberg, College of Business Clemson University, ifo Institute Munich, Stockholm University, Banque de France, University of Nantes, Uppsala University, World Bank, University of St.Gallen, Austrian Institute of Economic Research, Copenhagen Business School, Federal Reserve Bank of Minneapolis, NYU Stern School of Busines, University of Bonn, Mannheim University, University of Manchester, University College London, University of Lausanne, Arizona State University, University of Birmingham, Federal Reserve Bank of Chicago, European Central Bank, Bank for International Settlements, Basel, University of Maryland, Amsterdam School of Economics, Columbia University, Christian Albrechts University at Kiel, Princeton University, Stockholm School of Economics, University of Chicago Booth School of Business, University of Warwick, Leibniz University Hannover, University of Heidelberg, University of Copenhagen, Northwestern University, New York University, Federal Reserve Bank of Minneapolis, Indiana University, Karlsruhe Institute of Technology.

2.J.2 Wave 3

We identify the email addresses of all economists who published in the top 20 economics journals on JEL code “E: Macroeconomics and Monetary Economics” in the years 2015-2019. We consider the following journals: Journal of Political Economy, Quarterly Journal of Economics, Econometrica, Review of Economic Studies, American Economic Review, Journal of Economic Literature, Journal of Economic Perspectives, Journal of the European Economic Association, Journal of Financial Economics, Review of Financial Studies, Journal of Finance, Review of Economics and Statistics, International Economic Review, Journal of Monetary Economics, Review of Economic Dynamics, Economic Journal, American Economic Journal: Macroeconomics, American Economic Journal: Applied Economics, Journal of Economic Growth, and Brookings Papers an Economic Activity.

We also identify students from the top ten European and the top ten US economics departments according to the Shanghai 2020 ranking. The departments in the US are: Harvard, MIT, UChicago, Northwestern, Yale, Princeton, Berkeley, Stanford, Colombia, NYU. The departments in Europe are: LSE, Oxford, Cambridge, UCL, Toulouse, Warwick, Rotterdam, Bocconi, Zurich, Oslo.

We also invited PhD students from Bonn and Copenhagen (where two of the authors are based) as well as all respondents we reached out to in Wave 1 of our expert survey.

We sent a link to our study to all of these economists by email. We did not send any reminders. In total, we contacted 4,367 economists. 375 economists responded to our survey, corresponding to a response rate of 8.6%.

Chapter 3

Fighting Climate Change: The Role of Norms, Preferences, and Moral Values

Joint with Teodora Boneva, Felix Chopra, and Armin Falk

Abstract: We document individual willingness to fight climate change and its behavioral determinants in a large representative sample of US adults. Willingness to fight climate change – as measured through an incentivized donation decision – is highly heterogeneous across the population. Individual beliefs about social norms, economic preferences such as patience and altruism, as well as universal moral values positively predict climate preferences. Moreover, we document systematic misperceptions of prevalent social norms. Respondents vastly underestimate the prevalence of climate-friendly behaviors and norms among their fellow citizens. Providing respondents with correct information causally raises individual willingness to fight climate change as well as individual support for climate policies. The effects are strongest for individuals who are skeptical about the existence and threat of global warming.

Acknowledgements: We thank Chris Roth for helpful comments and discussions; Ana Bras Monteiro and Tobias Reinheimer for excellent research assistance; and Markus Antony for administrative support. **Funding**: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866. Funding by the Deutsche Forschungsgemeinschaft (DFG) through CRC TR 224 (Project A01, B03) is gratefully acknowledged. **Ethics approval**: The study obtained ethics approval from the German Association for Experimental Economic Research (#Xx5i4FQa, 02/09/2021). **Research transparency**: The main research questions, the survey design, and the sampling approach were pre-registered at the AEA RCT Registry (#AEARCTR-0007542). Data and code will be made available.

3.1 Introduction

Climate change is one of the greatest threats facing humanity today. Its social and economic implications range from increased mortality and violence to reduced human productivity and economic growth (IPCC, 2014; Carleton and Hsiang, 2016; Auffhammer, 2018). The estimated economic impacts are enormous. Studies indicate that climate change could lower global GDP by 23% by 2100 and further exacerbate existing inequalities (Burke, Hsiang, and Miguel, 2015; Diffenbaugh and Burke, 2019). While many countries around the world have committed to meeting the 1.5 or 2 degree targets set out in the Paris Agreement, progress towards these goals has been slow (UNEP, 2019). In fact, it has become increasingly likely that global temperatures may rise well above the 2 degree target throughout the course of this century, with potentially catastrophic impacts for both human society and ecosystems. Given the threat posed by climate change, it is important to understand what determines people's individual willingness to fight climate change, namely their willingness to engage in climate-friendly, sustainable, but potentially costly behavior. Understanding the determinants of these individual 'climate preferences' can help us to design effective policies against climate change that complement existing regulatory frameworks, such as carbon taxation.

In this paper, we shed light on the behavioral determinants of climate preferences. We explore the role of economic preferences, universal moral values, and beliefs about social norms. We also design a norm intervention to examine whether informing individuals about the prevalence of climate norms raises their willingness to fight climate change and their support for climate policies.

For this purpose, we administer a survey to a large representative sample of 8,000 US adults. We elicit individual willingness to fight climate change using an incentivized donation decision. More specifically, respondents are asked to divide $450 between themselves and a charitable organization that fights global warming. This decision captures the central trade-off that individuals face when deciding whether to take climate action, namely the notion that protecting the climate comes at a cost. To incentivize the decision, we implement the choices of a random subset of participants. The more money the respondents are willing to forgo and donate, the higher their willingness to fight climate change. To shed light on the potential determinants of climate preferences, we obtain detailed, individual-level information on perceived social norms, fundamental economic preferences, and moral values. We measure perceived social norms by asking respondents to estimate (i) the share of the US population that tries to fight global warming ('perceived behavior') and (ii) the share of the US population that thinks people in the US *should* try to fight global warming ('perceived norms'). To elicit economic preferences, we administer an experimentally validated survey to measure patience, willingness to take risks, altruism, trust, positive reciprocity, and negative reciprocity (Falk, Becker, Dohmen, Enke, Huffman, et al. (2018) and Falk, Becker, Dohmen, Huffman, and

Sunde (2018)). We further administer the Moral Foundations Questionnaire to obtain a measure of the relative importance of universal versus communal moral values (Haidt and Joseph (2004), Haidt (2012), Graham, Haidt, Koleva, Motyl, Iyer, et al. (2013), and Enke (2020)).

A natural question that arises is whether it is possible to raise individual willingness to fight climate change. While it is difficult to alter some behavioral determinants such as fundamental economic preferences or moral values, at least in the short run, beliefs about social norms are likely to be considerably more malleable. We therefore conduct a survey experiment to study the extent to which information provision can raise individual willingness to fight climate change. Respondents are randomized into a control condition or one of two treatments. The 'behavior treatment' provides respondents with truthful information about the proportion of the US population who try to fight global warming (62%), while the 'norms treatment' informs respondents about the true share of the US population who think that people in the US should try to fight global warming (79%). These low-cost information treatments have the potential to correct misperceptions about prevalent behaviors and norms and may shift individual willingness to fight global warming.

Several findings emerge from our study. First, we document large heterogeneity in individual willingness to fight climate change. In particular, climate preferences are systematically related to perceived social norms, economic preferences, as well as universal moral values. Conditional on a large set of covariates, perceived social norms strongly predict individual willingness to fight global warming. A one-standard-deviation increase in the perceived share of Americans trying to fight global warming is associated with a $12 higher donation amount, while a corresponding increase in the perceived share of Americans who think that people in the US *should* try to fight global warming is associated with a $14 higher donation. These results are consistent with individuals being 'conditional cooperators'. Put differently, respondents may be more willing to fight climate change if they believe that a higher proportion of their fellow citizens do the same. Among the economic preferences that we measure, patience, altruism, and positive reciprocity positively predict individual willingness to fight global warming. Similarly, universal moral values are positively associated with larger donations. Individuals with universal moral values are more willing to fight climate change compared to individuals who endorse communal, in-group-oriented values. The fight against climate change can be viewed as a global cooperation problem affecting present and future generations all around the world. It therefore is plausible that more patient and prosocial individuals as well as individuals with universal moral values more strongly value climate protection. Our finding that fundamental human traits, such as altruism, positive reciprocity, and moral universalism, are strong predictors of individual willingness to fight climate change helps us to understand the frequently observed cultural and political dissent on climate change (Dunlap, McCright, and Yarosh, 2016; Hornsey, Harris, and Fielding, 2018). In our data, economic preferences and universalism together explain about 40% of the large partisan gap in willingness to fight climate change.

Second, we document large heterogeneity in beliefs about prevalent behaviors and norms in the US. We find that respondents on average misperceive prevalent social norms. On average, respondents in our sample underestimate the true share of Americans who try to fight global warming as well as the true share of Americans who think that people in the US should try to fight global warming. This underestimation of climate norms is concerning because it could hamper individual willingness to fight climate change. Whether or not correcting these misperceptions can shift climate behavior is a question that we explore with the survey experiment.

Third, we find that both treatments positively affect individual willingness to fight climate change. Being informed about the true share of Americans who try to fight global warming raises donations by $12 (or 4.7%), while being informed about the true share of Americans who think that people in the US should try to fight global warming increases donations by $16 (or 6.3%). The effect sizes are strong considering the minimalist nature of the interventions. A heterogeneity analysis reveals that the positive treatment effects on the donation amount are primarily driven by the subgroup of respondents whose prior beliefs lie below the actual shares. Reassuringly, we do not observe a back-firing effect among respondents with prior beliefs above the actual shares. For them, the estimated treatment effects are also positive, albeit insignificant. We further explore whether the information treatments differentially affect individuals who are more or less skeptical about the existence and threat of human-caused climate change. We find that the information treatments are more effective for 'climate change deniers', who may have been surprised to learn that they hold minority views. The results are promising as they suggest that simple, low-cost informational interventions may be well-suited to reach skeptical subgroups of the population who are otherwise difficult to reach and convince.

Finally, we study whether the treatments causally affect individual support for climate policies (e.g. a carbon tax, subsidies for green energy, pollution regulation) and individual willingness to engage in political actions (e.g. volunteer time, attend a protest, contact government officials). Both treatments significantly raise individual support for climate policies. Again, the estimated treatment effects are stronger for the subgroup of the population who we classify as 'climate change deniers'.

Our findings have important implications for climate politics. Misperceptions of climate norms prevail in the US and can form a dangerous obstacle to climate action. However, at the same time, they can provide a unique opportunity to promote and accelerate climate-friendly behavior. A simple, easily scalable, and cost-effective intervention can correct these misperceptions and encourage climate-friendly behavior. This intervention is particularly effective for climate change skeptics, who are commonly difficult to reach but crucial for building up a broad alliance against climate change. Our results suggest that social norms should play a pivotal role in the policy response to climate change. Policies that foster social norms should complement formal regulations. For example, while carbon taxation is an effective tool to curb CO_2 emissions, muted public support for such environmental policies has so far been a significant political constraint. Fostering social norms might alleviate these political constraints by increasing support for environmental policies—even if they are individually costly.

Our study builds on and contributes to several strands of the literature. First, we contribute to the literature studying the role of social norms in human behavior (see, e.g., Durlauf and Young, 2001; Bowles, 2004; Young, 2008; Young, 2015; Nyborg, Anderies, Dannenberg, Lindahl, Schill, et al., 2016; Nyborg, 2018). We extend this literature and show that individual beliefs about prevalent climate behaviors and norms strongly predict individual willingness to fight climate change. Importantly, we document that Americans vastly *underestimate* the true share of their fellow citizens who try to fight or think that Americans should try to fight global warming. We show that correcting these misperceptions leads to a significant increase in individual willingness to fight climate change and increases individual support for climate-friendly public policies.[1]

Misperceptions of social norms have been documented in settings where social norms are in a phase of transition, giving rise to a phenomenon referred to as '*pluralistic ignorance*' (Allport, 1924; Miller and McFarland, 1987). The majority of a population may privately endorse a norm but incorrectly assume that it is not endorsed by others. This incorrect belief may discourage people from endorsing the norm in public, thereby confirming other people's pessimistic beliefs. For instance, Kuran (1991) argues that a misperception of others' attitudes delayed the collapse of the communist regime in the Soviet Union. More recently, Bursztyn, González, and Yanagizawa-Drott (2020) study the role of misperceived social norms regarding female labor force participation in Saudi Arabia. Our evidence suggests that pluralistic ignorance exists in the context of climate norms and that a low-cost intervention has the potential to significantly alter individual willingness to fight climate change. Thereby, we contribute to recent work which shows that misperceptions about others' behavior, traits, and attitudes are widespread (Bursztyn and Yang, 2021). For instance, research in psychology and political science documents that people tend to underestimate how many of their fellow citizens believe that climate change is real and dangerous (Leviston, Walker, and Morwinski, 2013; Geiger and Swim, 2016;

1. Related to our work are recent studies showing that informational interventions that raise people's awareness about their neighbors' energy consumption or water use causally affect energy or water demand (see, e.g., Allcott (2011), Costa and Kahn (2013), Ferraro and Price (2013), and Jachimowicz, Hauser, O'Brien, Sherman, and Galinsky (2018)). In contrast to these studies, we provide causal evidence that (misperceived) social norms play a role in determining individual willingness to fight climate change and support for public policies. Our study thus also differs from recent correlative analyses that find a positive association between norm perception and environmental behavior (Farrow, Grolleau, and Ibanez, 2017; Valkengoed and Steg, 2019).

Pearson, Schuldt, Romero-Canyas, Ballew, and Larson-Konar, 2018; Mildenberger and Tingley, 2019; Ballew, Rosenthal, Goldberg, Gustafson, Kotcher, et al., 2020).

Moreover, we contribute to the literature examining the relationship between economic preferences and human behavior. Fundamental economic preferences such as time preferences, risk preferences, or prosociality have been shown to predict a wide range of human behaviors (see, e.g., Barsky, Juster, Kimball, and Shapiro, 1997; Dohmen, Falk, Huffman, and Sunde, 2009; Dohmen, Falk, Huffman, Sunde, Schupp, et al., 2011; Falk, Becker, Dohmen, Enke, et al., 2018; Figlio, Giuliano, Özek, and Sapienza, 2019). They have also been shown to predict a set of specific pro-environmental behaviors such as individual willingness to save energy or invest in energy-efficient technology (see, e.g., Newell and Siikamki, 2015; Schleich, Gassmann, Meissner, and Faure, 2019; Fischbacher, Schudy, and Teyssier, 2021; Lades, Laffan, and Weber, 2021). In contrast to these studies, we examine the relationship between economic preferences and individual willingness to fight climate change – as measured through an incentivized donation decision – in a large, representative sample of US adults. The decision to give up money to protect the climate reflects a central trade-off that individuals face when deciding whether to engage in climate-friendly behavior. This allows us to abstract from ancillary factors that are likely to shape specific pro-environmental decisions but are context-specific (e.g., the riskiness of investments in energy-efficient technology).

Finally, we explore the relationship between universal moral values and individual willingness to fight climate change.[2] Recent advances in moral psychology posit that people's moral values can be partitioned into different moral foundations and that holding universal moral values predicts individual behaviors such as voting or support for polices such as environmental protection (Haidt and Joseph, 2004; Haidt, 2012; Graham et al., 2013; Enke, Rodríguez-Padilla, and Zimmermann, 2019; Enke, 2020; Welsch, 2020). We show that universal moral values predict climate preferences over and above what can be predicted by economic preferences such as social preferences. Holding universal moral values might be particularly relevant in the context of climate change, where local behavior has consequences for people around the globe.

3.2 Study Design

To study individual willingness to fight climate change and its behavioral determinants, it is important to obtain a reliable and inter-personally comparable measure of individual willingness to fight climate change as well as detailed information on its potential determinants, such as perceived social norms, fundamental economic

2. See Drews and Bergh (2016), Gifford (2011), or Swim, Clayton, Doherty, Gifford, Howard, et al. (2009) for broad reviews of other determinants of climate behavior and climate policy support.

preferences, and moral values. To make inferences about the US population, a large representative sample is required. Establishing a causal relationship between perceived social norms and climate behavior further requires exogenous variation in the perception of norms. This section explains how we design the sampling approach and survey to meet these requirements.

3.2.1 Sample and survey procedures

We collect survey data from a representative sample of 8,000 study participants in the US. To be eligible to participate in the study, respondents had to reside in the US and be at least 18 years old. The data collection was carried out in two waves. The first wave of data ($N = 2,000$) was collected in March 2021. This wave of data forms the basis for the descriptive analysis presented in this paper, and informs the treatments embedded into wave 2. The second wave of data ($N = 6,000$) was collected in April 2021 and it contains the information experiment that allows us to study the causal relationship between perceived social norms and individual willingness to fight climate change.[3]

We used a stratified sampling approach to ensure that the samples represent the adult US population in terms of gender, age, education, and region. Comparing our samples to data from the American Community Survey 2019, we note that the distribution of demographic characteristics in our samples closely matches the distribution of characteristics in a nationally representative sample (see Appendix Table 3.A.1).

The survey contains several modules. In the following, we explain how we measure individual willingness to fight climate change (Section 3.2.2) and proceed with describing our measures of potential determinants (Section 3.2.3). We then present the information intervention embedded into wave 2 and explain how we elicit posterior beliefs (Section 3.2.4). We also measure individual support for climate policies, political engagement, climate change skepticism and a range of background characteristics (Section 3.2.5). The exact wording of the main survey blocks is provided in Appendix 3.B.

3.2.2 Measuring individual willingness to fight climate change

To measure individual willingness to fight climate change, we use an incentivized donation paradigm. Respondents are asked to divide $450 between themselves and

3. To collect the data, we collaborated with the professional survey company *Pureprofile*, which is frequently used in social science research. All survey participants were part of the company's online panel and participated in the survey online. The online surveys were scripted in the survey software Qualtrics. In both waves, the median time to complete the survey was 18 minutes. Respondents could only participate in one of the two waves. We screen out participants who do not pass an attention check (see Appendix 3.B.1) or speed through the survey with a duration of less than three minutes. Both exclusion criteria are pre-registered.

atmosfair, a charitable organization that fights global warming and offsets CO_2 emissions.[4] The more money that a respondent is willing to donate, the higher their willingness to fight climate change. The measure is quantitative and inter-personally comparable, and it captures the central trade-off underlying most individual-level decisions to fight climate change: mitigating climate change comes at a cost, whether in terms of money, time, or convenience. The amount of $450 was chosen because, by donating the full amount, respondents could offset the annual CO_2 emissions of an average US citizen.[5] We explain this to respondents in order to put their contribution decision into context and render it meaningful and tangible.

Before respondents make their decision, the instructions provide further information on *atmosfair*. Participants are informed that the charity actively contributes to CO_2 mitigation by promoting, developing, and financing renewable energies worldwide. Further information is provided on the charity's annual expenditure dedicated to the fight against global warming ($12 million) as well as its low overhead costs (5%). To minimize rounding, respondents can indicate their responses using a slider ranging from $0 to $450.

The incentive scheme is probabilistic: 25 participants are chosen at random and their decisions are implemented accordingly. The use of high-stake incentives mitigates the problem of experimenter demand effects or social desirability bias that might be present in hypothetical decisions.

3.2.3 Measuring behavioral determinants

Perceived social norms. Social norms are behavioral rules that express the collectively shared understanding of what is typical and morally acceptable behavior. They set the standards of conduct, shape individual behavior, are decentrally enforced, and could thus create a potent momentum either in favor of or against climate action (Bicchieri, 2006; Krupka and Weber, 2013; Nyborg et al., 2016; Bursztyn and Jensen, 2017). We thus hypothesize that individual willingness to fight global warming is determined by individual perceptions of other people's behavior (*'perceived behavior'*) as well as individual perceptions of what other people believe should be

4. Throughout the survey, we use the term "global warming" instead of the preferred scientific term "climate change" as the former is less likely to be confused with short term or seasonal weather changes or ozone depletion, a misunderstanding that still occasionally arises (Lorenzoni, Leiserowitz, Franca Doria, Poortinga, and Pidgeon, 2006). To avoid confusion, we define global warming as follows at the beginning of the survey: "*Global warming means that the world's average temperature has considerably increased over the past 150 years and may increase more in the future.*" Throughout this text, we use the terms global warming and climate change interchangeably.

5. At the time of the survey, it cost about $28 to offset 1 ton of CO_2 emissions. The World Bank estimates that a typical US resident causes about 16 tons of CO_2 emissions per year.

done (*'perceived norms'*).[6] Beliefs about the choices that other people make reflect the perceived behavioral standard or norm in a community, which is particularly relevant when people condition their cooperation on the action of others ('conditional cooperation', Fischbacher, Gächter, and Fehr, 2001). Beliefs about what other people consider appropriate reflect the perceived moral rules or principles in a community. People might have a preference to adhere to the prevalent rules to protect their reputation or self-image (Bursztyn and Jensen, 2017; Falk, 2021).

Before eliciting respondents' perceptions about prevalent social norms, we first ask respondents two questions which allow us to establish prevalent behaviors and endorsement of norms in a representative sample of US adults. Specifically, we ask all respondents about their own behavior and endorsement of the norm to fight global warming: Do they "try to fight global warming" (yes/no)? Do they believe "people in the US should try to fight global warming" (yes/no)?

To measure *perceived* social norms, we then ask all respondents to estimate what proportion of the US population "try to fight global warming" ('perceived behavior') and what proportion think that "people in the US should try to fight global warming" ('perceived norms'). Before making their guesses, respondents are informed that we have gathered survey evidence on whether people try to fight global warming and whether they think that people in the US should try to fight global warming. More specifically, it is explained that we have surveyed a large sample of the US population and that the survey results "represent the views and attitudes of people in the United States". For ease of comprehension, respondents are not asked to estimate proportions but rather estimate the number of people to whom the statement applies out of 100 people we asked:

- *Out of 100 people we asked, how many stated that they try to fight global warming?*
- *Out of 100 people we asked, how many stated that they think that people in the United States should try to fight global warming?*

To determine whether individual perceptions are correct, we can compare participants' guesses with the actual shares of wave 1 respondents answering affirmatively to the questions whether they "try to fight global warming" and whether they think that "people in the US should try to fight global warming". We incentivize the guesses that respondents make to induce and reward careful and accurate responses. In particular, every respondent can earn a $1 bonus if their guess in a randomly-selected belief question differs at most by three from the true value.[7] The resulting

6. The former are sometimes referred to as descriptive norms or empirical beliefs, while the latter are also sometimes referred to as second-order normative beliefs, injunctive norms, or prescriptive norms (Cialdini, Reno, and Kallgren, 1990; Bicchieri, 2006).

7. The perceived behavior and the perceived norms question are the central but not the only belief questions in the survey. In total, we ask fifteen different belief questions, all of which are incentivized by the reward scheme. The additional belief questions are introduced in Section 3.2.4.

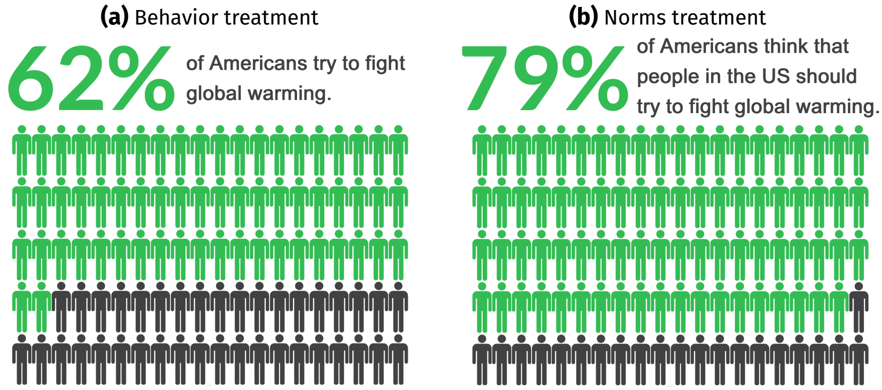
measures of perceived behaviors and perceived norms are simple, yet quantitative, incentivized, and inter-personally comparable. Together, they capture the two key facets of social norms that have been identified as key drivers of human behavior in many contexts.

Economic preferences. Economic preferences have been shown to predict a range of important decisions and they are likely to be important determinants of individual willingness to fight climate change. To explore the relationship between economic preferences and the propensity to fight global warming, we obtain detailed individual-level measures of economic preferences following the methodology used in the Global Preferences Survey (Falk, Becker, Dohmen, Enke, et al. (2018) and Falk, Becker, Dohmen, Huffman, et al. (2018)). This experimentally validated survey relies on a range of qualitative and quantitative survey items and allows us to construct preference measures for six fundamental preferences: *patience*, *willingness to take risks*, *altruism*, *trust*, *positive reciprocity*, and *negative reciprocity*. The latter two capture the willingness to reward kind or punish unkind actions, respectively. More information on the survey items and how the composite measures are computed can be found in Appendix 3.C. For ease of interpretation, each preference measure is standardized to have a mean of zero and a standard deviation of one.

Universal moral values. Moral universalism captures the tendency to extend altruistic and moral concerns to individuals who are socially distant (Singer, 2011; Crimston, Bain, Hornsey, and Bastian, 2016; Enke, 2020). Given the global nature of climate change, there are strong reasons to hypothesize that individual willingness to fight global warming is determined by the relative importance of universal versus communal moral values. Moral Foundations Theory (MFT) posits that people's moral concerns can be partitioned into five distinct foundations: care/harm, fairness/reciprocity, in-group/loyalty, authority/respect, and purity/sanctity. "Universal" values – captured by the care/harm and fairness/reciprocity foundations – apply irrespective of the people involved. "Communal" values – captured by the in-group/loyalty and authority/respect foundations – are tied to certain groups or relationships (Haidt and Joseph, 2004; Haidt, 2012; Graham et al., 2013; Enke, 2020). We administer the Moral Foundations Questionnaire (MFQ) to measure the distinct foundations and calculate the relative importance of universal moral values following the approach proposed by Enke (2020). More information on how the standardized measure is constructed can be found in Appendix 3.C.

3.2.4 Shifting perceived social norms

Given the threat posed by global warming, it is important to understand which interventions could increase individual willingness to fight climate change. While it is difficult to alter fundamental human traits such as altruism, patience, or moral values in the short term, beliefs about social norms are likely to be considerably more



Notes: Panels a and b provide a visual summary of the information provided to participants in the behavior and the norms treatments, respectively. The exact wording of the survey instructions is provided in Appendix 3.B.

Figure 3.2.1. Information treatments in wave 2

malleable.[8] As we will show in Section 3.3.2, respondents on average misperceive the prevalence of social norms in the US. Motivated by this finding, we embed an information experiment into wave 2. The exogenous variation induced by this experiment allows us to study whether the perceived prevalence of social norms causally affects individual willingness to fight global warming.

After eliciting respondents' beliefs about prevalent behaviors and norms, we provide randomly-selected participants with truthful information about the proportion of the US population who (i) "try to fight global warming" ('*behavior treatment*') or (ii) think that "people in the US should try to fight global warming" ('*norms treatment*'). Estimates of both shares are derived from wave 1. More specifically, we randomize respondents in wave 2 into one of three treatments. Appendix Figure 3.A.1 summarizes the structure of the experiment.

1. Behavior treatment In this treatment, respondents are informed about the share of the US population who "try to fight global warming". Respondents are first informed about the fact that "we recently surveyed 2,000 people in the United States and asked them whether they try to fight global warming. Respondents come from all parts of the population and their responses represent the views and attitudes of people in the United States." On the following page, respondents learn that 62% of Americans try to fight global warming. To ensure that participants pay attention, the

8. Economic preferences such as altruism and patience are also malleable, especially during the childhood period, and can be affected through educational interventions in the case of patience (Alan and Ertac, 2018) or through an enriched social environment in the case of altruism (Kosse, Deckers, Pinger, Schildberg-Hörisch, and Falk, 2019; Rao, 2019). While it is possible that such interventions can lead to an increased willingness to fight climate change, these interventions are more difficult to implement on a larger scale.

information is revealed piece by piece, and respondents need to spend a minimum of 5 seconds on the final screen before being able to proceed. A graph on the final screen expresses the information visually, making it salient and tangible (see Figure 3.2.1.a).

2. Norms treatment In an analogous manner, respondents in the norms treatment learn that 79% of Americans think that people in the US should try to fight global warming (see Figure 3.2.1.b).

3. Control No information is provided to participants in the control condition.

Subsequently, we elicit individual willingness to fight climate change with the incentivized donation decision (see Section 3.2.2), which constitutes our main outcome measure. This study design allows us to assess whether providing respondents with accurate information about prevalent behaviors or norms can shift individual climate behavior.

Respondents randomized into the behavior or norms treatment are likely to revise their beliefs about prevalent behaviors or norms in the US. Such a shift in beliefs may lead to a change in individual willingness to fight climate change. Since – as we will show – individuals systematically underestimate the share of Americans trying to fight global warming as well as the share who think that Americans should try to fight global warming, we posit that the information interventions are likely to increase individual willingness to fight climate change. We opt for the dual approach of shifting both perceived behavior and perceived norms, as both are regarded as central drivers of human behavior. However, conceptually, these two entities are closely related. A change in perceived behavior may also lead to a change in perceived norms and vice versa. We explore this question in further detail in Section 3.3.3.

To study belief revisions, we include a post-treatment module in which we elicit posterior beliefs. Respondents are asked to estimate what proportion of the US population engages in a set of concrete climate-friendly behaviors ('perceived behaviors') and what proportion of the US population thinks that one should engage in those behaviors ('perceived norms'). The set of concrete behaviors includes restricting meat consumption, avoiding flights, using environmentally-friendly alternatives to fossil-fueled cars, using green electricity, adapting shopping behavior to the carbon footprint of products, and politically supporting the fight against global warming. Guesses are incentivized using the same reward scheme as described in Section 3.2.3. To determine whether guesses are correct, we compare individual responses to the actual share of wave 1 respondents who report engaging in these behaviors or stating that they think one should engage in those behaviors. For the purpose of the analysis, we compute a perceived behavior index and a perceived norms index by calculating the average across the six climate-friendly behaviors/norms items. We then standardize each index to have a mean of zero and a standard deviation of one among control group respondents. Conceptually, individual perceptions about the prevalence of concrete behaviors/norms are strongly related to the more general behavior/norm of "trying to fight global warming". We can thus use those questions to test for and detect belief revisions without repeating our main questions, thereby mitigating experimenter demand effects and consistency bias in survey responses (Haaland et al., forthcoming).

3.2.5 Additional measures

Climate change skepticism. The public and political debate on climate change has been shaped by a denial of its existence, dangers, or human origin. This phenomenon is particularly relevant in the US where climate change skepticism is widespread and has often formed a key obstacle to effective responses against climate change (Dunlap and McCright, 2011; Leiserowitz, Maibach, Roser-Renouf, Smith, and Dawson, 2013). The subgroup of climate change deniers thus holds particular political relevance, and the survey includes a diverse set of items that allow us to measure respondents' skepticism. We ask respondents to indicate how much trust they have in climate science, whether they think scientists agree that global warming is happening, how worried they are about global warming, whether they think it will harm people in the US, and whether they think that climate change is human-caused (see Appendix 3.B). These questions are asked at the beginning of the survey to ensure that the responses are not affected by the information treatments. We use this information to explore the heterogeneity of treatment effects.

Policy support and political engagement. In addition to eliciting individual willingness to fight climate change, we collect detailed information on the extent to which individuals support different climate polices (e.g., a carbon tax, subsidies for green energy, pollution regulation) and are willing to engage politically (e.g., volunteer time, attend protest, contact government officials). We pose a total of 18 questions adapted from a detailed politics module developed as part of the Climate Change in the American Mind Project (Howe, Mildenberger, Marlon, and Leiserowitz, 2015). Respondents can express their policy support and individual political engagement on a four-point Likert scale (see Appendix 3.B), which we recode in our analysis to ensure that larger values indicate more policy support and political engagement. For ease of interpretation, we aggregate individual items into a policy support index (7 items), a political engagement index (11 items), and a joint index comprising all 18 items. Each index is standardized to have a mean of zero and a standard deviation of one among control group respondents. The questions are posed after the information treatments in wave 2, which allows us to study whether shifting beliefs about prevalent behaviors and norms causally affects policy support and willingness to engage politically.

Background characteristics. We collect detailed information on individual background characteristics. Those include age, gender, education, employment status, household income, the number of children, and whether the respondent thinks of themselves as being closer to the Republican or Democratic party. We use those variables as additional control variables in the analysis.

3.3 Results

3.3.1 Willingness to fight climate change and its determinants

To measure climate preferences, we use an incentivized donation decision in which respondents divide \$450 between themselves and a charitable organization that fights global warming. We use this measure to study how climate preferences are distributed across the population and examine which factors predict those preferences. For the purpose of this descriptive analysis, we focus on survey data collected in wave 1 ($N = 2,000$), which did not contain any treatment manipulation.

Appendix Figure 3.A.2 displays the distribution of individual willingness to fight global warming, as measured through the incentivized donation decision. On average, respondents are willing to donate \$225 of the \$450. There is a considerable degree of heterogeneity across respondents, with 6% donating \$0, 12% donating \$450, and the remaining 82% donating some value in between.

We explore which factors predict individual willingness to fight climate change. For this purpose, we regress the donation amount (in \$) on (i) individual beliefs about prevalent behaviors or norms, (ii) our measures of fundamental economic preferences (i.e., patience, risk-taking, altruism, positive reciprocity, negative reciprocity, and trust), (iii) universal moral values, and (iv) a range of background characteristics. Given that beliefs about prevalent behaviors and norms are conceptually related and highly correlated in our data ($\rho = 0.67$), we estimate two separate regression models, including one belief measure at a time. For the purpose of this analysis, the belief measures are standardized to have a mean zero and a standard deviation of one. The results are reported in columns 1 and 2 of Table 3.3.1, respectively.

First, perceived behaviors and norms are strong predictors of climate preferences. Controlling for the large set of covariates, a one-standard-deviation increase in perceived behavior is associated with a \$12 higher donation amount ($p < 0.001$), while a corresponding increase in perceived norms is associated with a \$14 higher donation ($p < 0.001$).[9] These results are consistent with norm perceptions playing an important role in determining individual willingness to fight global warming. This could, for example, be the case if individuals are 'conditional cooperators' or if they have a preference for complying with existing social norms. Whether or not this relationship can be interpreted as causal is a question we turn to in Section 3.3.3.

9. We note that both belief measures have a standard deviation of 22 percentage points. The coefficients can therefore also be interpreted as follows: A 10 percentage point increase in the behavior belief is associated with a \$5.50 higher donation amount, while a corresponding increase in the norms belief is associated with a \$6.50 higher donation amount.

Table 3.3.1. Determinants of climate change behavior

	Donation ($)	
	(1)	(2)
Perceived social norms		
Behavior belief	12.237***	
	(3.154)	
Norms belief		14.500***
		(3.058)
Economic preferences		
Altruism	51.267***	51.734***
	(3.477)	(3.448)
Patience	15.195***	15.192***
	(3.105)	(3.096)
Risk	-1.411	-0.792
	(3.373)	(3.354)
Positive reciprocity	9.571***	7.877**
	(3.239)	(3.258)
Negative reciprocity	-3.338	-2.540
	(3.214)	(3.185)
Trust	1.071	0.831
	(3.233)	(3.203)
Moral foundations		
Relative universalism	23.772***	23.420***
	(3.301)	(3.290)
Sociodemographics		
Democrat	45.143***	44.160***
	(6.241)	(6.246)
Age	0.685	0.702
	(1.035)	(1.034)
Age (squared)	-0.007	-0.006
	(0.011)	(0.011)
Female	16.943***	16.520***
	(6.367)	(6.331)
Log income	9.965***	9.895***
	(3.741)	(3.726)
College degree	-15.320**	-15.953**
	(6.522)	(6.504)
Employed	8.453	8.868
	(6.661)	(6.638)
Parent	4.659	4.695
	(6.498)	(6.478)
R^2	0.281	0.284
N	1,975	1,975
Mean of dep. var.	225.21	225.21

Notes: This table shows OLS regression estimates using respondents from wave 1, where the dependent variable is the amount donated to the charitable organization that fights global warming. Perceived social norms, economic preferences, and universal moral values are standardized. “Democrat”, “Female”, “College degree”, “Employed” and “Parent” are binary indicator variables. “Log income” is coded as the log of the income bracket’s midpoint. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Second, the results presented in Table 3.3.1 further reveal that climate donations are fundamentally related to economic preferences. Altruism and positive reciprocity – both of which are facets of prosociality – positively predict the donation amount. The magnitudes of the estimated coefficients are sizeable. For example, a one-standard-deviation increase in altruism is associated with a $52 higher donation amount. Similarly, patience positively predicts donation decisions. These patterns are plausible given that climate action benefits other people around the world as well as future generations. We find no statistically significant associations between climate preferences and risk preferences, negative reciprocity, or trust.

Third, we find a strong positive association between universal moral values and climate preferences. A one-standard-deviation increase in relative universalism – namely the extent to which individuals endorse universal moral values that apply equally to all humans rather than communal or ingroup-restricted values – is associated with a $23 higher donation amount. Climate change is a global problem and individuals whose moral values apply irrespective of the people involved are more likely to make larger donations, presumably because they are more likely to take the welfare of other people outside of their community into account.

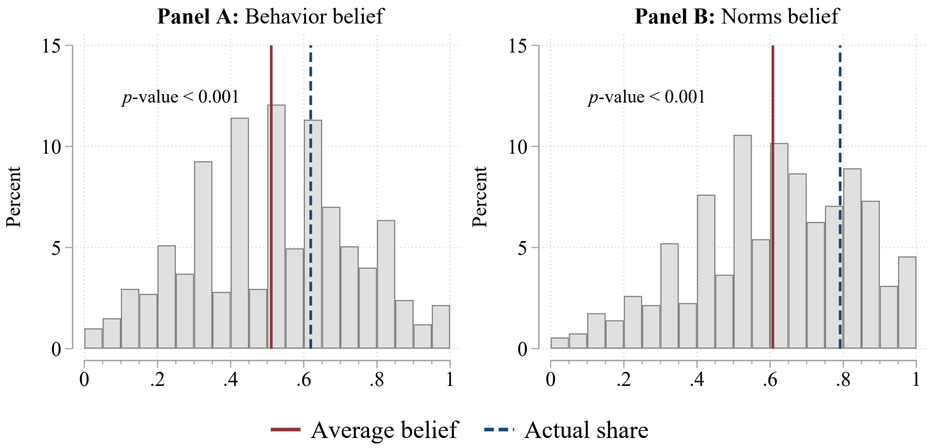
Finally, demographic characteristics also significantly predict individual willingness to fight climate change. Democrats on average contribute about $45 more than Republicans, female respondents about $16 more, and household income is also positively associated with the donation amount. However, higher education negatively predicts climate donations. Further analyses reveal that this effect is entirely driven by Republicans among whom a college education is associated with a $27 lower donation amount (see Appendix Table 3.A.2).[10]

Taken together, the results suggest that perceived social norms, economic preferences, and universal moral values are likely to shape individual willingness to fight climate change. Since climate action is commonly conceived as a global and intergenerational cooperation problem, it seems plausible that a higher willingness to fight climate change requires some degree of prosociality, patience, and universal moral values. Beliefs about prevalent behaviors and norms are also likely to be key determinants of individual willingness to fight global warming if individuals act as 'conditional cooperators' or have a preference to comply with existing social norms.

3.3.2 Misperceived social norms

Having established which factors are predictive of individual willingness to fight climate change, we now explore the distribution of beliefs about behaviors and norms

10. We are not the first to document a negative education gradient among Republicans (Hamilton, 2011; Newport and Dugan, 2015). It has been hypothesized that highly-educated individuals are cognitively better equipped to rationalize and internalize the views of their cultural community, which for Republicans might correspond to climate change skepticism (Kahan, Peters, Wittlin, Slovic, Ouellette, et al., 2012; but see Van Der Linden, Maibach, Cook, Leiserowitz, Ranney, et al., 2017).



Notes: This figure shows the distribution of perceived social norms in wave 1. Panel A shows the distribution of people's beliefs about the share of Americans who say that they try to fight global warming. Panel B shows the distribution of people's beliefs about the share of Americans who say that one *should* fight global warming. Each panel indicates the average belief across respondents (solid red) as well as the actual shares (dashed blue) as vertical lines.

Figure 3.3.1. Perceived social norms: fight global warming

in more detail. Given that these beliefs are potentially malleable, it holds particular importance to establish whether there are systematic misperceptions of prevalent behaviors and norms. For the purpose of this analysis, we again rely on the survey data collected in wave 1.

Figure 3.3.1 depicts the distribution of perceived social norms. Panel A displays perceived behavior, i.e., the distribution of individual beliefs about the share of the US population that tries to fight global warming. Panel B displays perceived norms, i.e., the distribution of beliefs about the share of Americans who think that people in the US should try to fight global warming. The average belief is indicated by a vertical red line, whereas the actual share is marked by a dotted blue line.

Figure 3.3.1 reveals a considerable degree of heterogeneity in individual beliefs. Both panels further reveal that respondents vastly misperceive the prevalence of climate-friendly behaviors and norms among their fellow citizens. On average, respondents believe that 51% of Americans try to fight global warming, while the actual share is 62% (p-value < 0.001). The majority of participants – namely 67% – underestimate how prevalent climate-friendly behavior is in the US. Similarly, respondents on average believe that 61% of Americans think that people in the US should try to fight global warming, while the actual share is 79% (p-value < 0.001). Again, most participants (76%) underestimate this share.[11] We find larger misper-

11. We also elicit beliefs about concrete climate change behaviors, e.g., restricting meat consumption, avoiding flights and cars, or consuming only green electricity. These measures are highly

ceptions among respondents who are older, have a lower income, have a lower education, or are Republicans (see Appendix Table 3.A.4).

Taken together, while the majority of Americans try to fight global warming and a vast majority agrees that people in the US should try to fight global warming, most Americans underestimate the degree to which other Americans engage in climate-friendly behaviors and share those normative views. This underestimation of climate norms is likely to hamper individual willingness to fight climate change.

3.3.3 Correcting misperceived social norms

As established in the previous sections, beliefs about prevalent behaviors and norms strongly predict individual willingness to fight climate change. At the same time, there are systematic misperceptions of the actual share of Americans fighting or thinking that one should fight climate change. Can information interventions that inform respondents about the true shares can affect individual willingness to fight climate change? The information experiment embedded in wave 2 allows us to study this question. Respondents are randomized into (i) a 'behavior treatment', in which they are informed that 62% of Americans try to fight global warming, (ii) a 'norms treatment', in which they are informed that 79% of Americans think that people in the US should try to fight global warming, or a (iii) a control group. Appendix Table 3.A.3 presents the balancing of characteristics across the three groups. We cannot reject the null hypothesis that the three groups differ in terms of observable characteristics and conclude that the randomization was successful. Appendix Figure 3.A.3 displays the wedge between wave 2 respondents' beliefs about prevalent behaviors and norms and the actual shares. As can be seen from both figures, wave 2 participants also vastly underestimate the true shares, providing us with an ideal opportunity to exogenously correct inaccurate perceptions. The average gap between the perceived and actual shares is 10 percentage points in the case of perceived behaviors and 17 percentage points in the case of perceived norms.

To estimate the causal impact of the information treatments, we regress willingness to fight climate change – as measured through the incentivized donation decision (in $) – on treatment indicators and a set of control variables.[12] The results are reported in column 1 of Table 3.3.2 and reveal that the impacts of the information treatments are sizeable and highly statistically significant. Being informed about the true share of Americans who try to fight global warming leads to a $12 increase in donations (p-value = 0.012), while being informed about the true share

correlated with the abstract measure (see Appendix Table 3.A.5). Moreover, Appendix Figures 3.A.4 and 3.A.5 show that we document similar norm misperceptions for these concrete behaviors.

12. The set of control variables includes controls for gender (indicator), age (continuous), log income, college degree (indicator), employment (indicator), party affiliation (indicator), and census region (three indicators). Appendix Table 3.A.6 presents results of the regressions without control variables. The estimated coefficients are very similar in magnitude and significance.

Table 3.3.2. Treatment effects on climate donations and posterior beliefs

	(1) Donation ($)	(2) Behavior belief (post.)	(3) Norms belief (post.)
Behavior treatment	11.725**	0.279***	0.235***
	(4.675)	(0.030)	(0.030)
Norms treatment	15.674***	0.370***	0.350***
	(4.701)	(0.031)	(0.030)
N	5,991	5,988	5,976
Control group mean	249.31	0	0
z-scored	No	Yes	Yes
Controls	Yes	Yes	Yes

Notes: This table shows OLS regression estimates using respondents from wave 2. The dependent variable is the donation to the climate charity (in $). It is regressed on binary indicators that take the value of 1 for respondents in the behavior treatment and norms treatment, respectively. "Behavior belief" is an index of six post-treatment beliefs about the share of Americans engaging in concrete climate-friendly behaviors to fight global warming. "Norms belief" is an index of six post-treatment beliefs about the share of Americans who say that one should engage in concrete climate-friendly behaviors to fight global warming. Both indices are standardized to have a mean of zero and a standard deviation of one in the control group. All regressions include controls for gender (indicator), age (continuous), log income, college degree (indicator), employment (indicator), party affiliation (indicator), and census region (three indicators). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

of Americans who think that people in the US should try to fight global warming increases donations by $16 (p-value < 0.001). The effects correspond to a relative increase of 4.7% and 6.3%, respectively. While the point estimate of the coefficient for the norms treatment is somewhat larger than the point estimate of the coefficient for the behavior treatment, we note that the two are not significantly different from each other (p-value = 0.39). Given that not all respondents misperceive prevalent behaviors and norms at the baseline and some respondents might not fully revise their beliefs in light of the information provided, both effect sizes suggest a powerful impact of perceived social norms on individual willingness to fight climate change.[13]

Using the posterior norm perception module, we provide evidence that the treatments indeed shift posterior beliefs in the way that one would expect. To study belief revisions, we regress the posterior beliefs about concrete climate-friendly behaviors and norms on the treatment indicators and the same set of control variables. As explained in Section 3.2.4, the set of concrete behaviors includes different actions such

13. We can derive the treatment effect per standardized change in beliefs under the assumption that respondents fully update their beliefs to the information provided, which implies an average belief increase of 0.47 standard deviations in the behavior treatment and 0.82 standard deviations in the norms treatment. The behavior treatment thus has a $24.8 effect and the norms treatment a $19.2 effect on climate donations per standardized belief change. Both figures likely underestimate the true effect because most respondents presumably only partially update their beliefs.

Table 3.3.3. Treatment effect heterogeneity: Prior above/below actual share

	Dependent variable: Donation ($)			
	Prior < actual share		Prior ≥ actual share	
	(1)	(2)	(3)	(4)
Behavior treatment	14.931**		5.231	
	(5.875)		(7.701)	
Norms treatment		19.111***		4.747
		(5.387)		(9.623)
N	2,579	3,054	1,399	946
Control group mean	243.09	241.67	260.69	273.71
Controls	Yes	Yes	Yes	Yes

Notes: This table shows OLS regression estimates using respondents from wave 2. The dependent variable is the donation to the climate charity (in $). It is regressed on binary indicators that take the value of 1 for respondents in the behavior treatment and norms treatment, respectively. We run separate analyses for respondents with prior norm perceptions strictly below the actual share (columns 1-2) and equal to or above the actual share (columns 3-4). We consider beliefs about others' behavior in the behavior treatment and beliefs about others' norms in the norms treatment. Given that the actual shares are different for the two beliefs, we do not pool all three treatment groups in this analysis. Instead, we only use respondents in the control condition and the behavior treatment in the analysis presented in columns 1 and 3, and only use respondents in the control condition and the norms treatment in the analysis presented in columns 2 and 4. All regressions include the set of controls described in Table 3.3.2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

as reducing meat consumption or avoiding flights. The two posterior belief indices are standardized, and the results are reported in columns 2 and 3 of Table 3.3.2, respectively. Both information treatments successfully shift beliefs, which are revised upwards by 0.24 to 0.37 standard deviations. We also observe spill-over effects. Information about prevalent behavior also shifts beliefs about prevalent norms and vice versa. As remarked earlier, the treatments should not be interpreted as separate manipulations of orthogonal concepts but rather as statistically independent yet conceptually-related treatments with a common effect: they both strengthen perceived social norms.

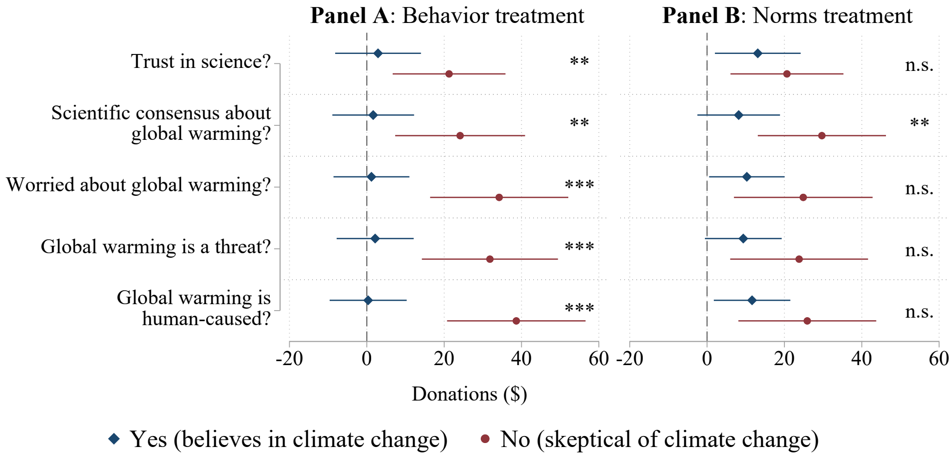
Treatment effect heterogeneity by prior. We explore heterogeneity in treatment effects across different subgroups. First, we examine whether the treatments are more effective for respondents whose priors are below the actual shares. Table 3.3.3 separately displays the treatment effects for respondents whose prior beliefs are below the true shares (Panel A) and those whose prior beliefs are equal to or above the true shares (Panel B). As can be seen from this table, the positive treatment effects that we document for the full sample are almost entirely driven by those individuals whose priors are below the actual shares. Among them, the behavior treatment increases donations by $15 (p-value = 0.011), whereas the norms treatment increases donations by $19 (p-value < 0.001). Reassuringly, we do not observe a back-firing effect. For respondents whose priors are equal to or above the actual shares, the estimated coefficients are positive albeit smaller in magnitude and insignificant.[14] However, we note that we cannot reject the null hypothesis that the treatment effect coefficients are the same for both subgroups.

Treatment effect heterogeneity by climate change skepticism. Next, we explore whether the information treatments lead to a stronger increase in individual willingness to fight climate change for respondents who are skeptical about the existence and threat of human-caused climate change. From a policy perspective, this subset of the population is particularly relevant as it is typically difficult to reach and convince that climate change matters.

Figure 3.3.2 compares the treatment effects across respondents who express skepticism about climate change and those who do not. The sample is split based on five indicators that capture different facets of climate change skepticism: having low trust in climate science, believing that the presence of climate change is still scientifically debated, not being worried about climate change, not perceiving it as a threat for the US, and believing that climate change is mainly the result of natural causes. For all indicators and both treatments, we observe that the point estimates of the treatment coefficients are larger in magnitude for climate change deniers. In the behavior treatment, most coefficients are also statistically different from each other across the two subgroups. For example, the behavior treatment increases donations by $24 for those who report not being worried about global warming and by $39 for those who do not believe that climate change is human-caused. By contrast, we do not find a statistically significant impact of the behavior treatment for respondents who do report being worried or who do believe that climate change is human-caused. These differences in effect sizes are statistically significant at the 5% level (see also Table 3.A.7). In the norms treatment, the differences are more muted.

Climate change deniers tend to have more pessimistic prior beliefs about the prevalence of climate norms in the US. However, we observe largely identical results even if we control for treatment heterogeneity by priors (see Table 3.A.8). Thus, the same information appears to have differential informational value for climate change deniers – even conditional on the same prior belief. Climate change deniers do not only have more scope to adjust their behavior. They might also be surprised to learn that their views are in fact minority views and that the majority of their

14. Appendix Figure 3.A.6 displays non-parametric estimates of the moderating role of pre-treatment beliefs for our information treatments (Xu, Hainmueller, Mummolo, and Liu, 2017; Hainmueller, Mummolo, and Xu, 2019). As can be seen from this figure, the effects of the behavior and the norms treatment are stronger among respondents with low pre-treatment beliefs. Moreover, both treatments have a weakly positive effect across the whole belief distribution.



Notes: This figure shows OLS estimates of the treatment effects of the behavior (Panel A) and the norms treatment (Panel B) on donations (in $) in different subsamples. We use respondents from wave 2 and include the set of controls described in Table 3.3.2. 95% confidence intervals are shown. Each panel shows treatment effects among respondents who are skeptical of climate change ("No") and those who believe in climate change ("Yes"), where we use disagreement with different statements as a proxy for skepticism: "Trust in science" means that the respondent trusts climate scientists "a lot" or "a great deal" (on a five-point Likert scale). "Scientific consensus about global warming" means that the respondent thinks that most scientists think that global warming is happening. "Worried about global warming" means that the respondent is "somewhat worried' or "very worried" about global warming (on a four-point Likert scale). "Global warming is a threat" means that the respondent thinks that global warming will do "a moderate amount" or "a great deal" of harm (on a four-point Likert scale). "Global warming is human-caused" means that the respondent thinks that global warming is caused by human activities. For each sample split, we indicate the level of significance of a test of equality of coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, n.s. $p \geq 0.10$.

Figure 3.3.2. Treatment effect heterogeneity by climate change skepticism

fellow citizens does take climate change seriously, as indicated by the large share of Americans who take action against it or think that this should be done.[15]

Treatment effects on policy support and political engagement. Do the positive treatment effects of the information treatments also carry over to the political domain? To study this question, we collect post-treatment information on policy support and political engagement (see Section 3.2.5). Columns 1 and 2 of Table 3.3.4 present the estimated treatment effects on the standardized indices of support for climate policies and willingness to engage in political actions. Column 3 presents the results for the standardized, joint index. We find that both treatments significantly increase support for climate policies. The behavior treatment significantly increases

15. It is unlikely that the much weaker treatment effect among respondents who believe in and are concerned about climate change can be attributed to a "ceiling effect". In the control treatment, the large majority of these climate change "believers" (about 73% to 75% depending on the question) can still increase their donation by at least $25.

Table 3.3.4. Treatment effects on support for policies and actions to fight global warming

	(1) Policies	(2) Actions	(3) All
Behavior treatment	0.088***	0.039	0.061**
	(0.026)	(0.027)	(0.026)
Norms treatment	0.066**	0.012	0.034
	(0.026)	(0.027)	(0.026)
N	5,999	5,994	5,993
z-scored	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: This table shows OLS regression estimates using respondents from wave 2. Dependent variables: "Policies" is an index measuring individual support for policies to fight climate change (7 items). "Actions" is an index measuring political engagement in different types of political activities (11 items). "All" is a joint index comprising all 18 items. All indices are constructed by taking the sum of all positively coded items and standardizing the sum to have a mean of zero and a standard deviation of one in the control group. The indices are regressed on binary indicators that take the value of 1 for respondents in the behavior treatment and norms treatment, respectively. All regressions include the set of controls described in Table 3.3.2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

policy support by 0.09 standard deviations, while the norms treatment significantly increases policy support by 0.07 standard deviations. The estimated coefficients are positive albeit insignificant when we consider willingness to engage in political actions as the outcome. When we use the joint index as the outcome, we find that the behavior treatment significantly increases the index by 0.06 standard deviations, while the norms treatment has an insignificant positive effect of 0.03.

Consistent with the results reported above, we also find that the estimated impacts of the treatments on policy support and political engagement tend to be stronger for the subgroup of climate change deniers. Appendix Figure 3.A.7 shows that both the behavior and the norms treatment significantly increase individual support for policies to flight global warming by 10 to 20 percent of a standard deviation among climate change deniers. By contrast, our information treatments have hardly any impact on policy support among respondents who believe in climate change.

Taken together, we conclude that providing people with accurate information not only has the potential to increase individual willingness to fight climate change – especially among climate change deniers – but that it can also increase individual support for climate policies.

3.4 Discussion

We document that fundamental human traits such as altruism, positive reciprocity, and moral universalism are strong predictors of individual willingness to fight climate change. This finding could prove fruitful in understanding the frequently observed cultural and political dissent on climate change (Dunlap, McCright, and Yarosh, 2016; Hornsey, Harris, and Fielding, 2018). Indeed, in our data, economic preferences and universalism together explain about 40% of the large $74 baseline donation gap between Republicans and Democrats (see Appendix Table 3.A.9). Likewise, they explain 25% of the gap in policy preferences. Both results suggest that the political divide on climate change can be partially attributed to deeply entrenched human traits. The important role of prosociality further illustrates that many individuals care about the well-being of others and therefore seem to partially internalize the positive externalities of climate action. The traditional economic model of purely self-interested agents facing an insurmountable collective action problem thus underestimates the scope for climate action. Indeed, our survey documents that many Americans are actually willing to act against global warming. 62% of Americans try to fight global warming, and 79% think that this should be done. Moreover, many respondents are willing to give up money to support the work of a climate charity.

Our finding that Americans vastly underestimate the prevalence of climate norms in the US holds particular political relevance. We show both correlationally and causally that perceived social norms are a key driver of individual willingness to fight climate change. The fact that climate norms are commonly underestimated in the US can thus form a dangerous obstacle to climate action. It could trap Americans in an equilibrium with low climate engagement: Individuals are discouraged by the (mis)perceived lack of support, and they abstain from taking actions themselves, which sustains the pessimistic beliefs held by others – a phenomenon that has been dubbed pluralistic ignorance (Allport, 1924; Miller and McFarland, 1987; Bursztyn, González, and Yanagizawa-Drott, 2020).

However, this diagnosis also implies a unique opportunity to promote and accelerate climate-friendly norms and behavior. We show that a simple, easily scalable, and cost-effective intervention – namely informing respondents about the actual prevalence of climate norms in the US – corrects these misperceptions and encourages climate-friendly behavior. Importantly, we find that this intervention is particularly effective for climate change deniers, namely the group of people who are commonly difficult to reach, but crucial for building up a broad alliance against climate change. Moreover, convincing those who remain skeptical of human-caused climate change is likely to have particularly high returns if these individuals still have ample scope to make their behavior more climate-friendly.

Arguably, the effect of a single, minimalist message as embodied in our information treatments is likely to dissipate with time. However, large-scale information campaigns that repeatedly announce and effectively communicate the actual prevalence of climate norms could correct existing misperceptions and permanently foster climate norms (Bicchieri, 2017). They could trigger a positive feedback loop where learning about the existing support of climate norms encourages Americans to take visible action against climate change, which encourages others to follow suit.

3.5 Conclusion

In this paper, we study the behavioral determinants of individual willingness to fight climate change in a large-scale, representative survey with 8,000 US adults. In a first step, we document that fundamental human traits – namely patience, altruism, positive reciprocity, and moral universalism – are strongly correlated with individual willingness to fight climate change, as measured in a donation decision. Beliefs about the climate behavior and norms of others also matter: Individuals who perceive stronger climate norms are willing to give up more money to support the climate charity. In a second step, we zoom in on perceived social norms, as they are malleable in the short term and can create a potent momentum either in favor of or against climate action. We find that Americans strongly underestimate the support of climate norms in the US. An information experiment shows that informing respondents about the true prevalence of climate norms in the US corrects these misperceptions and increases climate donations.

The widely-observed underestimation of climate norms in the US can form a dangerous obstacle to climate action, whereby moving forward it will be crucial to correct these misperceptions. Our results thus suggest that social norms should play a pivotal role in the policy response to climate change. Policies that foster social norms should complement formal regulations such as carbon taxation. Finally, we hope that the study also showcases an important role that economic and social science research will have to play in the warming years ahead. Its key responsibilities will include monitoring the perception of climate norms, detecting misperceptions early, and exploring how they can effectively be corrected.

References

Alan, Sule, and Seda Ertac. 2018. “Fostering Patience in the Classroom: Results from Randomized Educational Intervention.” *Journal of Political Economy* 126 (5): 1865–911. [229]

Allcott, Hunt. 2011. “Social norms and energy conservation.” *Journal of Public Economics* 95 (9-10): 1082–95. [223]

Allport, Floyd Henry. 1924. *Social Psychology*. Boston: Houghton Mifflin. [223, 242]

Auffhammer, Maximilian. 2018. “Quantifying economic damages from climate change.” *Journal of Economic Perspectives* 32 (4): 33–52. [220]

Ballew, Matthew T., Seth A. Rosenthal, Matthew H. Goldberg, Abel Gustafson, John E. Kotcher, Edward W. Maibach, and Anthony Leiserowitz. 2020. “Beliefs about others’ global warming beliefs: The role of party affiliation and opinion deviance.” *Journal of Environmental Psychology* 70: 101466. [224]

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro. 1997. “Preferences Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study.” *Quarterly Journal of Economics* 112 (2): 537–79. [224]

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York: Cambridge University Press. [226, 227]

Bicchieri, Cristina. 2017. *Norms in the Wild: How to Diagnose, Measure and Change Social Norms*. New York: Oxford University Press. [242]

Bowles, Samuel. 2004. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton: Princeton University Press. [223]

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. 2015. “Global non-linear effect of temperature on economic production.” *Nature* 527 (7577): 235–39. [220]

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott. 2020. “Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia.” *American Economic Review* 110 (10): 2297–3029. [223, 242]

Bursztyn, Leonardo, and Robert Jensen. 2017. “Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure.” *Annual Review of Economics* 9: 131–53. [226, 227]

Bursztyn, Leonardo, and David Y. Yang. 2021. “Misperceptions about Others.” *Working Paper*, [223]

Carleton, Tamma A., and Solomon M. Hsiang. 2016. “Social and economic impacts of climate.” *Science* 353 (6304): [220]

Cialdini, Robert B., Raymond R. Reno, and Carl a. Kallgren. 1990. “A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places.” *Journal of Personality and Social Psychology* 58 (6): 1015–26. [227]

Costa, Dora L., and Matthew E. Kahn. 2013. “Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment.” *Journal of the European Economic Association* 11 (3): 680–702. [223]

Crimston, Daniel, Paul G. Bain, Matthew J. Hornsey, and Brock Bastian. 2016. “Moral expansiveness: Examining variability in the extension of the moral world.” *Journal of Personality and Social Psychology* 111 (4): 636–53. [228]

Diffenbaugh, Noah S., and Marshall Burke. 2019. “Global warming has increased global economic inequality.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (20): 9808–13. [220]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2009. "Homo Reciprocans: Survey Evidence on Behavioural Outcomes." *Economic Journal* 119 (536): 592–612. [224]

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9 (3): 522–50. [224]

Drews, Stefan, and Jeroen C.J.M. van den Bergh. 2016. "What explains public support for climate policies? A review of empirical and experimental studies." *Climate Policy* 16 (7): 855–76. [224]

Dunlap, Riley E., and Aaron M. McCright. 2011. "Organized Climate Change Denial." In *The Oxford Handbook of Climate Change and Society*. Edited by John S. Dryzek, Richard B. Norgaard, and David Schlosberg. Oxford: Oxford University Press. [231]

Dunlap, Riley E., Aaron M. McCright, and Jerrod H. Yarosh. 2016. "The Political Divide on Climate Change: Partisan Polarization Widens in the U.S." *Environment: Science and Policy for Sustainable Development* 58 (5): 4–23. [221, 242]

Durlauf, Steven N., and H. Peyton Young. 2001. *Social Dynamics*. Cambridge: MIT Press. [223]

Enke, Benjamin. 2020. "Moral values and voting." *Journal of Political Economy* 128 (10): 3679–729. [221, 224, 228, 274]

Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann. 2019. "Moral Universalism and the Structure of Ideology." *Working Paper*, [224]

Falk, Armin. 2021. "Facing yourself - A note on self-image." *Journal of Economic Behavior & Organization* 186: 724–34. [227]

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics* 133 (4): 1645–92. [220, 224, 228, 272]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2018. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Working Paper*, [220, 228]

Farrow, Katherine, Gilles Grolleau, and Lisette Ibanez. 2017. "Social Norms and Pro-environmental Behavior: A Review of the Evidence." *Ecological Economics* 140: 1–13. [223]

Ferraro, Paul J., and Michael K. Price. 2013. "Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment." *Review of Economics and Statistics* 95 (1): 64–73. [223]

Figlio, David, Paola Giuliano, Umut Özek, and Paola Sapienza. 2019. "Long-Term Orientation and Educational Performance." *American Economic Journal: Economic Policy* 11 (4): 272–309. [224]

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters* 71 (3): 397–404. [227]

Fischbacher, Urs, Simeon Schudy, and Sabrina Teyssier. 2021. "Heterogeneous preferences and investments in energy saving measures." *Resource and Energy Economics* 63: 101202. [224]

Geiger, Nathaniel, and Janet Swim. 2016. "Climate of Silence: Pluralistic Ignorance as a Barrier to Climate Change Discussion." *Journal of Environmental Psychology* 47: 79–90. [223]

Gifford, Robert. 2011. "The Dragons of Inaction: Psychological Barriers that Limit Climate Change Mitigation and Adaptation." *American Psychologist* 66 (4): 290–302. [224]

Graham, Jesse, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean Wojcik, and Peter Ditto. 2013. "Moral Foundations Theory: The pragmatic validity of moral pluralism." *Advances in Experimental Social Psychology* 47: 55–130. [221, 224, 228]

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. No date. "Designing Information Provision Experiments." *Journal of Economic Literature*, (forthcoming): ().

Haidt, Jonathan. 2012. *The righteous mind: Why good people are divided by politics and religion*. New York: Vintage. [221, 224, 228]

Haidt, Jonathan, and Craig Joseph. 2004. "Intuitive ethics: How innately prepared intuitions generate culturally variable virtues." *Daedalus* 133 (4): 55–66. [221, 224, 228]

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27 (2): 163–92. [239, 261]

Hamilton, Lawrence C. 2011. "Education, politics and opinions about climate change evidence for interaction effects." *Climatic Change* 104 (2): 231–42. [234]

Hornsey, Matthew J., Emily A. Harris, and Kelly S. Fielding. 2018. "Relationships among conspiratorial beliefs, conservatism and climate scepticism across nations." *Nature Climate Change* 8 (7): 614–20. [221, 242]

Howe, Peter D., Matto Mildenberger, Jennifer R. Marlon, and Anthony Leiserowitz. 2015. "Geographic variation in opinions on climate change at a state and local scales in the USA." *Nature Climate Change* 5 (6): 596–603. [231, 270]

IPCC. 2014. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by Core Writing Team, R. K. Pauchari, and L. A. Meyer. [220]

Jachimowicz, Jon M., Oliver P. Hauser, Julia D. O'Brien, Erin Sherman, and Adam D. Galinsky. 2018. "The critical role of second-order normative beliefs in predicting energy conservation." *Nature Human Behaviour* 2 (10): 757–64. [223]

Kahan, Dan M., Ellen Peters, Maggie Wittlin, Paul Slovic, Lisa Larrimore Ouellette, Donald Braman, and Gregory Mandel. 2012. "The polarizing impact of science literacy and numeracy on perceived climate change risks." *Nature Climate Change* 2 (10): 732–35. [234]

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. 2019. "The Formation of Prosociality: Causal Evidence on the Role of Social Environment." *Journal of Political Economy* 128 (2): 434–67. [229]

Krupka, Erin L., and Roberto A. Weber. 2013. "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association* 11 (3): 495–524. [226]

Kuran, Timur. 1991. "The East European Revolution of 1989: Is it Surprising that We Were Surprised?" *American Economic Review* 81 (2): 121–25. [223]

Lades, Leonhard K., Kate Laffan, and Till O. Weber. 2021. "Do economic preferences predict pro-environmental behaviour?" *Ecological Economics* 183: 106977. [224]

Leiserowitz, Anthony A., Edward W. Maibach, Connie Roser-Renouf, Nicholas Smith, and Erica Dawson. 2013. "Climategate, Public Opinion, and the Loss of Trust." *American Behavioral Scientist* 57 (6): 818–37. [231]

Leviston, Z., I. Walker, and S. Morwinski. 2013. "Your opinion on climate change might not be as common as you think." *Nature Climate Change* 3 (4): 334–37. [223]

Lorenzoni, Irene, Anthony Leiserowitz, Miguel de Franca Doria, Wouter Poortinga, and Nick F. Pidgeon. 2006. "Cross-national comparisons of image associations with "global warming" and "climate change" among laypeople in the United States of America and Great Britain." *Journal of Risk Research* 9 (03): 265–81. [226]

Mildenberger, Matto, and Dustin Tingley. 2019. "Beliefs about Climate Beliefs: The Importance of Second-Order Opinions for Climate Politics." *British Journal of Political Science* 49 (4): 1279–307. [224]

Miller, Dale T., and Cathy McFarland. 1987. "Pluralistic Ignorance: When Similarity is Interpreted as Dissimilarity." *Journal of Personality and Social Psychology* 53 (2): 298–305. [223, 242]

Newell, Richard G., and Juha Siikamki. 2015. "Individual time preferences and energy efficiency." *American Economic Review* 105 (5): 196–200. [224]

Newport, Frank, and Andrew Dugan. 2015. "College-Educated Republicans Most Skeptical of Global Warming." *Gallup Report* March 26: [234]

Nyborg, Karine. 2018. "Social Norms and the Environment." *Annual Review of Resource Economics* 10: 405–23. [223]

Nyborg, Karine, John M. Anderies, Astrid Dannenberg, Therese Lindahl, Caroline Schill, Maja Schlüter, W. Neil Adger, Kenneth J. Arrow, Scott Barrett, Stephen Carpenter, F. Stuart Chapin, Anne Sophie Crépin, Gretchen Daily, Paul Ehrlich, Carl Folke, Wander Jager, Nils Kautsky, Simon A. Levin, Ole Jacob Madsen, Stephen Polasky, Marten Scheffer, Brian Walker, Elke U. Weber, James Wilen, Anastasios Xepapadeas, and Aart De Zeeuw. 2016. "Social norms as solutions." *Science* 354 (6308): 42–43. [223, 226]

Pearson, Adam R., Jonathon P. Schuldt, Rainer Romero-Canyas, Matthew T. Ballew, and Dylan Larson-Konar. 2018. "Diverse segments of the US public underestimate the environmental concerns of minority and low-income Americans." *Proceedings of the National Academy of Sciences* 115 (49): 12429–34. [223]

Rao, Gautam. 2019. "Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools." *American Economic Review* 109 (3): 774–809. [229]

Schleich, Joachim, Xavier Gassmann, Thomas Meissner, and Corinne Faure. 2019. "A large-scale test of the effects of time discounting, risk aversion, loss aversion, and present bias on household adoption of energy-efficient technologies." *Energy Economics* 80: 377–93. [224]

Singer, Peter. 2011. *The Expanding Circle: Ethics, Evolution, and Moral Progress*. Princeton University Press. [228]

Swim, Janet, Susan Clayton, Thomas Doherty, Robert Gifford, George Howard, Joseph Reser, Paul Stern, and Elke Weber. 2009. "Psychology & global climate change: Addressing a multifaceted phenomenon and set of challenges." *Report of the American Psychological Association Task Force on the Interface Between Psychology and Global Climate Change*, [224]

UNEP. 2019. *Emissions Gap Report 2019*. United Nations Environment Programme. [220]

Valkengoed, Anne M. van, and Linda Steg. 2019. "Meta-analyses of factors motivating climate change adaptation behaviour." *Nature Climate Change* 9 (2): 158–63. [223]

Van Der Linden, Sander, Edward Maibach, John Cook, Anthony Leiserowitz, Michael Ranney, Stephan Lewandowsky, Joseph Árvai, and Elke U. Weber. 2017. "Culture versus cognition is a false dilemma." *Nature Climate Change* 7 (7): 457. [234]

Welsch, Heinz. 2020. "Moral Foundations and Voluntary Public Good Provision: The Case of Climate Change." *Ecological Economics* 175: 106696. [224]

Xu, Yiqing, Jens Hainmueller, Jonathan Mummolo, and Licheng Liu. 2017. "INTERFLEX: Stata module to estimate multiplicative interaction models with diagnostics and visualization." *Statistical Software Components, Boston College Department of Economics*, (3). [239, 261]

Young, H. Peyton. 2008. "The New Palgrave Dictionary of Economics." In *The New Palgrave Dictionary of Economics*. Edited by Steven N. Durlauf and Lawrence E. Blume. London: Palgrave MacMillan. [223]

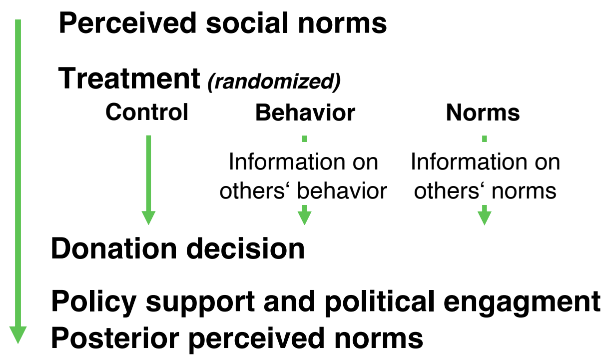
Young, H. Peyton. 2015. "The Evolution of Social Norms." *Annual Review of Economics* 7 (1): 359–87. [223]

Appendix 3.A Supplementary analyses

Table 3.A.1. Comparison of the sample to the US population

Variable	Wave 1	Wave 2	ACS (2019)
Female	51%	51%	51%
Age: 18-34	30%	30%	30%
Age: 35-54	32%	32%	32%
Age: 55+	38%	38%	38%
Education: Bachelor's degree or above	32%	31%	31%
Region: Northeast	17%	17%	17%
Region: Midwest	21%	21%	21%
Region: South	38%	38%	38%
Region: West	24%	24%	24%

Notes: Columns 1 and 2 display the summary statistics for the survey samples of waves 1 and 2, respectively. Column 3 displays summary statistics based on the American Community Survey 2019.



Notes: This figure provides an overview of the structure of the experiment. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Figure 3.A.1. Structure of experiment

Table 3.A.2. Education and individual willingness to fight global warming

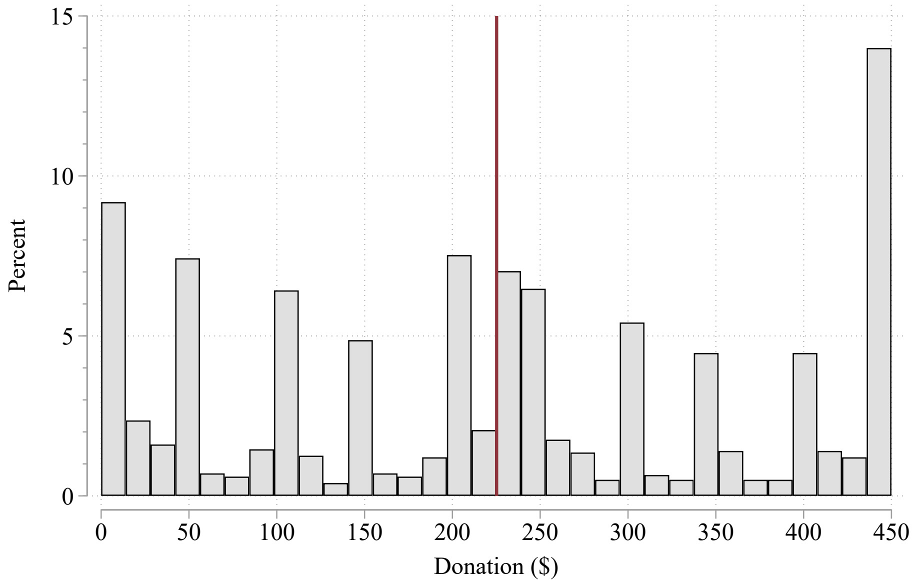
	Outcome: Donation ($)	
	(1)	(2)
Democrat x college degree	-6.838	-6.480
	(8.096)	(8.062)
Republican x college degree	-28.214***	-27.201***
	(10.320)	(10.429)
N	1,975	1,975
Control group mean	225.21	225.21
Demographic controls	Yes	Yes
Preferences and moral universalism	Yes	Yes
Normative belief	Behavior belief	Norms belief

Notes: This table shows OLS regression estimates where the dependent variable are donations (in $) using respondents from wave 1. All regressions specifications are identical to those in Table 3.3.1, including demographic controls, economic preferences, moral universalism as well as normative beliefs as covariates. However, we replaced the “College degree” indicator with a “Democrat x college degree” and a “Republican x college degree” indicator. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Table 3.A.3. Test of balance

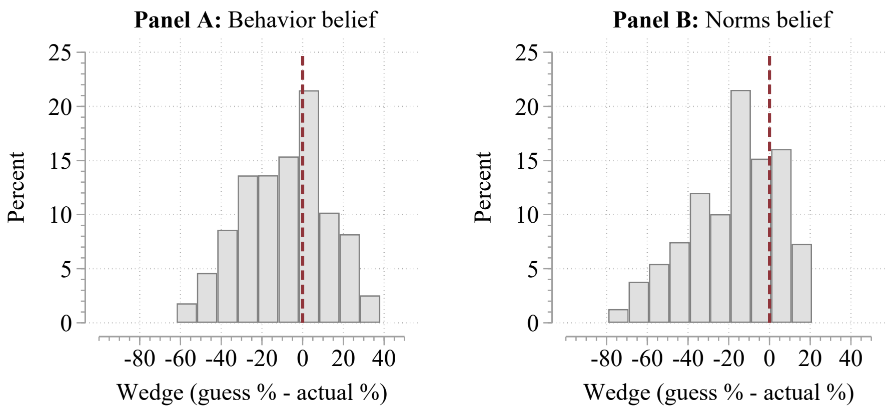
	Means (std. dev.)			Differences (*p*-values)		
	Control group (C)	Behavior treatment (T_1)	Norms treatment (T_2)	T_1 - C	T_2 - C	T_2 - T_1
Behavior belief	52.096	51.627	51.644	-0.470	-0.452	-0.017
	(21.339)	(21.213)	(21.391)	(0.486)	(0.503)	(0.980)
Norms belief	62.172	61.667	61.328	-0.505	-0.845	0.339
	(21.357)	(21.535)	(21.948)	(0.458)	(0.217)	(0.621)
Altruism	-0.008	-0.024	0.032	-0.016	0.040	-0.057*
	(0.982)	(0.984)	(1.032)	(0.600)	(0.206)	(0.076)
Patience	-0.020	0.005	0.015	0.025	0.035	-0.010
	(0.993)	(0.989)	(1.019)	(0.424)	(0.265)	(0.744)
Risk	-0.001	-0.005	0.006	-0.005	0.007	-0.011
	(0.989)	(1.011)	(1.000)	(0.887)	(0.827)	(0.719)
Pos. reciprocity	-0.018	0.021	-0.002	0.039	0.016	0.023
	(1.024)	(0.983)	(0.993)	(0.223)	(0.619)	(0.463)
Neg. reciprocity	-0.011	0.012	-0.001	0.023	0.011	0.013
	(0.999)	(0.978)	(1.023)	(0.455)	(0.733)	(0.692)
Trust	-0.028	0.017	0.010	0.045	0.038	0.007
	(1.001)	(1.000)	(0.999)	(0.156)	(0.229)	(0.825)
Rel. universalism	-0.027	0.021	0.006	0.047	0.032	0.015
	(0.987)	(1.020)	(0.993)	(0.138)	(0.303)	(0.639)
Age	48.114	47.350	47.847	-0.763	-0.266	-0.497
	(17.727)	(17.055)	(17.438)	(0.166)	(0.632)	(0.361)
Female	0.494	0.522	0.514	0.029*	0.020	0.008
	(0.500)	(0.500)	(0.500)	(0.071)	(0.202)	(0.593)
Log income	10.782	10.795	10.815	0.013	0.033	-0.020
	(0.882)	(0.879)	(0.858)	(0.645)	(0.236)	(0.471)
College degree	0.473	0.479	0.457	0.007	-0.015	0.022
	(0.499)	(0.500)	(0.498)	(0.676)	(0.335)	(0.166)
Employed	0.499	0.488	0.506	-0.012	0.007	-0.018
	(0.500)	(0.500)	(0.500)	(0.467)	(0.672)	(0.248)
Democrat	0.528	0.535	0.539	0.007	0.011	-0.003
	(0.499)	(0.499)	(0.499)	(0.640)	(0.497)	(0.833)
Northeast	0.170	0.165	0.174	-0.005	0.004	-0.009
	(0.376)	(0.372)	(0.380)	(0.692)	(0.717)	(0.447)
Midwest	0.204	0.211	0.216	0.007	0.012	-0.005
	(0.403)	(0.408)	(0.411)	(0.602)	(0.362)	(0.697)
South	0.390	0.385	0.365	-0.005	-0.025	0.020
	(0.488)	(0.487)	(0.482)	(0.743)	(0.105)	(0.196)
Parent	0.562	0.557	0.550	-0.005	-0.012	0.007
	(0.496)	(0.497)	(0.498)	(0.762)	(0.441)	(0.640)
p-value of joint *F*-test				0.426	0.684	0.425
Observations	1,987	1,995	2,018	3,982	4,005	4,013

Notes: Columns 1–3 show the means and standard deviations of respondent covariates in the different treatments of wave 2. Columns 4–6 show differences in means between the groups indicated in the column header together with *p*-values in parentheses. The *p*-values of the joint *F*-test are determined by regressing the treatment indicator on the vector of demographic controls. The F-test tests the joint hypothesis that none of the covariates predicts treatment assignment. *Covariates* “Behavior belief” and “Norms belief” are the perceived social norm measures, ranging from 0 to 100. Economic preferences (altruism, patience, risk, pos. reciprocity, neg. reciprocity, trust) and moral universalism (rel. universalism) are standardized. “Female”, “Employed”, “Democrat”, “Parent”, and the three census region dummies are binary indicators. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.



Notes: This figure shows the distribution of the monetary amounts donated to the climate charity in wave 1. The average donation is indicated by the vertical red line.

Figure 3.A.2. The distribution of individual willingness to fight global warming



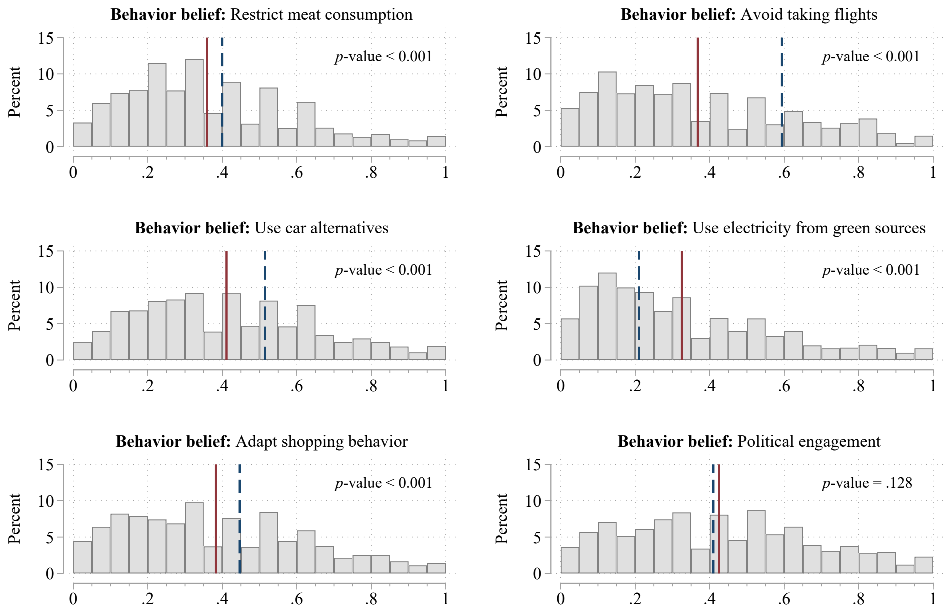
Notes: Using respondents from wave 2, this figure shows the distribution of the wedge between the respondent's perceived social norms and the actual shares in wave 1. Panel A shows people's belief about the share of Americans who say that they try to fight global warming. Panel B shows people's belief about the share of Americans who say that one should fight global warming. The red vertical line indicates the actual shares from wave 1.

Figure 3.A.3. Wedge in beliefs about social norms

Table 3.A.4. Determinants of norm misperceptions

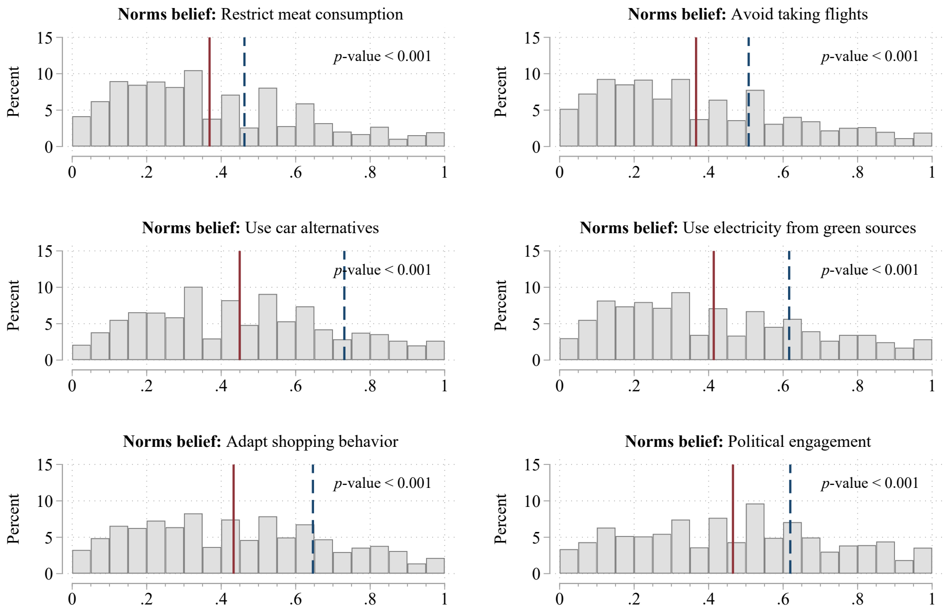
	Dependent variable: Absolute prediction error (in percentage points)			
	Behavior belief		Norms belief	
	(1) Full sample	(2) Underestimators only	(3) Full sample	(4) Underestimators only
Democrat	-1.869***	-1.997**	-3.130***	-3.343***
	(0.663)	(0.868)	(0.814)	(0.945)
Age	0.063***	0.077***	0.133***	0.138***
	(0.021)	(0.027)	(0.026)	(0.030)
Female	0.919	0.823	1.378*	1.277
	(0.665)	(0.866)	(0.805)	(0.930)
Log household income	-0.508	-0.556	-1.104**	-1.617***
	(0.423)	(0.531)	(0.540)	(0.624)
College degree or more	-0.956	-0.264	-2.299***	-2.947***
	(0.727)	(0.969)	(0.892)	(1.050)
Currently employed	1.024	0.781	0.601	1.014
	(0.727)	(0.947)	(0.903)	(1.054)
Parent	-0.046	-1.238	-0.828	-0.745
	(0.703)	(0.915)	(0.863)	(0.998)
Constant	23.107***	26.513***	30.344***	39.914***
	(4.581)	(5.684)	(5.786)	(6.683)
N	1,996	1,334	1,996	1,519
R^2	0.013	0.013	0.033	0.040

Notes: This table shows OLS regression estimates using respondents from wave 1. The dependent variable in each column is the absolute difference between the respondent's stated belief (behavior/norms) and the actual share. "Behavior belief" is the respondent's belief about the share of Americans who fight global warming. "Norms belief" is the respondent's belief about the share of Americans who think one should fight global warming. Columns 1 and 3 use the full sample, while columns 2 and 4 focus on the subset of respondents who underestimate the actual shares. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.



Notes: This figure shows the distribution of behavior beliefs in wave 1 for concrete climate-friendly behaviors. Each panel shows the distribution of people's beliefs about the share of Americans who say that they engage in the specific climate-friendly behavior indicated in the title of the panel. The solid red line indicates the average belief. The dashed blue line indicates the actual share of Americans engaging the behavior.

Figure 3.A.4. Perceived prevalence of concrete climate-friendly behaviors



Notes: This figure shows the distribution of norms beliefs in wave 1 for concrete climate-friendly behaviors. Each panel shows the distribution of people's beliefs about the share of Americans who say that one *should* engage in the specific climate-friendly behavior. The solid red line indicates the average belief. The dashed blue line indicates the actual share of Americans saying that one should engage in the behavior indicated in the title of the panel.

Figure 3.A.5. Perceived prevalence of norms for concrete climate-friendly behavior

Table 3.A.5. Relationship of abstract and specific perceived norm measures

	(1) Restrict meat consumption	(2) Avoid taking flights	(3) Use car alternatives	(4) Use green electricity	(5) Adapt shopping behavior	(6) Political engagement
Panel A: Behavior						
Behavior belief	0.477***	0.362***	0.471***	0.421***	0.480***	0.468***
	(0.021)	(0.022)	(0.021)	(0.021)	(0.020)	(0.020)
N	1,994	1,993	1,993	1,994	1,992	1,993
R^2	0.228	0.131	0.222	0.178	0.231	0.219
Panel B: Norms						
Norms belief	0.410***	0.340***	0.454***	0.416***	0.471***	0.448***
	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)
N	1,994	1,993	1,993	1,994	1,992	1,993
R^2	0.168	0.116	0.206	0.174	0.222	0.201

Notes: This table shows OLS regression estimates using respondents from wave 1. All coefficients can be interpreted as Pearson correlation coefficients. The dependent variables in Panel A are beliefs about the share of Americans who engage in the concrete climate-friendly behavior indicated in the column header. The dependent variables in Panel B are beliefs about the share of Americans who say that one should engage in the concrete climate-friendly behaviors. "Behavior belief" is the respondent's belief about the share of Americans who fight global warming. "Norms belief" is the respondent's belief about the share of Americans who think one should fight global warming. All beliefs are standardized to have a mean of zero and a standard deviation of one. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Table 3.A.6. Treatment effects on climate donations and posterior beliefs: No controls

	(1) Donation ($)	(2) Behavior belief (post.)	(3) Norms belief (post.)
Behavior treatment	12.852***	0.285***	0.244***
	(4.824)	(0.031)	(0.031)
Norms treatment	17.485***	0.374***	0.355***
	(4.857)	(0.031)	(0.031)
N	5,991	5,988	5,976
Control group mean	249.31	0	0
z-scored	No	Yes	Yes

Notes: This table shows OLS regression estimates using respondents from wave 2. "Behavior treatment" is a binary indicator taking value one for respondents who received information about the share of Americans who try to fight global warming. "Norms treatment" is a binary indicator taking value one for respondents who received information about the share of Americans who say that one should try to fight global warming. "Behavior belief" is an index of six post-treatment beliefs about the share of Americans engaging in concrete climate-friendly behaviors to fight global warming. "Norms belief" is an index of six post-treatment beliefs about the share of Americans who say that one should engage in concrete climate-friendly behaviors to fight global warming. Both indices are standardized to have a mean of zero and a standard deviation of one in the control group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Table 3.A.7. Treatment effect heterogeneity: Climate change “denier”

	Dependent variable: Donation ($)				
	Interactant:				
	(1) No trust in science	(2) No scientific consensus	(3) Not concerned	(4) Not a threat	(5) Caused by nature
Panel A: Behavior treatment					
Treatment (a)	2.733	1.335	1.004	1.895	0.122
	(5.661)	(5.392)	(5.007)	(5.085)	(5.082)
Treatment x Interactant (b)	18.268*	22.561**	33.200***	29.943***	38.333***
	(9.357)	(10.126)	(10.410)	(10.330)	(10.466)
Interactant	-91.364***	-82.718***	-140.489***	-128.326***	-127.592***
	(7.145)	(7.472)	(7.751)	(7.710)	(7.865)
Linear combination (a + b)	21.001***	23.896***	34.204***	31.837***	38.455***
	(7.444)	(8.568)	(9.121)	(8.981)	(9.144)
N	3,978	3,978	3,978	3,978	3,978
Controls	Yes	Yes	Yes	Yes	Yes
Panel B: Norms treatment					
Treatment (a)	13.000**	8.245	10.241**	9.397*	11.639**
	(5.667)	(5.460)	(4.987)	(5.069)	(5.053)
Treatment x Interactant (b)	7.751	21.274**	14.928	14.560	14.569
	(9.353)	(10.044)	(10.406)	(10.398)	(10.386)
Interactant	-89.976***	-80.385***	-139.925***	-127.516***	-128.427***
	(7.140)	(7.465)	(7.742)	(7.726)	(7.852)
Linear combination (a + b)	20.751***	29.519***	25.169***	23.957***	26.208***
	(7.442)	(8.431)	(9.136)	(9.084)	(9.082)
N	4,000	4,000	4,000	4,000	4,000
Controls	Yes	Yes	Yes	Yes	Yes

Notes: This table shows OLS regression from wave 2. The dependent variable is the donation to the climate charity ($). It is regressed on a treatment dummy for the behavior treatment (Panel A) and the norm treatment (Panel B), respectively, an interactant that varies across columns, and its interaction with the treatment dummy. Interactants are indicated by the column header. Each interactant is a binary variable taking value one. “No trust in science” means that the respondent trusts climate scientists “a moderate amount”, “a little” or not at all (on a five-point Likert scale). “No scientific consensus” means that the respondent thinks that most scientists think that global warming is not happening or that there is no consensus among scientists. “Not concerned” means that the respondent is “not very worried’ or “not at all worried” about global warming (on a four-point Likert scale). “Not a threat” means that the respondent thinks that global warming will do “only a little” or no harm at all (on a four-point Likert scale). “Caused by nature” means that the respondent thinks that global warming is caused by natural activities. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Table 3.A.8. Treatment effect heterogeneity: Climate change "denier" – Robustness to controlling for the interaction between treatment and prior beliefs

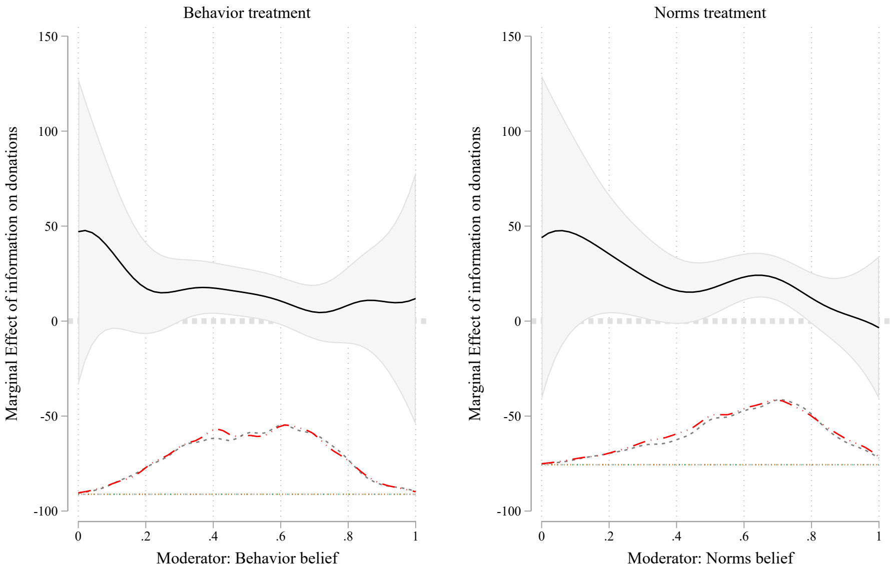
	Dependent variable: Donation ($)				
	Interactant:				
	(1) No trust in science	(2) No scientific consensus	(3) Not concerned	(4) Not a threat	(5) Caused by nature
Panel A: Behavior treatment					
Treatment (a)	9.683	9.228	12.670	13.362	13.353
	(13.391)	(13.147)	(12.671)	(12.741)	(12.617)
Treatment x Interactant (b)	17.090*	21.511**	32.559***	29.391***	38.440***
	(9.420)	(10.142)	(10.464)	(10.379)	(10.424)
Interactant	-89.111***	-80.718***	-138.804***	-126.678***	-127.584***
	(7.202)	(7.502)	(7.782)	(7.742)	(7.823)
Linear combination (a + b)	26.772**	30.739**	45.229***	42.753***	51.793***
	(13.242)	(13.961)	(13.740)	(13.726)	(13.967)
N	3,978	3,978	3,978	3,978	3,978
Controls	Yes	Yes	Yes	Yes	Yes
Treatment x Prior	Yes	Yes	Yes	Yes	Yes
Panel B: Norms treatment					
Treatment (a)	27.580*	18.851	22.250	25.774*	26.725*
	(15.650)	(15.657)	(15.042)	(15.231)	(14.881)
Treatment x Interactant (b)	5.596	18.748*	13.119	13.138	13.001
	(9.481)	(10.173)	(10.688)	(10.649)	(10.508)
Interactant	-84.081***	-74.126***	-134.167***	-121.945***	-123.874***
	(7.214)	(7.569)	(7.935)	(7.871)	(7.927)
Linear combination (a + b)	33.176**	37.599**	35.370**	38.912***	39.726***
	(14.660)	(15.099)	(14.536)	(14.745)	(14.640)
N	4,000	4,000	4,000	4,000	4,000
Controls	Yes	Yes	Yes	Yes	Yes
Treatment x Prior	Yes	Yes	Yes	Yes	Yes

Notes: This table shows OLS regression from wave 2. The dependent variable is the donation to the climate charity ($). It is regressed on a treatment dummy for the behavior treatment (Panel A) and the norm treatment (Panel B), respectively, an interactant that varies across columns, and its interaction with the treatment dummy. Interactants are indicated by the column header. Each interactant is a binary variable taking value one. "No trust in science" means that the respondent trusts climate scientists "a moderate amount", "a little" or not at all (on a five-point Likert scale). "No scientific consensus" means that the respondent thinks that most scientists think that global warming is not happening or that there is no consensus among scientists. "Not concerned" means that the respondent is "not very worried' or "not at all worried" about global warming (on a four-point Likert scale). "Not a threat" means that the respondent thinks that global warming will do "only a little" or no harm at all (on a four-point Likert scale). "Caused by nature" means that the respondent thinks that global warming is caused by natural activities. All regressions include the corresponding prior belief and the interaction between the treatment indicator and the prior belief. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Table 3.A.9. Preferences and universal values explain the partisan gap

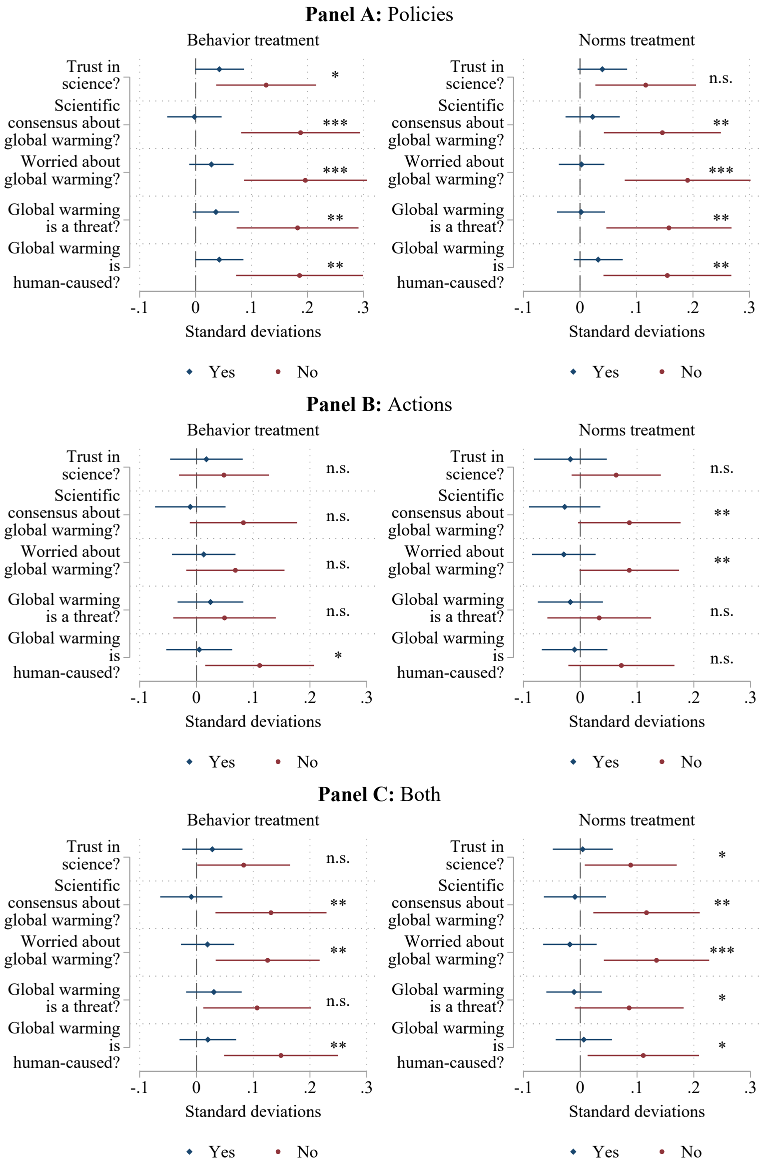
	Donation ($)		Policy support	
	(1)	(2)	(3)	(4)
Democrat	74.323***	46.084***	0.923***	0.709***
	(6.523)	(6.279)	(0.041)	(0.040)
N	1,993	1,976	1,993	1,979
R^2	0.086	0.275	0.221	0.337
Demographic controls	Yes	Yes	Yes	Yes
Preferences and moral universalism		Yes		Yes

Notes: This table shows OLS regression estimates using respondents from wave 1. "Democrat" is a binary indicator taking value one if respondents identify with the Democrat party. We include our standard set of demographic controls: gender (indicator), age (continuous), log income, college degree (indicator), employment (indicator), and census region (three indicators). The dependent variable in columns 1–2 are donations, whereas the dependent variable in columns 3–4 is our standardized index of support for policies to fight global warming. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.



Notes: This figure shows the results from a non-linear interaction analysis using the *interflex* package (Xu et al., 2017; Hainmueller, Mummolo, and Xu, 2019) and restricting the sample to respondents from wave 2. The left panel excludes respondents in the norms treatment, while the right panel excludes respondents in the behavior treatment. The dashed lines at the bottom of each panel plot the distribution of the pre-treatment belief. 95% confidence intervals using robust standard errors are shown. Both panels show results without including additional controls.

Figure 3.A.6. Treatment effect heterogeneity by perceived social norms: Non-parametric estimates



Notes: This figure shows treatment effects in different subsamples using respondents from wave 2. Panel A shows treatment effects on the policy support index, Panel B shows treatment effects of the action index, and Panel C shows treatment effects on the joint index. 95% confidence intervals are shown. Each panel shows estimates for the subsample of climate change deniers – e.g., those who have no trust in science or do not believe in human-caused global warming – and the subsample of respondents who are not skeptical of climate change. "Trust in science" means that the respondent trust climate scientists "a lot" or "a great deal" (on a five-point Likert scale). "Scientific consensus about global warming" means that the respondent thinks that most scientists think that global warming is happening. "Worried about global warming" means that the respondent is "somewhat worried' or "very worried" about global warming (on a four-point Likert scale). "Global warming is a threat" means that the respondent thinks that global warming will do "a moderate amount" or "a great deal" of harm (on a four-point Likert scale). "Global warming is human-caused" means that the respondent thinks that global warming is caused by human activities.

Figure 3.A.7. Heterogeneity by "climate change denier": Political outcomes

Appendix 3.B Questionnaire

This appendix presents the main survey blocks, following the order of exposition in the paper. The full questionnaire containing all questions administered as part of this study can be downloaded from https://osf.io/chvy6/.

3.B.1 Attention screener

The next question is about the following problem. In questionnaires like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This compromises the results of research studies. **To show that you are reading the survey carefully, please choose both "Very strongly interested" and "Not at all interested" as your answer to the next question.**
Given the above, how interested are you in politics?

1.Very strongly interested

2.Very interested

3.A little bit interested

4.Not very interested

5.Not at all interested

Only participants who select both (a) and (e) pass this attention screener.

3.B.2 Measuring individual willingness to fight climate change

A decision about money

Please pay special attention to the next question in which you will make a decision about money. We will randomly select 25 respondents. If you are among them, your decision will be a real decision. The decision will be implemented and you can receive up to $450.

Your decision

Here is the decision: You can divide $450 between yourself and a charitable organization that fights global warming. The amount that you keep for yourself will be added to your account. The amount that you donate will go to the award-winning charity *atmosfair*. *atmosfair* actively contributes to CO_2 mitigation by promoting, developing and financing renewable energies worldwide. In this way, a donation saves CO_2 that would otherwise be created by fossil fuels. *atmosfair* spends around $12 million per year to fight global warming and uses less than 5% of donated funds to cover administrative costs. You can find more information on *atmosfair* here.

It costs about $450 to offset the yearly CO_2 emissions of a typical US citizen. This number is calculated as follows: It costs about $28 to prevent 1 ton of CO_2 emissions. The World Bank estimates that a typical US citizen causes about 16 tons of CO_2 emissions per year.

How much of the $450 would you like to donate to *atmosfair*?

3.B.3 Introducing bonus scheme

Bonus payment possible

There are several questions in this survey, in which we will ask you to guess how other respondents answered a question. These questions are flagged with the sign:

You can earn a bonus of $1. This works as follows: We will randomly select one of the flagged questions. Your response to this question is considered as correct if it differs at most by three from the correct number you are asked to guess. If your response to this question is correct, $1 will be added to your account.

3.B.4 Measuring perceived social norms

Do you try to fight global warming?
[Yes/No]
Do you think that people in the United States should try to fight global warming?
[Yes/No]

[PAGE BREAK]

The questions on this page are bonus questions. This means that you can earn additional money if you answer them correctly.

As part of this research project, we recently surveyed many people in the United States and asked them the same questions. Respondents come from all parts of the population and their responses represent the views and attitudes of people in the United States.
What do you think? Out of 100 people we asked, how many stated that...

1.... they try to fight global warming?

2.... they think that people in the United States should try to fight global warming?

3.B.5 Treatments: Shifting perceived social norms

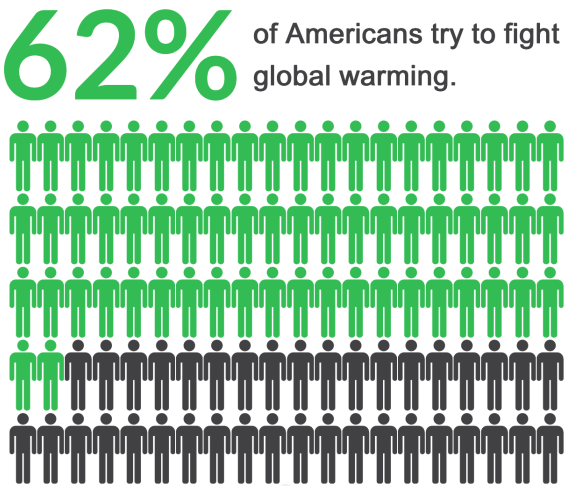
3.B.5.1 Behavior treatment

What do other people in the United States do?

We recently surveyed 2,000 people in the United States and asked them whether they try to fight global warming. Respondents come from all parts of the population and their responses represent the views and attitudes of people in the United States. On the next page, you will learn how they responded. Please read the information carefully.

[PAGE BREAK]

We asked 2,000 Americans: Do you try to fight global warming? Yes or no?
Here are the results:



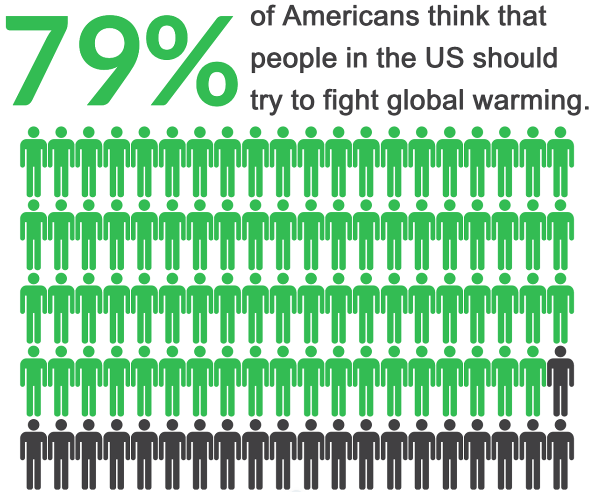
3.B.5.2 Norms treatment

What do other people in the United States think?

We recently surveyed 2,000 people in the United States and asked them whether they think people in the US should try to fight global warming. Respondents come from all parts of the population and their responses represent the views and attitudes of people in the United States. On the next page, you will learn how they responded. Please read the information carefully.

[PAGE BREAK]

We asked 2,000 Americans: Do you think that people in the United States should try to fight global warming? Yes or no?
Here are the results:



3.B.6 Measuring posterior beliefs

The questions on this page are bonus questions. This means that you can earn additional money if you answer them correctly.

As part of this research project, we recently surveyed many people in the United States and asked them the same questions. Respondents come from all parts of the population and their responses represent the views and attitudes of people in the United States.
We asked respondents to state whether they have taken different actions to fight global warming over the last year.
What do you think? Out of 100 people we asked, how many stated that...

1.... restrict their meat consumption?

2.... avoid taking flights?

3.... regularly use environmentally-friendly alternatives to their private car such as walking, cycling, taking public transport or car-sharing?

4.... receive electricity only from green/renewable sources (e.g., solar energy or wind power)?

5.... adapt their shopping behavior to the carbon footprint of products?

6.... politically support the fight against global warming, e.g. participate in a demonstration, sign a letter, or support a political organization?

[PAGE BREAK]

Do you think that people in the United states **should**...

1.... restrict their meat consumption?

2.... avoid taking flights?

3.... regularly use environmentally-friendly alternatives to their private car such as walking, cycling, taking public transport or car-sharing?

4.... receive electricity only from green/renewable sources (e.g., solar energy or wind power)?

5.... adapt their shopping behavior to the carbon footprint of products?

6.... politically support the fight against global warming, e.g. participate in a demonstration, sign a letter, or support a political organization?

[PAGE BREAK]

The questions on this page are bonus questions. This means that you can earn additional money if you answer them correctly.

BONUS

What do you think? Out of 100 people we asked the same questions, how many stated that they think that people in the United States should...

1.... restrict their meat consumption?

2.... avoid taking flights?

3.... regularly use environmentally-friendly alternatives to their private car such as walking, cycling, taking public transport or car-sharing?

4.... receive electricity only from green/renewable sources (e.g., solar energy or wind power)?

5.... adapt their shopping behavior to the carbon footprint of products?

6.... politically support the fight against global warming, e.g. participate in a demonstration, sign a letter, or support a political organization?

3.B.7 Measuring climate change skepticism

In general, how much do you trust scientists who do research on global warming?

1.A great deal

2.A lot

3.A moderate amount

4.A little

5.Not at all

Which comes closest to your own view?

1.Most scientists think global warming is happening.

2.There is a lot of disagreement among scientists about whether or not global warming is happening.

3.Most scientists think global warming is not happening.

How worried are you about global warming?

1.Very worried

2.Somewhat worried

3.Not very worried

4.Not at all worried

How much do you think global warming will harm people in the United States?

1.Not at all

2.Only a little

3.A moderate amount

4.A great deal

Do you think that global warming is mainly...?

1.a result of human activities

2.a result of natural causes

3.B.8 Measuring policy support and political engagement

Taken from the detailed politics module developed as part of the Climate Change in the American Mind Project (Howe et al., 2015).

Policy support

How much do you support or oppose the following policies?
Strongly support / Somewhat support / Somewhat oppose / Strongly oppose

1.Fund more research into renewable energy sources, such as solar and wind power.

2.Regulate carbon dioxide (the primary greenhouse gas) as a pollutant.

3.Set strict carbon dioxide emission limits on existing coal-fired power plants to reduce global warming and improve public health. Power plants would have to reduce their emissions and/or invest in renewable energy and energy efficiency. The cost of electricity to consumers and companies would likely increase.

4.Require fossil fuel companies to pay a carbon tax and use the money to reduce other taxes (such as income tax) by an equal amount.

5.Require electric utilities to produce at least 20% of their electricity from wind, solar, or other renewable energy sources, even if it costs the average household an extra $100 a year.

6.Provide tax rebates for people who purchase energy-efficient vehicles or solar panels.

How much do you agree or disagree with the following statements?
Strongly agree / Somewhat agree / Somewhat disagree / Strongly disagree

1.Schools should teach our children about the causes, consequences, and potential solutions to global warming.

Political engagement

How likely would you be to do each of the following things?
Definitely would / Probably would / Probably would not / Definitely would not

1.Vote for a candidate for public office because of their position on global warming.

2.Publicly display t-shirt, bumper sticker, button, wrist band, or sign about global warming.

3.Donate money to an organization working on global warming.

4.Volunteer your time to an organization working on global warming.

5.Write letters, email, or phone government officials about global warming.

6.Meet with an elected official or their staff about global warming.

7.Support an organization engaging in non-violent civil disobedience against corporate or government activities that make global warming worse.

8.Personally engage in non-violent civil disobedience (e.g., sit-ins, blockades, or trespassing) against corporate or government activities that make global warming worse.

9.Attend a political rally, speech, or organized protest about global warming.

10.Write a letter to the editor of a newspaper or magazine or call a live radio or TV show to express an opinion about global warming.

11.Share information about global warming on social media.

Appendix 3.C Construction of variables

3.C.1 Measuring economic preferences

We administer the Global Preferences Survey (GPS) and follow the methodology described in Falk, Becker, Dohmen, Enke, et al. (2018) to obtain detailed individual-level measures of economic preferences. More information on the construction of the variables can be found below.

1. *Patience.* The measure of patience (or time preference) is derived from the combination of responses to two survey measures, one with a quantitative and one with a qualitative format. The quantitative survey measure consists of a series of five interdependent hypothetical binary choices between immediate and delayed financial rewards. In each of the five questions, participants have to decide between receiving a payment today or a larger payment in 12 months. The qualitative measure of patience is given by the respondents' self-assessment regarding their willingness to wait on an eleven-point Likert scale, asking "how willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?".
2. *Risk Taking.* Risk preferences are also elicited through a series of related quantitative questions as well as one qualitative question. Just as with patience, the quantitative measure consists of a series of five binary choices. Choices are between a fixed lottery, in which the individual could win x or zero, and varying sure payments, y. The qualitative item asks for the respondents' self-assessment of their willingness to take risks on an eleven-point scale ("In general, how willing are you to take risks?").
3. *Positive Reciprocity.* Positive reciprocity is measured using one quantitative item and one qualitative question. First, respondents are presented a choice scenario in which they are asked to imagine that they got lost in an unfamiliar area and that a stranger – when asked for directions – offered to take them to their destination. Respondents are then asked which out of six presents (worth between 10 and 60 dollars) they would give to the stranger as a "thank you". Second, respondents are asked to provide a self-assessment about how willing they are to return a favor on an eleven-point Likert scale.
4. *Negative Reciprocity.* Negative reciprocity is elicited through three self-assessments. First, respondents are asked how willing they are to take revenge if they are treated very unjustly, even if doing so comes at a cost (Likert scale, 0-10). The second and third items probe respondents about their willingness to punish someone for unfair behavior, either towards themselves or a third person.
5. *Altruism.* Altruism is measured through a combination of one qualitative and one quantitative item, both of which are related to donations. The qualitative question asks respondents how willing they would be to give to good causes

without expecting anything in return on an eleven-point scale. The quantitative scenario depicts a situation in which the respondent unexpectedly receives 1,600 dollars and is asked to state how much of this amount they would donate.

6. *Trust*. The trust measure is based on one item, which asks respondents whether they assume that other people only have the best intentions (Likert scale, 0-10).

For each economic preference, the survey items are combined into a single preference measure. More specifically, each preference is computed by (i) calculating the z-scores of each survey item at the individual level and (ii) weighting these z-scores using the weights provided in Table 3.C.1. For ease of interpretation, each preference measure is standardized to have a mean of zero and a standard deviation of one.

Table 3.C.1. GPS Survey Items and Weights

Preference	Item description	Weight
Patience	Intertemporal choice sequence using staircase method	0.712
	Self-assessment: willingness to wait	0.288
Risk taking	Lottery choice sequence using stair case method	0.473
	Self-assessment: willingness to take risks in general	0.527
Positive reciprocity	Gift in exchange for help	0.515
	Self-assessment: willingness to return a favor	0.485
Negative reciprocity	Self-assessment: willingness to take revenge	0.374
	Self-assessment: willingness to punish unfair behavior toward self	0.313
	Self-assessment: willingness to punish unfair behavior toward others	0.313
Altruism	Donation decision	0.635
	Self-assessment: willingness to give to good causes	0.365
Trust	Self-assessment: people have only the best intentions	1

3.C.2 Measuring universal moral values

Moral Foundation Theory posits that people's moral concerns can be split into five foundations:

1. *Care/Harm*. This foundation measures the extent to which people care about the weak and try to keep others away from harm.
2. *Fairness/Reciprocity*. This measure captures the importance of equality, justice, rights and autonomy.
3. *In-group/Loyalty*. This foundation captures the extent to which people emphasize loyalty to the "in-group" (family, country) and how morally relevant betrayal is.
4. *Authority/Respect*. This foundation measures how important respect for authority, tradition and order is.
5. *Purity/Sanctity*. This measure captures the importance of ideas related to purity, disgust and traditional religious attitudes.

To obtain measures of the five foundations, we administer the Moral Foundations Questionnaire. In this survey, each moral foundation is measured using six different survey items. Respondents are either asked to assess the moral relevance of certain behaviors, or they are asked if they agree with certain moral value statements. All the questions are answered on a Likert scale (0–5). Table 3.C.2 provides an overview of the specific items that are included in each foundation. In order to construct the final scores, responses are summed.

To construct a measure of the relative importance of universal versus communal moral values, we follow the approach described in Enke (2020):

$$\begin{aligned} &\text{Relative importance of universal values} && (3.C.1)\\ &= \text{Universal values} - \text{Communal values} && (3.C.2)\\ &= \text{Harm/Care} + \text{Fairness/Reciprocity} - \text{In-group/Loyalty} - \text{Authority/Respect} && (3.C.3) \end{aligned}$$

To ease interpretation, the resulting measure is standardized to have a mean of zero and a standard deviation of one.

Table 3.C.2. Survey items: Moral Foundations Questionnaire

	Moral Relevance	Agreement with Statement
Harm/care	Emotional suffering	Compassion with suffering crucial virtue
	Care for weak and vulnerable	Hurt defenseless animal is the worst thing
	Cruelty	Never right to kill human being
Fairness/reciprocity	Treat people differently	Laws should treat everyone fairly
	Act unfairly	Justice most important requirement for society
	Deny rights	Morally wrong that rich children inherit a lot
In-group/loyalty	Show love for country	Proud of country's history
	Betray group	Be loyal to family even if done something wrong
	Lack of loyalty	Be team player, rather than express oneself
Authority/respect	Lack of respect for authority	Children need to learn respect for authority
	Conform to societal traditions	Men and women have different roles in society
	Cause disorder	Soldiers must obey even if disagree with order
Purity/sancity	Violate standards of purity	Not do things that are disgusting
	Do something disgusting	Call acts wrong if unnatural
	Act in a way that God would approve	Chastity is an important virtue

Note: For the items in column 1, respondents are asked to state to what extent these considerations are morally relevant (Likert scale from 0 to 5). For the items in column 2, respondents are asked to state whether they agree or disagree with the statements (Likert scale from 0 to 5).

Chapter 4

What's Worth Knowing? Economists' Opinions about Economics

Joint with Armin Falk

Abstract: We document economists' opinions about what is worth knowing and ask (i) which research objectives economic research should embrace and (ii) which topics it should study. Almost 10,000 economic researchers from all fields and ranks of the profession participated in our global survey. Detailed bibliometric data show that our sample represents the population of economic researchers who publish in English. We report three main findings. First, economists' opinions are substantially heterogeneous. Second, most researchers are dissatisfied with economics' current research topics and objectives. Third, on average, respondents think that economic research should become more policy-relevant, multidisciplinary, risky and disruptive, and pursue more diverse topics. We also find that disagreement with the status quo is more prevalent among female scholars and associated with lower job satisfaction and higher stress levels. Taken together, the results suggest that economics as a field does not appreciate and work on what economists collectively prefer.

Acknowledgements: We thank all scholars who participated in the survey. We thank Roland Bénabou, Teodora Boneva, Felix Chopra, Stefano DellaVigna, Paul Heidhues, Leander Heldring, Alex Imas, David Kaldewey, Yucheng Liang, Stephanie Majerowicz, Franz Ostrizek, Christopher Roth, Ran Spiegler, Andreas Stegmann, Lasse Stötzer, Román Zárate, Florian Zimmermann, and many conference participants for helpful comments and discussions; Sofia Badini, Iana Gerina, Anna Lane, Tobias Reinheimer, and Youpeng Zhang for excellent research assistance; the IDSC of IZA for data services; and Markus Antony and the IZA service team for administrative support. **Funding**: Funding by the Deutsche Forschungsgemeinschaft (DFG) through CRC TR 224 (Project A01) is gratefully acknowledged. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866. **Ethics approval**: The study was approved by the local ethics committee at the medical department of the University of Bonn. Approval date: April 11, 2020. Approval number: 119/20. **Pre-registration**: The main research questions, the survey design, and the sampling approach were pre-registered via the Open Science Framework (www.doi.org/10.17605/OSF.IO/Q5GR2).

4.1 Introduction

Science and research matter. They shape how we think about ourselves, how we live together, and how we design policies. What researchers work on, which topics they choose, and how they resolve trade-offs between different research objectives therefore holds central societal importance. However, as famously argued by Max Weber (1919), the question about what is "interesting" and "worth knowing" cannot be answered scientifically. Instead, researchers have to rely on intuition and subjective, value-driven assessments and beliefs. This paper provides evidence on how economists evaluate the current state of the profession in this respect. How do economists assess important research objectives such as policy relevance, causal identification, the role of multidisciplinarity, and the level of specialization? Does the distribution of actually-chosen research topics coincide with what economists believe to be desirable? Are the answers to these questions homogeneous or heterogeneous, and to what extent do economists' preferences differ from the current practice in economics? In other words, are economists "happy" with the current research objectives and topics in their profession?

To answer these questions, we conduct a large global survey among almost 10,000 professional academic economists. Our survey focuses on two main sets of questions. The first is concerned with ten trade-offs between fundamental research objectives, including policy relevance vs. causal identification, pure vs. applied theory, quantity vs. quality, and the level of specialization. We ask respondents to indicate whether they believe that the current state of research in economics is "about right", or whether they would prefer more or less of a specific research objective, respectively. The second set of questions relates to research topics in economics. Using the common JEL taxonomy, respondents indicate what their preferred distribution of topics would look like. We compare these shares with the actual distribution of topics. Finally, we investigate how potential dissatisfaction with the status quo relates to individual scholars' well-being. For this purpose, we ask respondents to rate how satisfied they are with their job in general, with the topics they work on, how stressful they perceive their job to be, and whether they think of academia as being "overly competitive".

Studying the views of the profession requires our sample to represent the full spectrum of economic researchers. To ensure this, we identified and invited all researchers who actively contribute to the international economics literature (published in English). Each author is matched with the bibliometric databases EconLit and Scopus to compile author-specific background data. We also gathered contact data of Ph.D. students at well-known graduate schools. In total, almost 10,000 scholars participated in our survey. Our sample is representative of the profession in terms of a wide range of observable characteristics, including gender, years since first publication, number of publications, centrality in the co-author network, number of Top Five publications, h-index, and main field.

Our three main findings can be summarized as follows. First, we document substantial heterogeneity in economists' preferences. This holds for both research topics as well as trade-offs concerning fundamental research objectives: Respondents assign largely varying importance to different JEL topics and support opposing views concerning research objectives.

Second, most respondents express dissatisfaction with the current state of economic research. Across the ten trade-offs under study, only 13% to 31% of respondents agree with the current practice in economics. On average, 79% of the respondents express a preference for deviating from the status quo. Likewise, economists on average prefer a distribution of research topics that markedly differs from the actual distribution of topics published in economics. Importantly, dissatisfaction does not simply reflect the views of unsuccessful or less experienced scholars. As we show in further analyses, discomfort with the field's current research objectives and topics is shared by its most distinguished and influential scholars, represented by economists with Top Five publications, editors of top journals, and referees for Top Five journals.

Third, despite the observed variation in preferences, a majority of economists actually agree on the direction of preferred change. In terms of research objectives, most economists express a preference for more policy-relevant, multidisciplinary, as well as risky and disruptive research. The shift towards increasing policy relevance is supported even if it comes at the cost of less causal identification, pure theory, basic research, and intrinsic interest. Economists also favor quality over quantity of publications and would prefer a lower level of specialization among researchers. For research topics, we document a preference for more diversity. For example, economists assign greater importance to currently less prominent topics such as *H Public Economics* or *N Economic History* and place less weight on the three most popular topics of *D Microeconomics*, *G Financial Economics*, and *L Industrial Organization*. We also show how individual characteristics relate to stated preferences. For instance, female authors place greater weight on policy relevance, while researchers mostly working in theory or methods value policy relevance less. Moreover, respondents strongly favor their own research topics.

Our results have various implications. First, the fact that economists display heterogeneous views about what constitutes "interesting" research objectives or topics reflects their pluralistic preferences. In fact, general agreement to the question "what is worth knowing?" is unlikely because one cannot *scientifically* provide such an answer. This was noted by Max Weber:

> *"Science further presupposes that what is yielded by scientific work is important in the sense that it is "worth being known." In this, obviously, are contained all our problems. For this presupposition cannot be proved by scientific means. It can only be interpreted with reference to its ultimate meaning, which we must reject or accept according to our ultimate position towards life."* (Weber, 1919/1946)

Weber's insight is empirically reflected in the observed heterogeneity of expressed preferences. We believe that it is an important insight to keep in mind when evaluating other researchers' work, whether as seminar participants, referees, or editors. We should acknowledge diversity and pluralism and other scholars' opinions and values. Our own views about "what is interesting" are valuable and irreplaceable, but also subjective.

Second, our findings about the systematic disagreement with economics' current research objectives and topics suggest that as a field we do not appreciate and work on what we collectively prefer. This speaks empirically to the recently-raised criticism about the research and publication process in economics. For example, critics have argued that economics favors "hard" methods over relevant questions, worships "mathiness", is too specialized, neglects critical topics of our times such as climate change or financial crises, and submits to a "tyranny" of top journals (e.g., Krugman, 2009; Colander, 2011; Shiller and Shiller, 2011; Romer, 2015; Oswald and Stern, 2019; Akerlof, 2020; Heckman and Moktan, 2020; Osterloh and Frey, 2020).

Third, turning to the individual scholars' well-being, we find that dissent with economics' research objectives and topics is associated with lower job satisfaction and higher stress levels. This could also have consequences for the diversity of scholars in economics (Bayer and Rouse, 2016; Buckles, 2019; Lundberg and Stearns, 2019; Lundberg, 2020). In particular, female economists are not only less satisfied with their job and report more job-related stress, but they also more strongly disagree with economics' current research objectives and topics. These results hold conditional on a large set of controls and suggest that the current under-representation of particular groups in economics could lead to an under-representation of their research preferences, rendering an academic career even less attractive to those who are disadvantaged. The findings thus suggest another reason why women are disadvantaged and remain under-represented in economics (Avilova and Goldin, 2018; Allgood, Badgett, Bayer, Bertrand, Black, et al., 2019; Card, DellaVigna, Funk, and Iriberri, 2020; Lundberg, 2020; Dupas, Modestino, Niederle, Wolfers, and The Seminar Dynamics Collective, 2021; Sarsons, Gërxhani, Reuben, and Schram, 2021).

More generally, our study adds to past research on (economic) research. Economists closely monitor the status quo of research in their own discipline, its topics and methods, the peer-review and publication process, as well as citation trajectories of articles, scholars, and entire fields (e.g., Card and DellaVigna, 2013; Hamermesh, 2018; Kleven, 2018; Angrist, Azoulay, Ellison, Hill, and Lu, 2020; Bowles and Carlin, 2020; Card and DellaVigna, 2020; Card, DellaVigna, et al., 2020; Currie, Kleven, and Zwiers, 2020; Goldin and Katz, 2020; Heckman and Moktan, 2020). Existing survey studies have documented economists' views on issues such as economic policy, reigning paradigms in the discipline, open science practices, or mental health (e.g., Andre, Pizzinelli, Roth, and Wohlfart, 2021; Bolotnyy et al., forthcoming; Colander, 2005; Frey, Humbert, and Schneider, 2010; Sapienza and Zingales, 2013; Swanson, Christensen, Littman, Birke, Miguel, et al., 2020). By contrast, our project studies economists' opinions about the current research practice in economics. We focus on the field's research objectives and topics, which have received little attention in past research. Moreover, our study is the first to give a voice to and represent the views of such a large and diverse group of economists.

The choice of research questions, topics, and objectives is arguably among the most important choices that a researcher faces. It reflects both freedom and responsibility. We hope that the results of our study stimulate and inform a debate about this important question to make progress in finding out *what is worth knowing*.[1]

The remainder of the paper is structured as follows. Section 4.2 presents the survey instrument, section 4.3 describes the sample and study population, section 4.4 describes the results, section 4.5 discusses the main findings, and section 4.6 concludes.

4.2 Survey

This study aims to document which research objectives and topics economists think should matter in economics and to compare their views with the current state of economic research. The survey is separated into two modules that are tailored to meet these objectives. Each respondent is randomly assigned to one module. The first module explores trade-offs between different research objectives, while the second focuses on research topics. Both modules contain several demographic questions, including career status, gender, nationality, and age. Both parts also include a block of questions on job satisfaction and stress. Below, we describe the main questions of each module in turn. Appendix 4.A contains their wording.[2]

4.2.1 Research objectives

The research objectives module explores whether economists think that economic research should embrace different research objectives than it does today. The module comprises ten questions that contrast and trade-off commonly-discussed research objectives, such as policy relevance versus researchers' intrinsic interest or more versus less specialization. Of course, these trade-offs are sometimes more and sometimes less severe, but in many cases economics can have more of one research goal only at the expense of the other. Respondents indicate whether, compared to the current state of economic research, they think economics should place more weight

1. A final remark seems to be in order: It would be inconsistent to study what economists consider worth being known without addressing whether this very question is actually worth being asked. Fortunately, we can once again refer to the judgment of thousands of economists. We asked a randomly selected quarter of our respondents whether they think that it is interesting to study how and on which topics economists think they should work. Almost all, 88%, think it is.

2. The full survey is available at https://osf.io/xwbdf/.

on one objective versus the other. Panel A of table 4.2.1 provides an overview of all ten questions. The questions can roughly be categorized into four blocks.[3]

Block 1 revolves around the **policy relevance and public importance of research**. Specifically, we ask how the societal relevance of a research project should be traded-off against a researcher's intrinsic interest and curiosity (question 1), against basic research (question 2), and against rigorous causal identification (question 3). The block also includes a question that asks whether economic theory should be "pure" and study general theoretical principles or "evidence-related" and focus on empirically observed, applied phenomena (question 4). The questions, thus, connect to the discussion about the role and importance of policy relevance in economic research. They also relate to George Akerlof's recent critique that economics often prioritizes "hard" research methods, including causal identification and technically advanced pure theory, over important research questions (Akerlof, 2020).

Block 2 deals with the **scope and breadth of economic research** and asks whether individual researchers should be more or less specialized (question 5) and whether their research should be more or less multidisciplinary (question 6). Here, multidisciplinarity means incorporating insights from other disciplines than economics in order to study economic questions. Both specialization and multidisciplinarity have frequently been discussed in economics (e.g., Shiller and Shiller, 2011; Fourcade, Ollion, and Algan, 2015).

Block 3 investigates the conflict between **productive tradition and risky innovation** (Kuhn, 1962; Foster, Rzhetsky, and Evans, 2015). Should economic research be more incremental and connect closely to the existing literature or more disruptive and propose new approaches (question 8)? Likewise, should economic research be less or more risky, where high risks projects have an uncertain impact, but may come with a higher expected impact (question 7)? The final question in this block investigates whether respondents prefer more papers of lower quality or fewer papers of higher quality (question 9).

Block 4 consists of a single question that relates to a longstanding debate about the **goal of theory in economics: prediction or explanation** (question 10). Is its goal to predict economic outcomes, irrespective of whether its theoretical assumptions and mechanism are empirically plausible (Friedman, 1953)? Or is its goal to understand and explain economic outcomes (Hausman, 2008)?

In each of the ten questions, respondents first read a brief description of the opposing research objectives. *Policy relevance*, for example, is described as "Research informs policy, with an impact on societal well-being." *Basic research* is described as "Research deals with fundamental and basic phenomena, laying the ground for more applied research. It has no immediate policy relevance." Then, participants indicate their view on a seven-point scale. Each scale is centered around the option

3. The order in which we present the questions here differs from their order in the survey, see appendix 4.A.

"Current state is about right". The other response options express dissatisfaction with the status quo and place increasing weight on one research objective versus the other. For instance, the question on *Basic research* versus *Policy relevance* has the response options "Much more", "Moderately more", and "Slightly more" policy relevance, "Current state is about right", as well as "Slightly more", "Moderately more", and "Much more" basic research. The question on specialization comes with the response options "Much less", "Moderately less", and "Slightly less" specialization, 'Current state is about right", as well as "Slightly more", "Moderately more", and "Much more" specialization. We test whether participants' assessments differ for the whole discipline of economics and their own field of expertise. Respondents are instructed to provide two answers: one for economics as a whole and one for their own primary JEL field.[4]

4.2.2 JEL topics

We ask the survey participants which share of papers should be written on which topic. Each respondent can allocate a total of 100 points between different research topics. The points represent all published research articles by economists in a given year so that each point corresponds to 1% of the total research output. Thus, respondents specify their preferred distribution of research topics in economics.

We use the Journal of Economic Literature's (JEL) subject descriptors to categorize research topics in economics. These so-called JEL codes have three layers and separate economics into 19 primary topics (or fields, 1st layer) with a total of 130 sub-topics (2nd layer) and 845 subject codes (3rd layer). Here, our main focus is on the 19 primary topics whose labels mostly align with commonly used field names such as *Public Economics* or *Industrial Organization*. Panel B of table 4.2.1 lists all primary JEL topics. We ignore the residual JEL category *Y Miscellaneous categories* which is typically not assigned to research articles. In the survey, respondents can explore the sub-topics and subject codes of each JEL topic to familiarize themselves with its content. The JEL classification system provides a unique opportunity to study topic choice in economics because it covers the whole discipline of economics and it is known to most economic researchers. Moreover, its stringent classification criteria are used to categorize most published research articles. This allows us to document the actual distribution of research topics in economics to which we can then compare the preferred distribution that we elicit in the survey.

4. Participants can assign themselves to one primary JEL field. The list of fields is slightly adjusted to separate *Theoretical Microeconomics* from *Empirical Microeconomics* and to distinguish the sub-fields of JEL category Z.

Table 4.2.1. Overview of research objective questions and JEL topics

Panel A: Research objective questions	
Block 1: Policy relevance and public importance of research	
1	Intrinsic interest vs. policy relevance
2	Basic research vs. policy relevance
3	Causal identification vs. importance
4	Pure theory vs. applied theory
Block 2: Scope and breadth of research	
5	Less vs. more specialization
6	Less vs. more multidisciplinarity
Block 3: Productive tradition or risky innovation	
7	Less vs. more risky research
8	Incremental vs. disruptive research
9	Quantity vs. quality
Block 4: Goal of theory: prediction or explanation	
10	Predictive theory vs. explanatory theory
Panel B: JEL topics	
A	General Economics and Teaching
B	History of Economic Thought, Methodology, and Heterodox Approaches
C	Mathematical and Quantitative Methods
D	Microeconomics
E	Macroeconomics and Monetary Economics
F	International Economics
G	Financial Economics
H	Public Economics
I	Health, Education, and Welfare
J	Labor and Demographic Economics
K	Law and Economics
L	Industrial Organization
M	Business Administration and Business Economics • Marketing • Accounting • Personnel Economics
N	Economic History
O	Economic Development, Innovation, Technological Change, and Growth
P	Economic Systems
Q	Agricultural and Natural Resource Economics • Environmental and Ecological Economics
R	Urban, Rural, Regional, Real Estate, and Transportation Economics
Z	Other Special Topics
Examples for JEL sub-topics: D6 Welfare Economics, D7 Analysis of Collective Decision Making	
Examples for JEL subject codes: D61 Allocative Efficiency • Cost–Benefit Analysis, D62 Externalities	

Notes: Panel A summarizes the ten research objective questions. Panel B presents the primary topics of the JEL classification system of the EconLit database (source: www.aeaweb.org/econlit/jelCodes.php).

4.3 Sample

Numerous researchers contribute to the economic literature and shape economic research objectives and topics. Here, our objective is to represent all strata of the economics profession and, hence, to give a voice to all active economic researchers, that is, all scholars who recently contributed to the international research exchange in economics. To meet this objective, we derive a large publication dataset that contains about 177,000 publications from the top 400 journals in economics, use these data to identify active contributors to the economic literature published in English, and invite all of them to the survey. This approach has three critical advantages: First, our study population is defined systematically in a data-driven way and encompasses all economic researchers who publish in English. Second, we are able to match detailed bibliometric background data to the survey responses. Third, we can use these data to quantify and control for selection into the sample. In particular, we can use post-stratification weights which ensure that our sample broadly represents the full spectrum of economic researchers. In this section, we describe how we compile the publication data (4.3.1) and identify the study population (4.3.2). We describe how we invite respondents and collect the survey data (4.3.3), and we characterize the sample of researchers that participated in the survey (4.3.4).

4.3.1 Publication data

We start from the publication database EconLit. It covers an extensive set of economic journals and, importantly, provides JEL codes for each published article which allows us to also study the actual distribution research topics in economics. The JEL codes are assigned in an independent and systematic review process by trained EconLit staff. This ensures maximal JEL code coverage and a consistent and systematic application of the classification criteria. We restrict our attention to published journal articles from 2009 to early December 2019, the time at which we downloaded the data. We exclude older articles because we are primarily interested in current economic research. We exclude working papers because their coverage is less systematic and JEL code information is often not available. We drop duplicate and non-research publications such as errata or memorials. Moreover, we only consider articles written in English, the lingua franca of economics and the language in which almost all high-impact research is published. Appendix 4.B documents the exact procedure.

EconLit, however, comes with two drawbacks: First, it does not contain information on articles' citations and, therefore, their scientific impact. Second, it includes more than 1,500 journals many of which have only a minuscule scientific impact or belong to neighboring fields such as business and management, statistics, or operations research. To circumvent these concerns, we concentrate on the 400 EconLit-indexed journals with the highest impact factor according to the Scopus 2018 Scimago Journal Ranking in the "Economics, Econometrics, and Finance" category. This restriction helps us to exclude journals that have hardly any influence on economic research at all and to zoom in on *economics* journals. Moreover, we are able to match 97.4% of these EconLit articles to Scopus's bibliometric database which includes information about article citations, journal rankings, and authors' background. We refine our final publication sample to the successfully matched articles, a total of 177,155 publications.

4.3.2 Study population

We use these publication data to identify the population of active English-publishing economic researchers. In a first step, we locate about 146,000 unique authors and gather further information about them.[5] We observe how many economic articles they published between 2009 and 2019, with whom they co-authored, to which JEL codes their articles are assigned, and how often their work is cited (as of December 2019). We use the co-author information to derive a discipline-wide co-author network from which we can derive how central and connected each author is. Moreover, we complement our data with Scopus's author information, including the authors' h-index, their total number of publications (with journal information and citations), the year of their first publication, and their institutional affiliation (as indicated in their publications). Finally, we predict the gender of each author from their names, using an algorithm of the commercial company Gender API (see Santamaría and Mihaljević, 2018). Appendix section 4.B.3 summarizes and describes all author covariates that will be used throughout the paper.

In a second step, we restrict the set of authors to *active economic* researchers. First, we exclude all scholars who did not publish an article in our publication data since 2015 (restriction 1). Second, we focus on scholars who publish at least 50% of their work in economics journals or have at least three articles in our sample (restriction 2). This step excludes researchers from neighboring fields who have little experience with the economic literature. Next, we exclude authors from non-academic institutions that have a very small publication output (restriction 3).[6] Those excluded are likely to be non-academic contributors or former academics who quit research. Finally, we consider only scholars for whom a valid email address can be found on-

5. We use Scopus's unique author identifiers, that are assigned to each article, to construct the author-level database. Scopus derives these identifiers with the help of an algorithm that tends to produce duplicates, that is, different author IDs for the same author. Thus, we combine separate author entries with identical first names, last names, and institutions. Further, we manually disambiguate all authors who have the same first and last name as an author who participated in the survey.

6. We consider an institution as non-academic if it contributed less than 20 articles to our publication sample and its name does not contain a keyword such as "school", "university", "research", or their counterparts in other languages. Authors who have at least three articles in our sample are exempted from this rule.

line (restriction 4).[7] Posting an email address online is a criterion for being active in research, but is also a precondition for the study: Only these scholars can be contacted and invited to the survey.

The procedure identifies 53,779 active economic researchers. Table 4.3.1 summarizes their characteristics. 26% of the population are female and about 75% work in Europe or Northern America. The average year of the first publication is 2007, which means that, on average, authors are active for 13-14 years at the time of the survey. On average, the authors write 4.8 articles in our publication sample with 5.8 unique co-authors, covering all JEL topics. In total, the average author has about 17.1 publications of which 75.9% fall into Scopus's economics category if we also count publications before 2009 and outside the top 400 EconLit journals. How successful are the authors? 12.1% are affiliated with one of the 50 leading research institutions (Shanghai Ranking), 6.1% published in a Top Five[8] journal since 2009, and the average h-index is 6.5.

Doctoral students. A limitation of our author population is that it does not contain junior researchers such as Ph.D. students who did not yet have the opportunity to publish their work. To partially offset this restraint, we derive a separate database of doctoral students. Specifically, we identify doctoral students in an economics program at one of the top 400 institutions (ranked according to total citations in our publication sample). We exclude institutions for which we could not find a central directory of student email addresses and students who are already part of the author population. This results in a population of 9,441 students from 219 institutions. 30.8% are female and 96.7% come from Europe or Northern America (see appendix table 4.C.1). Clearly, this group of students provides only a selected subset of Ph.D. students across the globe. Thus, we mainly use it to cross-verify the survey results among economic authors in a different population.

7. We gather most email addresses using Amazon's crowd-working platform Mechanical Turk. Each email address is collected at least twice by independent crowd-workers. We cross-verify all addresses. Conflicting cases are manually checked by crowd-workers and cross-verified once more. In a few cases, we also rely on corresponding author information from publications. We find an email address for 80% of the scholars who satisfy the other restrictions. Restricting the population to scholars with email address leads only to minor differences in the characteristics of the population (see appendix table 4.C.2). In later robustness analyses, we show that all results replicate with survey weights that match the characteristics of a population that also includes the scholars for whom no address could be found.

8. We consider the following journals as "Top Five": American Economic Review (but not Papers & Proceedings), The Quarterly Journal of Economics, Journal of Political Economy, Review of Economic Studies, and Econometrica. Publishing in these journals is commonly viewed as a primary indicator of academic success, although this practice has been strongly criticized (e.g., Heckman and Moktan, 2020).

Table 4.3.1. Characteristics of the study population and the sample

Variable	(1) Study population	(2) Unweighted sample	(3) Weighted sample
Gender, academic age			
Female	26.0%	23.1%	25.8%
Year of first publication (YYYY/MM)	2007/01	2006/01	2006/10
Number of papers			
Number of articles (in pub. sample)	4.8	5.6	4.9
Number of articles (overall)	17.1	18.3	16.2
Share of art. in econ. journals	75.9%	76.2%	76.8%
Co-author network (in pub. sample)			
Degree (number of unique co-authors)	5.8	6.5	5.7
Eigenvector centrality (index)	61.1%	65.6%	62.2%
Number of co-authors with Top Five pub.	0.5	0.8	0.5
Success			
Top 50 institution	12.1%	12.2%	12.5%
Publ. in Top Five Journal (in pub. sample)	6.1%	9.3%	6.1%
Num. of Top Five pub. (in pub. sample)	0.12	0.18	0.11
Average journal rank 1-400 (in pub. sample)	164.2	161.9	165.8
h-index	6.5	6.8	6.1
Continent			
Europe	40.4%	53.6%	40.5%
Northern America	33.9%	24.2%	33.9%
Asia	17.1%	13.4%	17.2%
Australia and New Zealand	4.3%	3.7%	3.3%
Latin America	2.7%	3.4%	3.3%
Africa	1.6%	1.7%	1.8%
Share of publications in JEL fields			
C Mathematical and Quantitative Methods	6.1%	6.3%	5.8%
D Microeconomics	13.1%	16.1%	13.5%
E Macroeconomics and Monetary Econ.	7.3%	7.4%	7.1%
F International Economics	4.4%	4.3%	4.2%
G Financial Economics	18.2%	11.3%	16.9%
H Public Economics	3.6%	4.3%	3.8%
J Labor and Demographic Economics	6.7%	9.8%	7.5%
L Industrial organization	8.3%	7.4%	8%
O Growth and Development Economics	8.5%	8.8%	9.2%
Q Agricultural and Environmental Econ.	7.1%	7.4%	7.4%
Other fields	16.6%	16.9%	16.6%
Sample size	53,779	7,794	7,794

Notes: Overview of covariates. Column 1: The eligible study population. Column 2: Respondents of the main sample, unweighted. Column 3: Weighted main sample (using post-stratification weights, see section 4.3.4). For a description of the covariates in the different rows see main text or appendix section 4.B.3.

4.3.3 Data collection

The survey was conducted online with the survey platform Qualtrics. We invited the full study population, 53,779 economic authors and 9,441 Ph.D. students, via email. The invitations were sent in random order from the 23rd of June 2020 to the 8th of July 2020.[9] To encourage participation among those who did not complete the survey, we sent a first reminder two weeks later and a second reminder in September 2020. We closed the survey on October 8th and drop all respondents who did not complete the main questions of their survey module.

9,921 researchers participated, yielding an overall response rate of 15.6%. Of those, 8,156 come from the population of economic *authors* (response rate: 15.2%), and 1,765 come from the *student* population (response rate: 17.8%). The main analyses rely on the data of 7,794 economic *authors* who completed the full survey. This restriction reduces changes in the sample size across different analysis steps due to missing data. Most respondents spent 9 to 25 minutes (25% and 75% percentile) to complete the survey, with a median response duration of 12 minutes.

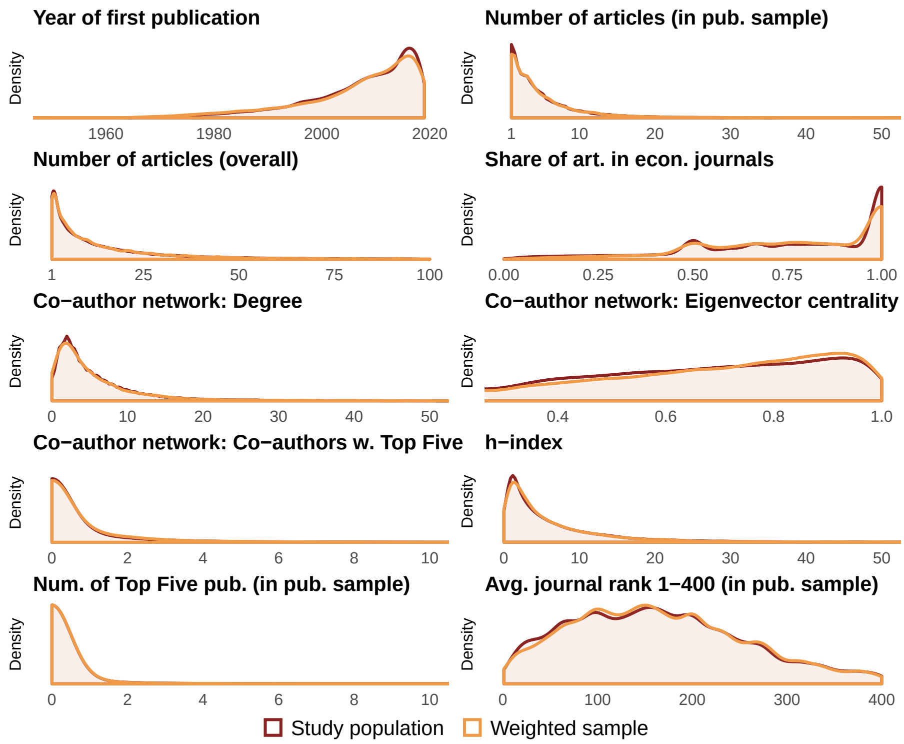
4.3.4 Sample characteristics

A unique feature of our study design is that we can observe and correct for selection into the sample on a diverse set of dimensions including gender, year of first publication (a proxy for "academic age"), continent of residence, publication success, research field, and position in the discipline-wide co-author network. This ensures that our main sample broadly represents the study population on a wide range of observable characteristics.

Column 2 of table 4.3.1 displays the characteristics of the unweighted main sample. By and large, it closely follows the characteristics of the study population. But we also observe evidence of selection into the sample. Participating researchers are on average slightly more experienced and successful than the average researcher in the study population. For instance, researchers in our sample have on average 0.8 more articles in our publication sample, 0.06 more Top Five publications, 0.7 more co-authors, and published their first publication 1 year earlier. Also, we observe slightly fewer female researchers in our sample (23% in the sample versus 26% in the population), more European researchers take part in the survey[10], and the participants publish relatively more papers in the JEL field *D Microeconomics* and *J Labor Economics* but less in *G Financial Economics* than the study population.

9. We also ran a small pilot invitation with 578 researchers on the 16th of July. Afterward, we introduced several small changes to the survey. 33 respondents saw the old survey version. We do not exclude their response data because the changes in the instructions were only minor.

10. The timing of the invitations, which were mostly sent between 2 PM and 9 PM CET, could have led to a higher response rate among Europe-based respondents.



Notes: Kernel density estimates for the distribution of covariates. Red: The eligible study population ($n = 53,779$). Yellow: The weighted main sample ($n = 7,794$). For a description of the covariates in the different sub-plots, see main text or appendix section 4.B.3.

Figure 4.3.1. Population and sample distributions of covariates

We calculate post-stratification weights to correct for these observed imbalances. Specifically, we use a raking algorithm and target the marginal distributions of gender (2 groups), the year of first publication (quartiles), the number of papers in our publication sample (quartiles), the h-index (quartiles), region (Europe, Northern America, Asia, Other), and the main research field (6 groups). The algorithm assigns greater weight to observations from under-represented groups. We follow the guidelines of the American National Election Study Weighting System (Pasek, Debell, and Krosnick, 2014). Appendix section 4.C.1 provides further details.

Column 3 of table 4.3.1 shows the characteristics of the weighted sample. The statistics illustrate that the weighting corrects for both targeted and untargeted imbalances. Across all covariates, the remaining differences between the weighted sample and the population are minor. Of course, table 4.3.1 displays only average values for many covariates which could conceal important differences in the variables' underlying distributions. Yet, figure 4.3.1, which contrasts the distributions of all continuous covariates in the population and the weighted sample, dispels this concern. In fact, the distributions overlap almost completely, indicating that our sample broadly represents the full spectrum of economic researchers.

The demographic module of our survey allows us to further characterize our sample (see appendix figure 4.C.1). About 90% of respondents engage in academic research (including 4.6% students). 8.5% describe themselves as “non-academic researcher”. 33.5% of the active academics are full professors, 28.2% have an associate professorship (or an equivalent position as reader or senior lecturer), and 22% are assistant professors (or lecturers). 88.9% of the respondents indicate that economics, econometrics, or finance is their primary academic discipline.

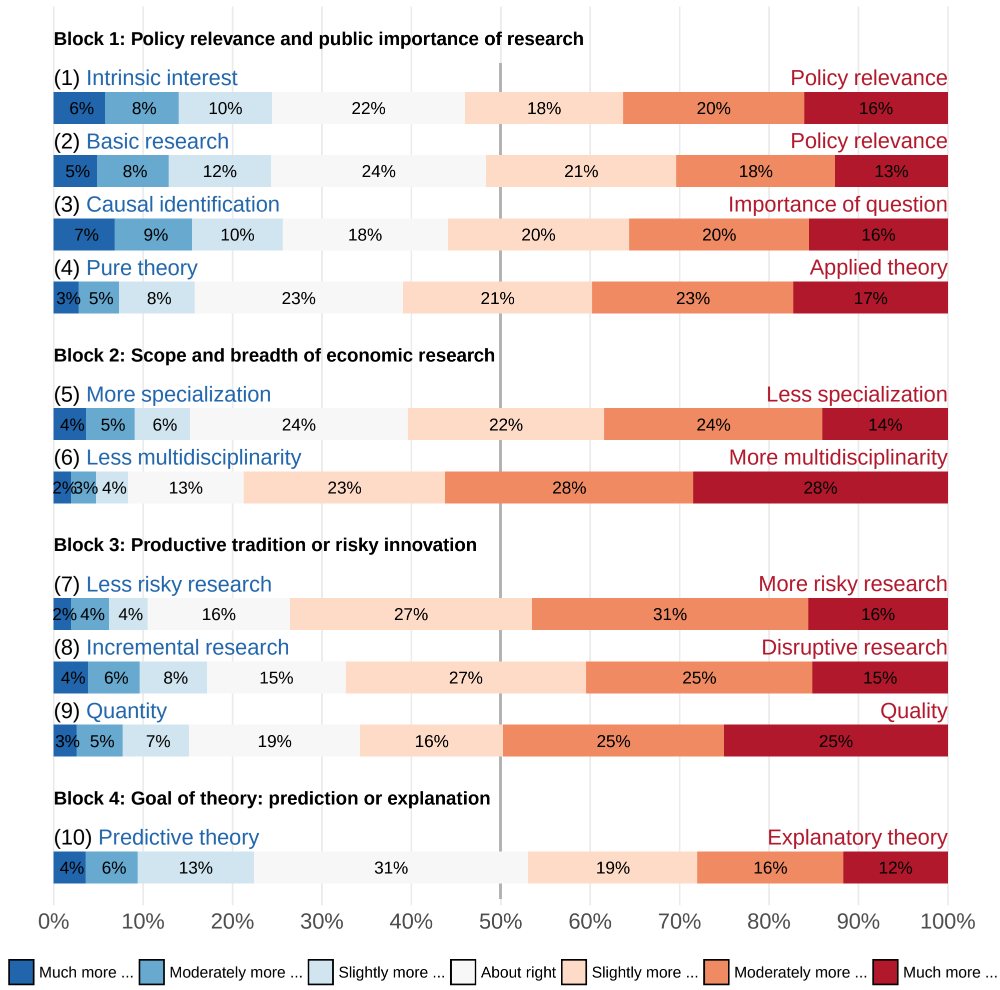
4.4 Results

In presenting our results, we first describe our findings with respect to research objectives before turning to the choice of topics. For both, objectives and topics, we discuss heterogeneity, aggregate outcomes, and determinants.

4.4.1 Research objectives

Heterogeneity of responses. Figure 4.4.1 displays the distribution of responses to the ten research objective questions. The questions ask respondents to trade off two opposing research objectives and indicate whether they think economic research should place more weight on one objective versus the other. The results reveal that economists’ opinions are vastly heterogeneous. Typically, both opposing research objectives as well as the neutral category (“Current state is about right”) attract significant support. For instance, 24% of the respondents advocate that intellectual, intrinsic interest should play a greater role in economic research relative to policy relevance than it does today, while 54% endorse the opposite view, and 22% are satisfied with the status quo (question 1). We observe heterogeneity not only in the direction but also in the magnitudes of the desired changes. For instance, 18% of economists believe that “slightly more”, 20% that “moderately more”, and 16% that “much more” policy relevance (vis-à-vis intrinsic interest) is needed. A similar picture emerges for most of the other questions.

Importantly, this dissent cannot simply be attributed to a generic inability of economic experts to agree on certain issues. For example, as we already noted in the introduction, a clear majority of economists (88.4%) support the purpose of this study and agree that studying how economists think economics should be conducted is interesting. Past research also shows that economists largely agree on factual issues such as the notion that higher government spending reduces unemployment or that carbon taxes are a more cost-effective environmental policy than mandatory car standards (Sapienza and Zingales, 2013; Andre et al., 2021). In other words, consensus among economic experts is possible, yet the question of which research objectives economics should pursue remains fundamentally disputed.



Notes: Distribution of survey responses to the ten research objective questions (weighted sample). The overarching question is: "In comparison with how research in economics is currently conducted, how should economists conduct research?" The labels at the top left and top right of each distribution summarize which two research objectives a question contrasts. The legend displays the available response categories. The full wording of the questions is available in appendix 4.A.

Figure 4.4.1. Distribution of survey responses to the research objective questions

Aggregate results. The aggregate results show that most economists express dissatisfaction with how research is currently conducted. Across the ten questions, only 13% to 31% (average: 20.6%) of respondents say that the current state of research is "about right". The large majority of economists thus prefer a deviation from the status quo. Note that we observe this pronounced dissatisfaction despite the fact that the answer category in support of the status quo is framed relatively moderately. Agreement with this category does not imply that the status quo is viewed as "exactly" right but only "about" right, leaving room for modest disagreement.

Despite the observed heterogeneity, we find that most economists actually agree on the preferred direction of change. In fact, for most objectives, more than half of the respondents agree about the direction in which economics should deviate from the status quo. First, economists favor more policy relevant research. 54% of the experts advocate a shift towards more policy relevance relative to intrinsic interest (question 1). This share significantly differs from 50% ($p<0.001$, t-test).[11] Likewise, 52% support a shift towards more policy relevance relative to basic research (question 2, $p=0.062$). For empirical work, 56% of economists favor working on more important research questions even if this comes at the cost of less causal identification (question 3, $p<0.001$). Moreover, for theoretical work, 61% would prefer more applied, evidence-related theory instead of pure theory (question 4, $p<0.001$).

Second, more than half of the respondents express a preference for a greater scope and breadth of economic research: Research should be less specialized (question 5, $p<0.001$) and more multidisciplinary (question 6, $p<0.001$), implying that economics should incorporate more insights from other disciplines to study economic questions. In fact, multidisciplinarity is the issue on which economists reach the most pronounced consensus, with almost 80% of respondents supporting a shift towards increasing multidisciplinarity.

Third, a majority endorses a shift towards more risky innovation instead of incremental, traditional research. Respondents say that economic research should be more risky (question 7, $p<0.001$), disruptive (question 8, $p<0.001$) and place a stronger focus on quality versus quantity (question 9, $p<0.001$).

The final question asks whether economic theory should place greater emphasis on predicting versus explaining outcomes (question 10). Here, the responses are more balanced. 47% of respondents indicate that they prefer a shift towards more explanation, 22% favor a shift towards more prediction, while 31% think that the status quo is about right, reflecting the largest fraction of neutral responses observed across all questions.

In short, the majority of economists agree on the direction of change. They favor a shift towards more policy-relevant and risky research with a broader scope and stronger multidisciplinary orientation.

11. See appendix table 4.D.1. We also show that average responses significantly differ from the neutral category.

We obtain virtually identical results with different weighting schemes: (i) weights that target a scholar population that also includes authors for whom no email address could be found, thus correcting for a potential differential availability of email contact data; (ii) identical weights for all authors; (iii) identical weights for all authors who say that economics is their primary academic discipline (89%); and finally (iv) identical weights that also include the full student sample (see appendix figure 4.D.1; appendix section 4.C.1 contains details on the weighting schemes). In particular, the responses of students largely mirror those of the authors (see also appendix figure 4.D.2). Thus, there appears to be no divide between the current population of publishing scholars and its next generation.

Do economists prefer different research objectives for their own field of expertise? To answer this question, we elicit respondents' opinions not only for economics as a whole but also for their main field. Appendix figure 4.D.3 compares the distribution of responses to both question types and documents largely identical results. Hence, economists express similar views about the state of the profession, irrespective of considering economics "as a whole" or their "own field", respectively. Appendix figure 4.D.4 disaggregates the field-specific responses and reports similar trends in each individual field. There are only a few exceptions. For instance, economists who identify either Microeconomic Theory, Economic History, Mathematical Methods, or Economic Thought/Heterodox Economics as their main field place less emphasis on policy relevance.

Predictors of responses. Next, we ask whether economists' opinions are systematically related to their characteristics. The rich author data allow us to regress the survey responses on basic demographic characteristics (gender, age, tenure, region), indicators of academic success (affiliation with top 50 institution, Top Five publication, h-index), and the share of theory and methods projects a researcher works on. We also account for the research topics respondents work on: We include (but for the sake of brevity do not report) the researchers' share of publications in each primary JEL topic and their share of publications in economics journals (see appendix section 4.B.3 for details about all covariates). We use the Benjamini-Hochberg procedure to correct all reported coefficients jointly for multiple hypotheses testing (Benjamini and Hochberg, 1995). Table 4.4.1 summarizes the results.[12]

Individual characteristics prove to be predictive of the views about research objectives. Most characteristics predict a consistent shift either towards or against the majority view (more policy relevance, broader scope, more risky innovation). For instance, female economists show on average greater support for policy relevance (question 1), multidisciplinarity (question 6), and disruptive research (question 8), in line with the majority view. By contrast, economists in Africa, Asia, and Latin

12. We obtain very similar results in ordered probit regressions and regressions with different weighting schemes. These analyses are available upon request.

Table 4.4.1. Predictors of preferred research objectives

	Response to research objective question (standardized)									
	Pol. relev. (vs. intrin. interest)	Pol. relev. (vs. basic research)	Importance (vs. causal ident.)	Applied theory (vs. pure)	Less specialization	More multidisciplinarity	More risky research	Disruptive research (vs. incremental	Quality (vs. quantity)	Explanation (vs. prediction)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Demographics										
Female	0.106**	0.076	0.065	0.072	0.018	0.167***	0.004	0.112**	0.034	0.047
	(0.041)	(0.043)	(0.043)	(0.042)	(0.044)	(0.042)	(0.044)	(0.043)	(0.043)	(0.043)
Age (in 10y)	0.028	0.013	−0.104***	−0.000	0.113***	−0.007	0.036	0.040	0.130***	0.009
	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)	(0.018)	(0.020)
Tenured	−0.044	−0.029	0.040	−0.038	−0.039	−0.033	−0.048	−0.051	−0.055	0.046
	(0.040)	(0.040)	(0.041)	(0.040)	(0.042)	(0.040)	(0.041)	(0.040)	(0.041)	(0.041)
Region (vs. NA/AUS/NZL)										
EUR	0.002	−0.053	0.013	−0.054	0.109**	−0.033	0.106**	0.076	0.194***	0.091*
	(0.040)	(0.041)	(0.041)	(0.039)	(0.041)	(0.040)	(0.041)	(0.042)	(0.040)	(0.041)
AF, AS, LA	−0.221***	−0.101	−0.195***	−0.165**	−0.284***	−0.132*	−0.339***	−0.234***	−0.101	0.030
	(0.058)	(0.059)	(0.058)	(0.058)	(0.061)	(0.058)	(0.062)	(0.059)	(0.059)	(0.058)
Success										
Top 50 inst.	0.037	0.001	−0.051	0.050	0.039	0.076	0.127*	0.109	−0.110	−0.082
	(0.053)	(0.052)	(0.057)	(0.050)	(0.054)	(0.052)	(0.055)	(0.054)	(0.059)	(0.056)
Top Five	−0.210***	−0.200***	−0.100	−0.158**	−0.116	−0.182**	0.001	−0.001	0.240***	−0.090
	(0.059)	(0.058)	(0.059)	(0.058)	(0.057)	(0.062)	(0.058)	(0.057)	(0.059)	(0.056)
h-index (in 10)	0.022	−0.001	0.125***	0.095***	0.018	−0.026	0.030	0.050	−0.076*	0.070*
	(0.033)	(0.034)	(0.034)	(0.030)	(0.032)	(0.034)	(0.033)	(0.030)	(0.034)	(0.033)
Project types (vs. empirics)										
Theory (in 10%)	−0.049***	−0.047***	0.008	−0.079***	−0.004	−0.009	−0.005	0.006	0.011	0.018**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Methods (in 10%)	−0.030**	−0.032**	−0.046***	−0.030**	−0.032**	−0.024*	−0.053***	−0.031**	0.010	0.003
	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)	(0.012)
JEL topic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	3,887	3,880	3,871	3,874	3,888	3,891	3,880	3,880	3,882	3,856
R^2	0.060	0.048	0.037	0.079	0.062	0.055	0.050	0.034	0.052	0.036

Notes: Weighted OLS regressions, robust standard errors in parentheses. The dependent variables are the *standardized* survey responses to the research objective questions, as indicated by the column labels. The explanatory variables include various author characteristics. Age and h-index are divided by 10, theory and methods are divided by 10%. All regressions control for the share of publications in each primary JEL topic as well as the share of publications in economics journals. p-values are adjusted for multiple hypotheses correction across all coefficients reported in this table, using the Benjamini-Hochberg-procedure. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. *Region abbreviations*: NA – Northern America, AUS – Australia, NZL – New Zealand, EUR – Europe, AF – Africa, AS – Asia, LA – Latin America.

America show weaker support of policy relevance (question 1, 3, 4) and disruptive research (question 7, 8), opposite to the majority view. Economists who have published a Top Five paper also tend to place less weight on policy relevance and multidisciplinarity but place more weight on quality. Likewise, theorists and methods researchers show a weaker preference for policy relevance, and the latter also tend to favor specialization and incremental research to a greater extent.

4.4.2 JEL topics

Aggregate results. Figure 4.4.2 compares the distribution of JEL topics in our publication sample (in blue) with the average survey response (in red). The former shows which fraction of papers is published in each JEL topic, which is derived from our publication data from the top 400 EconLit-indexed journals from January 2009 to December 2019.[13] It thus describes the state of economic research in the period before our survey was launched. We can directly compare it to the average survey responses, which show economists' average opinion on which share of papers should be written and published in each JEL topic.

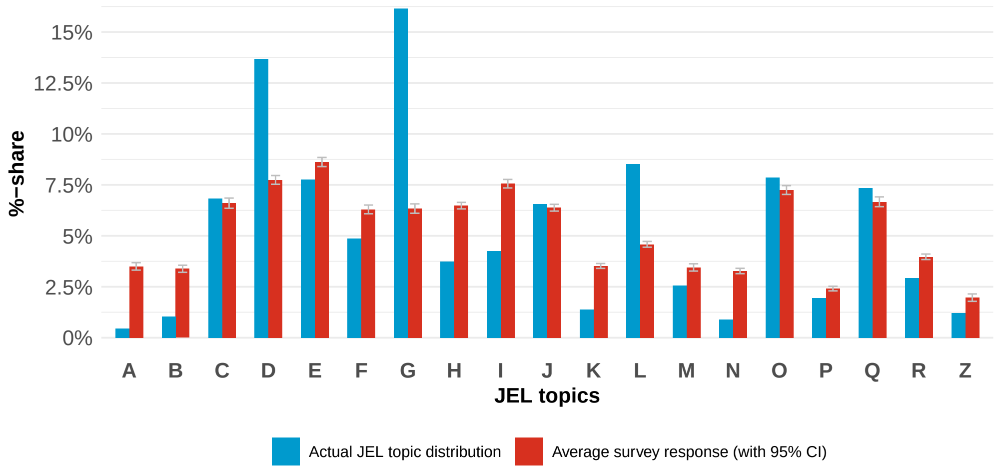
Qualitatively, we observe a similar ordinal ranking of JEL topics in the publication data and the average survey responses, as manifested in a sizable rank-order correlation of 0.76 ($p < 0.001$). JEL topics that dominate the research output in economics (such as *D Micro*, *E Macro*, or *G Finance*) also receive large weights in the survey. JEL topics that play a relatively minor role in economics today (such as *A General & Teaching*, *K Law and Economics*, or *N History*) also receive small weights in the survey.

Quantitatively, however, we observe sizeable discrepancies between the two distributions. Respondents on average spread the weights across the nineteen JEL categories more uniformly. For instance, the average weight that respondents assign to the field with most publications – *G Finance* – is 9.8 percentage points smaller than its actual share of publications (see figure 4.4.3). Respondents also place a much lower weight on the second and third most prominent fields, *D Micro* and *L Industrial Organization*. By contrast, respondents on average think that more work should be published in JEL fields that see relatively few publications in practice. In short, economists on average place more weight on minor JEL topics and less weight on the most common JEL topics. In other words, they favor a more diverse and pluralistic distribution of topics in economic research.[14]

A potential concern is that the results are overly sensitive to how we aggregate the survey responses and derive the actual distribution of JEL topics. Therefore, we

13. In practice, most papers are assigned to multiple JEL codes. We derive each paper's weight in topic j as the share of codes in j. For example, a paper with two codes in D and one code in L receives a weight of $\frac{2}{3}$ for D and a weight $\frac{1}{3}$ for L. In appendix 4.D.2, we show that the analyses are robust to using three alternative aggregation procedures.

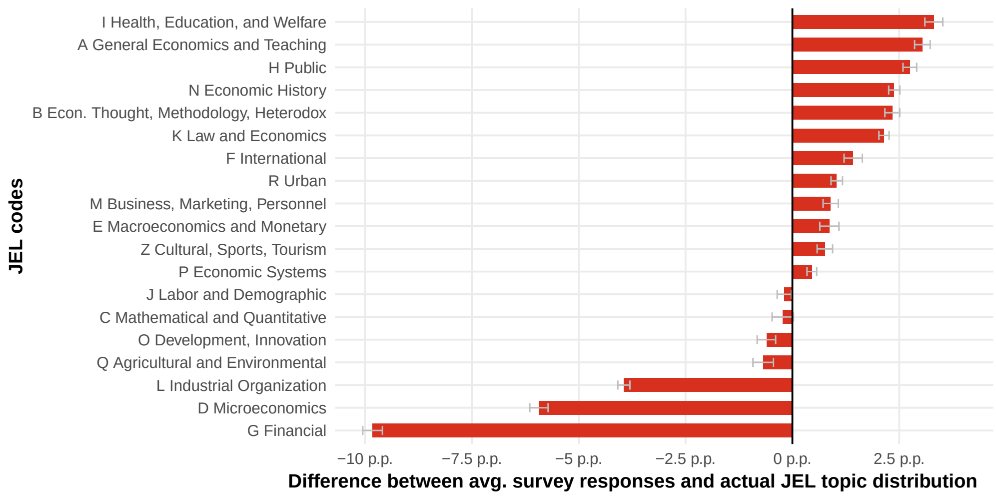
14. Appendix section 4.D.2 documents a similar phenomenon for the 130 JEL sub-topics.



Notes: Blue bars: Shares of JEL topics in our publication sample (EconLit publication data, top 400 journals, January 2009 - December 2019). Red bars: Weighted average survey responses with 95% confidence intervals.

JEL topics: **A** General Economics and Teaching, **B** Econ. Thought, Methodology, Heterodox, **C** Mathematical and Quantitative, **D** Microeconomics, **E** Macroeconomics and Monetary, **F** International, **G** Financial, **H** Public, **I** Health, Education, and Welfare, **J** Labor and Demographic, **K** Law and Economics, **L** Industrial Organization, **M** Business, Marketing, Personnel, **N** Economic History, **O** Development, Innovation, **P** Economic Systems, **Q** Agricultural and Environmental, **R** Urban, **Z** Cultural, Sports, Tourism.

Figure 4.4.2. Comparison of JEL topic distributions in econ. journals with survey responses



Notes: Differences between the red and blue bars from figure 4.4.2 with 95% confidence intervals.

Figure 4.4.3. Differences between the avg. preferred and the actual JEL topic distribution

conduct five additional tests to address these concerns. First, we explore the sensitivity of the survey results to different weighting schemes and include the responses from the student sample. Second, we exclude possibly "careless" participants whose response behavior suggests that they might not have paid sufficient attention to the survey. For instance, we exclude respondents who assign a positive weight to only a few topics, spend only little time on the JEL topics question, or show a low standard deviation of preferred topic shares, which indicates a potential uniformity bias in responses. Third, we derive the actual distribution of JEL topics only from papers that were published by an author of our study population. Fourth, one may argue that our set of top 400 EconLit journals still contains many outlets with negligibly low impact on economic research. We therefore also derive the JEL topic distribution of the top 200 and top 100 journals. Finally, given that the period 2009-2019 might be considered too long to study the *current* topics of economic research, we also calculate the topic distribution for the 2015-2019 and 2018-2019 periods and explore its time trends. We replicate our main conclusions in all of these sensitivity analyses (see appendix figures 4.D.6 and 4.D.7 and the discussion in appendix 4.D.2). In particular, we detect no sizeable time trends in the distribution of research topics over the last decade (see appendix figure 4.D.8). Thus, even a time lag between starting and publishing research projects – which could in principle separate current topic preferences and published research output – is unlikely to explain the results. Again, we observe virtually identical results in the author and student sample (appendix figure 4.D.9).

Relatedly, one may wonder how the survey responses compare to the topic distribution in Top Five journals. After all, these journals are considered "general interest journals" and aspire to publish the best economic research in all fields. Appendix figures 4.D.10 and 4.D.11 contrasts their topic distribution with the survey responses and the topic distribution in the top 400 journals. First of all, we notice that – compared to the full set of journals – Top Five journals publish more research in the fields *C Mathematical Methods*, *D Microeconomics*, and *J Labor and Demographic Economics*, but less research in the fields of *G Finance*, *O Development*, and *Q Environment and Agricultural Economics*. However, in comparison with economists' average survey responses, we can still conclude that the average economist would prefer a more diverse distribution of research topics. In particular, economists assign a 20.3 percentage points lower weight to *D Microeconomics*, the JEL topic that by far dominates Top Five publications (see appendix figures 4.D.10 and 4.D.11). It is also noteworthy that economists assign a 4.6 percentage points higher weight to *Q Environmental and Agricultural Economics*, mirroring the recent critique that top economic research is rather silent about climate change (Oswald and Stern, 2019).

The JEL topics module also asked respondents how economic research should be distributed across three broad project types: projects that predominantly focus on theory (formal and informal), empirics, or methods (e.g., econometrics or computational techniques). On average, economists think that about 48% of research should be empirical, 28% theoretical, and 24% should focus on methods (see appendix figure 4.D.12).

Heterogeneity. The average results conceal considerable heterogeneity in the responses and opinions of economists. Indeed, the small confidence intervals in figure 4.4.2 can be attributed to the large sample size, rather than a small dispersion of responses. Appendix figure 4.D.13 maps the distribution of responses for each JEL category. The shares assigned to most topics range from 0% to more than 10%.

Predictors of responses. The documented heterogeneity in preferred research topics is systematically related to respondents' characteristics. The strongest and most consistent predictor is the topic of the authors' own publications. Respondents favor their own fields. They assign an about 1 percentage point stronger weight to a JEL topic if they have a 10 percentage point higher share of publications in this topic (see appendix table 4.D.3). This corresponds to a weight increase of 0.19 standard deviations. Thus, a respondent who writes all publications on a single JEL topic would on average assign an about 10 percentage point (1.9 standard deviations) stronger weight to it.[15]

As before, we also explore a rich battery of other characteristics, including gender, age, region, and academic success. The most predictive characteristics are female gender, having published in a Top Five journal, and the share of one's work in economic theory and methods. For instance, female scholars place comparatively less weight on *E Macro* and *N History*, but more weight on *I Health, Education, Welfare*, *J Labor*, and *Q Environmental/Agricultural*. We refer the interested reader to appendix table 4.D.2, which summarizes the results.

4.5 Discussion

Investigating economists' opinions about economics in a large, representative survey, we document three main findings. First, economists' views about how economics should be done are vastly heterogeneous. Second, many economists express dissatisfaction with the current state of economic research. Third, despite the considerable heterogeneity in views, respondents on average agree on the preferred direction of change. They think that economic research should become (i) more policy-relevant, (ii) more multidisciplinary, (iii) more risky and disruptive, and (iv) pursue more diverse topics. In this section, we discuss these results.

The rich heterogeneity of opinions serves as a reminder that any statement about "right" or "interesting" research questions, objectives, and topics is inherently subjective. While there are often scientific criteria for what constitutes a good *answer*,

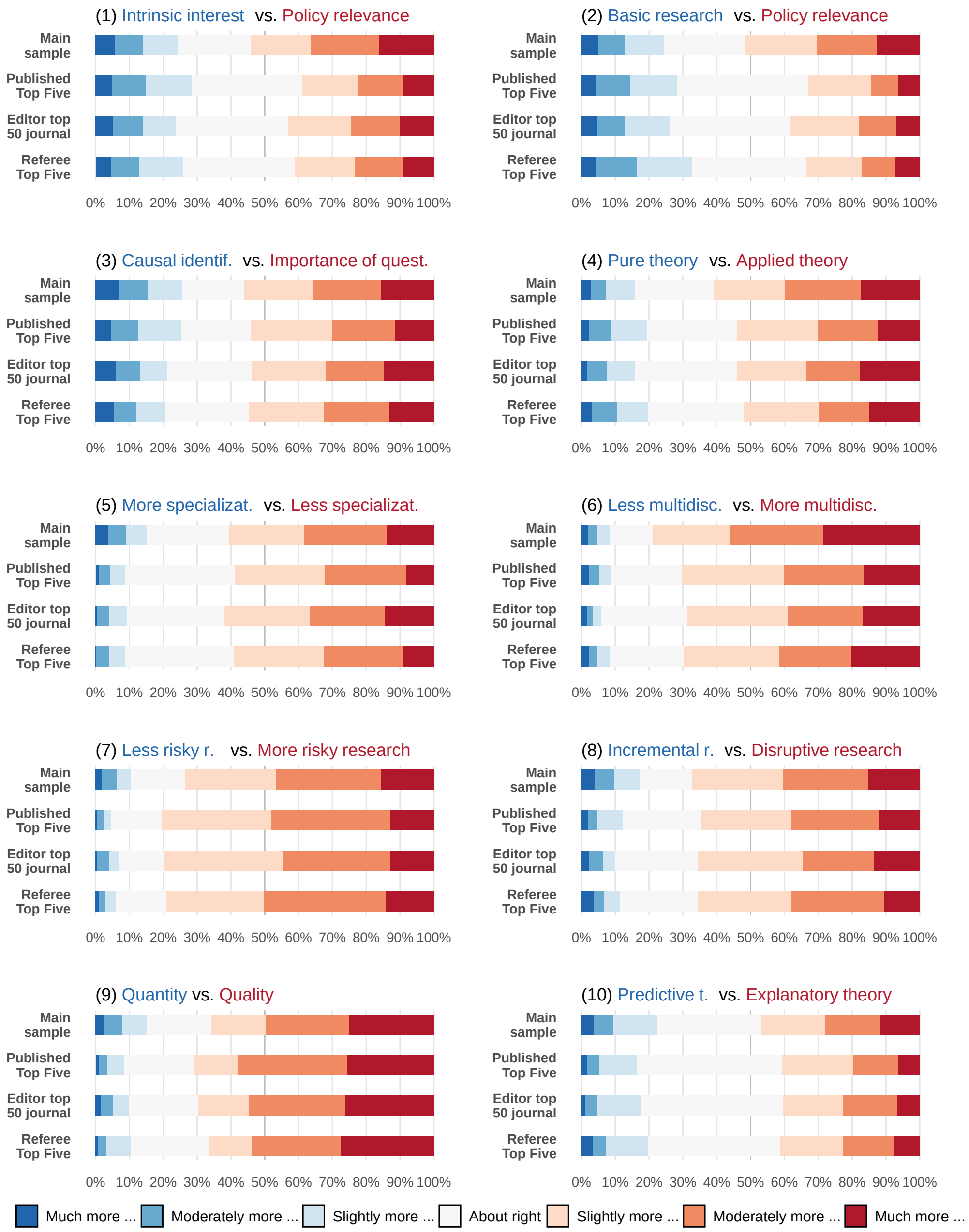
15. These results are robust to including controls and different weighting schemes (appendix table 4.D.3).

there are no objective guidelines for what constitutes a good *question*. The problem of problem choice eludes a clear, objective, scientific solution (Weber, 1919).

The documented mismatch between economists' views and current research practices in economics reveals that economists' research preferences are currently not reflected in their discipline's research output. Explanations for this mismatch are likely to be multifaceted and may range from researchers' strategic motives and career concerns (Reif, 1961; Frey, 2009; Akerlof, 2020), academic fads, fashions, and bandwagon effects (Sunstein, 2001; Bramoullé and Saint-Paul, 2010), to a "tyranny" of top journals (Heckman and Moktan, 2020). An empirical distinction of these explanations is beyond the study's design and purpose. Instead, we discuss potential implications of this mismatch.

We first ask whether the presented "majority" opinion is in fact "relevant". Science is not a democratic process and the majority opinion does not necessarily provide reliable guidance in academia. In practice, successful and highly reputed scholars typically have more influence on the discipline's research agendas, topics, and objectives (Bourdieu, 1975; Azoulay, Fons-Rosen, and Zivin, 2019). Their research is more visible and – as editors or referees – their judgments critically shape the publication process. One could argue that their experienced assessments indeed weigh more strongly than those of junior colleagues or scholars with a shorter academic track record. Top economists might see less need for change and therefore promote and reinforce the current status quo as authors, research leaders, referees, and editors.

However, this argument is firmly rejected by the data: Top economists widely share the discipline's discomfort with its research objectives and topics. To investigate this, we identify influential economists using three complementary approaches. First, we focus on economists who have published at least one article in a Top Five journal within our publication sample. Second, we locate editors and advisory board members at the top 50 EconLit-indexed economics journals between 2015 and 2020. Third, we identify scholars who have repeatedly refereed at Top Five journals between 2015 and 2020. Appendix section 4.D.3 contains further details. 6.1% of our weighted sample (population: 6.1%) have published a Top Five paper, 3.2% have served as a member of an editorial or advisory board at a top 50 journal (population: 3.6%), and 6.1% have repeatedly reviewed papers for the Top Fives (population: 4.9%). Figure 4.5.1 presents the distribution of their preferred research objectives and compares it to the views of the full sample. Aside from somewhat weaker support of policy relevance vis-à-vis intrinsic interest and basic research, the views of top economists mirror those of the field at large. In particular, they favor a shift towards more important research questions (at the costs of causal identification), less specialization, more multidisciplinary, and more risky research. Appendix figure 4.D.14 shows that their topic preferences are close to those of the full discipline as well.



Notes: Weighted distribution of survey responses to the ten research objective questions. The overarching question is: "In comparison with how research in economics is currently conducted, how should economists conduct research?" The results are displayed for the main sample and the (unweighted) subsets of authors with a Top Five publication (in our publication sample), editors at top 50 journals, and referees at Top Five journals.

Figure 4.5.1. Top economists' responses to the research objective questions

Second, we discuss whether recent trends in economic research are likely to reduce the future mismatch between the current research practice in economics and economists' views. Economics is a constantly evolving discipline and the change that many economists desire might already be on its way. We start with the research topics and derive the JEL topic distribution for each year from 2009 to 2019. We detect no consistent trend that, when extrapolated to the future, would move the distribution of research topics closer to economists' preferences (appendix figure 4.D.8). Thus, in terms of research topics, recent trends are unlikely to reduce the mismatch anytime soon.

Observing the development of research objectives is arguably more challenging, as objectives such as "policy relevance", "quality", or "disruptiveness" are difficult to quantify. Nonetheless, recent work assesses the evolution of multidisciplinarity, applied theory, and causal identification. These studies observe that, over the last decade, economics has become more multidisciplinary (Angrist et al., 2020; Buyalskaya, Gallo, and Camerer, 2021), theory has become less prevalent and more applied (Hamermesh, 2013; Angrist, Azoulay, Ellison, Hill, and Lu, 2017; Backhouse and Cherrier, 2017), and techniques of causal identification have become increasingly important (Currie, Kleven, and Zwiers, 2020).[16] We do not observe whether the shift towards identification has come at the cost of less policy relevance and research questions of lower public relevance (Akerlof, 2020). However, the trends in multidisciplinarity and applied theory have indeed brought the field closer to economists' preferred objectives. Thus, signs of progress are visible, but sustained change is needed to reduce the mismatch noticeably. For instance, multidisciplinarity is still the research objective for which we document the highest degree of dissatisfaction today, with almost 80% supporting a continued shift towards more multidisciplinary research.

Next, we turn from discipline-wide metrics to the individual researcher and investigate whether the widespread disagreement with the status quo has implications for the well-being of individual scholars. Do researchers who disagree with the current research objectives and topics show lower job satisfaction? To shed light on this, the survey asks respondents to rate (i) how satisfied they are with their job in general, (ii) with the topics that they work on, (iii) how stressful they find their job, and (iv) whether they perceive academia as "overly competitive". Table 4.5.1 regresses these standardized measures on a "satisfaction with economics" index score and a

16. Angrist et al. (2020) show that citations to other disciplines have increased in economics. Buyalskaya, Gallo, and Camerer (2021) observe that funding agencies, such as the NSF, have recognized the need to support interdisciplinary projects. Hamermesh (2013) and Angrist et al. (2017) document that less purely theoretical research is published in top journals, while Backhouse and Cherrier (2017) discuss that this development has been accompanied by a turn towards more applied theory. Currie, Kleven, and Zwiers (2020) use text-mining methods to show that publications increasingly mention causal identification techniques such as field experiments or regression discontinuity designs.

large set of demographic and bibliometric covariates. The index is a joint measure of economists' satisfaction with their discipline's research objectives and topics. We pool the samples from both survey modules to leverage maximal statistical power. The index is calculated as follows. In the research objectives module, the index measures how often and how strongly respondents agree with the status quo. We derive the sum of absolute deviations (in scale points) from the "about right" category and take its negative z-score. In the JEL topics module, the index measures how close the distribution that a respondent prefers is to the current topic distribution in economics. Here, we derive the sum of absolute deviations from the actual topic shares and take its negative z-score.

The results in table 4.5.1 show that a higher satisfaction with economics' research objectives and topics is paralleled by higher job satisfaction and less job-related stress. For instance, a one standard deviation increase in satisfaction with economic research is associated with a 0.07 standard deviation increase in general job satisfaction and a 0.13 standard deviation reduction in perceiving academia as being overly competitive. These results hold conditional on a rich vector of control variables, are robust to using different weighting schemes, and can be replicated in each survey module separately (appendix section 4.D.3). Hence, disagreeing with the current objectives and topics in economics is associated with a psychological and mental burden. As an aside, the results also reveal that tenured scholars report significantly higher job satisfaction, likewise economists who work for a leading research institution or have published in a Top Five journal.

The fact that researchers whose views and preferences align with the prevailing research practices are more satisfied could also have implications for the diversity of scholars in economics, in particular concerning gender (Bayer and Rouse, 2016; Avilova and Goldin, 2018; Buckles, 2019; Lundberg and Stearns, 2019; Lundberg, 2020). Indeed, column 5 of table 4.5.1 reveals that satisfaction with economic research substantially varies across demographic groups. It is highest for tenured scholars and economists who publish in Top Five journals.[17] It is lower for older, European-based, and female scholars.[18] Female economists are on average 0.07 standard deviations less satisfied with the current research objectives and topics in economics. One potential explanation is that under-represented groups such as women have comparatively less influence on the fields' research agendas so that their research preferences remain under-represented. In turn, disagreement with economics' practices could adversely affect who is willing to pursue an academic career. In this case,

17. We also find that editors at top 50 journals and referees at Top Five journals are more satisfied with the status quo (see appendix table 4.D.4). However, as documented above, these effects do not offset the overall dissatisfaction among top economists.

18. The results are robust to the use of different weighting schemes. We also find largely identical results if we estimate the regression separately for each survey module. For gender, the point estimates remain unchanged but lose significance due to the split sample size (see appendix table 4.D.9).

Table 4.5.1. Predictors of satisfaction

	Satisfaction (std.)				
	Own job	Own topics	Stress	Overly competitive	Satisfact. w/ econ.
	(1)	(2)	(3)	(4)	(5)
Satisfact. w/ econ.	0.072***	0.034**	−0.040***	−0.127***	
	(0.014)	(0.014)	(0.013)	(0.013)	
Female	−0.072**	0.027	0.216***	0.230***	−0.072**
	(0.032)	(0.031)	(0.030)	(0.029)	(0.031)
Age (in 10y)	0.025*	0.053***	−0.151***	−0.066***	−0.069***
	(0.014)	(0.013)	(0.013)	(0.013)	(0.014)
Tenured	0.153***	0.034	−0.026	−0.075**	0.068**
	(0.030)	(0.029)	(0.029)	(0.029)	(0.030)
Region: EUR	0.041	0.042	0.132***	0.114***	−0.096***
	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
Region: AF, AS, LA	−0.036	−0.104**	0.016	−0.024	−0.067
	(0.042)	(0.041)	(0.039)	(0.040)	(0.042)
Top 50 inst.	0.089**	0.080*	0.041	0.010	0.016
	(0.042)	(0.040)	(0.042)	(0.042)	(0.039)
Published Top Five	0.225***	0.175***	0.020	−0.143***	0.248***
	(0.042)	(0.043)	(0.045)	(0.047)	(0.043)
h-ind. (in 10)	0.113***	0.107***	−0.068***	−0.051**	0.010
	(0.020)	(0.023)	(0.024)	(0.023)	(0.024)
Method ctrl.	✓	✓	✓	✓	✓
Topic ctrl.	✓	✓	✓	✓	✓
Module FE	✓	✓	✓	✓	✓
Observations	7,489	7,493	7,487	7,493	7,497
R^2	0.046	0.037	0.076	0.065	0.048

Notes: Weighted OLS regressions, robust standard errors in parentheses. In each column, the dependent variable is a different, standardized survey measure of satisfaction: (1) job satisfaction, (2) satisfaction with own research topics, (3) job-related stress experiences, (4) perception of academia as overly competitive, and (5) the "satisfaction with economics" index score. Age and h-index are divided by 10. Method controls include the share of projects in theory and methods research respectively. Topic controls include the share of publications in each primary JEL topic as well as the share of publications in economics journals. p-values are adjusted for multiple hypotheses correction within the reported coefficients of each row, using the Benjamini-Hochberg-procedure. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

the under-representation of women would be self-reinforcing. Moreover, even conditional on satisfaction with economic research and a rich battery of controls, female economists show lower overall satisfaction. Their job satisfaction is 0.07 standard deviations lower, their reported stress is 0.2 standard deviations higher and they perceive academia as being overly competitive to a stronger extent. Taken together, these observations confirm the concern that economics is a male discipline (Lundberg, 2020; Wu, 2020; Dupas et al., 2021). Male researchers outnumber women (3:1, see table 4.3.1), are more satisfied with their job, less stressed, and agree with the field's research objectives and topics to a stronger extent.

We conclude that there are good reasons to be concerned about the mismatch between economists' views and the reality of economic research. For one, there is broad and systematic support for a change in economics' research objectives and topics, even among the discipline's most distinguished scholars. Moreover, the disagreement is associated with lower job satisfaction and is larger among female economists which may have consequences for diversity in economics.

4.6 Conclusion

We document economists' opinions about fundamental research objectives and topics in economics. Almost 10,000 economic researchers from all fields and ranks of the profession participate in our global survey. Detailed bibliometric data allow us to compare our sample to the population of economic scholars who publish in English and post-stratification weights ensure that our sample represents this population.

Our results reveal a strong degree of heterogeneity in economists' views and preferences regarding research objectives and topics. Most researchers disagree with the current state of economic research, including many of the field's most successful scholars. Respondents think that economic research should become more policy-relevant, multidisciplinary, risky and disruptive, and pursue more diverse topics. We also find that dissent with economics' research objectives and topics is associated with lower job satisfaction and is higher among female economists.

Our results serve as a reminder that our views about research questions, objectives, or topics are valuable and irreplaceable, but also inherently subjective. They further suggest that as a field we currently do not appreciate and work on what we collectively prefer. Since the choice of research questions and research objectives is arguably among the most important choices that a researcher makes, we hope that our results will contribute to an inclusive and open-minded debate about "what's worth knowing".

References

Akerlof, George A. 2020. "Sins of Omission and the Practice of Economics." *Journal of Economic Literature* 58 (2): 405–18. [280, 282, 300, 302]

Allgood, Sam, Lee Badgett, Amanda Bayer, Marianne Bertrand, Sandra E. Black, Nick Bloom, and Lisa D. Cook. 2019. "AEA Professional Climate Survey: Final Report." [280]

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. 2021. "Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples." *Working Paper*, [280, 291]

Angrist, Josh, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu. 2020. "Inside Job or Deep Impact? Extramural Citations and the Influence of Economic Scholarship." *Journal of Economic Literature* 58 (1): 3–52. [280, 302]

Angrist, Joshua, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu. 2017. "Economic Research Evolves: Fields and Styles." *American Economic Review: Papers & Proceedings* 107 (5): 293–97. [302]

Avilova, Tatyana, and Claudia Goldin. 2018. "What Can UWE Do for Economics?" *AEA Papers and Proceedings* 108: 186–90. [280, 303]

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin. 2019. "Does Science Advance One Funeral at a Time?" *American Economic Review* 109 (8): 2889–920. [300]

Backhouse, Roger E., and Béatrice Cherrier. 2017. "The Age of the Applied Economist: The Transformation of Economics since the 1970s." *History of Political Economy* 49: 1–33. [302]

Bayer, Amanda, and Cecilia Elena Rouse. 2016. "Diversity in the economics profession: A new attack on an old problem." *Journal of Economic Perspectives* 30 (4): 221–42. [280, 303]

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300. [294]

Bolotnyy, Valentin, Matthew Basilico, and Paul Barreira. 2021. "Graduate Student Mental Health: Lessons from American Economics Departments." *Journal of Economic Literature*, (forthcoming):

Bourdieu, Pierre. 1975. "The specifity of the scientific field and the social conditions of the progress of reason." *Social Science Information* 14 (6): 19–47. [300]

Bowles, Samuel, and Wendy Carlin. 2020. "What Students Learn in Economics 101: Time for a Change." *Journal of Economic Literature* 58 (1): 176–214. [280]

Bramoullé, Yann, and Gilles Saint-Paul. 2010. "Research cycles." *Journal of Economic Theory* 145 (5): 1890–920. [300]

Buckles, Kasey. 2019. "Fixing the Leaky Pipeline: Strategies for Making Economics Work for Women at Every Stage." *Journal of Economic Perspectives* 33 (1): 43–60. [280, 303]

Buyalskaya, Anastasia, Marcos Gallo, and Colin F. Camerer. 2021. "The golden age of social science." *Proceedings of the National Academy of Sciences of the United States of America* 118 (5): e2002923118. [302]

Card, David, and Stefano DellaVigna. 2013. "Nine Facts about Top Journals in Economics." *Journal of Economic Literature* 51 (1): 144–61. [280]

Card, David, and Stefano DellaVigna. 2020. "What do Editors Maximize? Evidence from Four Economics Journals." *Review of Economics and Statistics* 102 (1): 195–217. [280]

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. 2020. "Are Referees and Editors in Economics Gender Neutral?" *Quarterly Journal of Economics* 135 (1): 269–327. [280]

Colander, David. 2005. "The Making of an Economist Redux." *Journal of Economic Perspectives* 19 (1): 175–98. [280]

Colander, David. 2011. "How Economists Got It Wrong: A Nuanced Account." *Critical Review* 23 (1-2): 1–27. [280]

Currie, Janet, Henrik Kleven, and Esmée Zwiers. 2020. "Technology and Big Data Are Changing Economics: Mining Text to Track Methods." *AEA Papers and Proceedings* 110: 42–48. [280, 302]

Dupas, Pascaline, Alicia Sasser Modestino, Muriel Niederle, Justin Wolfers, and The Seminar Dynamics Collective. 2021. "Gender and the Dynamics of Economics Seminars." *Working Paper*, [280, 305]

Foster, Jacob G., Andrey Rzhetsky, and James A. Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review* 80 (5): 875–908. [282]

Fourcade, Marion, Etienne Ollion, and Yann Algan. 2015. "The Superiority of Economists." *Journal of Economic Perspectives* 29 (1): 89–114. [282]

Frey, Bruno S. 2009. "Economists in the PITS?" *International Review of Economics* 56: 335–46. [300]

Frey, Bruno S., Silke Humbert, and Friedrich Schneider. 2010. "What is economics? Attitudes and views of German economists." *Journal of Economic Methodology* 17 (3): 317–32. [280]

Friedman, Milton. 1953. "The Methodology of Positive Economics." In *Essays in Positive Economics*. Chicago: The University of Chicago Press. [282]

Goldin, Claudia, and Lawrence F. Katz. 2020. "The Incubator of Human Capital: The NBER and the Rise of the Human Capital Paradigm." *Working Paper*, [280]

Hamermesh, Daniel S. 2013. "Six Decades of Top Economics Publishing: Who and How?" *Journal of Economic Literature* 51 (1): 162–72. [302]

Hamermesh, Daniel S. 2018. "Citations In Economics: Measurement, Uses, and Impacts." *Journal of Economic Literature* 56 (1): 115–56. [280]

Hausman, Daniel M. 2008. "Why look under the hood?" In *The Philosophy of Economics: An Anthology*. 3rd edition. Cambridge University Press. [282]

Heckman, James J., and Sidarth Moktan. 2020. "Publishing and Promotion in Economics: The Tyranny of the Top Five." *Journal of Economic Literature* 58 (2): 419–70. [280, 287, 300]

Kleven, Henrik. 2018. "Language Trends in Public Econmics." *Working Paper*, [280]

Krugman, Paul. 2009. "How Did Economists Get It So Wrong?" In *New York Times (09/06/2009)*. [280]

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press. [282]

Lundberg, Shelly, editor. 2020. *Women in Economics*. London: CEPR Press. [280, 303, 305]

Lundberg, Shelly, and Jenna Stearns. 2019. "Women in Economics: Stalled Progress." *Journal of Economic Perspectives* 33 (1): 3–22. [280, 303]

Osterloh, Margit, and Bruno S. Frey. 2020. "How to avoid borrowed plumes in academia." *Research Policy* 49 (1): 103831. [280]

Oswald, Andrew, and Nicholas Stern. 2019. "Why does the economics of climate change matter so much – and why has the engagement of economists been so weak?" *Royal Economic Society Newsletter* October: [280, 298]

Pasek, Josh, Matthew Debell, and Jon A. Krosnick. 2014. "Standardizing and Democratizing Survey Weights: The ANES Weighting System and anesrake." *Working Paper*, [290, 319]

Reif, F. 1961. "The Competitive World of the Pure Scientist." *Science* 134 (3494): 1957–62. [300]

Romer, Paul M. 2015. "Mathiness in the Theory of Economic Growth." *American Economic Review: Papers & Proceedings* 105 (5): 89–93. [280]

Santamaría, Lucía, and Helena Mihaljević. 2018. "Comparison and benchmark of name-to-gender inference services." *PeerJ Computer Science* 2018 (7): 1–29. [286, 315]

Sapienza, Paola, and Luigi Zingales. 2013. "Economic Experts versus Average Americans." *American Economic Review* 103 (3): 636–42. [280, 291]

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." *Journal of Political Economy* 129 (1): 101–47. [280]

Shiller, Robert J., and Virginia M. Shiller. 2011. "Economists as Worldly Philosophers." *American Economic Review: Papers & Proceedings* 101 (3): 171–75. [280, 282]

Sunstein, Cass R. 2001. "On Academic Fads and Fashions." *Michigan Law Review* 99 (6): 1251–64. [300]

Swanson, Nicholas, Garret Christensen, Rebecca Littman, David Birke, Edward Miguel, Elizabeth Levy Paluck, and Zenan Wang. 2020. "Research Transparency Is on the Rise in Economics." *AEA Papers and Proceedings* 110: 61–65. [280]

Weber, Max. 1919. "Wissenschaft als Beruf." In *Geistige Arbeit als Beruf*. München, and Leipzig: Duncker & Humblot. [300]

Weber, Max. 1946. "Science as a Vocation." In *From Max Weber: Essays in Sociology*. Edited by H. H. Gerth and C. W. Mills. New York: Oxford University Press.

Wu, Alice H. 2020. "Gender Bias in Rumors among Professionals: An Identity-Based Interpretation." *Review of Economics and Statistics* 102 (5): 867–80. [305]

Appendix 4.A Instructions of main questions

This appendix provides extracts from the two main modules of the survey. The full survey is available at https://osf.io/xwbdf/.

4.A.1 Research objectives

Introductory instructions for a respondent who selected the field D Empirical Microeconomics

How should economists do research?

In the first part of the survey, we would like you to think about how economics as a research field should do research these days.

Please note: This part is *not* about how *you* personally should do research nor about how the field *actually does* research. Instead, we would like you to take a normative perspective and indicate how economic researchers should do research in general.

Please state your normative view about the optimal approach to economic research. You will face ten questions that describe trade-offs between different research strategies or styles. Of course, these trade-offs are sometimes more and sometimes less severe, but in many cases economics can have more of one research style only at the expense of the other.

Your task is to indicate whether you think that the field's current way of doing research is appropriate or whether you think that the field should place more weight on one research style versus the other.

The overarching question is: **In comparison with how research in economics is currently conducted, how *should* economists conduct research?**

Please give separate responses for

1. your primary research field: D Empirical Microeconomics and
2. the discipline of economics as a whole.

Exemplary layout for research objective question "policy relevance vs. intrinsic interest".

Policy relevance versus intrinsic/intellectual interest?
Policy relevance: Research informs policy, with an impact on societal well-being.
Intrinsic and intellectual interest: Research is intrinsically rewarding to the researcher who conducts the project due to his/her own curiosity and interest.

	Your primary JEL field*	Economics as a whole
Much more policy relevance	○	○
Moderately more policy relevance	○	○
Slightly more policy relevance	○	○
Current state is about right	○	○
Slightly more intrinsic/intellectual interest	○	○
Moderately more intrinsic/intellectual interest	○	○
Much more intrinsic/intellectual interest	○	○

*Your primary research field: D Empirical Microeconomics.

Response scale
Participants respond on a seven-point scale. Each scale is centered around the option "Current state is about right". The other response options express dissatisfaction with the status quo and place increasing weight on one research objective versus the other. For instance, the question on *Basic research* versus *Policy relevance* has the response options "Much more", "Moderately more", and "Slightly more" policy relevance, "Current state is about right", as well as "Slightly more", "Moderately more", and "Much more" basic research. The question on specialization comes with the response options "Much less", "Moderately less", and "Slightly less" specialization, 'Current state is about right", as well as "Slightly more", "Moderately more", and "Much more" specialization.

Wording of all research objective questions in original order

Less versus more specialization?
Specialization is defined as the extent to which each individual researcher focuses solely on one specific topic.

Less versus more risky research?
Some research projects are "safe bets" with a very foreseeable impact. Other research projects are of high risk with very uncertain impact. A higher risk may come with a higher expected impact.

More incremental versus more disruptive research?
Incremental: A research project that builds on and connects closely to the existing literature.
Disruptive: A research project that extends considerably beyond the existing literature and proposes new approaches.

Less versus more multidisciplinary research?
Multidisciplinary research incorporates insights from other disciplines than economics to study economic questions.

Quantity of papers versus quality of papers?
More papers of lower quality or fewer papers of higher quality?

Policy relevance versus intrinsic/intellectual interest?
Policy relevance: Research informs policy, with an impact on societal well-being.
Intrinsic and intellectual interest: Research is intrinsically rewarding to the researcher who conducts the project due to his/her own curiosity and interest.

Policy relevance versus basic research?
Policy relevance: Research informs policy, with an impact on societal well-being.
Basic research: Research deals with fundamental and basic phenomena, laying the ground for more applied research. It has no immediate policy relevance.

For empirical work: Causal identification versus importance of research question
Identification: Research identifies the phenomenon of interest credibly and causally, above and beyond establishing correlational patterns.
Importance: Research question is of general interest and/or has societal relevance.

For theoretical work: More pure theory versus more applied and evidence-related theory?

Pure theory: Studies general theoretical principles.
Applied and evidence-related theory: Studies an empirically-observed phenomenon theoretically. Organizes empirical evidence, matches its facts, and/or provides testable predictions.

For applied theoretical work: More emphasis on prediction versus explanation?
How should economists evaluate applied theoretical models?

- More focus on *predicting* outcomes.
- More focus on *explaining* outcomes (using plausible assumptions and plausible theoretical mechanisms).

4.A.2 JEL topics

Which topics should economists work on?

From our experience, this question will take you at the very least 1 minute to answer. It is the main question of this survey.

Please state your normative view about the optimal composition of research topics on which economists should be working.

For this matter, suppose you are endowed with **100 points, representing the total number of published research articles** produced by all economists in a given year. Hence, each point corresponds to 1% of the total research output.

Please allocate these 100 points between the nineteen research topic categories defined by the Journal of Economic Literature (JEL) classification system. The more points you allocate to a specific JEL category, the higher the published output concerning topics in this category should be.

For your convenience, you can click on each JEL code for further information on the JEL sub-categories.

Please allocate the 100 points across the categories on which you think economists should work and publish these days.

▸ **A** General Economics and Teaching	0
▸ **B** History of Economic Thought, Methodology, and Heterodox Approaches	0
▸ **C** Mathematical and Quantitative Methods	0
▸ **D** Microeconomics	0
▸ **E** Macroeconomics and Monetary Economics	0

List of JEL topics continues.

Appendix 4.B Publication and author data

4.B.1 Derivation of the publication data

This section documents step by step how the publication database is derived. We start from the EconLit publication database which we downloaded on the 4th of December 2019. We consider all publications in the 400 EconLit-indexed journals with the highest impact factor according to Scopus's 2018 Scimago Journal Ranking in the "Economics, Econometrics, and Finance" category. We restrict our attention to publications since 2009. Additionally, we impose the following restrictions:

1. Articles have English full text.
2. Information on authors is available.
3. To ensure that only genuine research articles are included in the final sample:
   - We concentrate only on articles that are classified as journal articles by EconLit.
   - We delete articles that have been assigned to the JEL category Y which includes book reviews, memorials, or other ancillary content.
   - Moreover, we exclude publications that contain keywords such as "erratum", "reply to", or "memorial" that were chosen to identify the most common ancillary publications. The full list of keywords is available upon request.
   - Finally, we exclude all articles with titles that appear more than twice in the database – an indicator for multiple comments on another research article, editorials, or other repeated ancillary publications.
4. Non-duplicate articles.

   To exclude duplicates, we keep only the first article with duplicated titles within each journal. If the title has no abstract information (an indicator for ancillary publications), we drop all within-journal duplicated titles.
5. Can be matched to a Scopus article.

   97.4% of all articles that satisfy the above conditions can be matched to a Scopus article.[19] The details of the matching algorithm are available upon request. The Scopus data were downloaded from Scopus API between December 5 and 12, 2019 via http://api.elsevier.com and http://www.scopus.com.

4.B.2 JEL code metrics

The EconLit data assign each article to one or (typically) more JEL codes. This section explains how we translate the three-digit JEL codes into primary JEL topics. We use four different metrics to describe the JEL topics of a paper. We use the *Weight*

19. A similar set of restrictions was applied to the Scopus data.

metric in our main specifications and run robustness checks with the three alternative metrics.

Example: Throughout this subsection, we consider an article with JEL Codes E21, E32, F34, and G51. Thus, the article has two codes in field E, one code in field F, and one code in field G.

Weight. An article's topic weight is the share of its JEL codes that belong to this topic. The above example article would be classified as E: 50%, F: 25%, G: 25%, all other fields: 0%. Each article has a total weight of 100%.

Indicator. An article's topic indicator is 1 if at least one JEL code belongs to the topic and zero otherwise. The above example article would be classified as E: 1, F: 1, G: 1, all other fields: 0.

Sum. An article's topic sum is the number of JEL codes that belong to the topic. The above example article would be classified as E: 2, F: 1, G: 1, all other fields: 0.

Primary. An article's primary topic is the JEL topic with the largest count of codes (see "Sum" above). This means that an article with a unique most frequent topic is fully (100%) assigned to this topic. If the maximum is not unique, which happens for about 3 out of 10 articles, we split the shares equally across the most frequent topics (e.g., 50%-50% if there are two most frequent topics). The above example article would be classified as E: 100%, all other fields: 0%.

Thus, the JEL code metrics differ in two respects: Whether they are sensitive to multiple JEL codes in a topic (*Weights*, *Sum* are, *Indicator* is not, *Primary* is an intermediate case) and whether each paper has the same total weight (this is only the case for *Weights* and *Primary*). In our main analysis, we use the *Weights* metric because we want to give equal total weight to each paper and view the occurrence of multiple JEL codes in one field as evidence that this topic is covered more extensively.

4.B.3 Author data: Covariates

This section summarizes and defines all author covariates that will be used throughout the paper.

Covariates derived from the publication data

Female. The gender of an author is estimated from their first and last name, using the commercial Gender API algorithm (see Santamaría and Mihaljević, 2018). The author names are taken from the Scopus publication data. The algorithm produces missing values for 2.4% of the study population. *Female* is a binary indicator that takes the value 1 if a respondent's name is classified as female.

Year of first publication. The Scopus author data contains the year of the author's first publication.

Number of articles (in sample). The number of articles in our publication sample that can be assigned to an author.

Number of articles (overall). The total number of journal publications that Scopus attributes to an author, capped at 200. This includes articles outside our publication sample, in particular articles that were published before 2009 or outside the top 400 EconLit-indexed journals.

Share of publications in economic journals. The share of an author's journal publications (see "Number of articles (overall)") that are published in a journal of Scopus's "Economics, Econometrics, or Finance" category.

Co-author network. The undirected, unweighted co-author network constructed from all co-author relationships observed in our publication sample. The network includes all authors, even those who are not part of the study population.

Degree (number of co-authors). The number of unique co-authors of an author in our publication sample.

Eigenvector centrality (index). An index of an author's eigenvector centrality in the co-author network. The index measures which share of authors has a lower eigenvector centrality. For instance, an index value of 70% means that the author's eigenvector centrality is larger than the centrality of 70% of all authors in the network.

Number of co-authors with Top Five publication. The number of co-authors of the author who have published at least one article in a Top Five journal in our publication sample (also see "Published in Top Five Journal").

Top 50 institution. A binary indicator that takes the value 1 if an author is affiliated with a top 50 research institution in economics. We derive the indicator from the Scopus author data which contain information about the institution with which the author was affiliated in their last publications. We match the institution names to the Shanghai Academic Ranking of World Universities in Economics 2020.

h-index. h-index, derived from the Scopus citation data of *all* publications of an author (as of December 2019, see "Number of articles (overall)").

Published in Top Five Journal (in sample). A binary indicator that takes the value 1 if the author published at least one article in a Top Five journal within our publication sample. The Top Five journals are the American Economic Review, The Quarterly Journal of Economics, the Journal of Political Economy, the Review of Economic Studies, and Econometrica. Publications in the Papers & Proceedings of the American Economic Review are not counted as Top Five publication.

Number of Top Five publications (in sample). The number of Top Five publications (see above) that an author published within our publication sample.

Average journal rank 1-400 (in sample). The average journal rank of an author's publications in our publication sample. The journal ranks range from 1-400. The journals are ranked according to the Scopus 2018 Scimago impact factor in the "Economics, Econometrics, and Finance" category. Higher ranked journals (numerically they have a lower rank) have a higher journal impact factor.

Continent. The Scopus author data contain information about the institution with which the author was affiliated in their last publications, including the country of the institution, which is available for 99.5% of the authors in the study population. This allows us to deduce the last known continent of residence of a researcher.

Share of publications in JEL topics. The average JEL topic weight of an author's articles (see appendix section 4.B.2).

Note: The author-average share of publications in a topic may differ from the paper-average share of publications in a topic. The author-average assigns equal weight to each author (irrespective of their number of publications), while the paper-average assigns equal weight to each paper. Therefore, we use author-averages only when we study heterogeneity in authors' survey responses. In contrast, when we analyze the field-wide distribution of JEL topics, we use paper-averages.

Covariates in the Ph.D. student sample

For the population of Ph.D. students, we only have data on their gender (derived as above), their continent of residence (derived as above), and the rank of their institution. An institution's rank is derived from the number of total citations that authors from the institution receive for articles that are in our publication sample.

Covariates derived from the survey data

Female. We also measure the gender of respondents in the survey. We use this more accurate measure in the heterogeneity analysis of survey responses.

Age. The age of respondents. Continuous.

Ph.D. student. An binary indicator that takes the value 1 if the respondent says they are a (doctoral) student.

Tenured. We ask respondents who are active in academic research whether they have tenure. *Tenured* is a binary indicator variable that takes the value 1 if the respondent says they have tenure.

Research type: theory/empirics/methods. We ask respondents which fraction of their research is predominantly theoretical, predominantly empirical, and predominantly methods research.

Appendix 4.C Sample

4.C.1 Weighting procedure

We follow Pasek, Debell, and Krosnick (2014) and use the R package anesrake to derive weights for the author sample. We target the following marginal distributions of the study population.

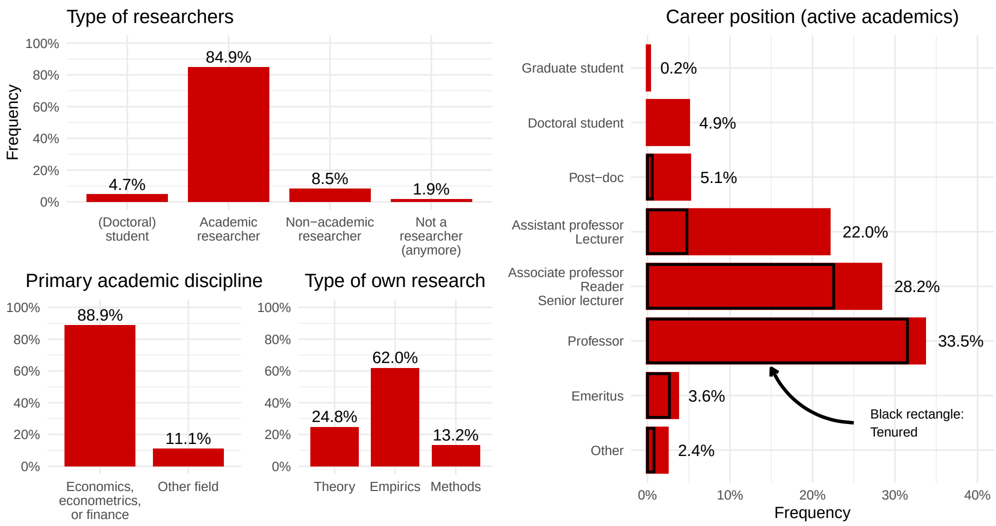
1. Gender: female versus male or missing (2 groups)
2. Year of first publication (quartiles)
3. Number of papers in our publication sample (quartiles)
4. h-index (quartiles)
5. Continent (Europe, Northern America, Asia, Other)
6. Main research field
   The main research topic of an author is the JEL field in which they have the highest share of publications. We consider the following six groups: D Microeconomics, E Macroeconomics, G Financial Economics, J Labor Economics, Other, and Multiple. The group "Multiple" contains authors who have multiple JEL fields with a maximal share of publications (e.g. two fields with a share of 50% each).

87% of the weights are between 0.5 and 2. The minimal weight is 0.3, and the maximal weight is 3.59. This indicates that no extreme weights occur.

Alternative weighting schemes. We use the following alternative weighting schemes in robustness checks throughout the paper.

- **Weighted, including no email** The sample is weighted to represent the population of authors which also includes the scholars for whom no email address could be found but who satisfy the other eligibility criteria described in section 4.3.2 of the main text. We use the same weighting approach as outlined above.
- **Unweighted** Identical weight (1) for all participating authors. This approach includes also the few respondents who started but did not complete the survey.
- **Unweighted, only economics** Identical weight (1) for all participating authors who say that their primary academic discipline is economics, econometrics, or finance.
- **Unweighted, with Ph.D.** Identical weight (1) for all participants, including participants from the Ph.D. student sample.

4.C.2 Characteristics of the main sample



Notes: Weighted survey responses.

Figure 4.C.1. Demographic characteristics of the weighted sample

4.C.3 Characteristics of the student sample

Table 4.C.1 presents the distribution of demographic characteristics in the population of invited Ph.D. students and the sample of participating students. See appendix section 4.B.3 for a description of the covariates.

Table 4.C.1. Characteristics of the population and the sample of Ph.D. students

Variable	**Population**	**Sample**
Female	30.8%	28.8%
Region: Europe	34.2%	50.3%
Region: Northern America	62.5%	46.5%
Region: Asia	2.1%	2.5%
Region: Australia and New Zealand	1.2%	0.7%
Rank of institution	124.8	126.0
Sample size	9441	1765

4.C.4 Selection into invitation and selection into completion

Table 4.C.2 summarizes and compares the characteristics of five different groups.

1. **Incl. no email**: The population of active economic researchers plus those for whom no email address could be found.

2. **Population**: The main study population.
3. **Participated**: The unweighted sample of participating authors, including those who do not complete the survey.
4. **Unweighted sample**: The unweighted main sample.
5. **Weighted sample**: The weighted main sample.

Columns 2, 4, and 5 equal columns 1 to 3 in table 4.3.1. Table 4.C.2 reveals that there are only a few differences between the main study population (column 2) and the population which also includes authors without email data (column 1). It also shows that the differences between the sample of participating authors (column 3) and the sample of authors who complete the survey (column 4) are negligible.

Table 4.C.2. Characteristics of economic researchers: From the email address collection to study completion

Variable	(1) Incl. no email	(2) Study population	(3) Partici-pated	(4) Unwgt. sample	(5) Weighted sample
Gender, academic age					
Female	27%	26%	23.3%	23.1%	25.8%
Year of first publication (YYYY/MM)	2008/01	2007/01	2006/01	2006/01	2006/1
Number of papers					
Number of articles (in pub. sample)	4.4	4.8	5.7	5.6	4.9
Number of articles (overall)	15.3	17.1	18.4	18.3	16.2
Share of art. in econ. journals	77.6%	75.9%	76%	76.2%	76.8%
Co-author network (in pub. sample)					
Degree (number of unique co-authors)	5.4	5.8	6.5	6.5	5.7
Eigenvector centrality (index)	59.3%	61.1%	65.6%	65.6%	62.2%
Number of co-authors with Top Five pub.	0.4	0.5	0.8	0.8	0.5
Success					
Top 50 institution	11.2%	12.1%	12.3%	12.2%	12.5%
Published in Top Five Journal (in pub. sample)	5.1%	6.1%	9.2%	9.3%	6.1%
Number of Top Five publications (in pub. sample)	0.10	0.12	0.17	0.18	0.11
Average journal rank 1-400 (in pub. sample)	170.8	164.2	161.6	161.9	165.8
h-index	5.8	6.5	6.8	6.8	6.1
Continent					
Europe	38.8%	40.4%	53.3%	53.6%	40.5%
Northern America	31.6%	33.9%	24.6%	24.2%	33.9%
Asia	20.6%	17.1%	13.4%	13.4%	17.2%
Australia and New Zealand	4.2%	4.3%	3.6%	3.7%	3.3%
Latin America	3%	2.7%	3.4%	3.4%	3.3%
Africa	1.9%	1.6%	1.7%	1.7%	1.8%
Share of publications in JEL fields					
C Mathematical and Quantitative Methods	6.1%	6.1%	6.4%	6.3%	5.8%
D Microeconomics	12.6%	13.1%	16%	16.1%	13.5%
E Macroeconomics and Monetary Economics	7.3%	7.3%	7.4%	7.4%	7.1%
F International Economics	4.4%	4.4%	4.3%	4.3%	4.2%
G Financial Economics	18.4%	18.2%	11.5%	11.3%	16.9%
H Public Economics	3.5%	3.6%	4.3%	4.3%	3.8%
J Labor and Demographic Economics	6.3%	6.7%	9.7%	9.8%	7.5%
L Industrial organization	8.4%	8.3%	7.4%	7.4%	8%
O Growth and Development Economics	9.1%	8.5%	8.8%	8.8%	9.2%
Q Agricultural and Environmental Economics	7.4%	7.1%	7.4%	7.4%	7.4%
Other fields	16.5%	16.6%	16.9%	16.9%	16.6%
Sample size	67,546	53,777	8,156	7,794	7,794

Notes: Overview of covariates. Column 1: The population of researchers before authors are excluded for whom no email address could be found. That is, all authors who satisfy restrictions 1 to 3 (see main text, section 4.3.2. Column 2: The eligible study population. Column 3: All respondents who participated in the survey, including those who did not complete it. Column 4: Respondents of the main sample, unweighted. Column 5: Weighted main sample. For a description of the covariates in the different rows see main text or appendix section 4.B.3.

Appendix 4.D Supplementary tables and figures

4.D.1 Research objectives

Aggregate results, statistical tests. Table 4.D.1 reports the majority shares of respondents who directionally agree on which research objective economics should place more weight on and tests whether these shares differ from 50%. It also reports the average response (in scale points) for each question and tests whether the means differ from the neutral "About right" category.

Aggregate results, robustness to different weighting schemes. Figure 4.D.1 shows that we obtain virtually identical results if we recalculate the distribution of survey responses with the different weighting schemes and sub-samples that are described in appendix section 4.C.1.

Responses of Ph.D. students. Figure 4.D.2 compares the average responses in the main sample with the responses in the sample of Ph.D. students.

Field-specific responses. Figure 4.D.3 compares the distribution of responses for economics as a whole and the respondents' own primary JEL field. It documents largely identical results. Appendix figure 4.D.4 disaggregates the field-specific responses and diagnoses similar trends in almost all fields.

Table 4.D.1. Majority shares and avg. responses to research objectives questions

(A) Share of respondents holding majority opinion, questions 1-5

	Pol. relev. (vs. intrin. interest)	Pol. relev. (vs. basic research)	Importance (vs. causal ident.)	Applied theory (vs. pure)	Less specialization
	(1)	(2)	(3)	(4)	(5)
Fraction "more"	0.540	0.516	0.559	0.609	0.604
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
p: fraction=0.50	<0.001	0.062	<0.001	<0.001	<0.001
Observations	4,028	4,018	4,008	4,009	4,030

(B) Share of respondents holding majority opinion, questions 6-10

	More multidisciplinarity	More risky research	Disruptive research (vs. incremental)	Quality (vs. quantity)	Explanation (vs. prediction)
	(1)	(2)	(3)	(4)	(5)
Fraction "more"	0.787	0.735	0.674	0.657	0.469
	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)
p: fraction=0.50	<0.001	<0.001	<0.001	<0.001	<0.001
Observations	4,034	4,022	4,022	4,022	3,993

(C) Average response (in scale points -3 to 3, mid-point: 0), questions 1-5

	Pol. relev. (vs. intrin. interest)	Pol. relev. (vs. basic research)	Importance (vs. causal ident.)	Applied theory (vs. pure)	Less specialization
	(1)	(2)	(3)	(4)	(5)
Mean response	0.621	0.526	0.591	0.920	0.848
	(0.030)	(0.028)	(0.031)	(0.027)	(0.027)
p: mean=0	<0.001	<0.001	<0.001	<0.001	<0.001
Observations	4,028	4,018	4,008	4,009	4,030

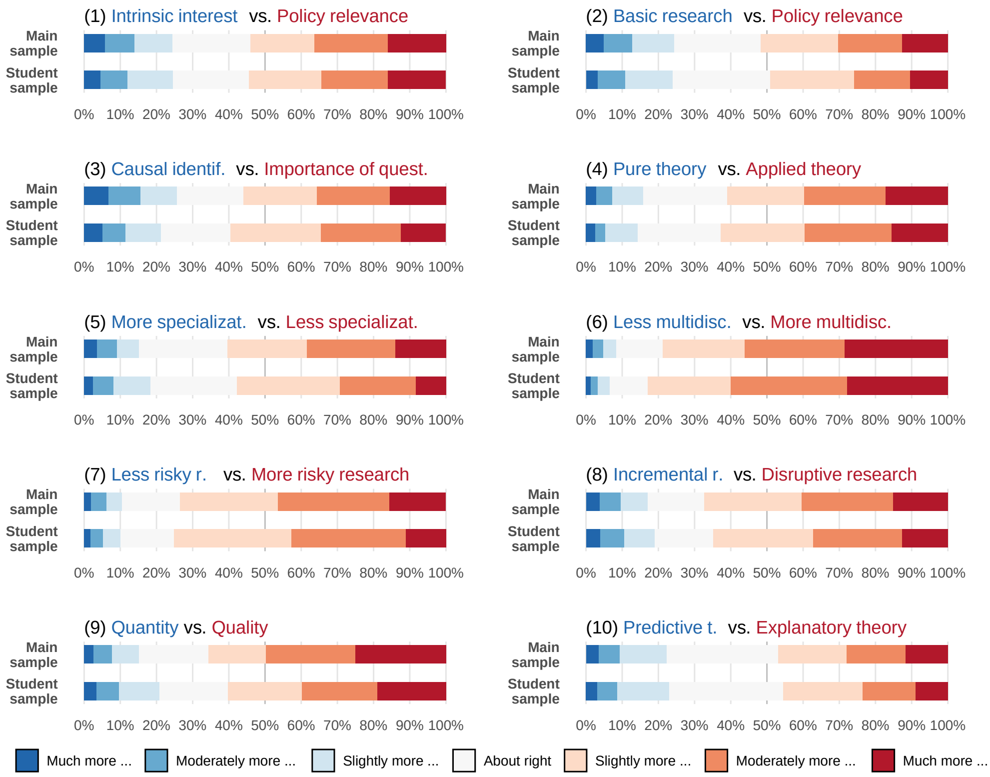
(D) Average response (in scale points -3 to 3, mid-point: 0), Questions 6-10

	More multidisciplinarity	More risky research	Disruptive research (vs. incremental)	Quality (vs. quantity)	Explanation (vs. prediction)
	(1)	(2)	(3)	(4)	(5)
Mean response	1.484	1.170	0.923	1.150	0.512
	(0.025)	(0.025)	(0.028)	(0.028)	(0.027)
p: mean=0	<0.001	<0.001	<0.001	<0.001	<0.001
Observations	4,034	4,022	4,022	4,022	3,993

Notes: Results are based on weighted OLS regressions on a constant (i.e. estimates of averages), robust standard errors in parentheses. The dependent variables are responses to the ten research objective questions. In panels (A) and (B), the independent variable is a binary indicator for endorsing the majority opinion summarized in the column titles ("Slightly more ...", "Moderately more ...", or "Much more ..." of the research objective stated in the column title). Estimates thus report the share of respondents who endorse the majority opinion. Panels (C) and (D) report the average response in scale points (scale ranges from -3 to 3, mid-point: 0). p-values are reported in the second row of each table and adjusted for multiple hypothesis testing within panels (A) and (B) (10 tests) as well as (C) and (D) (10 tests) respectively, using the Benjamini-Hochberg procedure. All tests are two-sided.

Notes: Survey responses to the ten research objectives questions. Different weighting schemes and samples are employed. *Main*: Main weighted survey sample. The other weighting schemes are described in appendix section. 4.C.1.

Figure 4.D.1. Robustness of responses to the research objectives questions

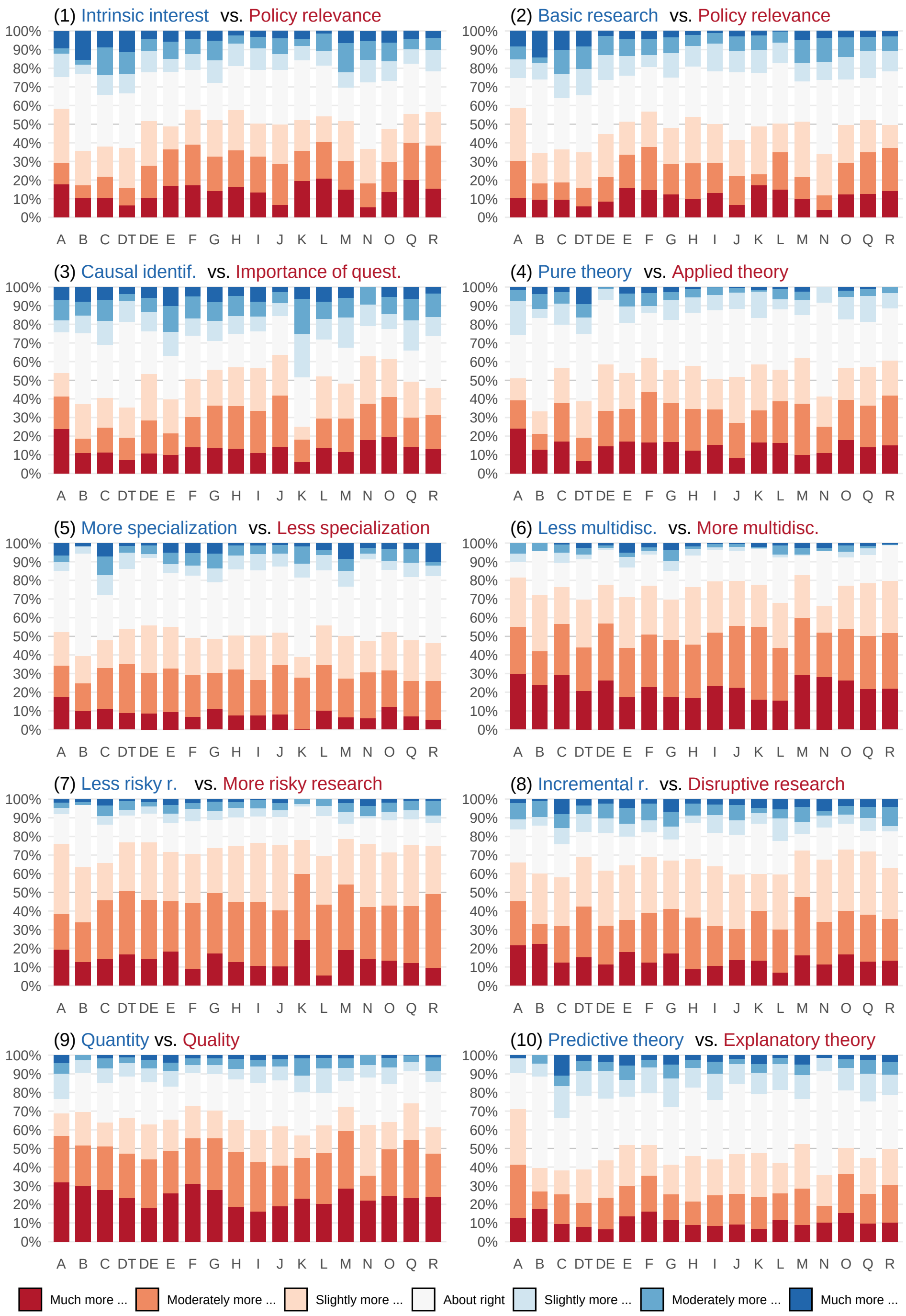


Notes: Survey responses to the ten research objectives questions in the (weighted) main sample and the (unweighted) sample of Ph.D. students.

Figure 4.D.2. Responses to the research objectives questions in the main sample and the Ph.D. student sample

Notes: Weighted distribution of survey responses. The figure compares responses for (i) economics as a whole and (ii) one's own primary JEL field.

Figure 4.D.3. Research objectives for (i) economics as a whole and (ii) one's own primary JEL field.



Notes: Weighted distribution of survey responses. The figure compares the responses for respondents' own primary JEL fields. We distinguish between *DT* Theoretical Microeconomics (including Game Theory) and *DE* Empirical Microeconomics. JEL fields with less than 50 respondents are not shown (P, Z).

Figure 4.D.4. Research objectives for each primary JEL field

4.D.2 JEL topics

Sub-topics. We ask the participants to reconsider three randomly selected topics to which they assigned positive weight and specify the importance of each of its sub-topics. For each JEL topic, respondents can allocate 100 points between its JEL sub-topics which represent published research articles within this field. Figure 4.D.5 compares the distribution of JEL sub-topics in our publication data (blue bars) with the average survey responses (pink bars).[20] On average, respondents prefer a more uniform topic distribution than can be observed in practice.

We replicate this finding if we proportionally adjust the survey weights to the share a respondent gives to the base category (red bars). For instance, if a respondent assigns a share of 15% to JEL topic *D Microeconomics* and a share of 5% to *K Law and Economics*, we multiply his or her survey weight by 0.15 when we derive the average survey responses for D's sub-topics and by 0.05 when we derive the average survey response for K's sub-topics.

These results have to be taken with a grain of salt because only respondents who assigned a positive weight to a primary JEL topic were asked to specify weights for its sub-topics. Moreover, it seems possible that respondents' understanding of the detailed JEL sub-topics does not always align with the EconLit guidelines.

Robustness. Figure 4.D.6 and figure 4.D.7 show that the conclusions from the comparison of the actual JEL topic distribution (blue bars) and average survey responses (red bars) can be replicated in several robustness checks. Specifically, we calculate the **actual JEL topic shares** in the following specifications:

- **Main**: Main estimate as described in main text.
- **JEL: Indicator**: Uses the *Indicator* metric to aggregate the publications' JEL topics (see 4.B.2).
- **JEL: Sum**: Uses the *Sum* metric to aggregate the publications' JEL topics (see 4.B.2).
- **JEL: Primary**: Uses the *Primary* metric to aggregate the publications' JEL topics (see 4.B.2).
- **Top 200**: Considers only publications in the set of top 200 journals.
- **Top 100**: Considers only publications in the set of top 100 journals.
- **Since 2015**: Considers only publications since 2015.
- **Since 2018**: Considers only publications since 2018.

20. Among the respondents who assign a positive weight to a given JEL topic, those who assign positive weights to fewer other topics have a higher chance to be asked about its sub-topics. Their views would be overrepresented if we used our standard survey weights. Here, we therefore adjust these weights for the differential sampling probabilities.

- **Authors**: Considers only publications by authors who are part of the author population, as specified in section 4.3.2 of the main text.

Moreover, we calculate the **average survey response for each JEL topic** for the following robustness specifications which are tailored to exclude possibly careless respondents:

- **Main**: Main estimate as described in the main text.
- **Wgt. no email**: Weighting scheme *Weighted, including no email*. See appendix section 4.C.1 for details about the weighting schemes.
- **Unwgt.**: Identical weight for all participating authors (weighting scheme: *Unweighted*).
- **Unwgt. econ**: Identical weight for all participating authors who say that their primary academic discipline is economics, econometrics, or finance (weighting scheme: *Unweighted, only economics*).
- **Unwgt. w/ Ph.D.**: Identical weight for all participants, including participants from the Ph.D. student sample (weighting scheme: *Unweighted, with Ph.D.*).
- **Robust 1**: Excludes respondents who assign positive weight only to few JEL categories, namely the 25% respondents who assign a positive weight to the fewest JEL topics.
- **Robust 2**: Excludes respondents who assign a very large weight to one category, namely the 25% respondents with the largest maximum assigned share.
- **Robust 3**: Excludes respondents who frequently assign the same share to different categories, namely the 25% respondents with the most duplicate share values.
- **Robust 4**: Excludes respondents who frequently "round" and assign multiples of 5 to the different JEL topics, namely the 25% respondents who use most rounded values.
- **Robust 5**: Excludes respondents with a low response variation, namely the 25% respondents with the lowest standard deviation of JEL shares.
- **Robust 6**: Excludes respondents with a low response duration for the JEL topics questions, namely the 25% respondents with the lowest response duration.

Time trends. Figure 4.D.8 shows the actual topic distribution in economics for each year from 2009 to 2019. The time trends are mostly so minuscule that the mismatch between research output and today's topic preferences is unlikely to dissipate in the future. For JEL codes F, M, and O the mismatch even grew in recent years. For JEL codes I and L, the mismatch became slightly smaller, but it would still take them about a decade to fully disappear if past time trends prove to be persistent.

Responses of Ph.D. students. Figure 4.D.9 compares the average responses in the main sample with the responses in the sample of Ph.D. students.

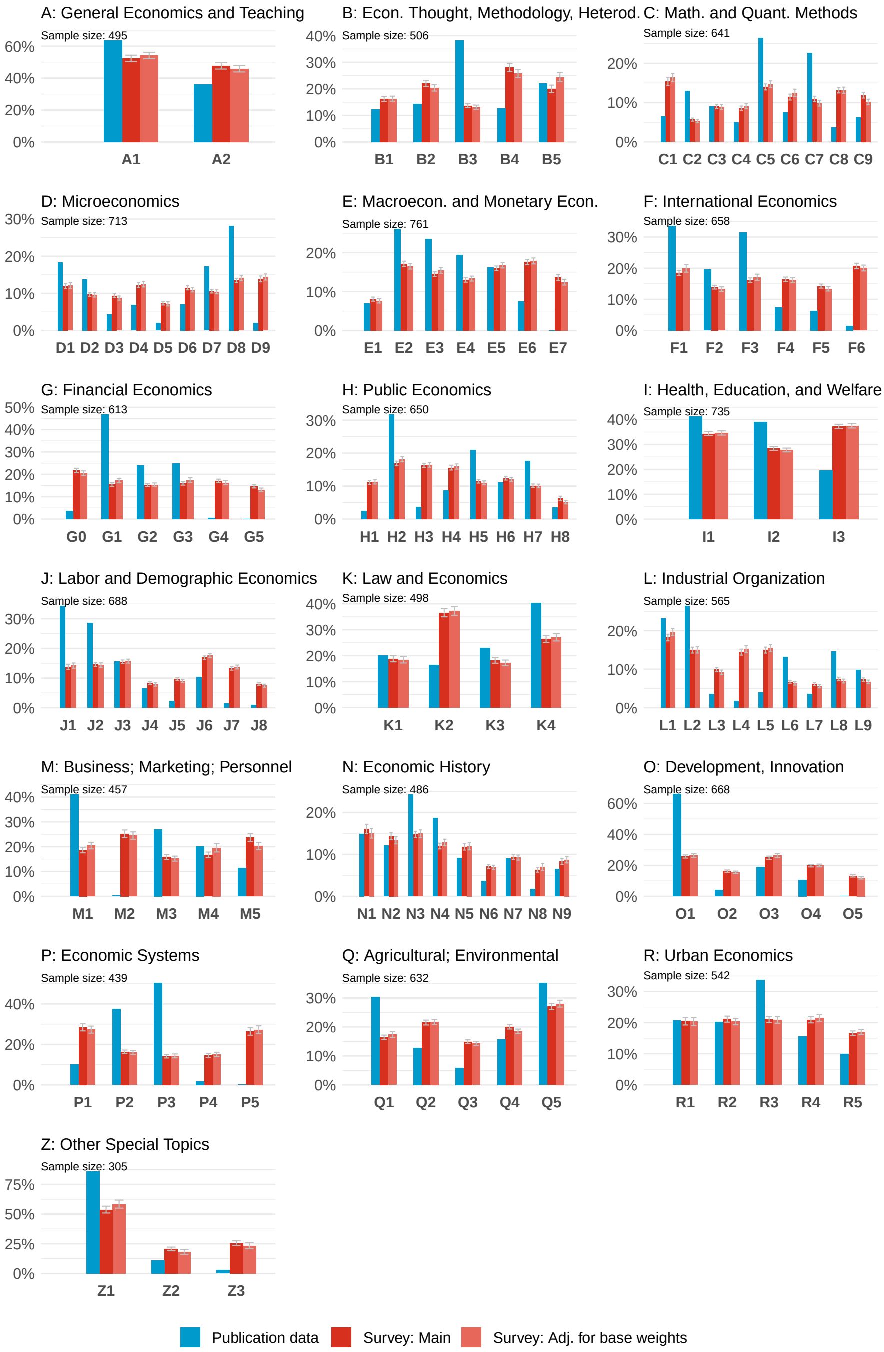
Comparison to Top Five journals. Figure 4.D.10 compares the distribution of JEL topics in Top Five articles of our publication sample (in blue) with the average survey response (in red). The former shows which fraction of papers was published in each JEL topic in a Top Five journal from January 2009 to December 2019. The latter shows economists' opinions on which share of papers should be written and published in each JEL topic. Figure 4.D.11 plots the differences between both distribution (average survey response – actually observed share) for each JEL topic. Again, we can draw the conclusion that the average economist would prefer a more diverse distribution of research topics.

Project types. Figure 4.D.12 plots the average response to the question how economic research should be distributed across three broad project types: theory (formal and informal), empirics, and methods (e.g. econometrics or computational techniques).

Heterogeneity. Figure 4.D.13 plots kernel density estimates of the response distribution for each JEL topic and reveals the large heterogeneity of economists' opinion about the importance of different JEL topics.

Predictors of responses. We explore the heterogeneity of survey responses by regressing the responses on a rich set of variables that cover basic demographic characteristics (gender, age, tenure, region), academic success (affiliation with top 50 institution, Top Five publication, h-index), and the share of theory and methods projects a researcher is working on. We run a separate regression for each JEL topic. We also account for any effect the researchers' own choice of research topics might have and include (but – for the sake of brevity – do not report) the share of publications in each primary JEL topic as well as the share of publications in economics journals (see appendix 4.B.3 for details). We use the Benjamini-Hochberg procedure to correct all reported coefficients for multiple hypotheses testing. Table 4.D.2 summarizes the results. To facilitate orientation, we report only the statistical significance of the coefficients. $+++/---$ indicates a p-value below 0.01, $++/--$ a p-value below 0.05, and $+/-$ a p-value below 0.10 for positive and negative coefficient respectively.

Bias for own research field. Table 4.D.3 shows that the topics of an author's publications strongly predict their perceived importance. We regress the weight assigned to a JEL topic on the share of an author's publications in the topic. This means we regress the weight assigned to *D* on the share of publications in *D* or the weight assigned to *E* on the share of publications in *E*). The dependent variable is the weight assigned to a JEL topic j by respondent i. The predictor is the share of own publications of respondent i in JEL topic j. The underlying data has a panel structure with about 3,600 respondents (dimension 1) and 19 JEL topics (dimension 2). All regressions include topic fixed effects. Respondent fixed effects are not necessary because each respondent's weights sum up to 1, that is, there are no level differences between respondents. We show that the results are robust to including controls (column 2) and different weighting schemes (column 3-5).

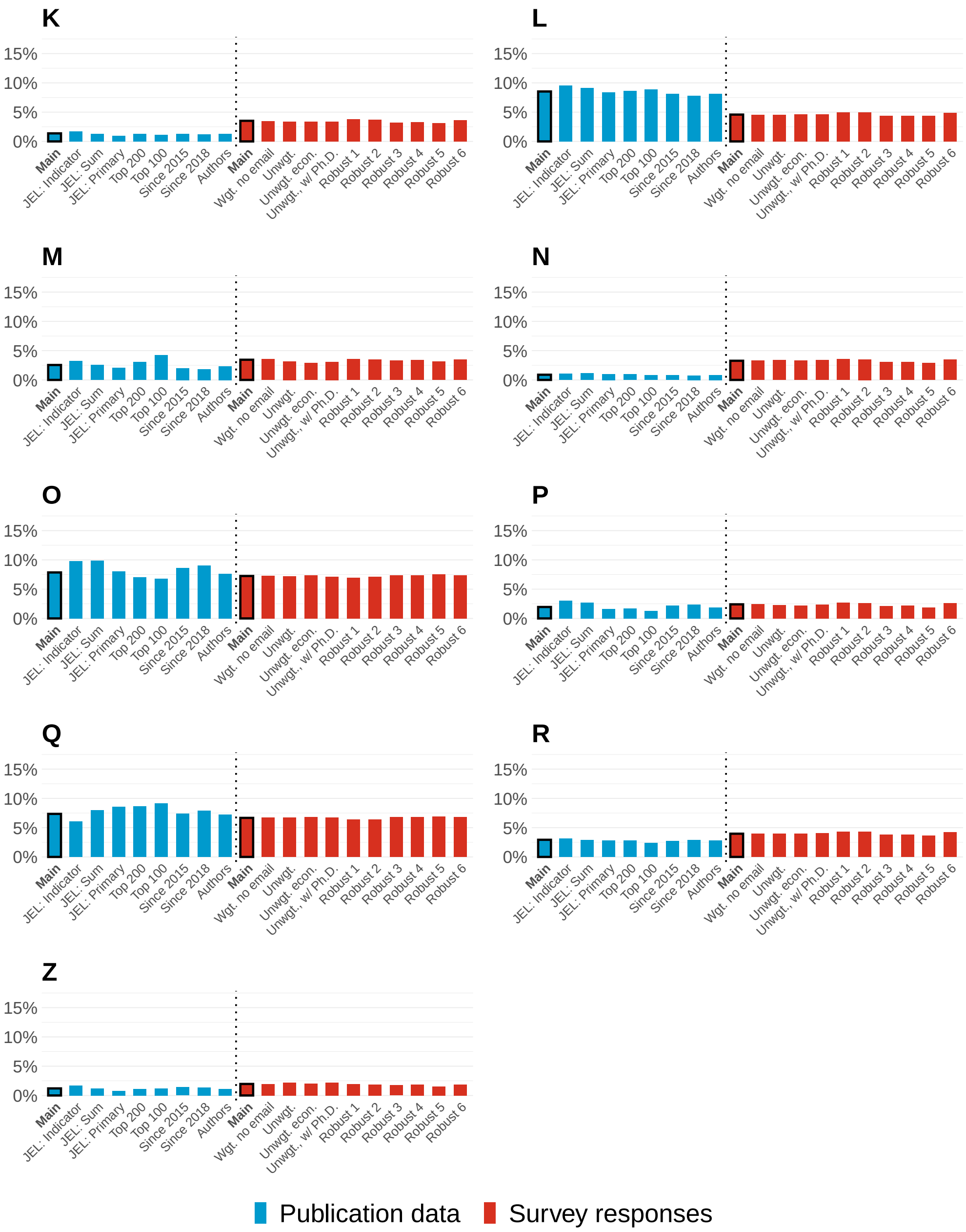


Notes: Blue bars: Share of JEL sub-topics in our publication data (EconLit publication data, top 400 journals, January 2009 - December 2019). Red bars: Weighted average survey response with 95% confidence interval. Orange bars: With weights adjusted for share assigned to main JEL-topic.

Figure 4.D.5. Comparison of actual JEL topic distribution and average survey responses for JEL sub-topics

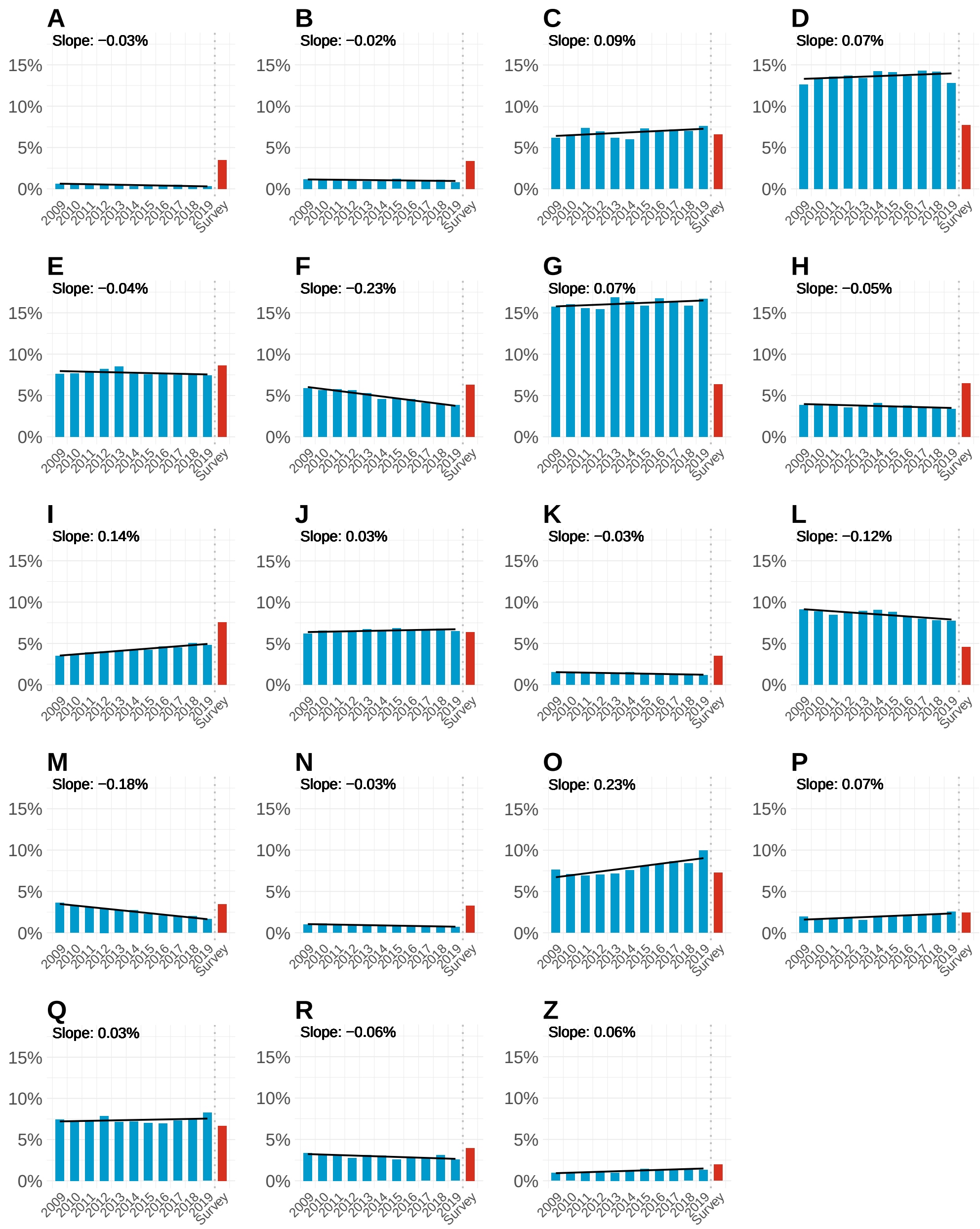
Notes: Black border: Main estimates. Blue bars: Share of JEL topics in our publication data (EconLit publication data, top 400 journals, January 2009 - December 2019). Red bars: Weighted average survey response with 95% confidence interval. Both distributions are calculated in different robustness specifications that are described in the discussion above.

Figure 4.D.6. Robustness of JEL topic distributions – part 1



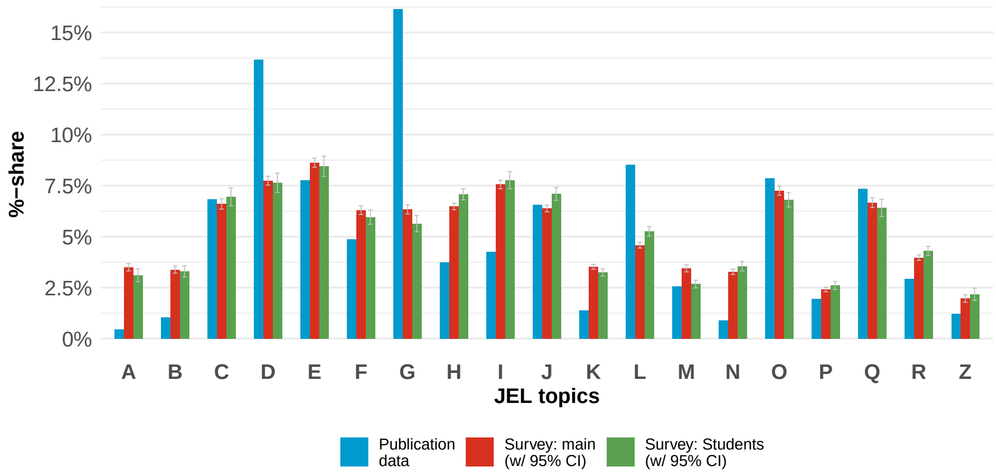
Notes: Black border: Main estimates. Blue bars: Share of JEL topics in our publication data (EconLit publication data, top 400 journals, January 2009 - December 2019). Red bars: Weighted average survey response with 95% confidence interval. Both distributions are calculated in different robustness specifications that are described in the discussion above.

Figure 4.D.7. Robustness of topic distributions – part 2



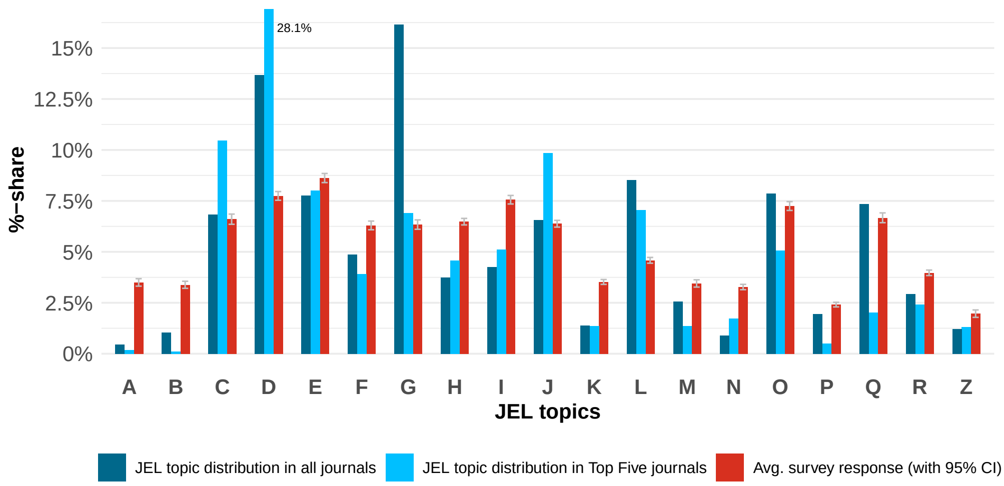
Notes: Blue bars: Share of JEL topics in our publication data (EconLit publication data, top 400 journals) for each year with linear time trend (slope reported). Red bars: Weighted average survey response.

Figure 4.D.8. Time trends in the topic distribution over the last decade



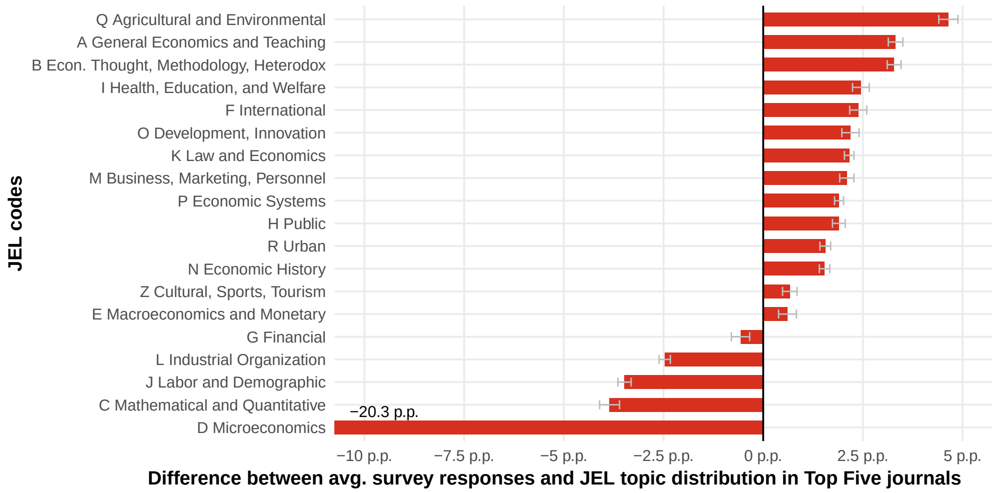
Notes: Blue bars: Share of JEL topics in our publication data (EconLit publication data, top 400 journals). Red bars: Weighted average survey responses in the main sample with 95% confidence intervals. Green bars: (Unweighted) average survey responses in the sample of Ph.D. students with 95% confidence intervals.

Figure 4.D.9. Preferred JEL topics in the main sample and the Ph.D. student sample



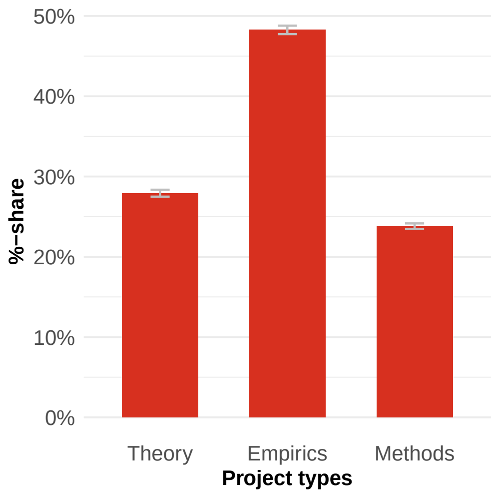
Notes: Dark blue bars: Share of JEL topics in top 400 EconLit-indexed journals. Light blue bars: Share of JEL topics in Top Five articles. EconLit publication data, January 2009 - December 2019. Red bars: Weighted average survey response with 95% confidence interval.

Figure 4.D.10. Comparison of JEL topic distribution in Top Five journals with survey responses



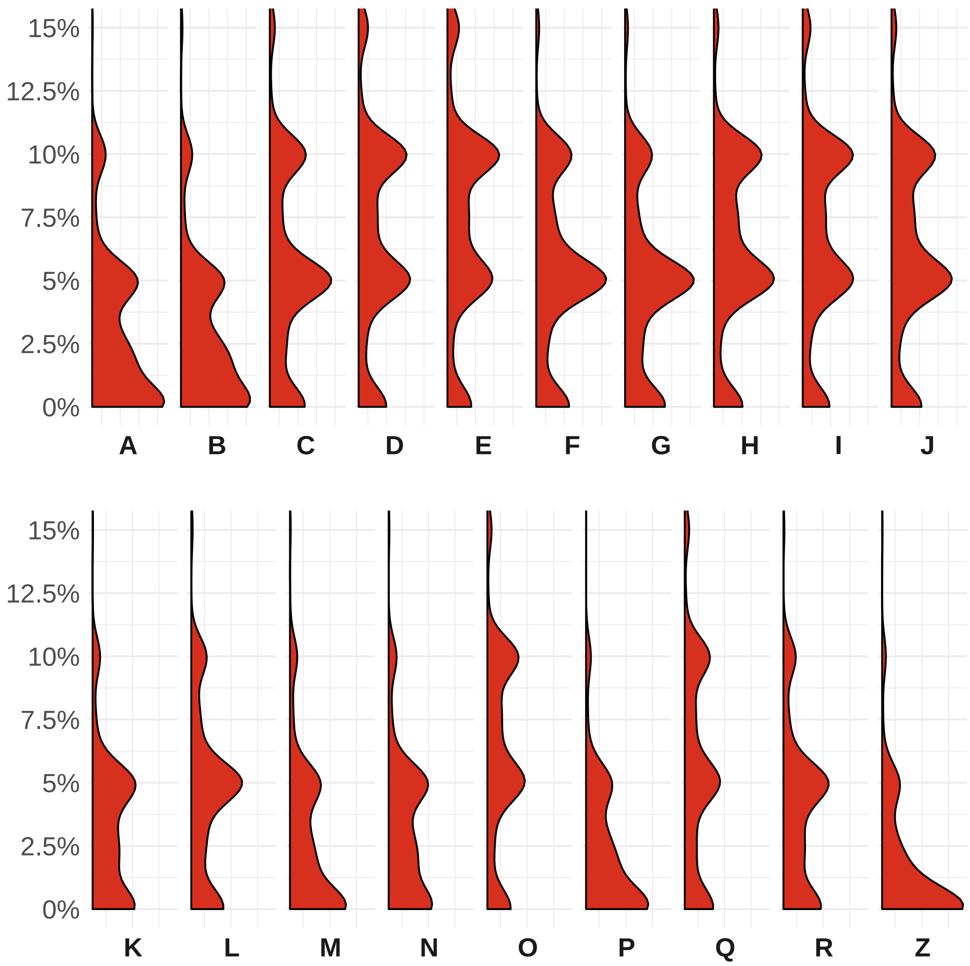
Notes: Differences between red bars and blue bars from the above figure 4.D.10 with 95% confidence intervals.

Figure 4.D.11. Differences between the average preferred and the actual JEL topic distribution



Notes: Weighted survey responses with 95% confidence intervals. Respondents were asked what share of economists' work should be predominantly theoretical, empirical, or focus on methods. "Please allocate 100 percentage points to the following three options according to what you think economists should work and publish on these days."

Figure 4.D.12. Comparison of respondents' preferred and actual distribution of project types



Notes: Weighted kernel density estimates, displayed from 0% to 15%.

Figure 4.D.13. Distribution of survey responses for each JEL topic

Table 4.D.2. Predictors of preferred JEL topics

	(1) A	(2) B	(3) C	(4) D	(5) E	(6) F	(7) G	(8) H	(9) I	(10) J
Demographics										
Female	.	.	.	–	– – –	.	.	.	+ + +	+ +
Age	.	+	.	.	.	.	.	.	.	–
Tenured	.	.	.	.	.	.	+	.	.	.
Region (vs. NA/AUS/NZL)										
EUR	.	.	.	.	.	.	.	.	.	.
AF, AS, LA	.	.	.	.	.	.	+ + +	.	.	.
Success										
Top 50 institution	.	.	.	.	.	.	.	.	.	.
Published Top Five	– – –	– – –	.	.	+ +	.	.	.	.	.
h-index	– –	.	.	.	.	.	.	.	.	.
Project types(vs. empirics)										
Theory	.	.	+	+ + +	+ +	.	.	– –	– –	– – –
Methods	.	.	+ + +	.	.	.	.	– – –	.	– –

	(11) K	(12) L	(13) M	(14) N	(15) O	(16) P	(17) Q	(18) R	(19) Z
Demographics									
Female	.	–	.	– – –	+	.	+ + +	.	.
Age	.	–	.	.	.	.	.	– –	.
Tenured	.	.	.	.	.	.	–	.	.
Region (vs. NA/AUS/NZL)									
EUR	–	– –	+	.	.	.	+	.	.
AF, AS, LA	.	– –	+ + +	–	.	.	–	.	.
Success									
Top 50 institution	.	.	.	.	.	– –	– – –	.	.
Published Top Five	.	.	.	.	.	– –	.	.	.
h-index	.	.	.	.	.	–	.	.	.
Project types (vs. empirics)									
Theory	.	+ +	–	.	.	.	–	.	.
Methods	– – –	– – –	.	– –	–	.	.	.	.

Notes: Results from weighted OLS regressions with robust standard errors. The dependent variable is the share assigned to the respective JEL topic of each column. The rows contain the explanatory variables of the regressions. We also control for (but do not report) the share of publications in each primary JEL topic as well as the share of publications in economics journals. We use the Benjamini-Hochberg procedure to correct *all* reported coefficients jointly for multiple hypotheses testing. +++/– – – indicates a p-value below 0.01, ++/– – a p-value below 0.05, and +/– a p-value below 0.10 for positive and negative coefficient respectively. Non-significant results are represented by a dot.

Table 4.D.3. Bias for own research field

	%-weight assigned to JEL topic				
	(1)	(2)	(3)	(4)	(5)
Own share (%)	0.106***	0.098***	0.101***	0.113***	0.112***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
Topic FE	✓	✓	✓	✓	✓
Controls	–	✓	–	–	–
Weights	Main	Main	Incl. no email	Unwgt.	Unwgt., econ.
Observations	70,699	68,191	70,699	75,639	63,859
R^2	0.149	0.173	0.143	0.151	0.170

Notes: Weighted OLS regressions, with standard errors (clustered on respondent level) in parantheses. The dependent variable is the %-weight assigned to a JEL topic j by respondent i. The predictor is the %-share of own publications of respondent i in JEL topic j. All regressions include topic fixed effects. Respondent fixed effects are not necessary because each respondent's weights sum up to 1. Column 2 interacts additional control variables with the topic fixed effects, namely gender, age, a tenure dummy, region (EUR and AF, AS, LA), a top-50-institution dummy, a published-Top-Five dummy, h-index, the share of research in theory and methods respectively, and the share of publications in economics. Columns 3-5 use different weighting schemes. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

4.D.3 Discussion

Top economists. We derive the following indicators for influential and successful scholars.

- **Top Five**: *Published Top Five* is a binary indicator that takes the value 1 if the author published at least one article in a Top Five journal within our publication sample (top 400 EconLit journals, 2009-2019, see main text section 4.3.1). The Top Five journals are the American Economic Review, The Quarterly Journal of Economics, the Journal of Political Economy, the Review of Economic Studies, and Econometrica. Publications in the Papers & Proceedings of the American Economic Review are not counted as Top Five publication.

  The *Published a Top Five* indicator is also used in other heterogeneity analyses of the paper.
- **Editor**: We compile a list of editors and advisory board members of the top 50 journals in economics from the years 2015-2020. We start from all EconLit-indexed journals and focus on the 50 outlets with the highest Scopus 2018 Scimago journal ranking. Most journals list their editors and board members in each printed issue. Since personnel turnovers are rare, we download the first issues of the years 2020, 2018, and 2016 and extract all available editor information. If an issue does not contain editor information, we check an earlier or older issue. Some journals do not announce their editors in print. Here, we derive information on their current editors and advisory board members from the journals' websites. In total, we find 2,818 editors and advisory board members.

  Based on the names, we match the editor data to our author database and manually disambiguate all cases in which multiple matches are found. In total, 93.1% of all editors can be matched to a scholar in our author data. The *Top 50 editor* dummy takes value 1 for successful matches, i.e. recent or current editors or advisory board members at the top 50 journals in economics.
- **Referees**: We compile a list of scholars who have repeatedly refereed at Top Five journals in the years 2015-2020. The American Economic Review, the Journal of Political Economy, and Econometrica publish a list of all referees yearly. The Quarterly Journal of Economics published a list of referees who reviewed four or more papers for 2018 and 2019, and the Review of Economic Studies published a list of recipients of an excellence in refereeing award in the years 2016 to 2019. We download these lists and extract the names of referees. We focus on referees that appear at least twice in the lists, that is, referees that review for at least two Top Five journals or in at least two years. In total, we find 4,229 Top Five referees.

  Based on the names, we match the referee data to our author database. In total, 69.0% of all referees can be matched to a unique scholar in our author data.

The *Top Five referee* dummy takes value 1 for successful matches, i.e. referees at Top Five journals.

Top economists' satisfaction with economics. Figure 4.D.14 shows that the topic preferences of top economists are very close to those of the full sample.

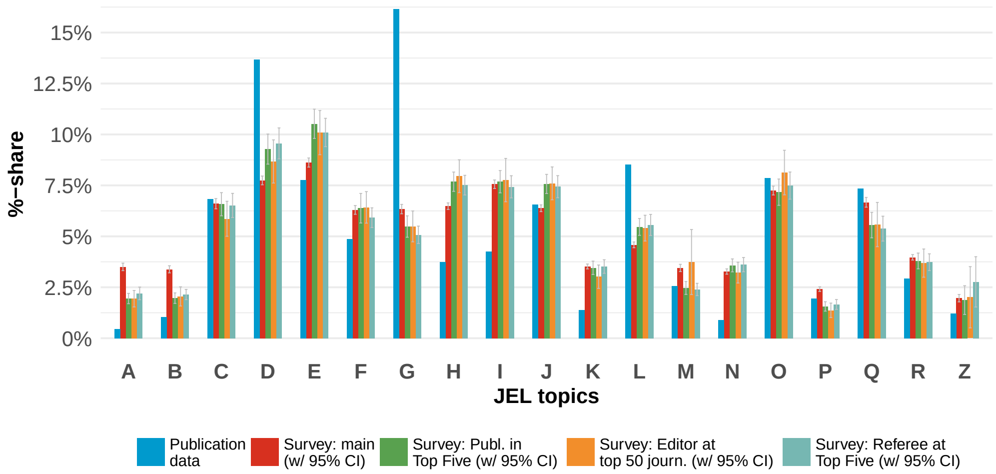
Table 4.D.4 shows that, similar to economists with a Top Five publication, editors at top 50 journals and referees at Top Five journals are more satisfied with the status quo in economics. It regresses the "satisfaction with economics" index on the three different "top economist" indicators.

The results are robust to using different weighting schemes. Moreover, similar results are obtained for each survey module and with the following alternative explanatory variables:

- Published articles in Top Five journal: Results are replicated with the number of Top Five publications.
- Editors at top journals: Results are replicated with editors at top 25 journals (155 cases) and top 10 journals (58 cases).
- Referees at Top Five journals: Results are replicated if we consider only referees that are mentioned at least five times in our list (i.e. referees with at least five different journal-year combinations).

Results of these analyses are available upon request.

Predictors of satisfaction – robustness. Tables 4.D.5 (satisfaction with own job), 4.D.6 (satisfaction with own research topics), 4.D.7 (stress), 4.D.8 (academia overly competitive), and 4.D.9 (satisfaction with economics) show that the analyses of satisfaction are robust to using different weighting schemes. Similar results are obtained for each survey module.



Notes: Blue bars: Shares of JEL topics in our publication sample (EconLit publication data, top 400 journals, January 2009 - December 2019). Red bars: Weighted average survey responses with 95% confidence intervals. Other bars: Unweighted average survey responses with 95% confidence intervals for different groups of top economists.

Figure 4.D.14. Comparison of JEL topic distributions in economics journals with survey responses in main sample and among top economists

Table 4.D.4. Top economists' satisfaction with economics

	Satisfaction with economics (std. index)			
	(1)	(2)	(3)	(4)
Top Five article	0.248***			0.153***
	(0.043)			(0.047)
Top 50 editor		0.245***		0.141**
		(0.061)		(0.063)
Top Five referee			0.253***	0.167***
			(0.048)	(0.053)
Author backgr.	✓	✓	✓	✓
Method ctrl.	✓	✓	✓	✓
Topic controls	✓	✓	✓	✓
Module FE	✓	✓	✓	✓
Observations	7,497	7,497	7,497	7,497
R^2	0.048	0.046	0.048	0.050

Notes: Weighted OLS regressions, robust standard errors in parentheses. The dependent variable is the "satisfaction with economics" index score. Higher values indicate higher satisfaction. The explanatory variable varies across panel: an indicator for having published in a Top Five journal (in our publication sample), for editors at top 50 journals, or referees at Top Five journals. Author background controls include gender, age, an indicator for having tenure, region dummies, an indicator for being at a top 50 institution, and h-index. Method controls include the share of projects in theory and methods research respectively. Topic controls include the share of publications in each primary JEL topic as well as the share of publications in economics journals. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 4.D.5. Predictors of satisfaction with own job – robustness

	Satisfaction with own job (std.)					
	(1)	(2)	(3)	(4)	(5)	(6)
Satisfact. w/ econ.	0.072***	0.067***	0.081***	0.079***	0.083***	0.063***
	(0.014)	(0.014)	(0.012)	(0.013)	(0.019)	(0.020)
Female	−0.072**	−0.071*	−0.084***	−0.079**	−0.063	−0.078
	(0.032)	(0.033)	(0.028)	(0.029)	(0.044)	(0.046)
Age (in 10y)	0.025*	0.027*	0.020	0.019	0.030	0.021
	(0.014)	(0.014)	(0.012)	(0.013)	(0.019)	(0.020)
Tenured	0.153***	0.159***	0.140***	0.144***	0.154***	0.152***
	(0.030)	(0.031)	(0.027)	(0.028)	(0.042)	(0.042)
Region: EUR	0.041	0.038	0.051*	0.066**	0.056	0.025
	(0.031)	(0.033)	(0.028)	(0.030)	(0.042)	(0.046)
Region: AF, AS, LA	−0.036	−0.042	−0.016	−0.003	−0.064	−0.013
	(0.042)	(0.043)	(0.037)	(0.040)	(0.058)	(0.060)
Top 50 inst.	0.089**	0.090*	0.092**	0.077**	0.155**	0.018
	(0.042)	(0.044)	(0.034)	(0.036)	(0.056)	(0.061)
Published Top Five	0.225***	0.232***	0.220***	0.226***	0.208***	0.259***
	(0.042)	(0.043)	(0.037)	(0.038)	(0.060)	(0.058)
h-index (in 10)	0.113***	0.111***	0.112***	0.113***	0.090***	0.135***
	(0.020)	(0.021)	(0.018)	(0.020)	(0.028)	(0.030)
Weights	Main	Wgt., no email	Unwgt.	Unwgt., only econ.	Only objectives	Only JEL
Method ctrl.	✓	✓	✓	✓	✓	✓
Topic ctrl.	✓	✓	✓	✓	✓	✓
Module FE	✓	✓	✓	✓	–	–
Observations	7,489	7,489	7,490	6,776	3,903	3,586
R^2	0.046	0.045	0.050	0.048	0.049	0.049

Notes: Weighted OLS regressions, robust standard errors in parentheses. The dependent variable is a standardized survey measure of job satisfaction ("All things considered, how satisfied or dissatisfied are you with your job in general?"). Higher values indicate higher satisfaction. Columns 1-4 employ different weighting schemes. Columns 5-6 estimate the regression for both survey modules separately. "Satisfact. w/ econ." is the satisfaction with economics index score (standardized). Age and h-index are divided by 10. Method controls include the share of projects in theory and methods research respectively. Topic controls include the share of publications in each primary JEL topic as well as the share of publications in economics journals. p-values are adjusted for multiple hypotheses correction within the reported coefficients of each row, using the Benjamini-Hochberg procedure. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 4.D.6. Predictors of satisfaction with own research topics – robustness

	Satisfaction with own topics (std.)					
	(1)	(2)	(3)	(4)	(5)	(6)
Satisfact. w/ econ.	0.034**	0.031**	0.043***	0.044***	−0.013	0.089***
	(0.014)	(0.015)	(0.013)	(0.014)	(0.017)	(0.023)
Female	0.027	0.022	0.026	0.027	0.027	0.023
	(0.031)	(0.032)	(0.027)	(0.028)	(0.041)	(0.045)
Age (in 10y)	0.053***	0.050***	0.053***	0.054***	0.081***	0.024
	(0.013)	(0.014)	(0.012)	(0.013)	(0.017)	(0.020)
Tenured	0.034	0.040	0.038	0.038	0.028	0.043
	(0.029)	(0.031)	(0.026)	(0.027)	(0.038)	(0.045)
Region: EUR	0.042	0.040	0.043	0.030	0.071	0.005
	(0.030)	(0.031)	(0.027)	(0.029)	(0.039)	(0.046)
Region: AF, AS, LA	−0.104**	−0.114**	−0.075*	−0.076*	−0.052	−0.153**
	(0.041)	(0.043)	(0.037)	(0.039)	(0.053)	(0.062)
Top 50 inst.	0.080*	0.090**	0.043	0.030	0.107*	0.038
	(0.040)	(0.041)	(0.035)	(0.037)	(0.052)	(0.061)
Published Top Five	0.175***	0.176***	0.160***	0.160***	0.185***	0.188***
	(0.043)	(0.043)	(0.039)	(0.040)	(0.060)	(0.062)
h-index (in 10)	0.107***	0.109***	0.106***	0.099***	0.061*	0.148***
	(0.023)	(0.023)	(0.019)	(0.021)	(0.029)	(0.035)
Weights	Main	Wgt., no email	Unwgt.	Unwgt., only econ.	Only objectives	Only JEL
Method ctrl.	✓	✓	✓	✓	✓	✓
Topic ctrl.	✓	✓	✓	✓	✓	✓
Module FE	✓	✓	✓	✓	–	–
Observations	7,493	7,493	7,494	6,777	3,905	3,588
R^2	0.037	0.038	0.039	0.039	0.041	0.048

Notes: Weighted OLS regressions, robust standard errors in parentheses. The dependent variable is a standardized survey measure of satisfaction with one's own research topics ("All things considered, how satisfied or dissatisfied are you with the topics on which you are working these days?"). Higher values indicate higher satisfaction. Columns 1-4 employ different weighting schemes. Columns 5-6 estimate the regression for both survey modules separately. "Satisfact. w/ econ." is the satisfaction with economics index score (standardized). Age and h-index are divided by 10. Method controls include the share of projects in theory and methods research respectively. Topic controls include the share of publications in each primary JEL topic as well as the share of publications in economics journals. p-values are adjusted for multiple hypotheses correction within the reported coefficients of each row, using the Benjamini-Hochberg procedure. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 4.D.7. Predictors of stress – robustness

	Stress (std.)					
	(1)	(2)	(3)	(4)	(5)	(6)
Satisfact. w/ econ.	−0.040***	−0.041***	−0.042***	−0.040***	−0.090***	0.012
	(0.013)	(0.014)	(0.012)	(0.013)	(0.018)	(0.019)
Female	0.216***	0.214***	0.211***	0.208***	0.251***	0.174***
	(0.030)	(0.031)	(0.028)	(0.029)	(0.042)	(0.044)
Age (in 10y)	−0.151***	−0.148***	−0.156***	−0.159***	−0.164***	−0.136***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.018)	(0.019)
Tenured	−0.026	−0.026	−0.026	−0.014	−0.031	−0.021
	(0.029)	(0.029)	(0.026)	(0.027)	(0.040)	(0.041)
Region: EUR	0.132***	0.126***	0.127***	0.144***	0.178***	0.077
	(0.030)	(0.030)	(0.027)	(0.029)	(0.041)	(0.043)
Region: AF, AS, LA	0.016	0.013	0.006	0.013	0.033	0.004
	(0.039)	(0.040)	(0.037)	(0.039)	(0.057)	(0.054)
Top 50 inst.	0.041	0.054	−0.001	0.018	−0.023	0.100
	(0.042)	(0.044)	(0.037)	(0.039)	(0.057)	(0.061)
Published Top Five	0.020	0.009	0.002	−0.002	0.088	−0.051
	(0.045)	(0.046)	(0.041)	(0.042)	(0.064)	(0.064)
h-index (in 10)	−0.068***	−0.067***	−0.055**	−0.056**	−0.080**	−0.061
	(0.024)	(0.024)	(0.021)	(0.023)	(0.033)	(0.031)
Weights	Main	Wgt., no email	Unwgt.	Unwgt., only econ.	Only objectives	Only JEL
Method ctrl.	✓	✓	✓	✓	✓	✓
Topic ctrl.	✓	✓	✓	✓	✓	✓
Module FE	✓	✓	✓	✓	–	–
Observations	7,487	7,487	7,488	6,772	3,901	3,586
R^2	0.076	0.074	0.075	0.075	0.099	0.070

Notes: Weighted OLS regressions, robust standard errors in parentheses. The dependent variable is a standardized survey measure of job-related stress experiences ("In general, how stressful do you find your job?"). Higher values indicate higher stress. Columns 1-4 employ different weighting schemes. Columns 5-6 estimate the regression for both survey modules separately. "Satisfact. w/ econ." is the satisfaction with economics index score (standardized). Age and h-index are divided by 10. Method controls include the share of projects in theory and methods research respectively. Topic controls include the share of publications in each primary JEL topic as well as the share of publications in economics journals. p-values are adjusted for multiple hypotheses correction within the reported coefficients of each row, using the Benjamini-Hochberg-procedure. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 4.D.8. Predictors of "Academia overly competitive" – robustness

	Agreement with "Academia overly competitive" (std.)					
	(1)	(2)	(3)	(4)	(5)	(6)
Satisfact. w/ econ.	−0.127***	−0.123***	−0.130***	−0.127***	−0.177***	−0.070***
	(0.013)	(0.014)	(0.012)	(0.013)	(0.019)	(0.019)
Female	0.230***	0.230***	0.250***	0.252***	0.232***	0.232***
	(0.029)	(0.030)	(0.026)	(0.028)	(0.041)	(0.042)
Age (in 10y)	−0.066***	−0.058***	−0.072***	−0.068***	−0.061***	−0.075***
	(0.013)	(0.014)	(0.012)	(0.013)	(0.018)	(0.020)
Tenured	−0.075**	−0.083***	−0.064**	−0.040	−0.112**	−0.028
	(0.029)	(0.030)	(0.026)	(0.027)	(0.040)	(0.042)
Region: EUR	0.114***	0.112***	0.116***	0.112***	0.080*	0.146***
	(0.030)	(0.030)	(0.028)	(0.030)	(0.040)	(0.044)
Region: AF, AS, LA	−0.024	−0.030	−0.017	−0.025	−0.084	0.044
	(0.040)	(0.041)	(0.038)	(0.040)	(0.056)	(0.056)
Top 50 inst.	0.010	0.003	0.045	0.062	0.042	−0.027
	(0.042)	(0.043)	(0.037)	(0.039)	(0.056)	(0.061)
Published Top Five	−0.143***	−0.135***	−0.184***	−0.183***	−0.108	−0.171**
	(0.047)	(0.047)	(0.044)	(0.045)	(0.065)	(0.067)
h-index (in 10)	−0.051**	−0.054**	−0.025	−0.043	−0.050	−0.056
	(0.023)	(0.024)	(0.021)	(0.023)	(0.032)	(0.033)
Weights	Main	Wgt., no email	Unwgt.	Unwgt., only econ.	Only objectives	Only JEL
Method ctrl.	✓	✓	✓	✓	✓	✓
Topic ctrl.	✓	✓	✓	✓	✓	✓
Module FE	✓	✓	✓	✓	–	–
Observations	7,493	7,493	7,494	6,778	3,905	3,588
R^2	0.065	0.062	0.063	0.063	0.082	0.060

Notes: Weighted OLS regressions, robust standard errors in parentheses. The dependent variable is a standardized survey measure of perceiving academia as overly competitive ("I would personally criticize academia for being overly competitive"). Higher values indicate larger agreement and hence lower satisfaction. Columns 1-4 employ different weighting schemes. Columns 5-6 estimate the regression for both survey modules separately. "Satisfaction econ." is the satisfaction with economics index score (standardized). Age and h-index are divided by 10. Method controls include the share of projects in theory and methods research respectively. Topic controls include the share of publications in each primary JEL topic as well as the share of publications in economics journals. p-values are adjusted for multiple hypotheses correction within the reported coefficients of each row, using the Benjamini-Hochberg procedure. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 4.D.9. Predictors of satisfaction with economics – robustness

	Satisfaction with economics (std. index)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	−0.072**	−0.069*	−0.068**	−0.055*	−0.081	−0.070
	(0.031)	(0.032)	(0.028)	(0.029)	(0.042)	(0.045)
Age (in 10y)	−0.069***	−0.068***	−0.065***	−0.064***	−0.084***	−0.051**
	(0.014)	(0.015)	(0.013)	(0.014)	(0.021)	(0.020)
Tenured	0.068**	0.060*	0.066**	0.061**	0.067	0.070
	(0.030)	(0.031)	(0.026)	(0.028)	(0.041)	(0.043)
Region: EUR	−0.096***	−0.084**	−0.091***	−0.105***	−0.130***	−0.057
	(0.030)	(0.031)	(0.027)	(0.028)	(0.042)	(0.043)
Region: AF, AS, LA	−0.067	−0.054	−0.077*	−0.088**	−0.007	−0.115
	(0.042)	(0.043)	(0.038)	(0.040)	(0.059)	(0.058)
Top 50 inst.	0.016	0.013	0.018	0.030	−0.012	0.045
	(0.039)	(0.041)	(0.034)	(0.036)	(0.056)	(0.055)
Published Top Five	0.248***	0.250***	0.232***	0.215***	0.304***	0.185***
	(0.043)	(0.043)	(0.038)	(0.039)	(0.063)	(0.057)
h-index (in 10)	0.010	0.011	0.002	−0.003	−0.006	0.022
	(0.024)	(0.024)	(0.021)	(0.024)	(0.037)	(0.029)
Weights	Main	Wgt., no email	Unwgt.	Unwgt., only econ.	Only objectives	Only JEL
Method ctrl.	✓	✓	✓	✓	✓	✓
Topic ctrl.	✓	✓	✓	✓	✓	✓
Module FE	✓	✓	✓	✓	–	–
Observations	7,497	7,497	7,498	6,781	3,908	3,589
R^2	0.048	0.045	0.048	0.048	0.060	0.045

Notes: Weighted OLS regressions, robust standard errors in parentheses. The dependent variable is the "Satisfaction with econimics" index score. Higher values indicate higher satisfaction. Columns 1-4 employ different weighting schemes. Columns 5-6 estimate the regression for both survey modules separately. Age and h-index are divided by 10. Method controls include the share of projects in theory and methods research respectively. Topic controls include the share of publications in each primary JEL topic as well as the share of publications in economics journals. p-values are adjusted for multiple hypotheses correction within the reported coefficients of each row, using the Benjamini-Hochberg-procedure. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.