The Annual Proceedings of

The Wealth and Well-Being of Nations

2022-2023

Volume XIV: Experimental Methods and the Wealth and Well-Being of Nations

John A. List

Diep Phan, Editor

The Miller Upton Program at Beloit College

The Wealth and Well-Being of Nations was established to honor Miller Upton, Beloit College's sixth president. This annual forum provides our students and the wider community the opportunity to engage with some of the leading intellectual figures of our time. The forum is complemented by a suite of programs that enhance student and faculty engagement in the ideas and institutions that lay at the foundation of free and prosperous societies.

Senior Seminar on The Wealth and Well-Being of Nations:

Each year, seniors in the Department of Economics & Business participate in a semester-long course that is built around the ideas and influence of that year's Upton Scholar. By the time the Upton Scholar arrives in October, students will have read several of his or her books and research by other scholars that has been influenced by these writings. This advanced preparation provides students the rare opportunity to engage with a leading intellectual figure on a substantive and scholarly level.

Endowed Student Internship Awards:

A portion of the Miller Upton Memorial Endowments supports exceptional students pursuing high-impact internship experiences. Students are encouraged to pursue internships with for-profit firms and non-profit research organizations dedicated to advancing the wealth and well-being of nations.

Student Research Colloquium and Speaker Series:

The department has initiated a research colloquium that gives students the opportunity to read and discuss seminal articles aimed at deepening their understanding of the market process. Students also develop original analysis that applies economic ideas to novel contexts. Colloquium participants receive close mentoring as they craft an article with the eventual goal of publication in a newspaper, magazine, or academic journal. The themes of the research colloquium and annual forum are supported with a speaker series featuring the next generation of scholars working on questions central to our understanding of the nature and causes of wealth and well-being.

Annual Proceedings of The Wealth and Well-Being of Nations:

The keynote address presented by the Upton Scholar is an important contribution to the public discourse on the nature and causes of wealth and well-being. Further, the annual forum includes presentations by noted scholars who expand upon or challenge the work of the Upton Scholar. These presentations are assembled in the *Annual Proceedings of the Wealth and Well-Being of Nations*, which serves as an important intellectual resource for students, alumni, and leaders within higher education.

The Annual Proceedings of the Wealth and Well-Being of Nations

2022-2023

Volume XIV

Diep Phan
Editor

Jennifer Kodl
Managing Editor

The Annual Proceedings of the Wealth and Well-Being of Nations, 2022-2023.

Printed in the United States of America

ISBN 979-8-218-26736-0

Contents

Contributors

Diep Phan is the Elbert H. Neese, Jr. Professor of Economics at Beloit College.

John A. List, the 2022 Miller Upton Scholar, is the Kenneth C. Griffin Distinguished Service Professor in Economics at the University of Chicago. His research focuses on questions in microeconomics, with a particular emphasis on using field experiments to address both positive and normative issues.

Anya Samek is an Associate Professor of Economics and Strategy at the Rady School of Management at University of California-San Diego. She is also a Faculty Research Fellow at the National Bureau of Economic Research and an affiliate of the FAIR Centre at the Norwegian School of Economics.

Sally Sadoff is a Professor of Economics and Strategy at the University of California-San Diego.

Sam Lindquist is a Predoctoral Research Fellow at the Rady School of Management at University of California-San Diego.

Alec Brandon is an Assistant Professor at Carey Business School at Johns Hopkins University.

Phil Chen is an Assistant Professor of Political Science at University of Denver, formerly Assistant Professor of Political Science at Beloit College.

Phil Shields is a Professor of Philosophy and Religious Studies at Beloit College.

Edward Vezosa is a 2023 graduate of the Department of Economics & Business at Beloit College.

Quint Studer is founder of Pensacola's Studer Community Institute, a non-profit organization focused on improving the community's quality of life. He also founded Vibrant Community Partners, which coached communities in building a blueprint for achieving growth and excellence.

Introduction

Diep Phan[1]

Introduction

As the Elbert Neese Professor of Economics, it is my honor to introduce the fourteenth Annual Proceedings of the Wealth and Well-Being of Nations, a collection of papers based on the talks at the 2022 Miller Upton Forum on Experimental Methods and the Wealth and Well-Being of Nations, featuring Professor John A. List as the Upton Scholar.

The Miller Upton Forum on the Wealth and Well-Being of Nations is the centerpiece of a suite of programs named in honor of Beloit College's sixth president, Miller Upton. The programs are inspired by Miller's unflagging dedication to the ideals of a liberal society: political freedom, the rule of law, and peace and prosperity through the voluntary exchange of goods and ideas. Since its in inauguration in 2008, the forum has examined a variety of critical factors that are thought to determine nations' prosperity, such as social institutions and the rule of law (2008 Inaugural Forum), property rights (2009 Forum), entrepreneurship (2010 and 2015 Forums), self-governance (2011 Forum), institutional change (2012 Forum), economic freedom (2013 Forum), energy and climate change issues (2014 and 2016 Forums), and human capital (2019 Forum). The eleventh forum in 2018 turned to an imperative and timely topic--economic globalization--which encompasses international trade, international finance, and international migration. The thirteenth forum in 2021 returned to the globalization topic, but with a focus on international migration.

Compared with all previous forums, the 2022 forum is quite a deviation, as its theme--experimental methods--isn't a topic but a methodology. As such, the presentations in this forum span a wide variety of topics, but they all share one thing in common: the use of experimental method. As usual, the forum is

1 Diep Phan is the Elbert H. Neese, Jr. Professor of Economics at Beloit College.

interdisciplinary by design, so it had the participation of not only distinguished economists, but also a political scientist, a psychologist, and a philosopher.

Professor John A. List is the Kenneth C. Griffin Distinguished Service Professor in Economics at the University of Chicago. He received his Ph.D. in economics at the University of Wyoming, then became a professor at the University of Central Florida, University of Arizona, and University of Maryland, and since 2005, he has been at the University of Chicago. Over the years, he has held many distinguished positions, chairs, professorship, including being an economic adviser at the White House. List has also held various positions in the private sector; he was the chief economist at Uber, then Lyft, and now Walmart.

List's research is impressive in terms of both quantity and quality, and he might be the most famous economist that we have invited to campus. According to RePEc (Research Papers in Economics), the abstracts of List's papers were the most frequently viewed of any economist in the world during the 12 months leading into the Fall 2022 Upton Forum. He also ranked 1st in the number of downloads over these same 12 months, according to RePEc. In terms of research topics, List's interests vary widely. He has done work in charitable giving, labor market, health, education, early childhood education, gender and racial issues, sports economics, and many more. In terms of methodology, List is well known for his pioneering work using experimental methods, especially natural field experiments.

But what makes List special to Beloit is not just his academic fame, but the fact that he's a Wisconsinite. He was born and grew up in Sun Prairie, Wisconsin. He attended Sun Prairie High School and completed his bachelor's degree at the University of Wisconsin -- Stevens Point, majoring in economics. Before he was John List the economist, he had come to the city of Beloit to play golf, and to Beloit College to play basketball. At some point, he employed a Beloit econ alum as his research assistant, and he is a friend since high school of Stopher Bartol '88, who now serves on the board of trustees of the college.

For readers unfamiliar with experimental methods, it is helpful to give a quick overview of this methodology. Imagine that one needs to study the effect of obtaining a college education on the wage of someone named Barry Allen. In theory, it's simple: just compare Barry's wage with college education ($w_{college}$) and Barry's wage without college education ($w_{no\_college}$). In practice, there is only one Barry Allen who either attends college or does not, so only either $w_{college}$ or $w_{no\_college}$ is observed, but not both. That is, the counterfactuals are not observed, and hence

these two wages cannot be compared.[2]

However, a close substitute to observing the counterfactuals would be randomization. For example, a randomized control trial is a type of experiment typically run to measure the effect of a treatment, such as the effect of a drug on a patient's health outcome, or the effect of participation in a childhood education program on a person's schooling and labor market outcomes in later life. In such a trial, people or participants are randomly assigned into two groups: those in the treatment group receive the real treatment (the drug, or participation in the childhood education program), and those in the control group do not receive the real treatment (perhaps they receive a sugar pill, or they get information brochures instead of participating in the childhood education program). Because the two groups are not systematically different in anyway (due to randomization) except the treatment, then any observed difference in outcomes between the two groups must be due to the treatment; that is, the treatment effect can be measured by comparing the average outcomes between the two groups.

Randomized control trials are the gold standard for identifying and measuring causal effects and have been very common in medical and life sciences and physical sciences for a long time. But it was not until recent decades that it became popular in social sciences. In the economics discipline, economists have been running lab experiments for many years to test certain economic theories. For a long time, they did not run field experiments in which real people living their real lives in the real world are experimented on, because it was thought that it would be too hard or even impossible to do. That was the case until the 1990s, when people like John List came along. List was creative and pushed the boundaries of the field. Today it is widely recognized that natural field experiments can be done in economics, or in social sciences in general, and the method has enjoyed an immense increase in popularity.

The Upton Forum's topic varies each year, but the overarching theme is always the wealth and well-being of nations. Though economists still debate a great deal on which policies or strategies a country can pursue to promote its wealth

2 The counterfactuals could perhaps be observed if one lived in the fictional world of DC Comic's series The Flash. In that fictional world, infinite number of parallel universes co-exist in the same space but never collide, known as the multiverse. In such a multiverse, one can imagine that there exist two universes: one in which Barry Allen attends college, and another that is the same as the first one in every way, except that Barry Allen does not attend college. In that case, the counterfactual is observed, and it would be possible to measure the effect of attending college on Barry's wage. But of course, such a multiverse only exists in DC Comics, not real life, which brings us to John List's pioneering work in field experiments.

and well-being, they have reached a consensus that a nation's wealth and well-being can only be promoted with sufficient economic progress, which in turn is driven by accumulation of or productive investments in physical and human capital and technology. Central to this accumulation and investment process, especially regarding technology, is the market for ideas, and this is where List's talk at the forum comes in. The first essay (chapter 2) in this volume is based on that talk, which in turn is based on his most recent book "**The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale**." His talk and his book are about the market for science and for ideas, or more precisely, how we can leverage the market to promote good ideas to scale, and bad ideas to be revised so they are scalable in the future. Though his talk (and the book) is about scaling ideas, much of the science behind it is from experimental economics, and from List's decades of work running experiments in a wide variety of contexts.

In the essay, List told the story of how he realized the need for and arrived at the idea of writing the book. He then summarized the five vital signs to tell whether an idea is scalable, which come from the first half of the book: (i) false positive; (ii) know your audience; (iii) is it the chefs or is it the ingredients? (iv) unintended consequences and spillovers; and (v) the cost trap. He also briefly mentioned the second half of the book, in which he suggests ways to make ideas more scalable: the importance of structuring the right incentives, the power of marginal thinking and sunk cost fallacy. These concepts should be familiar to readers of the Upton Forum's Annual Proceedings who are economists or were economics major in college; they are the cornerstones of the economic way of thinking and are brilliantly applied by List in the context of running experiments, and trying out and scaling ideas.

At the 2022 Miller Upton Forum, Beloit College was honored to have the participation of three other experimental economists who are long time collaborators and co-authors of the Upton Scholar John List. Chapters 3 through 5 in this volume are based on their talks at the panel "**The Ideas and Influence of John List**" at the forum.

Anya Samek received her PhD in Economics from Purdue University in 2010, where she also did her master's and bachelor's degrees in economics. After Purdue, she was a post doc at the University of Chicago for two years where she worked with John List, and currently she is at the Rady School of Management at the University of California - San Diego. In her essay titled "**Toward an Understanding of John List: Field Experiments from Children to Charities**," Samek shared

many valuable lessons about conducting field experiments during her time working with John List. She pointed out two explanations for why economists should conduct field experiments with children: economists must understand children themselves for immediate purposes, and children inform our understanding of adults' behaviors. She highlighted initiatives that have been done to facilitate field experiments in charitable giving, and confirmed the importance of field experiments in fundraising, by presenting evidence showing that fundraising professionals are not good at predicting the results of many field experiments. Finally, she ended the essay offering directions for future research related to conducting field experiments on children and charitable giving. Overall, the essay gave an excellent overview into one area of the Upton Scholar's work, befitting the panel titled "The Ideas and Influence of John List" at the forum.

Sally Sadoff received her PhD in Economics from the University of Chicago in 2010. Currently, she is an associate professor of economics and strategy at the Rady School of Management at the University of California - San Diego. In her research, Sadoff studies labor, health, and education topics using experimental method. Chapter 4 of this volume, titled "**Understanding the interaction of sleep, social media, and mental health for productivity and performance: the role of field experiments,**" is written by Sadoff and her co-author Sam Lindquist. In the paper, the authors presented a simple theoretical framework to illustrate how sleep, social media, and mental health are highly inter-related, and how they all matter for the wealth and well-being of nations, in the sense that they affect subjective well-being and productivity. They provided three comprehensive reviews of the three literatures on these topics (mental health, social media, sleep), then synthesized them to arrive at an important conclusion: there is substantial heterogeneity in findings regarding the determinants and impacts of these factors. They concluded with a discussion of the policy implications yielded by the literature, and offered some guidance for the design of field experiments to help inform our understanding of the causal linkages between sleep, social media and mental health, and their impact on productivity.

Alec Brandon received his PhD in Economics from the University of Chicago in 2020 and is currently an assistant professor at the Johns Hopkins Carey Business School. Formerly, for two years in 2011-12 he was the associate director at the Becker Friedman Institute, a hub for cutting-edge research at the University of Chicago cutting across many departments, programs, and disciplines. Brandon's research focuses on the application of experimental methods to learn about the

impact of public policies and behavioral mechanisms driving those impacts. In his essay titled "**Does Self-Selection Diminish the Influence of Experimental Research**?" Brandon seeks to answer the question: why do experimental economists conduct framed field experiments, which randomize treatment amongst a self-selected sample, instead of natural field experiments, which randomize treatment amongst an unknowing sample? Evidence based data on framed and natural field experiments published in the most influential economic journals between 2005 and 2019 points to the following narrative: experimental economists conduct framed field experiments because there is increase in scientific influence (framed field experiments receive more citations per year than their natural counterparts). Furthermore, framed field experiments that are more transparent about their recruitment of samples enjoy a substantial premium in scientific influence.

By design, the Miller Upton Forum is interdisciplinary, so it has often invited the participation of scholars beyond the economics discipline. At the panel titled "**Experimental Methods in Social Sciences**," the Beloit community had a unique opportunity to hear perspectives from a political scientist and a philosopher regarding how their respective discipline view and employ the experimental method. Chapters 6 and 7 in this volume are based on their talks at this panel at the forum.

Phil Chen is an assistant professor of political science at the University of Denver. Prior to that, he was in the political science department at Beloit College. He received his PhD in American Politics and Political Methodology from the University of Minnesota in 2015. His work focuses on race, gender, and other social identities and how attitudes and beliefs about these identities shape American political behavior and public opinion. Chen has published in numerous political science and psychology journals and is currently working on a co-authored book about how racial and gender representation affects trust in the federal court system.

At the Miller Upton Forum in fall 2022, Chen gave a talk titled "**Racialized Courts: How racial attitudes shape perceptions of the American judicial system**." In it, he offered the audience an example of how the experimental method is used in political psychology to understand political issues, such as how public opinions about the courts are formed, or how the public evaluate the courts' decisions. Chen designed and conducted three survey experiments whereby respondents are given a small amount of information about a court case and are then asked to express their trust, legitimacy and perceptions of bias of the courts.

To test whether people evaluate the courts based on whether a decision's policy content accords with their racial worldview or based on the race of the judge, these factors were randomly varied within the surveys. The results from three different experiments (run across two different surveys in the summer and fall of 2021) confirm expectations: the race of the judge makes little difference in how respondents evaluate the court, but the direction of the decision is deeply important. Chen's study shows the power of experiments in the social sciences to challenge conventional wisdom. *"Contrary to popular belief, support for the federal courts is not nearly as stable as thought. Instead, this support is deeply racialized. Not only that, but that racialized support is contingent on the content and substance of court decisions and proposals. When a policy is racially egalitarian, racially conservative individuals exhibit lower levels of support for the courts, and vice versa."*

While the Upton Forum's overarching theme is the wealth and well-being of nations, the underlying philosophy behind it is an embrace of open inquiry and exchange of ideas. Indeed, president Miller Upton once said *"Dialogue is difficult … but it should be possible on a college campus where people … have mutual respect for one another and where there is … the possibility of finding some area of common ground and of mutual modification of beliefs."* For that reason, the essay by Phil Shields, a philosopher and an ethicist who challenges the field of experimental economics and the use of experimental method in studying human behavior, is a welcome addition to the forum.

Phil Shields received his PhD from the University of Chicago in 1991 and is currently Professor of Philosophy at Beloit College. He has held the Hales Professorship in Ethics since 2007. In his essay titled "**Maybe we can, but should we?**" Shields provided a formidable argument that questions the use of field experiments in social sciences: objectifying humans and reducing them as mere cause-and-effect mechanism comes at a human cost, because it diminishes human agency, dignity and well-being. Furthermore, it raises the risk that *"the resulting knowledge will be 'weaponized' to exploit people with or without their awareness."* Shields boldly critiques the work of John List and his co-author Uri Gneezy in their book ***The Why Axis***, based on several grounds, one being the concern over the extent to which we ought to employ extrinsic incentives to manipulate human behavior as opposed to cultivating intrinsic incentives. He concluded his essay with a powerful analogy: *"What science cannot account for, then, can be seen every time we raise and educate a child. While parents employ a vast array of causal behaviors to enculturate the child into our normative practices, they also talk to the child,*

from the beginning, as a participant, and the child soon engages the meaning of these practices on their own terms. The resulting realm of meaning enables children to mean what they say, to do something for its own sake, and even at times to do the right thing for the right reason. In these ways, the child expresses agency, as opposed to merely being determined by causes and being coerced into doing what parents want."

Chapter 8 in this volume is special, because it is authored by Edward Verzosa'23, now an alum of the college, but he was a senior student when the 2022 Upton Forum was taking place. A double major in economics and political science, Verzosa's economics senior thesis proposed a survey experiment to answer the question "Does Negative Campaigning Drive Polarization?" Having one of the best proposals in the economics senior seminar, Eddie was awarded funding by the Upton Program, which enabled him to continue his research as an honors project and implement the experiment in spring 2023. He successfully presented his research at the Spring 2023 Student Symposium. Verzosa shared that he saw this experiment as a stepping stone toward his career objective of influencing policy through research to address the polarization issue on a national scale, thus inspiring him to pursue a Master's degree in Economics.

Verzosa is interested in studying the rise of political polarization in the United States. He made an observation that there has also been an increase in the use of negative campaigning and attack advertisements. He wondered if it's not a mere correlation, and perhaps one could be causing the other? Using the tools learned in the economics senior seminar, he designed a survey experiment to pin down the causal relationship between the two. The survey was carried out during March 13-28 2023 and consisted of 994 participants -- a representative sample of the US population. The results indicate that negative campaigning does not significantly increase polarization, but rather may create a "backlash" effect against the sponsor. Similar to Chen's study (chapter 6), Verzosa's paper demonstrates how the experimental method can be applied in social sciences to answer policy-relevant questions.

The last chapter in this volume is written by Quint Studer, founder of Pensacola's Studer Community Institute, a nonprofit organization focused on improving the community's quality of life. He is a businessman, a visionary, and entrepreneur, and a mentor to many. He currently serves as the Entrepreneur-in-Residence at the University of West Florida. He is also a prolific writer, and writes a weekly employee development column that runs every Sunday in the Pensacola News Journal. He has authored seven books, including "Building a Vibrant Com-

munity: How Citizen-Powered Change is Reshaping America," which his talk at the forum and the last chapter is based on.

In his essay titled "**Understanding Your Community's Potential**," Studer points out the important factors one needs to build a foundation for vitalizing a community: start with a "burning platform"; pull together a guiding coalition; learn the basics of change management; diagnose the problem correctly before applying a solution treat; and don't underestimate the power of community branding and messaging. He then lists the four critical areas that improve quality of life and work together as "gears" to drive a community forward: a vibrant downtown, early learning and education, economic development, and civic engagement. Readers connected to the Beloit community may find that the most interesting aspect of Studer's essay is his presentation and discussion of the Quality of Life Survey that the city of Beloit conducted in early 2022. Results were fairly positive, but also revealed important areas that the city needs to work on.

Studer's essay, based on his book, is the culmination of his many years of experience working with the community to help revitalize Pensacola. It's imperative to point out that the lessons outlined in the essay were derived from an experimental approach, and it's interesting to compare List's and Studer's experimental approaches. Like List, Studer's experimental approach emphasizes the testing of an idea before scaling it or applying it to another context. Both List and Studer advocate for replication to fine tune an idea. The difference is that, while List uses randomization to draw causal inferences, Studer uses a trial-and-error approach on a small sample of communities, and use observations based on that small sample (instead of randomization) to draw conclusions. And of course, the other difference is that Studer's experiment was all about community development and revitalization, while List's experiments spanned a very large number of topics. Beloit College was very fortunate to have these two prominent thinkers and speakers at the 2022 Upton Forum.

Acknowledgements and Thanks

I wish to thank all the speakers who delivered talks at the 2022 Miller Upton Forum and contributed papers to this volume. I also wish to extend my thanks to all those who made the Upton Forum possible, in particular the alumni, donors, and former colleagues who helped raise funds and created the Upton Forum. I would like to especially thank Ms. Jennifer Kodl for all her excellent work in organizing the forum and editing the Annual Proceedings in all fourteen years that the program has been in existence. And finally, I would like to thank my colleagues and students at Beloit College and members of the local community, whose participation in the forum is critical to its continuing success.

On the topic of expressing thanks, below I quote Ajani Joseph'23 who attended the economic senior seminar and Upton Forum in fall 2022. He perfectly articulated the essence of the Miller Upon Forum:

> *"Since my freshman year, I have attended every Upton Forum. Every year, I understood more and therefore enjoyed each one more than the previous one. Though I always have high expectations for the Miller Upton Forum, this year's [forum] far exceeded what I thought it would be … dialogues from different perspectives were welcomed, and we were able to see different views of the experimental method from various departments. This is quite consistent with the breadth and depth that we emphasize here at Beloit College. Overall, I am a better student and individual having attended the forum, and the experience encapsulated the definition of a liberal arts education."*

The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale

John A. List [1, 2]

Introduction

First, I'd like to thank you all for choosing me as the Upon Scholar. My time lecturing at the 2022 Miller Upton Forum has been incredible. While teaching my lectures, the energy in the room and the energy the students brought has been open and honest. The community at Beloit College is a family environment. I want to commend the economics department. This is a place where we can have a debate and learn about science. We are all trying to make the world a better place, and the beautiful thing about Beloit College is that we can openly debate our ideas.

This is why the Miller Upton Forum holds such an important place in society. At least since Adam Smith's 1776 treatise on The Wealth of Nations, humankind has recognized the import of the economic science to promoting the wealth and well-being of nations. Indeed, how the market promotes wealth and well-being and how market forces can operate to reduce poverty and promote human development, from birth, remain key tenets of the Upton Forum. Today I want to push the Upton Forum in a slightly different direction than previous Upton Scholar recipients. I want to discuss the market for science, and how we can

1 John A. List, the 2022 Miller Upton Scholar, is the Kenneth C. Griffin Distinguished Service Professor in Economics at the University of Chicago. His research focuses on questions in microeconomics, with a particular emphasis on using field experiments to address both positive and normative issues.

2 This article is an adapted version of John A. List's National Association for Business Economics (NABE) Adam Smith Award Lecture. The original adaptation of this lecture can be found in *Business Economics,* List (2023).

leverage that market to promote good ideas to scale and bad ideas to be revised so they are scalable in the future.

I want to begin by talking a little bit about field experiments because a lot of you may not even know this, but field experiments are something I've been doing now for over 30 years. It is essentially using the world as my lab. Let me make this a little bit of a two-way street, and let me ask you, have you, in the last five years, used an Uber or Lyft? Have you voted in at least one of the past two presidential elections? How many of you have flown United Airlines, or have used Google to search, in the past three years? Has anyone not said yes at least once? All of you have been a subject in at least one of my field experiments. It's not creepy in that I cannot pinpoint your name. I respect privacy.

What I can say is I know what are good incentives to get you to give to a charitable cause. I know what are good incentives to keep you committed to a charitable cause. Why? Because we've been using field experiments now for 25 years in the economics of charity. I can talk about what motivates people to go and vote. Again, without compromising privacy. So a big reason why I am able to discuss the following is because of all of you, helping me learn about the real world through economics.

Now let's talk about *The Voltage Effect* (2022).

The Voltage Effect

The *Voltage Effect* (2022) is essentially about scaling. It's about the second half of my career. We talk a little bit about the first half, which is using field experiments to learn about the world. The second half, or understanding scaling, is actually how you change the world using science. I think, as social scientists, we have let down the broader community by asking and answering the wrong questions. That's what I think we will learn today and throughout the first several chapters of *The Voltage Effect* (2022).

Now, where I want to start is, how did we get to the Voltage Effect? I want to give you a little bit of background about the roots of the book. Of course, a book like *The Voltage Effect* (2022) has many roots, but one very important root happened in Chicago Heights. How many of you have heard of the community called Chicago Heights? It's a community that is about 30 miles south of Chicago. This is a community wherein 95% of households are on federal food stamps. This is a community that the manufacturing jobs have left behind. They called me in

2007, and they said, "John, we need your help." As a humane human, it's not really about answering that question, "yes" or "no." The question really is, where do we start?

Now, where we started was building our own pre-K. We wanted to start with early childhood, and we wanted to start because we believe, and I think most of us in this room believe, that the opportunity gap is something that continues to stifle not only equity but efficiency. You might be wondering right now, "What in the world does an economist know about building a pre-K?" Nothing. Zero. I don't know how to hire teachers. I don't know about bus lines. I don't know about federally-subsidized lunches or milk. I do now, but I didn't then. What I had was a community that wanted to open up and needed help. Right away, I realized they don't have resources. So I went to Ken and Anne Griffin, and they gave us $20 million to start the Chicago Heights Early Childhood Center.

We started building in 2008, and we opened in 2010. Three, four, and five-year-olds, thousands of them. I had three goals. One, I wanted to help Chicago Heights kids. Two, I wanted to learn about the education production function, using randomization to explore executive function – cognitive – skills. I wanted to write academic papers to teach the rest of the world about what we were learning about human capital skill formation. Then three, I wanted to create a curriculum that we could scale, so every child in America and the rest of the world could use our scientifically-created curriculum.

Dataset after data set came in 2011, 2012, 2013, and 2014. We're killing it. We are moving kids from the South Side of Chicago to the North Side of Chicago within six months. Looking at cognitive scores from a battery of tests, one called Woodcock-Johnson, looking at executive function skills. Many were created by us because it is a wild west in the executive function skill space. We were moving kids from the 20th percentile to the 60th within six months. Through CHECC, we're helping Chicago Heights kids. These kids are now sophomores and juniors in high school, and we're still tracking them.

Second goal: write academic papers. Yes. We were writing dozens and dozens of academic papers about what we'd learned. The most recent one just came out in the Journal of Political Economy on how social preferences are affected by our early childhood program.

Now we get to the third goal. What do you think happens at the third goal? Remember, the third goal was, I want to scale that program to change the world. Here's what happens: The Slap in the Face. "John, your program had

an impressive benefit profile, but don't expect it to happen at scale." This was around 2014 or 2015. At that point, I had been doing field experiments for close to 25 years. I had never been met with that criticism. Ever. I started doing field experiments at baseball card conventions. It'd be big halls, where people would walk in and buy, sell, and trade. In this case, people say, "Well, those are weird people. Your results will never generalize to other people." That was the argument. It was never about scaling. So I pushed back and said, "Why?" And they said, "Your idea does not have the silver bullet." What in the world does that mean? It doesn't have "the silver bullet"? Does anything? Maybe, I thought, they were talking about "Well, it doesn't have the Lebron James, or the Michael Jordan", but it felt like art to me. So I pushed them. I said, "I just don't understand." They said, "Look, all the experts come to us and say they have great interventions. And the ones that we choose, we scale those, and it's never close to what they promise." Now that right there caused me to pause. Chicago Heights is my petri dish, and I want to scale it up. What is the science behind the expectation to scale? What is the model that people are thinking about when they go from the small to the large?

That's when Dana comes in. Dana and I were sitting right down the street from Uber in 2016. I was a chief economist. Dana had also been thinking about scaling a lot. Dana works in early childhood. She's a surgeon. Dana is telling me, "Look, John. You need to take on this scaling thing. We need to add science to scaling." We sat there and talked at length about the work that was out there. I talked to a lot of businesses and a lot of governments. Where I ended up was, this was a state of play in scaling. You had some great idea, the innovation– right? "The pearls behind the swine," if you're into Biblical phrases. The innovation is great. Then, we move fast and break things. We throw spaghetti against the wall, and whatever sticks, you kick it. It's a gut feeling. Fake it till you make it. It's all art. This is simply art. I started to think, "Oh, my gosh, what if we started to work on an economic model, develop the economic theory, add data, and throw the economic science at the problem of scaling?" So that's what we did.

We started to write academic papers. Dana and I, and several other co-authors. We've now published dozens. I'm guessing not many have read one of our academic papers on scaling. When we write academic papers, we're lucky if two people read them. The editor and one of the three referees, and then we win. We wrote these academic papers, which is great. That's what we should be doing as academics. The problem is, the number of secrets and wisdom locked in academic

journals is enormous. We need more translators. We have to unlock our ideas and knowledge, and our stuff, you don't even want to read it. It's a bunch of math. It's a bunch of jargon. It's economese. If you don't speak economese, you can't read economese, and it's all wrapped up in a parochial paper that is written for a very narrow audience. There's too much of that. There is too much wisdom locked up, so I decided to write the book. I decided to translate our work to an audience that hopefully will read it and understand it.

The hardest part about the book business is getting people to open the book up. My worst enemy is a person who walks up and says, "I love your book. It's sitting next to me on my nightstand." And I say, "What chapter are you on?" "Oh, I'm about halfway through chapter one." And I say, "Well, when did you put it down?" "Oh, about three weeks ago." It's done. I'm done. There will be a new book that replaces mine, and they'll get halfway through the intro. It's still a tough business, but that's why I decided to write *The Voltage Effect* (2022). So, let's talk about *The Voltage Effect* (2022) now.

Let's ask a question; when the policymakers said, "Look, all these interventionists come to us and say they have something great," what do you think? Were they right or wrong? Well, that's what I call *the voltage effect*. It's the first law of scaling. The value proposition, whether it's the benefit-cost profile – however you want to define it, will change whenever you move from the small to the big, and in a predictable way.

This is where the engineers might be mad at me right now for my language. They want me to title my book, "The Wattage Effect," so then, instead of two people reading it, one would. Nobody wants to buy a book called "The Wattage Effect." So give me some artistic liberty. Here's a little bone for the engineers: high voltage is when you scale something up, and you reach a lot of different people, a lot of different situations. That's how I think about high voltage.

What about that "silver bullet" thing? Were they correct there? They got the first one right. They get the second one exactly wrong. This is not a best-shot problem. It's not a, have Michael Jordan, and you're good. This is an Anna Karenina Problem: scalable ideas are all alike. Each un-scalable idea is un-scalable in its own way. The first half of the book documents the five vital signs, which really represents the DNA of ideas that have a shot to scale. I'm not talking about execution yet. I'm talking about, does your idea have the DNA to even have a chance? In the second half of the book, we'll talk about four little behavioral economic secrets about execution.

Let me quickly go over these five vital signs. Vital sign number one, and it's amazing how many people fall prey to this. It was simply a false positive. This is ubiquitous among governments. I start off chapter one talking about Nancy Reagan, and her DARE program, which is one big fat false positive. Today, I want to talk about Tommy Lasorda. Remember Tommy Lasorda? Tommy Lasorda was the CEO of Chrysler. Tommy came to us at the University of Chicago and said, "We have a weight problem amongst our line workers. And it's costing us money: health care expenses, presenteeism, and absenteeism." He said, "Can you help us?" So, we put together a weight-loss program and we tested it in one plant. We killed it. Tommy immediately wanted to scale to the other 31 plants. Just like we all do when we get an initial set of results and we're really excited about it. We said, "Tom, let's just wait. Let's try it again in that plant. Let's replicate it, and then let's go to three new plants." Guess what happened when we did that? No result, no result, no result. It was simply a false positive. Two months later, Tommy got fired from Chrysler, and we ended up scaling a different plan with the new CEO. That's false positives.

Vital sign number two: know your audience. Do you know who Commander Spock is? He never gets it wrong. Ever. Let's put him on the side for a minute, and discuss smart thermostats. Engineers promised us that, if every household had a smart thermostat installed, we would take a big chunk out of the climate change problem. That's what their model said. So we gave it a go. We went to California and several thousand households signed up for a free, smart thermostat. We sent half of them the smart thermostat and kept the other half back. Then we observed them for three months, six months, nine months, and twelve months.

Guess how much energy they saved? Zero. The paper just came out two weeks ago in the NBER. What happened? The engineers assumed the end-user was Commander Spock. Folks, the end user is more like Homer Simpson. The person gets the smart thermostat, goes in, and fiddles with the pre-sets. They fiddle with the defaults and exactly undo all of the good stuff. As Dana knows, this is exactly what I did to ours. Dana threw me the 28-page manual. I got it. I looked at it, and when you get to my age, it seems like they write smaller and smaller. I looked at it and threw it back to her, and said, "Honey, I have this. Let me take care of it." What did I do? I undid all the presets. Know your audience.

Chapter three. Is it the chef or is it the ingredients? I think this chapter contains the richest part of the first half of the book because it combines horizontal scaling with vertical scaling. It talks about understanding, in the petri dish, what

are the most important features, or the non-negotiable features, that your idea has to have at scale?

After thermostats, I start out talking about restaurants. We all know a restaurant that kills it. It has $1 million in EBITDA. Then, they say, "Why not have ten or 20? We'll have ten million or 12 million of EBITDA, plus we'll make it to economies of scale." So they give it a go. I'm here to tell you today that, if that initial success was due to a unique chef, the idea will never scale. Unique humans don't scale. We try to teach unique skills. Then it doesn't work. If you can systematize it, you have a shot, but it's hard to systematize it. Think about Uber and Lyft. If we needed Danica Patrick, Jeff Gordon, or Michael Schumacher type of drivers, how's that going to work? It's not going to work until we systematize it.

We often do efficacy tests in the petri dish, because we want to give our idea its best shot. As academics, we are rewarded for big treatment effects. You get published in big journals and written up in the *New York Times* if you have big results. We do efficacy tests, and then we forget to tell everyone else that it was an efficacy test when we write it up. We are asking and answering the wrong question if we want to change the world with our science. We're asking the question "Can I create a program that can work in the petri dish, with the best inputs and a souped-up sample?" We really should be asking, "Can I create policy-based evidence?" What that means, if I'm talking about CHECC, is that I found out that I need good teachers at CHECC, in Chicago Heights. I only had to hire 30 teachers for my program. What would happen if I had to hire 30,000 good teachers, and kept the budget the same? What would happen to teacher quality? I didn't test for that. I did not oversample marginal teachers. I did not test the inputs that I needed to employ at scale.

In chapter three, I call that policy-based evidence. I think it's a very important reason why we've been working on development economics, inner-city schools, and discrimination for decades. We are asking and answering the wrong question. For business types, really what I'm talking about is situationally-congruent evidence. What will you face at scale?

Chapter four: unintended consequences and spillovers. Do you remember Ralph Nader? He wrote a popular book in '65. What was he complaining about? Highway safety. He complains so much and had such a deep impact that, by the time 1968 rolled around, the federal government mandated that every new automobile had to have a seat belt installed. Now young people are rolling their eyes, because they're, like, "Oh, my goodness, that was back in the cave-person days.

What do they mean, no seat belt?" That's what I'm talking about. No seat belt.

My colleague, Sam Peltzman, did an economic analysis that was published in 1975 that assessed how many lives were saved by that seat belt law.[3] Guess what Sam estimated in his 1975 paper? Zero. What happened? People wearing seat belts drove more aggressively. Even though they died a little bit less often, they hit people without seat belts, and they died a little bit more often.

That's a kind of spillover at the individual level. Another kind of spillover happened to me when I was the chief economist at Uber. Turn back the clock to January 27th, 2017. President Trump issued an executive order on immigration. Remember that night? People went nuts. Taxi cab drivers around JFK went on strike. Uber, whenever something like that happens in a market, a market disruption, you know what Uber does? They turn off surge charges. They don't want to be viewed as price gaugers. So, they turned off surge charges.

Nevertheless, a taxi cab driver thought that Uber turned off surge charges to try to break up the strike. He went on a Twitter rampage that ended with the hashtag #deleteuber. And overnight, the market share that Lyft had– remember, the main competitor to Uber is Lyft– Lyft had 5% market share Saturday morning. Sunday evening, they had 30% market share. Overnight, the hashtag, #deleteuber, gave Lyft a lifeline. Travis Kalanick came to me and my team. Travis said, "Look, John, your team is responsible for getting the drivers back." What does an economist do when somebody says, "Your job is to get the drivers back." What do you think my big idea was? Cash. Pay them more. I said, "Let's introduce tipping." Drivers want tipping. A lot of you might not remember this, but back then, there were tin cups in a lot of people's backseats. Customers didn't like it. Drivers thought they were not receiving the tips that they wanted. Everyone wanted tipping. All the executives said, no, right away. That was in early February 2017. I went door to door, saying, "We need tipping. All the drivers want tipping." I eventually won that battle. When you win a battle like that at Uber, guess what the booty is? The booty is, your team gets to roll out tipping.

A field experimentalist being told, "You get to roll out tipping." so what do I do? I do it as a massive field experiment where I started working in the summer of 2017 on little pilots. I did one right here in Chicago, where I took 5% of the drivers in the market and said, "You get to receive tips. The other 95% don't." Then I observe them. They made more money. They worked more. Win, win, but when

3 https://www.journals.uchicago.edu/doi/abs/10.1086/260352.

we rolled it out that October to all drivers in Chicago, guess what happened? All the drivers worked more. New drivers came in. They came in so much, and they worked so much more, that the wage effect that I observed in the summer went away completely. The market came to a new equilibrium where drivers would work more but drove around with empty cars more often. Their hourly wage was exactly the same as it was, pre-tipping. That's a market-wide spillover, that sometimes our ideas have. So, chapter four is about spillovers.

Chapter five is going to be left behind closed doors. I am an economist after all, and I understand incentives. So I don't want to give everything away. Let me quickly talk about the last part of the book.

For those that said yes to using Uber and/or Lyft, let me ask another question. How many of you tip your driver every time on Uber or Lyft? Guess what? Only 1% of customers tip on every trip. One percent. Three out of five people never, ever tip. Ever. But, guess what happens when I take those three out of five people who never tip on Uber or Lyft, and put them in a yellow cab, where they have to settle up afterward, face to face? Now guess how many of them tip? Ninety-five percent. Three out of five tip zero. Take those three out, they always tip face-to-face. Those are non-financial incentives– social pressure, social image, and social norms. These types of incentives are great incentives that scale. I talk about those and other behavioral economic incentives in chapter six.

Chapter seven hearkened me back to the classroom. We always tell our students, "Think on the margin. Be marginal thinkers. Not average thinkers." They see how to do it mathematically, but when they go out to the real world, they have no idea how to apply it. In this chapter about marginal thinking, let me tell you a little bit about what happened at Lyft.

I was the chief economist at Lyft for four years. The driver acquisition team came to me and said, "Look, Logan Green, the CEO and founder of Lyft, has given us money to bring in new drivers." I said, "Okay. Tell me what you have." They said, "Well when we place ads on Facebook, it costs us $300 on average to bring in the last thousand drivers." Okay. Then they said, the last thousand drivers using Google ads actually cost $400 each. I said, "What do you plan on doing?" They said, "Well, of course, we're going to place the ads on Facebook." I said, "Well, let's take a little thinner slice of the data. Tell me a little bit about, say, the last 25 drivers." They said, "Well, we don't have that just yet." I said, "Send it to me tonight."

Here's what they sent me. They said, "Professor, the last 25 drivers on Facebook ads actually cost us $1,000 each, and on Google, it cost us $500 each." They said, "We get your point. We wish we could go back in time and take money from Facebook to Google. Going forward now, for a while, we're going to use Google ads." Big data is great for precision, but when you add big data from a different regime, it causes you to think on the margin less often.

Chapter eight: winners quit. There are a lot of old-schoolers in the room that may have liked me up until now, but I'm telling you, we do not quit enough. We don't quit enough for two reasons: one, society tells us that quitting is repugnant. I was born and raised in the land of Vince Lombardi. "Winners never quit, quitters never win." Type in, "Quitting inspirational posters" in Google, and you will have enough posters to fill every museum in the world. Right? We don't quit. It's repugnant.

Another reason why we don't quit enough is because we neglect our opportunity cost of time. It's just human nature. Now that's a lot of economese, so let me quickly spell that out for you. I did a big survey on recent job quitters. I said, "Why did you quit your job?" Reason number one: I lost the meaning of work. Reason number two: I didn't get the promotion I deserved. Reason number three: I didn't get the pay raise I thought I deserved. Dot, dot, dot, all the way down to, I didn't like my cubicle anymore. Every reason was, my current lot in life got soiled. It was never, my opportunity set got better. That's why I left Lyft. I love Lyft. But my opportunity set got better. Walmart came to me, so I left. There's science in this chapter. The inspirational quotes on posters? That's art.

Chapter nine is about scaling culture. This was fun because I've been working on the gender pay gap for a long time. I've worked on inclusiveness, equity, et cetera. I learned a lot at Uber. As you might imagine, a lot of stuff went down at Uber. When I moved to Lyft, I really didn't appreciate culture until I lived it. I saw Uber's culture. I saw Lyft's culture. Now, I see a very different culture at Walmart. It all starts at the top. The people who are working day to day, doing data science, or marketing, or management, or what have you, they're following the cues and the leadership from the group of executives.

It happens in selection. So I talk here a lot about selection. Once you hire a rotten apple, it's really hard to change a rotten apple to a good apple. I talk about incentives to try to do that, but a lot of times, if you can't contract everything, the rotten apple's still rotten. So, that's chapter nine.

That's *The Voltage Effect* (2022). Thank you so much for your attention.

Epilogue

Optimizing the wealth and well-being of citizens is a meritorious goal. When markets must be perturbed to maximize that well being, science should be a key contributor. The goal of scientific research is to create a stock of knowledge on important issues, be them of concern to theorists, practitioners, or policymakers. Adding to this stock requires one to take a step back and examine how priors should change in light of empirical findings. What conclusions can we draw upon observing a statistically significant result? More generally, what should we consider to be standards of evidence, and what is the framework of proper inference given our research data? When is the program scalable? In this lecture, I have proposed a framework whereby we can begin to use science to determine the nature and extent of scalable ideas. Only then will we truly be on a path to fulfill the promise the Upton Forum strives to achieve.

References

List, John A.. 2022. "*The Voltage Effect*." London: Penguin Business.
.2023. "The Voltage Effect." *Business Economics.*
https://doi.org/10.1057/s11369-022-00298-9

Toward an Understanding of John List: Field Experiments from Children to Charities

Anya Samek [1, 2]

1. My introduction to field experiments

I met John List when I was a Ph.D. student at Purdue University in 2008. John had come down from the University of Chicago to give a talk about one of his recent field experiments in which he provided employees with incentives to lose weight. Up until that point, I had not had much exposure to field experiments – at Purdue, the focus in the experimental economics group had been on using the laboratory alone. The seminar was memorable for me because it helped me see how experiments could be used in a new context, one which had direct policy and practical relevance.

About a year later, I heard John give a keynote at a conference of the Economic Science Association (ESA). Field experiments were still relatively new then, and I could see that many in the audience were as impressed by both John and his research as I was. I was on the job market that year, and John offered me a postdoctoral scholar position in his group at the University of Chicago. I saw this as my opportunity to get the training in field experiments that I had not gotten during my PhD. This was the beginning of my decade-long collaboration with John.

In this article, I highlight two areas of John's research using field experiments (much of it joint with me) that illuminate the nature and causes of wealth and

1 Anya Samek is an Associate Professor of Economics and Strategy at the Rady School of Management at University of California-San Diego. She is also a Faculty Research Fellow at the National Bureau of Economic Research and an affiliate of the FAIR Centre at the Norwegian School of Economics.
2 I thank Shruti Jha and Adeline Sutton for valuable research assistance. I thank participants of the Beloit College Upton Forum 2022 for helpful comments.

well-being of nations. These two areas are children and charity. Over the past two decades, experimental economics has seen a large increase in the use of field experiments, and John has been at the forefront of this work. Field experiments bridge the stylized environment of the laboratory and the context-rich environment of the outside world.

The study of children can help explain the nature and causes of nations' wealth and well-being in several ways. For example, we can explore the background characteristics, schooling, and parenting inputs that shape child behavior and outcomes, which in turn affect wealth and wellbeing. Further, we can evaluate the impact of interventions that seek to improve wealth and well-being of children in adulthood (List, Samek, and Suskind, 2018). For example, we can consider how to improve proxies for higher educational attainment, which may in turn affect wealth and well-being (Tamborini et al., 2015).

The study of charitable giving can help us understand wealth and well-being since the way in which individuals and corporations give affects the distribution of resources in society, which can lead to changes in wealth and well-being.

In what follows, Section 2 provides an overview of what I have learned about field experiments during my time working with John. Section 3 discusses the broader impact of our field experiments with children. Section 4 discusses the aims and outcomes of our work using field experiments in charitable giving. Section 5 concludes with directions for future research in these two areas.

2. Learning by doing: Tips on implementation

I defended my Ph.D. dissertation in April of 2010 and moved to the University of Chicago to begin my position as a postdoctoral scholar. I was working at the Chicago Heights Early Childhood Center (CHECC), a large-scale field experiment testing the impact of different early childhood interventions on reducing the academic achievement gap (Fryer et al., 2015; Fryer et al., 2020). Over the following four years, CHECC would randomize over 2,000 low- and moderate-socio-economic-status (SES) households to preschool and parenting programs.

One goal of CHECC was to study the short- and long-term impact of early childhood programs on children's developmental trajectories. CHECC joined a long literature exploring these questions, including follow-up studies of the Perry Preschool Project and the Abecedarian Project (Conti et al., 2016). Another goal was to treat CHECC as a "living lab" and work with child and parent participants

of the program to understand the development of economic preferences and economic decision-making of young children.

I had spent the prior 5 years of my Ph.D. at Purdue in a quiet laboratory watching undergraduates make decisions at computer terminals. I would spend the next 5 years in busy school hallways trying to understand how best to implement field experiments with young children. I learned everything I now know about conducting field experiments in those school hallways and in meetings with John. I discuss my take-aways on implementation below:

No two field experiments are the same: Most laboratory experiments follow a similar protocol, from the initial subject recruitment to the subject payouts. In the field, however, the population, organizational partner or context will dictate the protocols that should be followed. When we started CHECC, we conducted pilots to understand what does and does not work, and what children do and do not understand, in various protocols. Specific to the study of children in economics experiments, we have now written an article in which we discuss the commonly used methods for eliciting economic preferences and studying the decision-making of children ages 3-17 years old (List et al., 2022). In that paper, we also discuss the developmental trajectories of children to help inform the types of elicitation tools that are most appropriate for each age group.

Things go wrong: In the laboratory, since protocols are standard, it is unusual for mistakes to happen. In the field, on the other hand, it is unusual for things to go exactly as planned. Hence, I always develop contingency plans, and create protocols with multiple points of data collection. For example, we need to make backup plans for if the internet goes down, if there is a winter storm that causes a school cancellation or if a field partner backs out at the last minute. We also need to devise protocols with the least room for mistakes, e.g., we pre-print labels with subject IDs to minimize user error when recording information.

Communication is key: Being able to communicate well is important. Researchers need to be able to speak both with field partners and with subjects who are not university students. Understanding how to convey ideas that involve economic jargon to laypeople makes the difference between being able to run your field experiment or not. For example, most laypeople are not willing to participate in "experiments" but would be fine participating in "pilots" or "studies." Changing the terminology can help to get ideas across. I discuss my further ideas for communicating well in Samek (2019a) and Carroll and Samek (2018, with reference to grocery stores). John has an excellent paper that incorporates commu-

nicating with field partners as well (List, 2011).[3]

Pay attention to the data analysis: John has also led the way from the perspective of bringing a rigorous data analysis methodology to our field. In the past decade, across several papers, John argues for the importance of conducting a power test (List et al., 2011) adjusting statistics for multiple hypothesis testing (List et al., 2019), scaling (Al-Ubaydli et al., 2019), replication (Maniadis et al., 2014; Maniadis et al., 2017; Brandon & List, 2015) and generalizability (Levitt & List, 2007; Al-Ubaydli & List, 2013).[4] Taking into account these considerations is important in order to ensure that the conclusions from the experiment are justified by the data.

On doing everything: To date, John has written 259 peer-reviewed articles, spanning topics such as education, pricing behavior, discrimination in the marketplace, the valuation of non-marketed goods and services, public goods provision, and charitable giving. I have also adapted John's eagerness to learn by conducting experiments on topics as varied as child development and education, health and health-promoting behaviors, charitable giving and financial literacy. This approach has both advantages and disadvantages. One advantage is that it allows us to study underlying causes of behavior across contexts – e.g., the same motive likely drives individuals to eat less healthy (Sadoff et al., 2020) and to procrastinate enrolling in health insurance (Samek et al., 2022). Another is that it can be more interesting to work in different domains. However, there are also disadvantages. For example, it can be time-consuming to keep up with literature across domains, and this means that every new project requires substantial preparation.

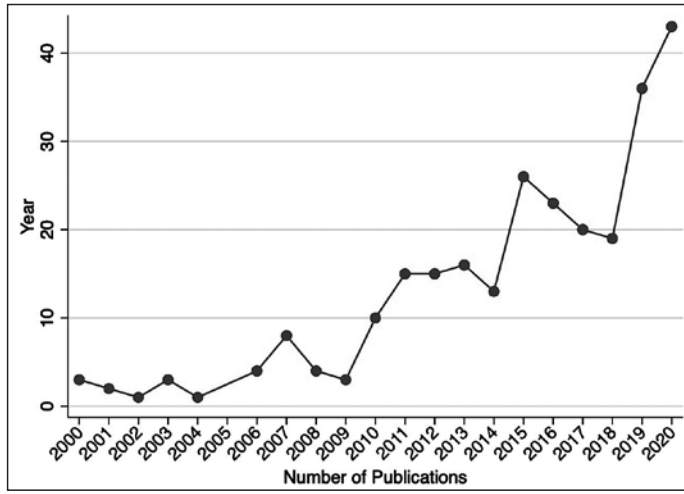
3. Field experiments at CHECC: What have we learned?

There has been an increase in interest in conducting economics experiments with children, and part of that increase can be attributed to publications using subjects from CHECC starting in 2010. Figure 1 is reprinted from List et al. (2022) and shows the increase in economics experiments with children.

3 More broadly, it is my opinion that communicating well is a skill that we as a profession tend to under-invest in. Communicating your research well to other academics both in talks and in written form is also important.

4 John was also a collaborator on Benjamin et al. (2017) on redefining statistical significance.

Figure 1: Growth in Economic Experiments with Children



Notes: This figure is re-printed from List et al. (2022). It shows the number of publications that use economics experiments each year from 2000 to 2021. CHECC started in 2010 and much of this increase can be attributed to experiments that use CHECC child subjects.

Why should economists conduct field experiments with children? In List et al. (2022) we classify our explanation into two silos. The first silo is that economists must understand children themselves for immediate purposes. For example, we may wish to know how educational interventions affect children (Sadoff, 2014; Lavecchia et al., 2016). We may also wish to understand the "sensitive periods" for timing interventions (Breitkopf et al., 2022) and understand who is helped most by interventions (e.g., as in Datar et al., 2022). The second silo is that children can inform our understanding of adults, who make the primary decisions in most markets. Experiments that fall into this silo can be thought of as complementary to methods such as neuroeconomics, which correlate decision-making with brain activity.

3.1 Evidence for the First Silo

In the spirit of the first silo, CHECC has taught us about the short- and medium- term impact of early childhood programs on schooling trajectories of participants. This is important since we know that schooling outcomes impact educational attainment, which may in turn affect wealth and well-being (Assari, 2019).

In Fryer et al. (2015), John and co-authors show that an intervention that teaches parents how to teach to their children and gives parents incentives to teach to their children results in improved executive functions of the treatment group children relative to the control group children. In Fryer et al. (2020), we show that preschool programs that target children directly improve children's' cognitive – or academic - skills. The children most helped by the preschool are those who have the lowest cognitive and executive functioning skills at baseline, suggesting that CHECC programs have the capacity to reduce academic achievement gaps. Finally, CHECC impacts brain activity as measured by electroencephalography (EEG). In Ye et al. (2022), children assigned to the treatment had greater brain activity related to executive functioning than children in the control group, and this brain activity was predictive of executive functioning skills for up to three years after the intervention.

CHECC was designed to allow us to consider the potential spillovers onto untreated children. In List et al. (2020), the authors document large spillover effects on control children who live in neighborhoods near treated children. Specifically, spillover effect on non-cognitive functions operates through the child's social network, while parental investment is an important channel through which cognitive spillover effects operate. Together with a broader literature on parental and peer effects, this suggests that human capital accumulation is not undertaken in isolation and is inherently a social activity (see also, e.g., Cochran and Brassard, 1979; Coleman, 1988; Corsaro, 2005).

Another spillover that we can observe is effects on untreated siblings of treated children. In Chuan et al. (2021) we find that immediate incentives increase the willingness of parents to invest in the target child, but they decrease investments in untreated children. We conclude that policymakers should be careful when implementing programs aimed at reducing societal inequalities, since they can inadvertently create intra-household inequalities in the sense of changing relative parental investments.

John and I have also worked in the schools surrounding CHECC to understand what types of interventions affect the nutritional intake of children. This work has mostly taken the form of testing the impact of nudges and incentives. In List and Samek (2015a), we show that incentives increase the likelihood that children choose a healthier dessert relative to the less healthy dessert in an after-school program. In List and Samek (2015b) we show that incentives also increase the likelihood of choosing a healthier milk in the school lunch-line.

I was inspired by John to continue this line of research separately as well. In Samek (2019b) I show that asking children to set goals increases the likelihood of choosing the healthier milk in the school lunch line. In Angelucci et al. (2019), my co-authors highlight important spillover effects of incentives in food choice. This work highlights a key feature of this first silo: we can use field experiments to understand the formation of health capital, which also has direct relevance for well-being.

3.2 Evidence for the Second Silo

In the spirit of the second silo, the CHECC "living lab" involved 36 co-authors and 14 publications to date and explored questions such as the causes and consequences of heterogeneities in social, time and risk preferences of young children, the developmental trajectories of child preferences and the impact of educational interventions on child preferences. In List et al. (2022), we provide a summary of the findings of these papers and the related literature on the development of economic preferences among children. I do not replicate this summary below. Instead, I provide some highlights that I think are of relevance to the nature and causes of wealth and well-being.

One important consideration when evaluating the nature and causes of wealth and well-being is the extent of intergenerational transmission of behaviors and preferences, since these affect field outcomes. As such, at CHECC we have taken advantage of the fact that parents also participate in our programs to evaluate the correlations of behavior of children and their parents. Among very young children ages 3-5 years old, we do not find a correlation between parent patience and child preferences (Andreoni et al., 2017). Related work on patience and intergenerational transmission is mixed, with some papers showing a correlation between child preferences and parent preferences and others not finding one (Bettinger and Slonim, 2007; Kosse and Pfeiffer, 2012).

I was inspired by John to extend my work on children to another project

with different collaborators using adolescents, and there we do find a correlation between parent and adolescent patience and risk preferences (Samek et al., 2019). In the related literature, there does not seem to be a correlation of altruism between children and parents (Bettinger and Slonim, 2006; Ben-Ner et al., 2017). There does seem to be a correlation of risk preferences between children and parents (Alan et al., 2017; Chowdhury et al., 2018; Samek et al., 2019). We interpret the above results to mean that parents do transmit their economic preferences to their children, but evidence is stronger when comparing parents to their older – or even adult, e.g., see Kimball et al. (2009), Dohmen et al. (2011) -- children.

Our work in which we observe parent decisions when the parent knows that their decisions will be visible to their child provides evidence that parents do seek to transmit their economic preferences to their children. In Houser et al. (2016), we experimentally vary whether the child is in the room with the parent when the parent has an opportunity to cheat in an experimental task. We find that the presence of daughters – but not sons – affects willingness to cheat. In particular, parents cheat less in front of daughters than in front of sons and cheat less in front of daughters than when they are alone. In Ben-Ner et al. (2017) we conduct a series of dictator games with parents and children separately and experimentally vary whether parents are told that their child will see the parent's decision prior to making their own decision. We find that some sub-sets of parents (fathers, and parents of generous children) model generous behavior when they know that their child will see their decision.

Another question in this literature focuses on the developmental trajectories of economic preferences among children. For example, in Andreoni et al. (2019), we study the evolution of time preferences of young children. We show that 3-year-old children are slightly more patient than 4-5-year-old children, but that patience increases with age from 5-12 years old. Sutter et al. (2019) provides a broad summary of how time, risk and social preferences develop with age.

Finally, we consider the role of economic preferences on shaping field outcomes. In Castillo et al. (2022), we show that child time preferences measured at the early age of 3-5 years old predict disciplinary referrals from grades 3-8. Further, time preferences continue to be predictive of disciplinary referrals even when controlling for early life cognitive skills and executive functions. This paper suggests that time preferences are a separate factor that shapes child outcomes, and highlights the importance of considering time preferences when exploring the causes of certain outcomes.

4. Field experiments in charitable giving

In 2015, John and I launched the Science of Philanthropy Initiative (SPI) (together with Michael Price). The dual goals of this initiative were to reduce the frictions for researchers who wish to conduct field experiments with charitable organizations and to highlight for charitable fundraisers the importance of conducting field experiments.

We sought to meet the first goal of reducing frictions in field experiment research by hiring a dedicated project manager, whose role it was to meet with and interview prospective charities and match them to interested researchers, including those within and outside of our networks. We also solicited grant applications from junior and senior researchers so that we could fund some of this work.

Finally, we conducted an annual conference for both academics and practitioners (i.e., fundraising professionals). During this conference, we presented information to academics about how best to communicate with field partners. This assistance with communication was important to us since, as I noted earlier, communicating well is one of my tips for conducting field experiments. We also incorporated networking opportunities to allow for natural collaborations to emerge between academics and practitioners.

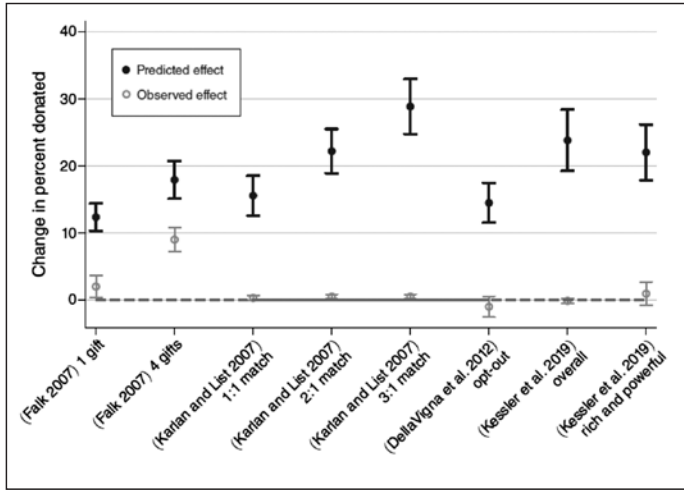
We sought to meet the second goal of highlighting the importance of conducting field experiments to charities through invitations to our annual meeting, webinars and providing resources about the results of our experiments on a dedicated website (e.g., see Jasper and Samek, 2014). The charitable sector represents an important component of economic activity, with over $300 billion individual donations to charity in the United States each year (Giving USA Foundation, 2022). Field experiments have been conducted that explore matching grants (Eckel and Grossman, 2003; Meier, 2007) , challenge gifts (List and Lucking-Reiley, 2002; Rondeau and List, 2008) , donor gifts or premiums (Eckel et al., 2017), suggested amounts (Edwards and List, 2014; Reiley and Samek, 2015) , lotteries (Landry et al., 2006), recognition (Soetevent, 2005), information about peers (Shang and Croson, 2009), signals of support (Kessler and Milkman, 2016), pay-what-you-want pricing (Gneezy et al., 2010), overhead aversion (Gneezy, et al., 2014), directed giving (Eckel et al., 2017), and the "gift" of an opt- out (Kamdar et al., 2015).

We believe that field experiments are important to fundraising in part due to my recent finding that fundraising professionals are not good at predicting the

results of many field experiments. In Samek and Longfield (2023), my co-author Chuck Longfield and I write about a survey we conducted with fundraising professionals. In this survey, we described the protocols of several published articles that used field experiments in a fundraising campaign to evaluate the impact of a solicitation technique. The solicitation techniques ranged from thank-you calls, donor gifts, matching grants, information about the pay status of the solicitor and donor agency. We gave the fundraising professionals information about the donation rates in the control group and asked them to predict the donation rates in the treatment group(s).

The results are summarized in Figure 2, reprinted from Samek and Longfield (2023). The gray bars show the difference in the percentage of the potential donors who donate when comparing the treatment and control groups. For example, the gray bars that are near zero in the matching (1:1. 2:1, and 3:1), opt-out and overall experiments, suggests that the fundraising technique used in those experiments did not have a substantial impact on the probability of donating. However, the black bars that are mostly above 10% suggest that the fundraising professionals believed that the solicitation techniques would result in a 10-percentage point or more increase in donation rates. This substantial over-prediction across almost all categories of fundraising techniques that we study suggests that field experiments would provide useful new information to fundraising professionals.

Figure 2: Fundraisers Over-Estimate Impact of Fundraising Techniques



Note: This figure is re-printed from Samek and Longfield (2023). The horizontal axis labels represent the experiment in the articles, that we asked fundraising professionals to predict the effects of. These include impact on donation rates of sending one or four postcards as a gift (1 gift or 4 gifts) to prospective donors (Falk, 2007); matching gifts with different match rates (Karlan and List, 2007); opportunities to opt-out of solicitation (DellaVigna et al., 2012); and being given a sense of agency over the use of donated funds to a sample of donors (overall) and a subsample of rich and powerful donors (Kessler et al., 2019).

5. Conclusion

Understanding the nature and causes of wealth and well-being of nations is an ongoing endeavor. I end this article with questions that have yet to be answered and directions for future research. Field experiments are critical for understanding the nature and causes of wealth and well-being of nations. This is because field experiments allow us to answer causal questions – i.e., we can ask what types of interventions can improve wealth and well-being. We can use field experiments across many contexts, including with charities and children as I describe in this article.

With respect to children, while all the work with children would be considered field experiments, a large portion of the work has generated correlational and not causal insights. For example, researchers have studied associations between child economic preferences and SES and child economic preferences and parent economic preferences. However, few papers have used randomization to evaluate the causal impact of parents or institutions on behaviors and economic preferences. A few notable exceptions are Capellen et al. (2019) and Kosse et al. (2020), which evaluate the impact of schooling and mentoring programs on social preferences. More work is needed that includes such randomization and evaluates the impact of institutions on economic preferences.

With respect to charity, we need a unified framework that will help fundraisers understand which fundraising techniques are most effective and the level of increase that they can expect from various techniques. The fundraising experiments to date have evaluated economically interesting questions but have not done so in a way that would allow for such a framework to emerge. For example, each of these experiments is in a different context with donors who potentially have different motivations from one another. Ongoing collaborations between academics and charities, though difficult, can create more overlap across studies.

Further, a framework that seeks to understand the factors that affect the external validity of different charitable contexts and the extent to which each context can be generalized would be helpful.

Broadly, we need to continue conducting field experiments that collect data on long-term outcomes or explore the impact of longer-term interventions in behavioral economics. Many of the studies with both children and charities have used a one-shot intervention and/or a single evaluation point. Studies have been implemented in the education field (e.g., see Jackson et al., 2020; Algan et al, 2022). Interventions in the health domain have previously been fairly short (e.g., a few days to a few weeks as in Samek and List, 2015a, 2015b). While longer-term interventions are costly, they are potentially more meaningful and more likely to affect economically relevant outcomes.

There is continued increase in interest in field experiments with children and charities. I have no doubt that this work will continue to be important and look forward to learning more as the fields mature.

References

Alan, S., Baydar, N., Boneva, T., Crossley, T. F., and Ertac, S. (2017). Transmission of Risk Preferences from Mothers to Daughters. *Journal of Economic Behavior & Organization*, 134, 60–77.

Algan, Y., Beasley, E., Cote, S., Park, J., Tremblay, R. E., and Vitaro, F. (2022). The impact of childhood social skills and self-control training on economic and noneconomic outcomes: Evidence from a randomized experiment using administrative data. *American Economic Review*, 112(8), 2553-2579.

Al-Ubaydli, O., & List, J. A. (2013). On the generalizability of experimental results in economics: With a response to camerer. In G. Frechette & A. Schotter (Eds.), Methods of modern experimental economics. *Oxford University Press*.

Al-Ubaydli, O., List, J., and Suskind, D. (2019). The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments. NBER Working Paper 25848. Cambridge, MA: *National Bureau of Economic Research*. https://doi.org/10.3386/w25848

Andreoni, J., Samek A., List, J., and Sprenger, C. (2017). Field Experiments on the Development of Time Preferences. *The Field Experiments Website*, no. 00615.

Andreoni, J., Kuhn, M. A., List, J. A., Samek, A., Sokal, K., & Sprenger, C. (2019). Toward an understanding of the development of time preferences: Evidence from field experiments. *Journal of Public Economics*, 177, 104039. https://doi.org/10.1016/j.jpubeco.2019.06.007

Angelucci, M., Prina, S., Royer, H., and Samek, A. (2019) Incentives and Unintended Consequences: Spillover Effects in Food Choice. *American Economic Journal: Economic Policy*, 11 (4): 66-95. https://doi.org/10.1257/pol.20170588

Assari S. (2019). Race, Education Attainment, and Happiness in the United States. *International Journal of Epidemiologic Research*, 6(2), 76–82. https://doi.org/10.15171/ijer.2019.14

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., et al. (2017). *Redefine Statistical Significance. Nature and Human Behavior*, 1, 0189. doi:10.1038/s41562

Ben-Ner, A., List, J., Putterman, L., & Samek, A. (2017). Learned generosity? An artefactual field experiment with parents and their children. *Journal of Economic Behavior and Organization*, 143, 28-44. https://doi.org/10.1016/j.jebo.2017.07.030

Bettinger, E. & Slonim, R., (2006) Using experimental economics to measure

the effects of a natural educational experiment on altruism. *Journal of Public Economics*, Elsevier, 90(8-9), 1625-1648

Bettinger, E., and Slonim, R. (2007) Patience among children. *Journal of Public Economics* 91.1, 343-363

Brandon, A., & List, J. A. (2015). Markets for replication. *Proceedings of the National Academy of Sciences*, 112(50), 15267-15268.

Breitkopf, L., Chowdhury, S., Kamhöfer, D., Schildberg-Horisch, H., Sutter, M. 2022. Sensitive Periods in the Formation of Socio-Emotional Skills: Evidence from a Randomized Trial. Working Paper.

Cappelen, A., List, J., Samek, A., & Tungodden, B. (2020). The Effect of Early Childhood Education on Social Preferences. *The Journal of Political Economy*, 128(7), 2739–2758. https://doi.org/10.1086/706858

Carroll, K.A., & Samek, A. (2018). Field experiments on food choice in grocery stores: A 'how-to' guide. Food Policy. https://doi.org/10.1016/j.foodpol.2018.03.004

Chowdhury, S., Sutter, M., and Zimmermann, K. (2018). Evaluating Intergenerational Persistence of Economic Preferences: A Large Scale Experiment with Families in Bangladesh. SSRN Electronic Journal.

Chuan, A., List, J., and Samek, A. (2021). Do Financial Incentives Aimed at Decreasing Interhousehold Inequality Increase Intrahousehold Inequality? *Journal of Public Economics*, 196, 104382.

Cochran, M.M. and Brassard, J.A. (1979). Child development and personal social networks. Child Development 50(3), 601–616.

Coleman, J.S. (1988). Social capital in the creation of human capital. *American Journal of Sociology* 94, S95–S120.

Conti, G., Heckman, J., & Pinto, R. (2016). The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour. Economic journal (London, England), 126(596), F28–F65. https://doi.org/10.1111/ecoj.12420

Corsaro, W.A. (2005). The Sociology of Childhood, 2nd edition. London: Pine Forge Press.

Datar, A., Nicosia, N., and Samek, A. (2022) Heterogeneity in Place Effects on Health: The Case of Time Preferences and Adolescent Obesity. NBER Working Paper No. w29935. https://doi.org/10.3386/w29935

DellaVigna, S., List, J., and Malmendier. U. (2012). Testing for Altruism and Social Pressure in Charitable Giving. *Quarterly Journal of Economics*, 127 (1), 1–56

Dohmen, T., Falk A., Huffman, D., and Sunde, U. (2011). The Intergenerational

Transmission of Risk and Trust Attitudes. *The Review of Economic Studies* 79 (2), 645–77.

Eckel, C. C., and Grossman, P.J. (2003). Rebate versus Matching: Does How We Subsidize Charitable Contributions Matter?. *Journal of Public Economics*, 87 (3–4), 681–701.

Eckel, C. C., Herberich, D. H., and Meer, J. (2017). A Field Experiment on Directed Giving at a Public University. *Journal of Behavioral and Experimental Economics*, 66, 66–71.

Eckel, C. C., Herberich, D.H., and Meer, J. (2017). A Field Experiment on Directed Giving at a Public University. *Journal of Behavioral and Experimental Economics*, 66, 66–71.

Edwards, J. T., and List, J. (2014). Toward an Understanding of Why Suggestions Work in Charitable Fundraising: Theory and Evidence from a Natural Field Experiment. *Journal of Public Economics*, 114, 1–13.

Falk, A. (2007). Gift Exchange in the Field. *Econometrica*, 75 (5), 1501–11.

Fryer Jr, R. G., Levitt, S. D., and List, J., (2015). Parental Incentives and Early Childhood Achievement: A Field Experiment in Chicago Heights. *National Bureau of Economic Research*, no. w21477. https://doi.org/10.3386/w21477

Fryer Jr, R. G., Levitt, S. D., List, J., and Samek, S. (2020). Towards an Understanding of What Works in Preschool Education. Work in Progress.

Giving USA Foundation (2022). Giving USA 2022: The annual report on philanthropy for the year 2021.

Gneezy, A., Gneezy, U., Nelson, L., and Brown, A. (2010). Shared Social Responsibility: A Field Experiment in Pay-What-You-Want Pricing and Charitable Giving. Science, 329 (5989), 325– 27.

Gneezy, U., Keenan, E., and Gneezy, A. (2014). Avoiding Overhead Aversion in Charity. Science, 346 (6209), 632–35.

Houser, D., List, J., Piovesan, M, Samek, A., Winter, J. (2016). Dishonesty: from parents to children. *European Economic Review*, 82, 242-254

Jackson, C. K., Porter, S. C., Easton, J. Q., Blanchard, A., and Kiguel, S. (2020). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review*: Insights, 2(4), 491-508.

Jasper, C. and Samek, A. (2015). Increasing charitable giving in the developed world. *Oxford Review of Economic Policy*, 30, 680-696. https://doi.org/10.1093/oxrep/gru032.

Kamdar, A., Levitt, S., List, J. Mullaney, B., and Syverson, C. (2015). Once and Done: Leveraging Behavioral Economics to Increase Charitable Contributions. SPI WorkingPaper 025.

Karlan, D., and List, J. (2007). Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment. *American Economic Review*, 97 (5), 1774–93.

Kessler, J. B., and Milkman, K. (2016). Identity in Charitable Giving. Management Science, 64 (2), 845–59.

Kessler, J. B., Milkman, K. L., and Zhang, C. Y. (2019). Getting the Rich and Powerful to Give. Management Science, 65 (9), 4049–62. https://doi.org/10.1287/mnsc.2018.3142.

Kimball, M. S., Sahm, C. R., and Shapiro, M. D. (2009). Risk preferences in the PSID: individual imputations and family covariation. *American Economic Review* 99 (2), 363-68.

Kosse, F., and Pfeiffer, F. (2012). Impatience among Preschool Children and Their Mothers. Economics Letters 115 (3), 493–95

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., and Falk, A. (2020). The Formation of Prosociality: Causal Evidence on the Role of Social Environment. *Journal of Political Economy*, University of Chicago Press, 128(2), 434-467.

Landry, C.E., Lange, A., List, J. Price, M, and Rupp, N. (2006). Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment. *Quarterly Journal of Economics*, 121 (2), 747–82.

Lavecchia, A., Liu, H., and Oreopoulos, P. (2016) Chapter 1 - Behavioral Economics of Education: Progress and Possibilities. *Handbook of the Economics of Education*. 5, 1–74.

List, J., and Lucking-Reiley, D. (2002). The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign. *Journal of Political Economy*, 110 (1), 215–33.

List, J., and Samek, A. (2015a). The behavioralist as nutritionist: leveraging behavioral economics to improve child food choice and consumption. *Journal of Health Economics*, 39, 135–146. https://doi.org/10.1016/j.jhealeco.2014.11.002

List, J., and Samek, A. (2015b). A Field Experiment on the Impact of Incentives on Milk Choice in the Lunchroom. *Public Finance Review*, 45(1), 44–67. https://doi.org/10.1177/1091142115602061

List, J., Momeni, F., & and Zenou, Y. (2020) The Social Side of Early Human Capital Formation: Using a Field Experiment to Estimate the Causal Impact of Neighborhoods. *National Bureau of Economic Research*. Working Paper 28283. https://doi.org/10.3386/w28283

List, J., Petrie, R., & Samek, A. (2022, in press). How Experiments with Children

Inform Economics. *Journal of Economic Literature*.

List, J., Sadoff, S., & Wagner, M. (2011). So You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design. *Experimental Economics.* https://doi.org/10.1007/s10683-011-9275-7

List, J., Samek, A., & Suskind, D. (2018). Combining behavioral economics and field experiments to reimagine early childhood education. *Behavioural Public Policy*, 2(1), 1-21. https://doi.org/10.1017/bpp.2017.6

List, J., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4), 773–793. https://doi.org/10.1007/s10683-018-09597-5

List, John, A. (2011). Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *Journal of Economic Perspectives*, 25 (3), 3-16. https://doi.org/10.1257/jep.25.3.3

Maniadis, Z., Tufano, F., & List, J. A. (2014). One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *American Economic Review*, 104(1), 277-290. http://dx.doi.org/10.1257/aer.104.1.277

Maniadis, Z., Tufano, F., & List, J. A. (2017). To Replicate or Not to Replicate? Exploring Reproducibility in Economics Through the Lens of a Model and a Pilot Study. *The Economic Journal*, 127, F209-F235. doi:10.1111/ecoj.12527

Marco Castillo, John List, Ragan Petrie, Anya Samek. (2022, Revise & Resubmit). Drivers of Behavior at an Early Age: Evidence from a Longitudinal Field Experiment. *Journal of Political Economy.*

Meier, S. (2007). Do Subsidies Increase Charitable Giving in the Long Run? Matching Donations in a Field Experiment. *Journal of the European Economic Association* 5 (6), 1203–22.

Reiley, D., and Samek, A. (2015). How Do Suggested Donations Affect Charitable Gifts? Evidence from a Field Experiment in Public Broadcasting. SPI Working Paper 142.

Rondeau, D., and List, J. (2008). Matching and Challenge Gifts to Charity: Evidence from Laboratory and Natural Field Experiments. *Experimental Economics*, 11 (3), 253–67.

Sadoff, S., Samek, A., Sprenger, C. (2020). Dynamic Inconsistency in Food Choice: Experimental Evidence from Two Food Deserts. *The Review of Economic Studies*. 87(4), 1954–1988. https://doi.org/10.1093/restud/rdz030

Sadoff, Sally. (2014). The Role of Experimentation in Education Policy. *Oxford Review of Economic Policy*. 30 (4), 597–620.

Samek, A. (2019a). Advantages and Disadvantages of Field Experiments. In

A. Schram & A Ule (Eds.), *Handbook of Research Methods and Applications in Experimental Economics* (pp. 105-120). Edward Elgar. https://doi.org/10.4337/9781788110563

Samek, A. (2019b). Gifts and goals: Behavioral nudges to improve child food choice at school. *Journal of Economic Behavior & Organization*, 164, 1-12. https://doi.org/10.1016/j.jebo.2019.05.008

Samek, A. and Longfield, C. (2023, forthcoming). Do Thank-You Calls Increase Charitable Giving? Expert Forecasts and Field Experimental Evidence. *American Economic Journal: Applied Economics.*

Samek, A., Sprenger, C., and Martinez, S. (2022, work in progress). Procrastination in Health Insurance Choice.

Samek. A., Gray, A., Datar, A., and Nicosia, N. (2019) Adolescent Time and Risk Preferences: Measurement, Determinants and Field Consequences. CESR-Schaeffer Working Paper No. 2019-003. https://dx.doi.org/10.2139/ssrn.3463495

Shang, Jen, and Rachel Croson. (2009). A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods. *Economic Journal*, 119 (540), 1422– 39.

Soetevent, A. R (2005). Anonymity in Giving in a Natural Context: A Field Experiment in 30 Churches. *Journal of Public Economics*, 89 (11–12), 2301–23.

Steven D. Levitt and John A. List. "Viewpoint: On the generalizability of lab behavior to the field." *Canadian Journal of Economics*, vol. 40, no. 2, 2007, pp. 347-370.

Sutter, M., Zoller, C., and Glätzle-Rützler, D.(2019). Economic behavior of children and adolescents – A first survey of experimental economics results. European Economic Review, Elsevier, 111(C), 98-121. https://doi.org/10.1016/j.euroecorev.2018.09.004

Tamborini, C. R., Kim, C., & Sakamoto, A. (2015). Education and lifetime earnings in the United States. Demography, 52(4), 1383-1407. https://doi.org/10.1007/s13524-015-0407-0

Ye, K. J., Samek, A., Yoder, K. J., Decety, J., Hortaçsu, A. and List, J. (2022, in press) Early Childhood Programs Change Test Scores But Do They Change Brain Activity?

Understanding the interactions of sleep, social media and mental health for productivity and performance: The role of field experiments

Samuel Lindquist[1] **and Sally Sadoff**[2]

1. Introduction

The study of the wealth and well-being of nations has generally focused on markets. Prior annual proceedings of the Upton Forum have examined the development and functioning of markets, cultural and institutional structures that support markets, and key inputs into markets including energy and human capital. More recently, there is a growing recognition that individual behaviors that take place largely outside of traditional markets are critical for productivity, performance and well-being. And relatedly, personal mental health and subjective well-being are important to understand both as direct measures of welfare, and as critical factors for labor market and human capital outcomes.

In this vein, we examine the role of sleep, social media and mental health for performance and productivity. We focus on these factors for several reasons, which we discuss in more detail throughout the paper. First, there is evidence — largely from naturally occurring data — that they matter for the wealth and well-being of nations. Second, these factors are highly interrelated. This raises a

1 Samuel Lindquist is a Predoctoral Research Fellow at the Rady School of Management, UC San Diego.
2 Sally Sadoff is a Professor of Economics and Strategy at the Rady School of Mangement, UC San Diego.

host of challenges for identifying causality in naturally occurring data, which field experiments can help address. Third, the existing experimental evidence on these factors is limited. Our goal in this paper is to provide guidance for the design of field experiments that can more fully inform our understanding of the causal linkages between sleep, social media and mental health, and their impact on productivity and performance.

Our approach builds on John List's pioneering research in field experiments. His far-reaching work has demonstrated the potential for field experiments to inform a broad range of economic questions in natural contexts. His field experiments are particularly influential because they unite tests of theory – more generally associated with lab experiments – with tests of policy relevant interventions. That is, economic frameworks motivate the experimental design so that the results can inform the parameters of interest. Finally, these frameworks often incorporate both standard theory and behavioral models in order to inform economic theory more broadly.

The remainder of the paper is structured as follows. In Section 2, we develop a framework for the interactions of sleep, social media and mental health for productivity and performance. We then discuss how the framework can guide the design, analysis and interpretation of experiments. In Section 3, we review the evidence on sleep, social media and mental health through the lens of our framework. Section 4 concludes with a discussion of promising avenues for future research.

2. Framework

Sleep, social media use, and mental health are interlinked phenomena, making their impact on productivity and performance difficult to disentangle from one another. We develop a simple framework to illustrate the identification challenge and help guide empirical design and analysis.

Let Y be productivity or academic performance, H be mental health, M be social media usage and S be sleep. We assume that Y is an additively separable function of H, M and S; H is a function of M and S; M is a function of H and S; and S is a function of H and M:

$$Y = f(H(M,S),M(H,S),S(H,M)) \qquad (1)$$
$$= f_1(H(M,S)) + f_2(M(H,S)) + f_3(S(H,M))$$

The causal impact on Y from a change in mental health, social media use or sleep, respectively, is:

$$\frac{\partial Y}{\partial H} = \frac{df_1}{dH} + \frac{df_2}{dM}\frac{\partial M}{\partial H} + \frac{df_3}{dS}\frac{\partial S}{\partial H} \quad (2)$$

$$\frac{\partial Y}{\partial M} = \frac{df_1}{dH}\frac{\partial H}{\partial M} + \frac{df_2}{dM} + \frac{df_3}{dS}\frac{\partial S}{\partial M} \quad (3)$$

$$\frac{\partial Y}{\partial S} = \frac{df_1}{dH}\frac{\partial H}{\partial S} + \frac{df_2}{dM}\frac{\partial M}{\partial S} + \frac{df_3}{dS} \quad (4)$$

Equations 2, 3, and 4 decompose the effects of mental health, social media use, and sleep, respectively, on Y. Considering equation 4 for the sake of illustration, we see that the impact of sleep on productivity can be broken down into an unmoderated effect, $\frac{df_3}{dS}$, its effect through the channel of social media, $\frac{df_2}{dM}\frac{\partial M}{\partial S}$, and its effect through the channel of mental health, $\frac{df_1}{dH}\frac{\partial H}{\partial S}$. The framework can be easily adapted to include additional channels. For example, if someone sleeps more, they may have more energy to exercise. Increased exercise may then have an impact on physical and mental health, as well as productivity and performance. Or, if someone sleeps more they may have less time to work or study, which could lower earnings or academic performance. We focus on sleep, social media and mental health because we believe these are three areas of increasing importance with important interactions that are not well understood.

This framework highlights the empirical challenge of identifying causal effects in naturally occurring data. First, naturally occurring data largely provides correlations. For example, there may be a negative correlation between social media use and mental health – i.e., lower levels of mental well-being among those with higher usage of social media. However, from a correlation alone it is not possible to determine whether social media usage is changing mental health (e.g., using social media decreases well-being); mental health is driving social media usage (e.g., when people are depressed they stay home and spend more time on social media); or, the correlation largely reflects an omitted channel such as the effect of sleep on both social media and mental health (e.g., when people don't get enough sleep, they feel more depressed and they have less energy, so they spend more time on social media).

Experiments can help address the identification challenge by exogenously varying one factor and examining the impact on the other factors. For example,

as we discuss in more detail below, an experiment may randomly assign limited access to social media and measure the impact on subjective well-being. Related work takes advantage of naturally occurring quasi-experiments that arguably exogenously vary one factor. For example, Braghieri et al. (2022) uses the staggered introduction of Facebook across U.S. colleges as an instrument for increased social media usage and finds that the introduction of Facebook increased instances of depression and anxiety.

A second challenge is that naturally occurring data may not include measures of the relevant channels. For example, time use data may include measures of social media use and sleep but not mental health. Relatedly, quasi-experimental data from naturally occurring experiments may have measures of the outcome of interest but not the factors of interest. For example, Braghieri et al. (2022) does not include direct measures of Facebook usage (or other related channels such as sleep). These data allow you to estimate the effect of the presence of Facebook on mental health, but are more limited in estimating the direct impact of actual Facebook usage on mental health. Furthermore, they cannot identify the role of interactions with other factors. For example, when Facebook is introduced on a college campus, students may stay up later on social media and get less sleep, which may contribute to the observed decline in mental health.

The framework also highlights the challenge of using laboratory (lab) experiments to estimate the full impact of changes in social media usage and sleep. Lab experiments that exogenously vary one factor over a short time frame in a highly controlled environment may limit the role of interactions with other channels. For example, in a sleep study examining the effects of sleep deprivation, participants may spend several nights in a lab with proscribed sleep time and then perform tasks to measure cognitive ability. This kind of study largely isolates the short run effects of decreased sleep on cognitive performance. It does not allow for the measurement of longer run impacts that may take time to emerge, such as the impact on mental health or overall labor market engagement or academic performance. And because participants are in a sleep lab, there is limited scope for changes in performance via outside channels such as social media habits.

Field experiments can address the limitations of lab experiments by measuring the impact of exogenous movements in one factor on various channels. The framework can help guide the design of those experiments. For example, a field experiment that exogenously varies sleep and measures the impact on performance should also measure the impact on time use (including social media

usage) and mental health. Field experiments can also examine the extent to which causal relationships are symmetric. For example, decreasing social media use may improve mental health. But improving mental health may not naturally lead to decreases in social media usage.

The framework can also guide the analysis and interpretation of both experimental and non-experimental data. Outside of the lab it is difficult, if not impossible, to assign individuals to a specific level of social media use, sleep or mental health. Estimating the impact of these factors relies on instruments that cause changes in these factors. In the case of sleep, for example, in non-experimental data, the instruments include differences in sunset times or class start times that affect sleep habits. In randomized field experiments, the instruments include goal setting, sleep aids and incentives to increase sleep. All of these are "encouragement" designs in which sleep itself is not directly assigned and so there may be varying degrees of compliance with the intervention. For example, in a field experiment, Giuntella et al. (in preparation) randomly assign participants to a treatment group in which they receive incentives for sleeping at least 7 hours a night on weekdays. These incentives increase sleep but there is imperfect compliance – i.e., participants in the treatment group do not meet the 7 hour threshold on every weeknight. The experiment yields an estimated intent-to-treat (ITT) impact of the intervention (i.e., incentives) on academic performance. This estimate can be converted into an estimate of the impact of increased sleep on academic performance, $\frac{\partial Y}{\partial S}$, but requires additional assumptions, including that the intervention only affected academic performance via sleep. And, this estimate does not isolate the "direct" effect of sleep on academic performance, $\frac{df_3}{dS}$, from the effects via other channels including mental health, $\frac{df_1}{dH}\frac{\partial H}{\partial S}$, and social media use, $\frac{df_2}{dM}\frac{\partial M}{\partial S}$. It is important to recognize these limitations when interpreting the estimated treatment effects of the intervention.

By measuring each of the potentially relevant channels, we can generate an understanding of which matter most for the question of interest. Experiments can then directly incorporate the most relevant channels into the design. For example, Giuntella et al. (in preparation) find evidence that changes in sleep lead to changes in social media use but less evidence that changes in sleep strongly affect mental health. Future work could test interventions aimed at both sleep and social media within the same experiment to better disentangle their interacted effects on

academic performance. More broadly, the framework highlights both the role of field experiments and their limitations in identifying the interacted causal impact of sleep, social media and mental health on productivity and performance.

3. Reviewing the Evidence

3.1 Mental Health

There is growing recognition of the association between mental health and economic outcomes. Greenberg et al. (2021) estimate the economic burden of depression alone to be over $300 billion dollars in the U.S., with workplace costs accounting for about 60% of the total. They estimate that both the overall burden and the share due to the workplace have increased over time. At the individual level, mental health disorders can carry large earnings penalties. A study of workers in Denmark found that depressed people made 34% less, people with bipolar disorder made 38% less, and those with schizophrenia made 74% less (Biasi et al., 2021). At the same time, research into effective treatments for mental illness is well established, with abundant evidence for the beneficial effects of targeted psychotherapeutic and pharmacological treatment programs, and sometimes, most effectively, a combination of both (Cuijpers et al., 2009; Kamenov et al., 2017; Smith & Glass, 1977).

Building on this work, a nascent literature examines the causal impact of mental health interventions on labor market outcomes, economic well-being, human capital investment and criminal activity. In addition, these studies explore novel forms of psychotherapy, such as therapy delivered in group settings by non-professionals, which have the potential to be delivered at a fraction of the cost and at scale. Lund et al. (2020) conduct a meta-analysis of the impacts of mental health interventions in low and middle-income countries. They find that, on average, psychological interventions improve economic outcomes like earnings and investment in middle-income countries, with more mixed effects in low-income countries. We discuss examples of interventions in low, middle and high income countries in more detail below.

A strand of this literature focuses on the impact of psychotherapy on criminally-engaged or at-risk populations. Little et al. (1993) found that Cognitive Behavioral Therapy (CBT) reduced rates of recidivism by 32%, from 54.9% in the control group to 37.1% in treatment. Heller et al. (2017) studied two separate

CBT programs conducted in the Chicago area. The first, known as Becoming A Man (BAM), targeted adolescent students at risk of criminal involvement. The program was mainly composed of exercises which prompted participants to slow down their thinking and reflect on the beliefs and assumptions they held, with the goal that it would help them make the context-appropriate decision in more consequential (and perhaps dangerous) settings. BAM reduced overall crime committed by 28 to 35%, and violent crime even more so, by 45 to 50%. Perhaps more impressively, early evidence suggests it even increased graduation rates by 12 to 19%. The other CBT program they studied, which was conducted in a juvenile detention center, reduced recidivism by 21%. Although the impact on recidivism is smaller than that in Little et al. (1993), it reinforces the important role therapy could play for incarcerated individuals and those at risk for criminal activity.

Blattman et al. (2017) studied an 8 week-long CBT program, with additional $200 cash transfers to some, administered to criminally-engaged men in Monrovia, Liberia. The therapeutic course of the program had significant short-run effects. Participants engaged in fewer thefts, sold fewer drugs, and were less likely to carry a weapon. The authors argue that therapy helped these men to reinvent their own self-perception and weigh the long-term benefits of an activity more heavily than the short-term gain. Indeed, participants who received therapy responded more positively to questions regarding their self-identity and mental health two to five weeks out from the start of sessions. Therapy also affected these individuals' time preferences: when asked whether they would rather receive, say, 1,000 Liberian Dollars now or 1,100 Liberian Dollars in two weeks, they were more likely than the control group to wait the two weeks in order to receive the larger sum. However, these effects did not persist in one-year follow-up surveys. On the other hand, those who received a $200 cash transfer on top of CBT still engaged in less criminal activity a year out from the intervention.

In sum, the research strongly suggests that psychotherapy and counseling programs can affect one's decision to engage in criminal activity, with mixed evidence on how long the impacts last. It appears that mixing counseling and psychotherapy with an additional monetary or labor component can produce stronger results. Alongside the outcomes reported in Blattman et al. (2017), new evidence from a Chicago-based program providing 18 months of CBT alongside employment and other social supports offers additional evidence for this (Bhatt et al., 2023). Individuals enrolled in the program through the referral of a community outreach worker (as opposed to through a referral algorithm

created by the researchers) were 79% less likely to be arrested for shooting and/or homicide and were 45% less likely to be a victim of violent crime. Individuals enrolled in the program algorithmically saw no such reductions.

Another strand of research focuses on therapeutic outcomes for individuals already suffering from mental health issues such as depression. Bhat et al. (2022) studied two programs conducted in the Indian state of Goa. The first was known as the Healthy Activity Program (HAP), in which a counselor met with depressed individuals weekly for 6-8 weeks. The primary goal of counseling was "behavioral activation", to encourage depressed individuals to schedule and engage in activities which they enjoyed. This program significantly reduced depression. As far out as five years from the beginning of treatment, participants were 11 percentage points less likely to be depressed than a control group. Similar to the BAM program in Heller et al. (2017), the HAP program changed how participants thought about their decisions. For example, it made them less overconfident, measured by how much they overvalued their performance relative to peers in an activity. The result was a more even-handed perception of self, with lower magnitudes of both under- and over-confidence.

The second program these authors studied attempted to address depression in pregnant women through the same technique of behavioral activation. However, unlike the HAP program, this program only marginally decreased perinatal depression in the short term and had no long-term effects. Furthermore, it did not alter patterns of thought as the HAP program did. Neither of these programs impacted employment or consumption.

These findings overlap somewhat with that of Angelucci and Bennett (2022), which explored the impacts of pharmacotherapy and group therapy sessions in a group of 1,000 depressed adults (86% of whom were women) in the Indian state of Karnataka. The intervention reduced depression. It also increased investment in children, increasing both school enrollment and attendance. Interestingly, the program did not increase employment or earnings. In fact, those just receiving pharmacotherapy worked over five fewer hours per week during their treatment of approximately four months. And those who received both pharmacotherapy and counseling worked over three hours less per week after the program ended. This is interesting given that an explicit goal of counseling was to help participants with work-related challenges and give them personalized advice on employment and other money-making opportunities.

Expanding out from specific populations like the criminally at-risk and depressed, a third strand of this literature studies the effect of psychotherapy

and cash transfers on both mental health and economic well-being in general populations. Barker et al. (2021) studied the impacts of CBT on human capital in rural Ghana. Like all of the programs previously discussed, it had significant positive effects on mental health. Interestingly, there was no differential impact between those with high and low baseline mental distress, suggesting CBT can help those who do not suffer from mental illness just as much as those who do. The program also increased perceived economic well-being: participants reported higher economic status, and projected a higher economic status in five years, compared to a control group. It's important to note that these results were measured just two to three months out from a twelve week course, a much shorter time span than some of the previous studies discussed above.

A similar study treated a general population in rural Kenya with therapy sessions, cash transfers, or both, for five weeks (Haushofer et al., 2020). Individuals who received just therapy showed no differences from the control group 12 months out from the intervention. There was no lasting impact on their mental health or economic outcomes such as asset holdings, consumption, or revenue. Indeed, the treatment arm that received both therapy and cash transfers had quite similar mental health and economic outcomes to the arm that just received cash transfers. Taken together, these results suggest that, while CBT can be beneficial for individuals without mental illness, its effects on those with mental illness are longer lasting.

The evidence reviewed suggests treating mental health issues can improve important outcomes in the household such as investment in children, but evidence on how these strides might extend to the workplace is mixed. Why did the improvements in mental health exhibited in these studies fail to materialize into tangible economic improvements like labor market productivity, and in some cases, even decrease time worked? It could be in part mechanical – time spent on the mental health intervention could serve as a substitute for time spent in the labor market. The lack of increases in earnings or time worked could also be due to the significant constraints on female employment where some of the studies took place (Bhat et al., 2022). Future work could examine the role of heterogeneity in both social context and individual mental health in understanding the differential impacts of mental health interventions.

Another important general insight from these studies is that effective psychotherapy need not be delivered by professional psychologists, psychiatrists, or therapists. Rather, it can be delivered by trained community members and

at a fraction of the cost. Programs that pair this type of therapy with a lump-sum cash transfer are remarkably cost-effective, and tend to result in stronger effects than either therapy or cash transfers in isolation. Recent studies in this vein test the impact of low-cost mindfulness interventions either delivered through group therapy or online apps (Cassar et al., 2022; Shreekumar & Vautrey, 2022). Shreekumar and Vautrey (2022) find that the $13-per-month app "Headspace" reduces rates of depression by 0.46 standard deviations (SDs), anxiety by 0.38 SDs, and stress by 0.47 SDs, reductions which are comparable to those usually achieved by professional therapists. Cassar et al. (2022) find that a university course on mindfulness also reduced levels of depression, anxiety, and stress. They find mixed effects on academic performance. In the short-run, the course decreased students' grades by 0.26 SDs. The authors suggest that a possible reason for this is that students enrolled in the course spent more time on self-care related activities such as sleeping and relaxing, possibly at the expense of time studying. By contrast, in the longer-run, they find evidence that the intervention increased grades by 0.28 SDs, a result mainly driven by the individuals who engaged in mindfulness practices outside of the course.

To our knowledge, these studies have not examined the interactions of mental health with social media use and sleep. Future work could estimate the impact of mental health interventions on these channels in order to better understand whether they are contributing to (or perhaps dampening) the effects of mental health improvements on performance. As evidenced by Cassar et al. (2022), an additional promising avenue for future work is to examine these questions in the context of academic performance, where the relationship between components of our mental health—like stress—and academic achievement can be complicated and are not well-understood.

3.2 Social Media

There has been rapid growth in research examining the impact of social media (see Aridor et al. (forthcoming) for a comprehensive review). Prior work has examined the political ramifications of social media. This includes studies of communication amongst protesters engaged in pro-democratic movements, such as during the Arab Spring (Howard et al., 2011; Tufekci, 2017); as well as research on self-siloing into groups with like-minded ideological views and higher consumption of misinformation (Allcott et al., 2019; Gentzkow & Shapiro, 2011).

Another strand of research has focused on the relationship between social media and mental health. Multiple meta-analyses examine the interplay between depression and social media usage. Some find a positive correlation between social media use and depression (Appel et al., 2016; Keles et al., 2020). Yet others emphasize that results are split between positive, negative, and null results (Odgers & Jensen, 2020). Furthermore, even when a positive correlation between social media and mental health is found, it may not be economically significant. For example, Odgers and Jensen (2020) emphasize that social media usage explains "less than 0.5% of the variance in symptoms [of mental health in studies] with poor adjustment for relevant confounding factors and estimates that are virtually always derived from correlation designs." Similarly, an 8-year-long longitudinal study on adolescents found that as they grew, and consequently spent more cumulative time on social media, their mental health did not change as a result. They only found an association between social media and mental health when looking across individuals, instead of within individuals across time. This suggests that social media does not cause poor mental health. Rather, that mental health, or some unknown third factor, is in fact driving social media use (Coyne et al., 2020).

To address concerns with causal identification, a number of lab and field experiments have examined the impact of social media use on measures of mental health and well-being. We summarize the design of these studies in Table 1 and, when possible, report the estimated impacts in Figure 1.

The lab experiments tend to induce more social media use by having participants in the treatment condition use some type of social media. The time span for use in lab experiments is rather short, from 7 to 20 minutes (Engeln et al., 2020; Sagioglou & Greitemeyer, 2014). These experiments generally find increases in feelings of subjective well-being, while not significantly moving feelings of depression or negative affect.

Unlike lab experiments which induced an increase in social media use, field experiments attempt to limit social media use. They do so over a much longer period of time, from one day to three quarters of a school year (Collis & Eggers, 2022; Przybylski et al., 2021). These studies generally find either null effects or that limiting social media usage decreases feelings of depression and increases one's feelings of subjective well-being. As discussed above, field experiments are able to capture longer-term effects in natural environments which allow multiple channels to interact.

It is important to note that we are working with a small sample of studies on this topic, and therefore the high variation in treatment effects is due in part to chance. The studies also vary significantly in their design. As already noted, some studies focus on the acute effects of social media while others focus on their chronic ones. Some focus on a particular population (most commonly University students) while others recruit from a general population of social media users. Some treatments include full abstinence from social media, while others limit it to a cutoff (for example, 10 minutes a day) or encourage reductions in usage. Some limit usage of just one site (most commonly Facebook), others limit all social media use. These dimensions are likely crucial in determining whether social media is a net positive or negative experience. The studies also vary in how they measure the outcomes, with some using validated scales and others a single question asking, for example, if someone feels depressed.

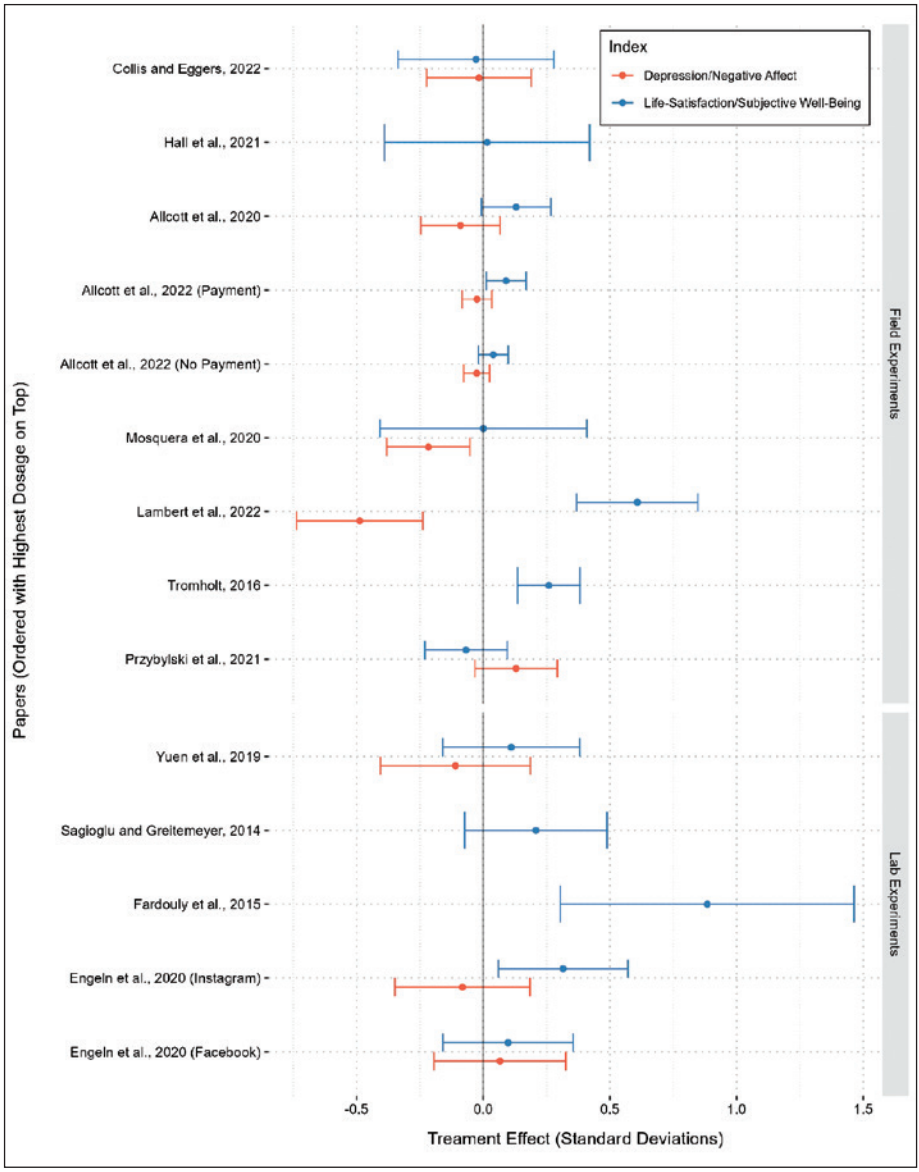
An especially important dimension along which the interventions vary is in how much of a change they induce on time spent on social media (summarized in column 3 of Table 1). For example, studies that focus on chronic effects span in intensity from, on the low end, not using Facebook for five days (Vanman et al., 2018) to, in the extreme, using social media for only ten minutes a day for three quarters of a school year (Collis & Eggers, 2022). Interestingly, we do not find a strong relationship between treatment intensity and treatment effect when comparing across studies. This may contradict the "Goldilocks" theory of social media use. This is the idea that some amount of social media use is beneficial but as use increases the drawbacks overtake the benefits and it becomes a net harm (Dienlin & Johannes, 2022; Przybylski & Weinstein, 2017). Looking across studies of different treatment intensities we find no evidence of this relationship. If this were the case, we would expect a U-shaped curve in treatment effects in Figure 1, as papers are in order from highest (at the top) to lowest treatment intensity.

These results suggest that the effects of social media are heterogeneous across people and contexts, as well as levels and types of usage. It can be difficult to separately identify the role of these dimensions in naturally occurring data because they are often correlated with one another. For example, consider the case of the effects of social media on young people, a topic that has rightly received considerable attention (Auxier & Anderson, 2021; Wells et al., 2021). Social media usage is more statistically associated with negative mental health outcomes in young people than in adults, and within that group, young women more so than young men (Orben et al., 2022). However, young people use social media more

than adults, and young women use social media more than young men (Vogels et al., 2022). These overlapping features make studying social media's particular harm on young people difficult: is it that it has a differential impact on young people? Or that it has a differential impact on women, who make up the majority of social media users? Or is it that spending more time on social media, as young people do, is harmful? Or maybe a combination of all three? Future work in field experiments could focus on causally identifying the heterogeneous effects of social media by gender, age and other critical demographic groups. As discussed above, these studies should gather rich data capturing the multidimensional impacts of social media, which may not be available in naturally occurring data.

Another nascent frontier for research on social media is its effect on labor market outcomes and productivity. Early evidence on this topic suggests social media may have significant negative impacts on both. For example, Marotta and Acquisti (2017) looks at the effect of the social media blocking app "Freedom" on how Amazon Mechanical Turk workers complete tasks. Those assigned to have social media blocked completed 8 more tasks per hour and earned $0.80 more than a control group. As Gen Z, the first generation to grow up using social media, enters the workforce, research on this topic will become even more pressing. Relatedly, even though many of the studies have been conducted with students, there is sparse evidence on how social media interacts with sleep and academic performance. While there is quasi-experimental evidence that social media use reduces self-perceived academic performance (Braghieri et al., 2022), the longest-run social media intervention to date found no measurable impact. Collis and Eggers (2022), which reduced usage to ten minutes a day for three quarters of a school year, found that the intervention had no impact on academic success. There is also little evidence on the impact of social media interventions on sleep which, as we discuss below, is potentially important for understanding the interacted effects on performance.

Figure 1: Effect of Limiting Social Media on Depression and Life-Satisfaction.



Notes: Dosages were determined by length of treatment and, if multiple treatments had the same duration, intensity of treatment (e.g., limiting versus abstaining social media). Error bars are 95% confidence intervals.

Table 1: Field Experiments on Effects of Social Media on Mental Health

Paper	Sample	Treatment(s)	Social media usage	Mental health & other outcomes
Tromholt (2016)	Facebook Users (N = 1,095)	A week off Facebook	87% full abstinence (over 60 minutes at baseline)	Envy, life satisfaction
Hunt et al. (2018)	University Students (N = 143)	Three weeks of social media limited to 10 minutes a day	19 minutes avg. less social media from 41 minutes baseline [1]	Loneliness, depression
Vanman et al. (2018)	Facebook Users (N = 138)	Five days off Facebook	120 minutes avg. less Facebook from 171 minute baseline	SWB, Stress
Allcott et al. (2020)	Facebook Users (N = 2,897)	Four weeks off Facebook	59.9 minutes avg. less Facebook from 74.5 minutes baseline	Depression, political polarization, SWB, WTA
Mosquera et al. (2020)	University Students (N = 1,765)	2 weeks off Facebook	95% full abstinence (112 minutes at base line)	WTP, depression, SWB, news polarization
Hall et al. (2021)	University & CC Students (N = 130)	4 weeks off all social media	20% full abstinence (68.76 minutes at baseline)	Loneliness, SWB, Life Quality
van Wezel et al. (2021)	University Students (N = 76)	7 days of 50% reduced Facebook	13.5 minutes avg. less Facebook from 25.6 minutes baseline	SWB, FOMO
Przybylski et al. (2021)	University Students (N = 297)	1 day off social media	50% full abstinence	Positive and negative affect, self-esteem, life satisfaction
Lambert et al. (2022)	Social media users (N = 154)	7 days off Instagram, Tiktok, Facebook, and Twitter	79 minutes avg. less social media from 71 minutes baseline [2]	SWB, depression, anxiety
Allcott et al. (2022)	Facebook Users (N = 2,126)	a) Three weeks with goal setting app ("limit") b) $50 for every hour reduction in daily avg. over three weeks ("bonus")	20 minutes avg. less screentime from 153 baseline 60 minutes avg. less screentime from 153 baseline	SWB, addiction, depression, concentration
Collis and Eggers (2022)	University Students (N = 122)	Three quarters of social media limited to 10 minutes a day	15.45 minutes avg. less social media from 24.5 baseline	Academic performance, SWB

[1] Baseline and change in use estimated from Hunt et al. (2018) Figure 1
[2] Baseline usage was based on self-report data which is frequently underestimated

Table 2: Lab Experiments on Effects of Social Media on Mental Health

Paper	Sample	Treatment(s)	Mental health & other outcome(s)
Sagioglou and Greitemeyer (2014)	Mechanical Turk Users (*N* = 263)	Facebook for 20 minutes	Positive and negative affect, meaningful activity
Verduyn et al. (2015)	University Students (*N* = 67)	10 mins active (vs passive) Facebook Use	SWB, envy, active vs. passive use
Fardouly et al. (2015)	University Students (*N* = 122)	Facebook for 10 minutes	Negative affect, body dissatisfaction
Yuen et al. (2019)	University Students (*N* = 312)	Facebook for 20 minutes	Positive and negative affect, envy, meaningful activity
Engeln et al. (2020)	Female University Students (*N* = 308)	a) Facebook for 7 minutes b) Instagram for 7 minutes	Positive and negative affect, body dissatisfaction, social comparison

3.3 Sleep

Based on lab studies, sleep scientists have long been aware that lack of adequate sleep can have detrimental impacts on attention, memory, cognition, and mood (Banks & Dinges, 2007; Killgore, 2010). Lack of sleep is a worldwide problem, with more than 1 in 3 US adults sleeping less than 7 hours a night, the recommended minimum (Liu et al., 2016). Evidence suggests the problem may be substantially worse in low-income countries. Experiments conducted in India and Sri Lanka revealed that adults average 5.6 and 6.4 hours of sleep, respectively, each night (Bessone et al., 2021; Schokman et al., 2018).

Sleep, or lack thereof, has clear economic implications. Due to decreased cognitive function and "cyberloafing", the sleep-deprived are less productive, with one study from Australia estimating the cost of sleep deprivation at 0.8% of the country's economy (Hillman et al., 2006; Mullainathan, 2014). However, until recently, sleep has failed to attract much attention from the economic field.

Prior work has demonstrated that sleep and working hours are inversely correlated: those who sleep more work less (Basner et al., 2007; Biddle & Hamermesh, 1990; Pfeifer, 2015). However, it is difficult to draw causal conclusions from these data, as much of it is founded on self-reported sleep time,

which usually overestimates actual time spent sleeping. More importantly, the presence of omitted variables which effect sleep and work time simultaneously cannot be ruled out. As the authors mention in Biddle and Hamermesh (1990), the pioneering work on sleep in economics, "we have not strictly established the direction of causation, if any, between sleep time and labor supply. It may be that variations in individuals' sleep time are beyond their control and that their labor supply changes in response to these variations."

To disentangle causation from correlation, economists have exploited quasi-experimental settings to estimate the impact of sleep on labor market outcomes (Carrell et al., 2011; Gibson & Shrader, 2018; Giuntella et al., 2017; Giuntella & Mazzonna, 2019; Jagnani, 2021; Jin & Ziebarth, 2020). The earlier the sun sets, the earlier people go to sleep. Thus time zone border areas, which have similar levels of natural light but are an hour apart on the clock, provide plausibly random variation in sleep time. These studies find that individuals living in the later, Eastern side of a time zone border have lower wages, with estimates ranging from .44% to 3%.

More recently, researchers have begun to use field experiments to examine the impact of interventions aimed at increasing sleep (Avery et al., 2022; Bessone et al., 2021; Giuntella et al., in preparation). Bessone et al. (2021) encouraged some of their 452 study participants in Chennai, India to sleep more, and paid others for sleeping longer than baseline. Although these encouragements and financial incentives did cause people to sleep more, their findings did not align perfectly with evidence from lab-based sleep studies. For example, contrary to previous evidence and expert predictions, the authors found that sleeping longer at nighttime had no effect on cognitive skills or subjective well-being. However, individuals who took a thirty minute nap at work in the afternoon scored higher on cognition tests, reported more positively on their well-being, and were more productive at work. The authors are not able to identify whether naps are more effective because of their timing or because they take place in a higher quality sleep environment than nighttime sleep. The null effects of nighttime sleep could be due to low quality of sleep, the added time needed to be in bed in order to sleep which substitutes for time working, or because baseline sleep is so far from recommended amounts that small increases do not have a significant impact.

In the U.S., Giuntella et al. (in preparation) encouraged college students to sleep more through app reminders on their phone as well as payments for sleeping at least 7 hours a night. Unlike Bessone et al. (2021), they found increases in

performance, with significant GPA increases amongst those who received payments. Important for informing the framework discussed above, Giuntella et al. (in preparation) additionally find that the sleep intervention leads to lower social media usage but does not have discernible impacts on mental health.

Table 3 summarizes the field and quasi-experimental evidence on the impacts of sleep on outcomes related to productivity and performance.

A clear finding shared in studies is that both encouragement and financial incentives work in getting people to sleep more, and perhaps unsurprisingly, payment works better than encouragement. And the financial incentive need not be large. Bessone et al. (2021) paid participants a maximum of 120 rupees ($1.70) and Giuntella et al. (in preparation) paid $4.75 for sleeping at least 7 hours. However, there are mixed findings on the effect of sleep on labor market and academic outcomes. Future work could explore the extent to which sleep environments, sleep timing, baseline sleep and demographic differences may help explain the heterogeneous impacts of increased (or decreased) sleep.

4. Discussion

As noted above, the small but growing literature on sleep, social media and mental health demonstrates substantial heterogeneity in findings. Future work could explore the extent to which the differential treatment effects are explained by differences in the interventions (e.g., intensity, length of time), the population (demographics, types of participants who select into the study, social context) and baseline levels of the behavior of interest.

From a public policy perspective, it is important to understand the cost-effectiveness of these interventions in terms of their effects on productivity and performance. Is it more cost-effective to receive some form of psychotherapy, detox from social media, or get more sleep? The answer to this depends not only on the elasticity of the targeted behavior (e.g., are people more responsive to incentives for sleep or incentives to decrease social media?), but also on the cross-elasticities of those behaviors with the other channels of interest (e.g., what are the downstream effects on mental health and performance of improving sleep vs. decreasing social media use?).

Taking for example the early evidence from sleep and social media interventions, it appears that sleep interventions, on top of improving academic performance and health, also decrease social media use because individuals substitute screen time

Table 3: Literature on the Effect of Sleep on Productivity and Performance

Paper	Sample	Treatment	Sleep Outcomes	Productivity/Performance Outcomes
Panel A: ***Field Experiments***				
Bessone et al. (2021)	Low-income Indian adults (N = 452)	a) Information, encouragement, payments	27 minutes more sleep	No improvements to cognition, productivity, or SWB
		b) Information, encouragement, payments, half hour nap at workplace	8 minutes more sleep	Increases in productivity (0.04 s.d., s.e. = 0.02), SWB (0.08 s.d., s.e. = 0.03), and cognition (0.10 s.d., s.e. = 0.05)
Avery et al. (2022)	British & American university students (N = 508)	$3.75 (£2.5) for going to bed by 1 am and $3.75 (£2.5) for sleeping between 7-9 hours	9 minutes more sleep	-.59 less screen time and .42 hours more studying
Panel B: ***Quasi-Experiments***				
Carrell et al. (2011)	American University students	Class starting 50 minutes later	Not reported	.116 s.d. increase in average course grades
Giuntella et al. (2017)	Chinese individuals	Sun setting 30 minutes later (within time zone)	17 minutes less sleep among employed	3% reduction in mental capabilities, 10.5% increase in depression score
Gibson and Shrader (2018)	American individuals	Sun setting 1 hour later (within time zone)	23 minutes less sleep	4% decrease in earnings among employed
Heissel and Norris (2018)	American K-12 students	Class starting 1 hour later (across time zones)	6-13 minutes more Sleep	Increases math scores by .082 s.d., s.e. = .025 (.009 s.d., s.e. = .035), reading scores by .057 s.d., s.e. = .023 (.061 s.d., s.e. = .036) for adolescents (children)
Giuntella and Mazzonna (2019)	American individuals	Sun setting 1 hour later (across time zones)	20 minutes less sleep among employed	4.4% decrease in earnings, 5.5 p.p. increase in obesity rates
Jin and Ziebarth (2020)	German hospital admissions	End of day-light savings time	1.56 minutes more sleep	Decreases all-cause hospital admissions (2.6%), heart attacks (.41%), and heart attacks (.05%)
Jagnani (2021)	Indian K-12 students	Seasonal variation in sunset time	Sun setting 1 hour later leads to 30 minutes less sleep	Sun setting 10 minutes earlier leads to 0.1 s.d. (s.e. = .05) lower test scores, 0.04 (s.e. = .01) less school years
Lusher et al. (2019)	Vietnamese university students	Class starting 1 hour later	4.3 minutes more self-reported sleep	.009 s.d. (s.e. = .005) grade increase for morning courses, null effect for all courses
Groen and Pabilonia (2019)	American high schoolers	Class starting 1 hour later	38 minutes more sleep	.156 s.d. (s.e. = .082) reading score increase, null effects on math, health, employment

Notes: Please refer to Table 4 for the location of sleep outcomes and productivity and performance outcomes within each paper.

for more sleep (Avery et al., 2022; Giuntella et al., in preparation). For example, to take the most recent and largest scale social media intervention, Allcott et al. (2022) paid participants $2.50 for each hour of reduced time on Facebook, Instagram, Twitter, Snapchat, YouTube, or web browsers. As a result, people averaged 56 minutes less screen time each day, meaning they were paid about $2.34. On the other hand, Avery et al. (2022) paid participants $7.50 if they both went to bed between 10 PM and 1 AM and slept 7 to 9 hours. As a result, the portion of people sleeping less than six hours fell by nearly 4%. They also spent about 36 less minutes on social media. So while moving sleep somewhat, their effect on social media was significant and comparable to interventions which explicitly targeted it. While the intervention was more expensive ($7.50 per day versus $2.50), it killed two birds with one stone. Future field experiments in this area should directly compare these different types of interventions in a single population to better understand which are the most cost-effective levers to pull in order to improve health, wealth and well-being.

References

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–76.

Allcott, H., Gentzkow, M., & Song, L. (2022). Digital addiction. *American Economic Review*, *112*(7), 2424–63.

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, *6*(2), 2053168019848554.

Angelucci, M., & Bennett, D. (2022). The economic impact of depression treatment in india: Evidence from community-based provision of pharmacotherapy. *American Economic Review (Forthcoming)*.

Appel, H., Gerlach, A. L., & Crusius, J. (2016). The interplay between facebook use, social comparison, envy, and depression. *Current opinion in psychology*, *9*, 44–49.

Aridor, G., Jim´enez-Dur´an, R., Levy, R., & Song, L. (forthcoming). The economics of social media.

Auxier, B., & Anderson, M. (2021). Social media use in 2021. *Pew Research Center*. https: //pewresearch-org-preprod.go-vip.co/internet/2021/04/07/social-media-use-in2021/

Avery, M., Giuntella, O., & Jiao, P. (2022). Why don't we sleep enough? a field experiment among college students. *Review of Economics and Statistics, Forthcoming*.

Banks, S., & Dinges, D. F. (2007). Behavioral and physiological consequences of sleep restriction. *Journal of clinical sleep medicine*, *3*(5), 519–528.

Barker, N., Bryan, G. T., Karlan, D., Ofori-Atta, A., & Udry, C. R. (2021). Mental health therapy as a core strategy for increasing human capital: Evidence from ghana.

Basner, M., Fomberstein, K. M., Razavi, F. M., Banks, S., William, J. H., Rosa, R. R., & Dinges, D. F. (2007). American time use survey: Sleep time and its relationship to waking activities. *Sleep*, *30*(9), 1085–1095.

Bessone, P., Rao, G., Schilbach, F., Schofield, H., & Toma, M. (2021). The economic consequences of increasing sleep among the urban poor. *The Quarterly Journal of Economics*, *136*(3), 1887–1941.

Bhat, B., de Quidt, J., Haushofer, J., Patel, V. H., Rao, G., Schilbach, F., & Vautrey, P.-L. P. (2022). The long-run effects of psychotherapy on depression, beliefs, and economic outcomes.

Bhatt, M. P., Heller, S. B., Kapustin, M., Bertrand, M., & Blattman, C. (2023). Predicting and preventing gun violence: An experimental evaluation of readi chicago. *National Bureau of Economic Research*, (30852). https://doi.

org/10.3386/w30852

Biasi, B., Dahl, M. S., & Moser, P. (2021). Career effects of mental health. *National Bureau of Economic Research Working Paper*.

Biddle, J. E., & Hamermesh, D. S. (1990). Sleep and the allocation of time. *Journal of Political Economy*, *98*(5, Part 1), 922–943.

Blattman, C., Jamison, J. C., & Sheridan, M. (2017). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review*, *107*(4), 1165–1206.

Braghieri, L., Levy, R., & Makarin, A. (2022). Social media and mental health. *American Economic Review*, *112*(11), 3660–93.

Carrell, S. E., Maghakian, T., & West, J. E. (2011). A's from zzzz's? the causal effect of school start time on the academic achievement of adolescents. *American Economic Journal: Economic Policy*, *3*(3), 62–81.

Cassar, L., Fischer, M., & Valero, V. (2022). Keep calm and carry on: The short- vs. long-run effects of mindfulness meditation on (academic) performance.

Collis, A., & Eggers, F. (2022). Effects of restricting social media usage on wellbeing and performance: A randomized control trial among students. *PloS one*, *17*(8), e0272416.

Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2020). Does time spent using social media impact mental health?: An eight year longitudinal study. *Computers in Human Behavior*, *104*, 106160.

Cuijpers, P., van Straten, A., Warmerdam, L., & Andersson, G. (2009). Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: A meta-analysis. *Depression and anxiety*, *26*(3), 279–288.

Dienlin, T., & Johannes, N. (2022). The impact of digital technology use on adolescent well-being. *Dialogues in clinical neuroscience*.

Engeln, R., Loach, R., Imundo, M. N., & Zola, A. (2020). Compared to facebook, instagram use causes more appearance comparison and lower body satisfaction in college women. *Body Image*, *34*, 38–45.

Fardouly, J., Diedrichs, P. C., Vartanian, L. R., & Halliwell, E. (2015). Social comparisons on social media: The impact of facebook on young women's body image concerns and mood. *Body image*, *13*, 38–45.

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, *126*(4), 1799–1839.

Gibson, M., & Shrader, J. (2018). Time use and labor productivity: The returns to sleep. *Review of Economics and Statistics*, *100*(5), 783–798.

Giuntella, O., Han, W., & Mazzonna, F. (2017). Circadian rhythms, sleep, and

cognitive skills: Evidence from an unsleeping giant. *Demography*, *54*(5), 1715–1742.

Giuntella, O., & Mazzonna, F. (2019). Sunset time and the economic effects of social jetlag: Evidence from us time zone borders. *Journal of health economics*, *65*, 210–226.

Giuntella, O., Saccardo, S., & Sadoff, S. (in preparation). Improving educational achievement through better sleep habits: The effect of technology-based behavioral interventions.

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Simes, M., Berman, R., Koenigsberg, S. H., & Kessler, R. C. (2021). The economic burden of adults with major depressive disorder in the united states (2010 and 2018). *Pharmacoeconomics*, *39*(6), 653–665.

Groen, J. A., & Pabilonia, S. W. (2019). Snooze or lose: High school start times and academic achievement. *Economics of Education Review*, *72*, 204–218.

Hall, J. A., Xing, C., Ross, E. M., & Johnson, R. M. (2021). Experimentally manipulating social media abstinence: Results of a four-week diary study. *Media Psychology*, *24*(2), 259–275.

Haushofer, J., Mudida, R., & Shapiro, J. P. (2020). The comparative impact of cash transfers and a psychotherapy program on psychological and economic well-being.

Heissel, J. A., & Norris, S. (2018). Rise and shine the effect of school start times on academic performance from childhood through puberty. *Journal of Human Resources*, *53*(4), 957–992.

Heller, S. B., Shah, A. K., Guryan, J., Ludwig, J., Mullainathan, S., & Pollack, H. A. (2017). Thinking, fast and slow? some field experiments to reduce crime and dropout in chicago. *The Quarterly Journal of Economics*, *132*(1), 1–54.

Hillman, D. R., Murphy, A. S., Antic, R., & Pezzullo, L. (2006). The economic cost of sleep disorders. *Sleep*, *29*(3), 299–305.

Howard, P. N., Duffy, A., Freelon, D., Hussain, M. M., Mari, W., & Maziad, M. (2011). Opening closed regimes: What was the role of social media during the arab spring? *Available at SSRN 2595096*.

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more fomo: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology*, *37*(10), 751–768.

Jagnani, M. (2021). Children's sleep and human capital production. *The Review of Economics and Statistics*, 1–45.

Jin, L., & Ziebarth, N. R. (2020). Sleep, health, and human capital: Evidence from daylight saving time. *Journal of Economic Behavior & Organization*,

170, 174–192.

Kamenov, K., Twomey, C., Cabello, M., Prina, A. M., & Ayuso-Mateos, J. L. (2017). The efficacy of psychotherapy, pharmacotherapy and their combination on functioning and quality of life in depression: A meta-analysis. *Psychological medicine*, *47*(3), 414–425.

Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, *25*(1), 79–93.

Killgore, W. D. (2010). Effects of sleep deprivation on cognition. *Progress in brain research*, *185*, 105–129.

Lambert, J., Barnstable, G., Minter, E., Cooper, J., & McEwan, D. (2022). Taking a oneweek break from social media improves well-being, depression, and anxiety: A randomized controlled trial. *Cyberpsychology, Behavior, and Social Networking*, *25*(5), 287–293.

Little, G. L., Robinson, K. D., & Burnette, K. D. (1993). Cognitive behavioral treatment of felony drug offenders: A five-year recidivism report. *Psychological Reports*, *73*(3 suppl), 1089–1090.

Liu, Y., Wheaton, A. G., Chapman, D. P., Cunningham, T. J., Lu, H., & Croft, J. B. (2016). Prevalence of healthy sleep duration among adults—united states, 2014. *Morbidity and Mortality Weekly Report*, *65*(6), 137–141.

Lund, C., Orkin, K., Witte, M., Davies, T., Haushofer, J., Bass, J., Bolton, P., Murray, S., Murray, L., Tol, W., et al. (2020). Economic impacts of mental health interventions in low and middle-income countries: A systematic review and meta-analysis.

Lusher, L., Yasenov, V., & Luong, P. (2019). Does schedule irregularity affect productivity?

evidence from random assignment into college classes. *Labour Economics*, *60*, 115– 128.

Marotta, V., & Acquisti, A. (2017). Online distractions, website blockers, and economic productivity: A randomized field experiment. *Preliminary Draft*.

Mosquera, R., Odunowo, M., McNamara, T., Guo, X., & Petrie, R. (2020). The economic effects of facebook. *Experimental Economics*, *23*(2), 575–602.

Mullainathan, S. (2014). Get some sleep, and wake up the g.d.p. *The New York Times*.

https://www.nytimes.com/2014/02/02/business/get-some-sleep-and-wake-up-thegdp.html

Odgers, C. L., & Jensen, M. R. (2020). Annual research review: Adolescent mental health in the digital age: Facts, fears, and future directions. *Journal of*

Child Psychology and Psychiatry, *61*(3), 336–348.

Orben, A., Przybylski, A. K., Blakemore, S.-J., & Kievit, R. A. (2022). Windows of developmental sensitivity to social media. *Nature communications*, *13*(1), 1–10.

Pfeifer, C. (2015). Unfair wage perceptions and sleep: Evidence from german survey data. *IZA Discussion Paper*.

Przybylski, A. K., Nguyen, T.-v. T., Law, W., & Weinstein, N. (2021). Does taking a short break from social media have a positive effect on well-being? evidence from three preregistered field experiments. *Journal of Technology in Behavioral Science*, *6*(3), 507–514.

Przybylski, A. K., & Weinstein, N. (2017). A large-scale test of the goldilocks hypothesis: Quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychological science*, *28*(2), 204–215.

Sagioglou, C., & Greitemeyer, T. (2014). Facebook's emotional consequences: Why facebook causes a decrease in mood and why people still use it. *Computers in Human Behavior*, *35*, 359–363.

Schokman, A., Bin, Y. S., Simonelli, G., Pye, J., Morris, R., Sumathipala, A., Siribaddana, S. H., Hotopf, M., Rijsdijk, F., Jayaweera, K., et al. (2018). Agreement between subjective and objective measures of sleep duration in a low-middle income country setting. *Sleep health*, *4*(6), 543–550.

Shreekumar, A., & Vautrey, P.-L. (2022). *Managing emotions: The effects of online mindfulness meditation on mental health and economic behavior* (tech. rep.). Tech. Rep., MIT.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American psychologist*, *32*(9), 752.

Tromholt, M. (2016). The facebook experiment: Quitting facebook leads to higher levels of well-being. *Cyberpsychology, behavior, and social networking*, *19*(11), 661–666.

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.

Vanman, E. J., Baker, R., & Tobin, S. J. (2018). The burden of online friends: The effects of giving up facebook on stress and well-being. *The Journal of social psychology*, *158*(4), 496–508.

van Wezel, M. M., Abrahamse, E. L., & Abeele, M. M. V. (2021). Does a 7-day restriction on the use of social media improve cognitive functioning and emotional well-being?

results from a randomized controlled trial. *Addictive behaviors reports*, *14*, 100365.

Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., Ybarra, O.,

Jonides, J., & Kross, E. (2015). Passive facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General*, *144*(2), 480.

Vogels, E., Gelles-Watnick, R., & Massarat, N. (2022). Teens, social media and technology. *Pew Research Center*. https://www.pewresearch.org/internet/2022/08/10/teenssocial-media-and-technology-2022/

Wells, G., Horwitz, J., & Seetharam, D. (2021). Facebook knows instagram is toxic for teen girls, company documents show. *Wall Street Journal*. https://www.wsj.com/ articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documentsshow-11631620739?mod=article inline

Yuen, E. K., Koterba, E. A., Stasio, M. J., Patrick, R. B., Gangi, C., Ash, P., Barakat, K., Greene, V., Hamilton, W., & Mansour, B. (2019). The effects of facebook on mood in emerging adults. *Psychology of Popular Media Culture*, *8*(3), 198.

5. *Appendix*

Table 4: Reference Locations – The Effect of Sleep on Productivity and Performance

Panel A: *Field Experiments*		
Paper	Sleep Outcomes	Performance Outcomes
Bessone et al. (2021)	Appendix Table A6	Table 4
Avery et al. (2022)	Table 3	Table 3
Panel B: *Quasi-Experiments*		
Carrell et al. (2011)	—	Table 4
Giuntella et al. (2017)	Table 3	Table 5
Gibson and Shrader (2018)	Table 1	Table 3
Heissel and Norris (2018)	Appendix Table A7	Table 4
Giuntella and Mazzonna (2019)	Table 1	Tables 4 & 6
Jin and Ziebarth (2020)	Table 1	Table 2
Jagnani (2021)	Table 1	Table 2
Lusher et al. (2019)	Table 6	Table 5
Groen and Pabilonia (2019)	Table 9	Table 5

Does Self-Selection Diminish the Influence of Experimental Research?

Alec Brandon [1, 2]

1. Introduction

Economists conduct framed field experiments to reproduce the results of a natural field experiment without having to overcome the challenges of conducting a natural field experiment. The reason that framed field experiments may fail to reproduce the results of a natural field experiment is self-selection bias. Whereas natural field experiments randomly assign a treatment amongst an unknowing sample, framed field experiments actively recruit their sample (Harrison and List, 2004). The recruitment stage in framed field experiments, where subjects decide on whether to self-select into the experiment, is well understood to potentially bias results (Heckman, 1979).

This study considers whether, relative to natural field experiments, the potential for self-selection bias in the results of framed field experiments diminishes their scientific influence. If the only objective of economists is to understand the causal relationship between stimuli, such as prices changes or policy interventions, and outcomes of interest, then the influence of framed field experiments should be smaller than their natural counterparts. However, if instead economists have different objectives, such as considering the potential existence of interesting causal relationships, then framed field experiments could enjoy even greater influence than their natural counterparts.

To determine whether self-selection diminishes the influence of framed field experiments, I construct a dataset of the 113 natural and framed field experiments published in the most influential economics journals between 2005

1 Alec Brandon is an Assistant Professor at Carey Business School at Johns Hopkins University.
2 This paper benefited from conversations with John List, Sally Sadoff, Anya Samek, and Joe Seidel.

and 2019. This dataset captures whether a publication used a natural or framed field experiment and other variables that could confound a simple comparison of each type of experiment's average influence. I merge onto this data a measure of each publication's influence: The number of citations received per year since its publication as measured by Google Scholar.

My main finding is that framed field experiments enjoy a substantial advantage in influence relative to natural field experiments. While the average framed field experiment in my data receives 55.6 citations per year, natural field experiments receive 17 to 30 percent fewer citations per year. Even though these differences are not estimated precisely enough to reject a null hypothesis of no difference, the differences are consistently found across six different specifications of control variables.

In a follow-up analysis, I consider whether framed field experiments are responding to concerns about bias by transparently reporting their recruitment procedures. Even though the state of transparency in the literature is remarkably poor, the studies that are transparent enjoy increased influence and this increased level of influence is estimated precisely enough to reject a null hypothesis of no difference. Collectively these findings suggest that economists are influenced by more than the degree of unbiasedness in empirical findings.

The remainder of this study is organized as follows. Section 2 describes the data. Section 3 presents the empirical strategies used to compare the influence of framed and natural field experiments. Estimates from this empirical strategy are presented in Section 4 and Section 5 concludes.

2. Data

The data in this study were collected by the author. This section describes collection of the data, the variables that were observed, and construction of the outcome of interest.

2.1 Universe of Data Considered

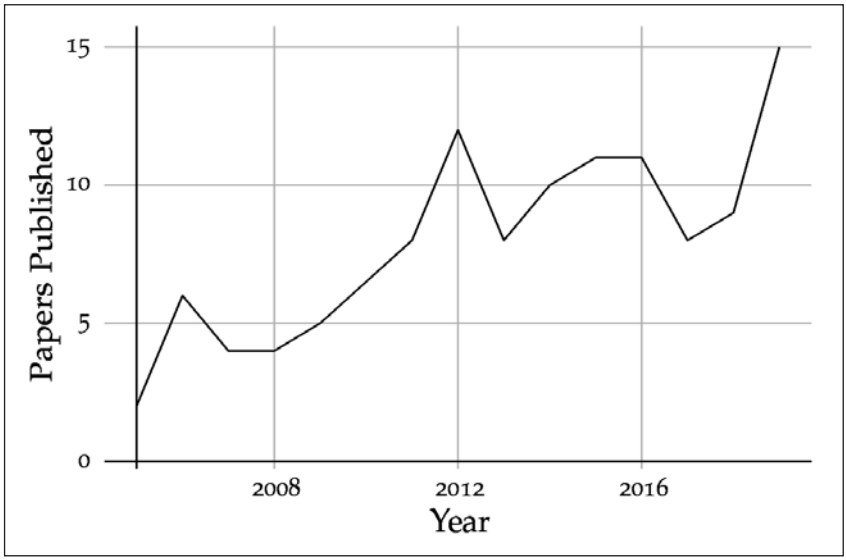
Given that the focus of this study is scientific influence, I consider the so-called "top five" economics journals. The journals that constitute the top five are the *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*. These journals are referred to as the top five because of their influence over the profession. See Heckman and Moktan (2020) for a discussion of the top five journals influence in economics.

I consider papers published in the aforementioned top five journals between 2005 and 2019. I selected 2005 as the starting point because it was the first year after the publication of Harrison and List (2004), which introduced the economics profession to field experiments and provided the taxonomy upon which I further restricted my data.

Reviewing each issue of the top five journals, I found 113 papers that used data from a framed or natural field experiment. Harrison and List (2004) characterize a field experiment as "framed" when a sample is recruited and then randomly assigned to a treatment or control group. In their typology, a field experiment is "natural" when the sample is recruited and randomized without knowledge of the recruitment and randomization. Lab and artefactual field experiments were not considered because there is no analogous design without a self-selected sample against which to compare their relatively influence.

Figure 1 plots the rate at which framed and natural field experiments were published in top five economics journals. For the first five years of this data, just two to five papers were published per year. This publication rate increases dramatically to eight to twelve publications per year between 2011 and 2018. Then at the end of the data there is one final bump to fifteen publications in the year 2019.

Figure 1: Framed and Natural Field Experiments by Year of Publication



Note: This figure plots the number of framed and natural field experiments published in top five economics journals for each year between 2005 to 2019.

This pattern is broadly similar to trends in the experimental literature more generally. Nikiforakis and Slonim (2019); Reuben et al. (2022) show a similar pattern for lab and field experiments in a broader set of journals and over a longer time horizon. However, one interesting point of divergence can be found. While the trend does not die off in Figure 1, Nikiforakis and Slonim (2019); Reuben et al. (2022) find that publications per year starts to fall for lab and field experiments more generally. The divergence between the trend reported in Figure 1 and in Nikiforakis and Slonim (2019); Reuben et al. (2022) is likely attributable to higher publication standards for lab and artefactual field experiments starting between 2010 and 2014.

In the Appendix, I plot the rate at which framed and natural field experiments were published in each of the top five journals. Figure A1 presents this plot, which indicates that the *Quarterly Journal of Economics* led the charge in publishing framed and natural field experiments, with the *American Economic Review* catching up to the *Quarterly Journal of Economics* by 2012. By the end of the sample, the *Journal of Political Economy*, *Econometrica*, and to a lesser extent *Review of Economic Studies* has caught up with the competition.

2.2 Independent Variables Collected

From each of the 113 papers summarized in Section 2.1, the following variables were collected. First, I determined whether a field experiment was framed or natural. A field experiment was determined to be framed if subjects were knowingly recruited to have a random chance of receiving a treatment. Alternatively, a field experiment was deemed natural if subjects were unaware of their participation in an experiment with a randomized treatment. This classification follows the definitions introduced in Harrison and List (2004).

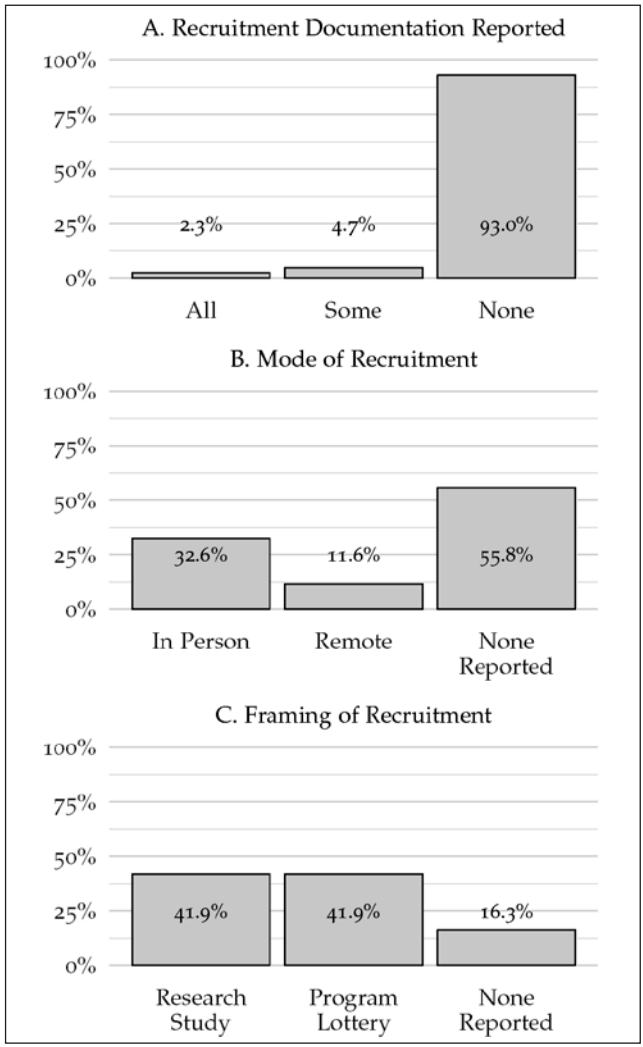
Second, the location in which the experiment was conducted was recorded. To keep this geographic variable parsimonious, I only recorded whether the experiment was conducted in the United States. Experiments conducted outside of the United States were overwhelmingly conducted in developing countries. This reflects the explosion in experiments conducted in developing countries at the start of the 21st century. See Banerjee and Duflo (2009) for a discussion of this work.

Third, I recorded the market in which the outcome of interest was collected. These markets were agriculture, charitable giving, crime, energy, education, finance, health, housing, industrial organization, labor, public, and trade. When an experiment collected outcomes of interest from more than one of these markets, I assigned the primary market targeted by the experimental treatment. For example, Heckman et al. (2013) estimates effects of the Perry Preschool program on educational outcomes as well as labor and criminal outcomes, but it is denoted as educational because that was the primary outcome of interest in the experiment.

Disaggregated patterns of publication for framed and natural field experiments are presented in a series of appendix figures. Figure A2 separately plots the rate at which framed and natural field experiments have been published. Even though they are more challenging to execute, natural field experiments appear more frequently than their framed counterpart in nearly every year. Figure A3 plots publication rates by whether the experiment was run in the United States or in a foreign country, with foreign countries typically representing development experiments. This figure reveals that after 2009 experiments conducted in foreign countries were much more prevalent, reflecting the rise of experiments in developing countries. Figure A4 plots publication rates by the market in which the outcome of interest was measured. Education and health have the most striking peaks in this figure, with education towering over the other markets in 2011 and maintaining that dominance until 2019 when it is surpassed by health.

For studies determined to be framed field experiments, I collected three additional variables. First, I determined whether a study provided any documentation of the recruitment procedures used. Second, I collected whether the study describes the way in which recruitment was framed and, if so, whether subjects were recruited to participate in a study or in a program lottery. Third, I determined whether the study describes the mode of recruitment. That is, whether recruitment was conducted in-person or remotely via the internet or a mail service.

Figure 2: Documentation of Recruitment in Framed Field Experiments



Note: This figure presents the relative frequency of recruitment reporting conventions and methods for experimental evaluation papers published in the top five economics journals between 2005 and 2019.

Figure 2 summarizes the variables collected on framed field experiments. This summary indicates that studies dedicate very little attention to detailing their recruitment procedures. 93 percent of studies provide no documentation of subject recruitment, 56 percent fail to report whether recruitment was conducted in-person or remotely, and 16 percent fail to report how randomization was explained to subjects. This lack of transparency is troubling because it is well understood that when recruitment interacts with a subject's anticipated response to an intervention, framed field experiments do not identify the average treatment effect of the intervention (Heckman and Smith, 1995).

2.3 Outcome of Interest

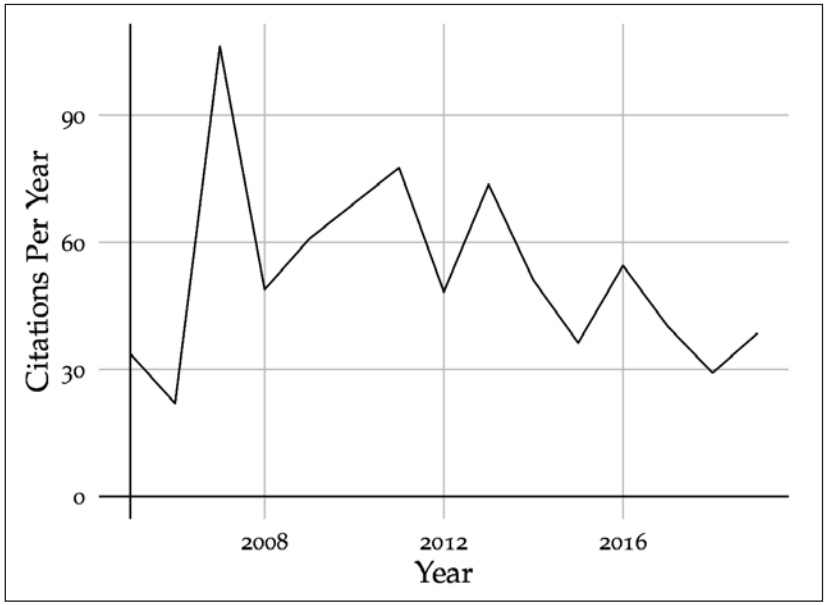
The objective of this investigation is to understand whether self-selection alters the influence of an experimental study. To quantify the influence of a study, I collected the number of Google Scholar citations received by each of the 113 studies. These citation counts were collected in November of 2019. To account for the effect of publication timing on accumulated citations, I normalize the number of citations by the number of years between 2019 and the year in which a study was published. That is, the outcome of interest is Google Scholar citations per year since publication. See Hamermesh (2018) for a discussion of the use of citations as a measure of scientific influence.

For each year in which a study was published, Figure 3 plots the average number of citations per year for natural and framed field experiments. Starting in 2005 and 2006, this figure shows that publications published early in the data have received 22 to 34 citations per year. Then in 2007 this figure jumps to 106 citations per year only to slowly revert back to 29 to 30 citations per year by 2017 to 2019. Digging into the data reveals that the spike in citations per year in 2007 was driven by the publication of a small number of very influential studies. These include a framed field experiment on neighborhood effects with the Moving to Opportunity program in Kling et al. (2007) and a natural field experiment on incentive effects in charitable giving in Karlan and List (2007).

On average, natural field experiments received 45.6 citations per year

and framed field experiments received 55.6 citations per year. Next we consider empirical strategies that can account for potential confounds that could explain why framed field experiments are more highly cited on average than natural field experiments.

Figure 3: Citations per Year for Framed and Natural Field Experiments by Year of Publication



Note: This figure plots the average number of citations per year received by framed and natural field experiments published in top five economics journals for each year between 2005 to 2019.

3. *Empirical Strategies*

In this section, I describe my empirical strategy for measuring the average difference in citations for natural field experiments relative to their framed counterparts. I then describe an empirical strategy for measuring the average difference in citations for framed field experiments with more transparent recruitment procedures. Throughout, I emphasize that the empirical strategies that I employ do not guarantee estimates with a causal interpretation.

3.1 Citation Differences for Framed and Natural Field Experiments

Consider the sample of studies described in Section 2. Let i index each study, Y_i denote the number of citations per year observed in November 2019, and X_i indicate whether study i used data from a natural field experiment. That is $X_i = 1$ when study i used data from a natural field experiment and $X_i = 0$ if it uses data from a framed field experiment.

Furthermore, let O_i denote different combinations of the observables described in Section 2.2 (other than X_i). When O_i includes more than one observable then I imagine it is a vector, whereas if it only includes one of the observables described in Section 2.2, then it is a scalar. These observables are intended to control for some of the potential differences between studies that use natural and framed field experiments. For example, studies on charitable giving in my data typically use data from natural field experiments, whereas studies on education in my data used a mixture of the two. If education research is more highly cited than research on charitable giving, then failing to control for education or charitable giving would confound the effect of a natural field experiment with the topic that the field experiment examined.

To measure the difference in influence of natural field experiments relative to framed field experiments, I estimate the following model,

$$Y_i = \alpha X_i + O_I'\lambda + U_i \qquad (1)$$

where U_i captures unobservable determinants of citations per year. The parameter of interest in equation 1 is α, which measures the difference in citations per year for a natural field experiment relative to a framed field experiment. The parameter (or vector of parameters) λ captures the effect of the observables on citations per year. Given that this is a nuisance parameter, I do not report estimates of λ. Estimates of α are obtained with ordinary least squares and inference on these estimates is conducted with standard errors that are robust to heteroskedasticity.

Before I report estimates of α it is worth emphasizing that a causal interpretation is likely inappropriate. A causal interpretation of α would justified with data where a study's use of a framed or natural field experiment was randomized. While such data would not be impossible to generate, it would be extraordinarily difficult. As a result, I attempt to approximate an experiment by controlling for as many observables as possible and hoping that different combinations of these

observables yields similar results. Nonetheless, the set of potential observables that may influence citations vastly exceeds the set of observables I recorded and I will refer to estimates of0.5 ptas estimated citation differences for framed and natural field experiments, not the effect of a natural field experiment on citations.

3.2 Citation Differences for Transparency in Framed Field Experiments

The empirical strategy developed in Section 3.1 measures the citation difference from a natural relative to a framed field experiment. Another parameter of interest is the citation difference from increased transparency of recruitment in framed field experiments. Toward this end, let i index each framed field experiment and T_i denote different measures of transparency. Working off the data summarized in Figure 2, I construct three measures of transparency. First, $T_i = 1$ if any recruitment documentation was reported and otherwise $T_i = 0$. Second, $T_i = 1$ if the mode of recruitment is described and otherwise $T_i = 0$. Third, $T_i = 1$ if the framing of randomization is described and otherwise $T_i = 0$.

To measure the difference in influence of more transparency in framed field experiment recruitment, I estimate the following model,

$$Yi = \beta T_I + O_i'\lambda + Vi, \qquad (2)$$

where V_i captures unobservable determinants of citations per year for framed field experiments and O_i is the same vector of observables described in Section 3.1. The parameter of interest in equation 2 is β, which measures the difference in citations per year for more transparent framed field experiments. I.e., the difference in citations when $T_i = 1$ relative to $T_i = 0$ on citations per year for framed field experiments. Estimates of β are obtained with ordinary least squares and inference is conducted with standard errors robust to heteroskedasticity.

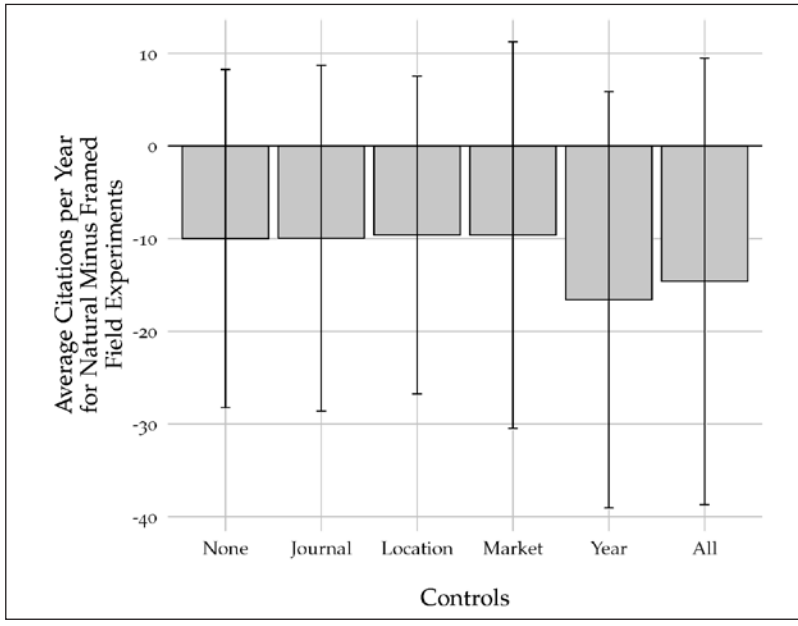
4. Results

This section reports the estimated differences in citations for framed and natural field experiments and for more transparent reporting of recruitment with framed field experiments.

4.1 Estimated Citation Differences for Framed and Natural Field Experiments

Figure 4 reports estimates of α from equation 1. These estimates can be interpreted as the average difference in citations for a natural field experiment relative to a framed field experiment. Each column of Figure 4 employs a different combination of control variables. The first column reports the unadjusted difference in average citations for natural and framed field experiments, while the subsequent columns add different controls. The second column controls for the journal in which a study was published, the third column controls for whether the study relied on data from an experiment conducted in the United States, the fourth column controls for the market in which the experiment was run, and the fifth column controls for the year in which the study was published. The sixth and final column in Figure 4 controls for journal, location, market, and year.

Figure 4: Average Difference in Citations for Natural Field Experiments



Note: This figure reports estimates of α from equation 1. These estimates can be interpreted as the average difference in citations per year for natural relative to framed field experiments. On average, framed field experiments in the data received 55.6 citations per year. Each estimate uses different combinations of controls. The first column features no controls, the second column controls for the journal in which the study was published, the third column controls for whether the study relied on data from an experiment conducted in the United States, the fourth column controls for the market in which the experiment was conducted, the fifth column controls for the year that the study was published, and the sixth column controls for journal, location, market, and year. Brackets in the figure indicate 95 percent confidence intervals, which are constructed with heteroskedastic robust standard errors.

Across every specification of controls in Figure 4 we see that, on average, natural field experiments are cited less frequently than framed field experiments. While the exact difference varies from specification to specification, the differences range from 9.6 to 16.6. To put these differences in citations per year into perspective, the average number of citations received per year for framed field experiments in the sample is 55.6. That is, on average, natural field experiments receive 17.3 to 30.0 percent fewer citations per year than framed field experiments.

Of course, the 95 percent confidence intervals in Figure 4 indicate that the estimates cannot reject the null hypothesis of no difference with high levels of certainty. Nonetheless, the direction of the estimates of α is surprising. It is well understood that self-selection can bias the estimated treatment effects obtained with field experiments. Accordingly, natural field experiments are thought to obtain estimated treatment effects that are less biased and, as a result, more influential. Yet the point estimates in Figure 4 indicate the opposite trend. Furthermore, this trend persists regardless of the controls variables used. Next I report estimates on whether the influence of framed field experiments depends on the transparency of their recruitment procedures.

4.2 Estimated Citation Differences for Transparency in Framed Field Experiments

The estimates in Section 3.1 are discouraging. Researchers conduct natural field experiments at great cost to overcome the self-selection bias that can confound the effects estimated with framed field experiments. Yet Figure 4 shows that

there is no evidence that overcoming self-selection bias leads to more influence.

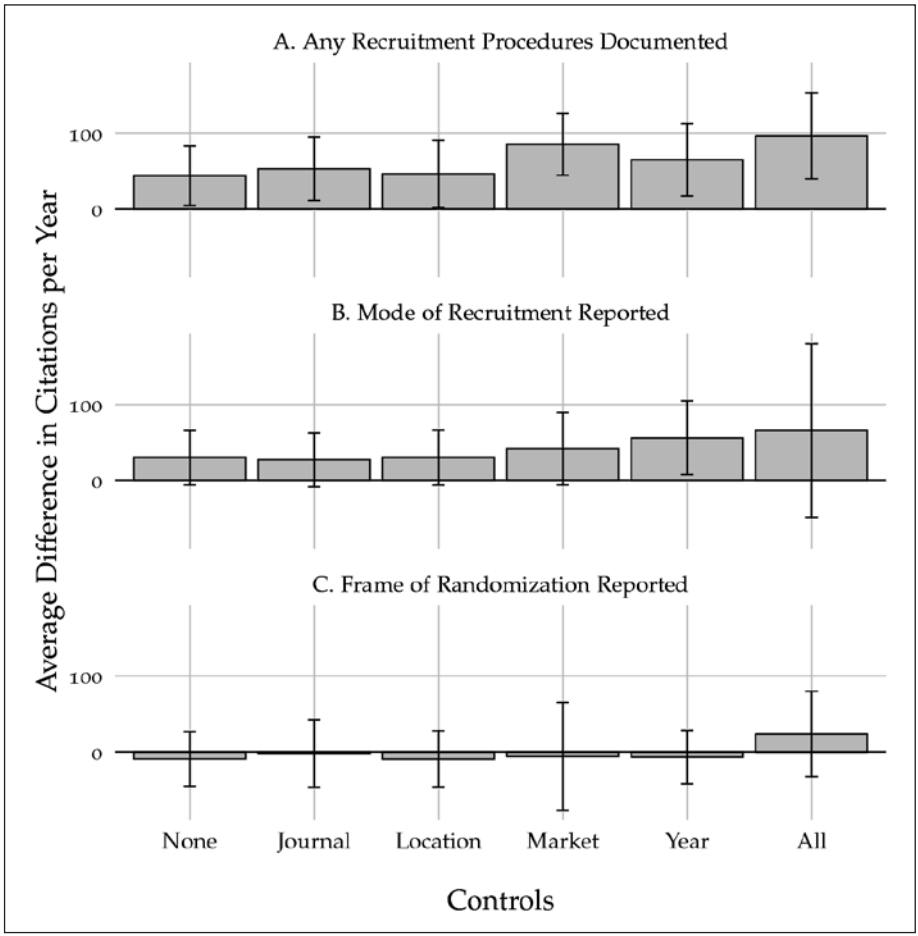
In this section I focus on the influence of framed field experiments. One way that framed field experiments could diminish concerns with self-selection bias is with transparency about the manner in which subjects were recruited to self-select into a study. While the data summarized in Figure 2 indicates that recruitment procedures are very opaque, it could be the case that opacity comes with a penalty on a paper's influence. To evaluate this possibility, I report estimates of equation 2 with the same combination of controls considered in Section 3.1.

Figure 5 shows that transparency can increase the influence of the findings obtained with framed field experiments. The top panel of Figure 5 shows that framed field experiments that documented any of their recruitment procedures have received 44 to 96 citations per year. On a baseline number of citations per year of 55.6, these estimates indicate an increase in citations per year of 79 to 173 percent. Furthermore, the 95percent confidence intervals on the top panel of Figure 5 show that these estimated differences in citations per year are precise enough to reject the null hypothesis of no difference in citations from transparency.

Moving to the middle panel of Figure 5, the definition of transparency is changed to indicate when a study provides any description of the mode in which recruitment took place. For example, if a study described that recruitment was conducted in-person or remotely via a mailer then it would be defined as transparent, whereas if a study had no such description then it would be defined as not transparent. The estimates reported in the middle panel of Figure 5 show that reporting the mode of recruitment increases citations per year by 27 to 66. Converted to percent changes, these estimates reflect 49 to 119 percent increases in citations per year. Yet these estimates are too noisy to reliably reject the null hypothesis of no average difference in citations for most of the specifications of controls.

The bottom panel of Figure 5 reports the average difference in citations when a framed field experiment describes the way in which randomization was framed for subjects. Unlike the top two panels, in this panel the sign of the difference average citations is ambiguous. Some specifications of control indicate that reporting the framing of randomization reduced citations per year by 10 and other specifications indicate that reporting the framing increased citations per year by 24. Across these different specifications, none of the estimated average differences are precise enough to reject a null hypothesis of no average difference.

Figure 5: Average Difference in Citations for Transparent Framed Field Experiments



Note: This figure reports estimates of $\boldsymbol{\beta}$ from equation 2. These estimates can be interpreted as the average difference in citations per year for framed field experiments that are more transparent about their recruitment procedures. The top panel defines a study as more transparent if any documentation of recruitment procedures are available, the middle panel defines a study as more transparent if the mode of recruitment is discussed in the paper, and the bottom panel defines a study as more transparent if it describes the framing of randomization. On average, framed field experiments in the data received 55.6 citations per year. Each estimate uses different combinations of controls. The first column features

no controls, the second column controls for the journal in which the study was published, the third column controls for whether the study relied on data from an experiment conducted in the United States, the fourth column controls for the market in which the experiment was conducted, the fifth column controls for the year that the study was published, and the sixth column controls for journal, location, market, and year. Brackets in the figure indicate 95 percent confidence intervals, which are constructed with heteroskedastic robust standard errors.

5. Conclusion

Why do experimental economists conduct framed field experiments instead of natural field experiments? Unlike with natural field experiments, the results of a framed field experiment can call pray to self-selection bias. This study presents evidence that is consistent with the following explanation: Experimental economists conduct framed field experiments because there is no increase in scientific influence that follows from the more difficult task of conducting a natural field experiment. While this is a discouraging finding, it is counterbalanced by a more promising one. Even though framed field experiments are remarkably opaque about their recruitment of samples, the framed field experiments that are transparent enjoy a substantial premium in scientific influence.

References

Banerjee, Abhijit and Esther Duflo, "The Experimental Approach to Development Economics," *Annual Review of Economics,* 2009*, 1 (1)*, 151-178.

Hamermesh, Daniel S., "Citations in Economics: Measurement, Uses, and Impacts," *Journal of Economic Literature,* 2018*, 56* (1), 115-156.

Harrison, Glenn W. and John A. List, "Field Experiments," *Journal of Economic Literature,* 2004, *42* (4), 1009–1055.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

– and Jeffrey A. Smith, "Assessing the Case for Social Experiments," *Journal of Economic Perspectives,* 1995, *9* (2), 85-110.

– and Sidharth Moktan, "Publishing and Promotion in Economics: The Tyranny of the Top Five," *Journal of Economic Literature*, 2020, *58* (2), 419-470.

– , Rodrigo Pinto, and Peter Savelyev, "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," American Economic Review, 2013, 103 (6), 2052–2086.

– , Rodrigo Pinto, and Peter Savelyev, "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 2013, *103* (6), 2052–2086.

Karlan, Dean and John A. List, "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment," *American Economic Review*, 2007, *97 (5)*, 1774-1793.

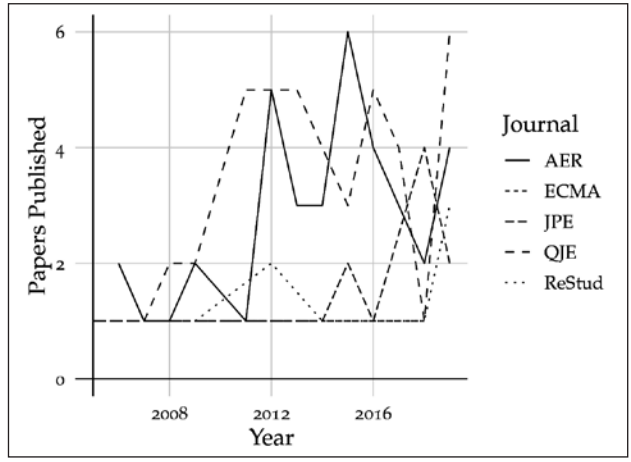
Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, [2007], *75* (1), 83-119.

Nikiforakis, Nikos and Robert Slonim, "Editors' Preface: Trends in experimental economics (1975-2018)," *Journal of the Economic Science Association*, 2019, 5, 143-148.

Reuben, Ernesto, Sherry Xin Li, Sigrid Suentens, Andrej Svorencik, Theodore Turocy, and Vasileios Kotsidis, "Trends in the publication of experimental economics articles," *Journal of the Economic Science Association*, 2022, *8*, 1–15.

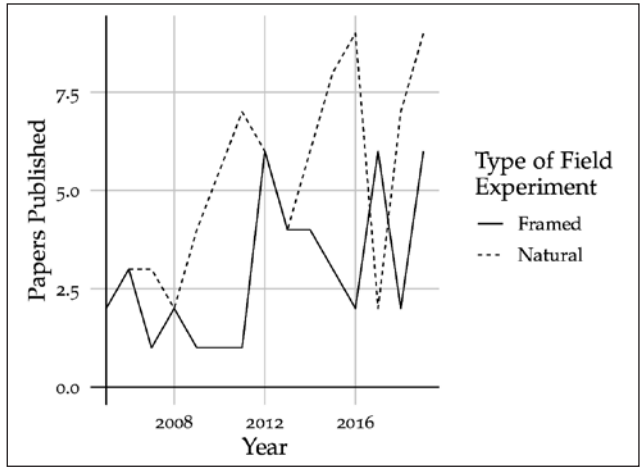
Appendix

Figure A1: Framed and Natural Field Experiments by Year and Journal



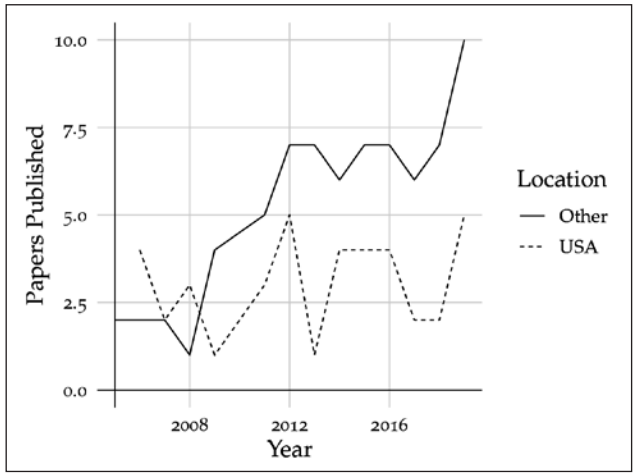
Note: This figure plots the number of framed and natural field experiments published in each of the top five economics journals for each year between 2005 to 2019.

Figure A2: Framed or Natural Field Experiments by Year of Publication



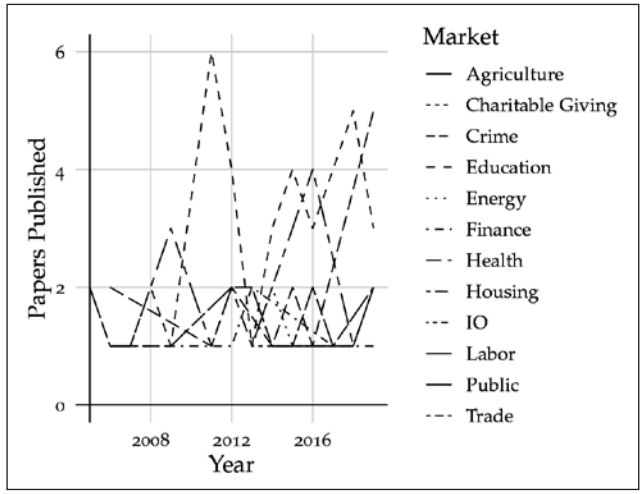
Note: This figure plots the number of framed and natural field experiments published in the top five economics journals for each year between 2005 to 2019.

Figure A3: Framed and Natural Field Experiments by Year and Location of Experiment



Note: This figure plots the number of framed and natural field experiments published in the top five economics journals for each year between 2005 to 2019 by whether the experiment was conducted in or outside the United States.

Figure A4: Framed and Natural Field Experiments by Year and Market



Note: This figure plots the number of framed and natural field experiments published in the top five economics journals for each year between 2005 to 2019 by whether the market in which the experiment was conducted.

Racialized Courts: How racial attitudes shape perceptions of the American judicial system

Philip Chen[1]

Over the past several years, experiments (in a variety of forms) have taken hold of the social scientific world. We are lucky to have shared the Upton Forum stage with Professor John List, an innovative and dedicated economist who has frequently relied on experimental methods in his work (Harrison and List 2004; Levitt and List 2007, 2009). In the classical sense of experimental methods, these techniques are used to uncover causal relationships between variables (McDermott 2002). While certainly useful for breaking down the "black box" of causality (in certain circumstances, a la Green, Ha, and Bullock 2010), experiments also allow for substantive tests of alternative theories and conventional wisdom.

In this paper, we test just such an alternative theory about public opinion of the U.S. court system. Canonically, support for the courts (usually measured in relation to the Supreme Court) is stable and relatively unchanging in the public (Gibson and Caldeira 2009; Gibson, Caldeira, and Baird 1998). Indeed, the idea that the public is generally supportive of the courts and view them as legitimate spawned decades of work about how the courts are (or are not) responsive to public opinion (Casillas, Enns, and Wohlfarth 2011; Epstein and Martin 2010; Giles, Blackstone, and Vining Jr 2008; Hall 2014).

In spite of this work, however, there is good reason to doubt that the courts, and in particular, the modern-day Supreme Court, is truly insulated from the whims of public opinion. We use a variety of experiments to show that, when the public learns about decisions made by various US courts, they do not maintain

1 Phil Chen is an Assistant Professor of Political Science at the University of Denver, formerly Assistant Professor of Political Science at Beloit College.

high levels of support for the courts, but instead bring their racial attitudes to bear on these evaluations.

Growing evidence shows that opinion about the courts and the behavior of the courts are becoming polarized along partisan and ideological lines (Bartels and Johnston 2013, 2020; Epstein et al. 2007; McKenzie 2012). This alone should give us pause as to whether the courts are truly able to maintain legitimacy in the face of increasingly partisan and acrimonious decisions. But the courts face increasing concerns centered around issues of race. In the 233-year history of the US Supreme Court, 116 justices have served on the court. Only 4 (3.4%) have not been White.[2] The federal court system as a whole is not significantly better. Of the 813 active Article III judges (as of early 2021), 599 (about 74%) were White. It should be obvious to most observers that the US court system is a white-dominated institution.

This pattern, on its own, would be concerning for individuals concerned about the representativeness of the court system. Unfortunately, the concerns do not stop simply at an unrepresentative court. Prior to Barack Obama's election in 2008, racial attitudes did not correspond closely with partisan attitudes. Thus, if the courts polarized along partisan lines, we wouldn't have expected that racial attitudes would present a similar challenge to court legitimacy. With Obama's election, however, the public began to sort themselves into parties along racial attitudinal lines (Tesler 2013, 2016; Tesler and Sears 2010). This means that those with more egalitarian racial attitudes became much more likely to identify as Democrats, while those with more anti-egalitarian racial attitudes became much more likely to identify as Republicans.

This presents a problem for the court system. Indeed, when you ask people how they feel about the courts in the abstract, they are generally supportive of the courts. Yet the vast majority of people do not think about the courts in the abstract. Particularly in a post-Dobbs world, most people, when they think about the courts at all, think about them because of a specific policy or decision that was announced.[3] Furthermore, the courts, almost by definition, must address controversial issues concerning constitutional protections for racial and ethnic minorities. This means that, despite theories to the contrary, it is highly likely that

2 Current Justices Thomas, Sotomayor, and Brown Jackson, as well as former Justice Marshall.

3 For those who are unfamiliar, Dobbs v. Jackson Women's Health Organization was the 2022 Supreme Court decision that overturned the precedent set in Roe v. Wade which protected a woman's right to an abortion.

support for the courts will split along racial attitudinal lines as the courts make decisions on cases involving race and ethnicity.

This leads to an expectation that, rather than evaluating courts in the abstract, the *substantive content* of decisions should prove to be most important for citizen evaluations of the courts. While conventional wisdom among many politicians remains that citizens look at how descriptively representative the courts are (that is, do I see people who share my ethnic, racial, or gender identity on the court?), we posit that the public is actually more sophisticated than this. Instead, members of the public evaluate whether a decision accords with their racial outlook on the world.

This results in three expectations. First, in the absence of specific outcome information about a court decision, the public will rely on racial attitudes to evaluate the courts when the courts make specific efforts on race or ethnicity. However, in the more common case of available policy information, anti-egalitarian racial attitudes will correlate with less positive evaluations of the courts when they affirm the rights and status of marginalized groups (and vice versa). Related to this, the race or ethnicity of the judge involved should have little effect on opinion. Thus, rather than a story about racial discrimination based on judge characteristics, we expect that court opinion is tied to the policy significance of the decision, coupled with an individual's own racial beliefs.

We tested these expectations using several different surveys with a variety of experiments embedded within the surveys. Within these surveys were two commonly used measure of anti-egalitarian attitudes. The first, and most commonly used, measure is known as "racial resentment" or "symbolic racism." This measure came into fashion in the 1980s as old-fashioned, biological racism became less common (Bobo and Kluegel 1997; Virtanen and Huddy 1998). Racially resentful beliefs are founded not on ideas of biological inferiority, but on the idea that Blacks do not reflect the core American values of individualism and equal opportunity (Kinder and Sanders 1996; Kinder and Sears 1981; Myrdal 1944; Sears and Henry 2003). In particular, those high in racial resentment endorse (to varying degrees) four beliefs about Black people, namely that they no longer face significant discrimination in the US, that the disadvantages faced by Black people are the result of not working hard enough, that they are demanding too much change too fast, and that Black people are getting more than they deserve, relative to White people. The now standard way to measure racial resentment asks respondents to agree or disagree with the following four statements:

1. Irish, Italian, Jewish, and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors.
2. Generations of slavery and discrimination have created conditions that make it difficult for blacks to work their way out of the lower class.
3. Over the past few years, blacks have gotten less than they deserve.
4. It's really a matter of some people not trying hard enough; if blacks would only try harder they could be just as well off as whites.

This measure of racial resentment has proven itself useful across numerous predictive environments, including policy beliefs (Bloeser and Williams 2022; Gilens 1995, 1999; Rouhani et al. 2022; Wallsten et al. 2017), candidate support (Kinder and Sanders 1996; Mendelberg 2001; Tesler and Sears 2010), and even non-racial policies that are simply associated with high-profile Black politicians with a phenomenon known as racial spillover (Benegal 2018; Benegal and Motta 2022; Chen and Mohanty 2018; Farhart and Chen 2022; Luttig and Motta 2017; Sheagley, Chen, and Farhart 2017; Tesler and Sears 2010).

Despite this usefulness, racial resentment only captures anti-Black racism. To address policies that may be coded racialized or ethnicized but are not anti-Black, Ramirez and Peterson (2020) developed the Latina/o Racism-Ethnicism (LRE) scale, which relies on similar theoretical foundations to racial resentment, but more specifically in relation to Latinx individuals. This measure was included in some of the surveys and involves the following four statements:

1. The Irish, Italians, Jews and many other ethnic groups immigrated to the United States legally. Latinos and Hispanics should do the same without any special favors.
2. Anti-immigration sentiment and racism have created conditions that make it difficult for Latinos and Hispanics to succeed in America.
3. Latinos and Hispanics would be more welcome in the United States if they would try harder to learn English and adopt US customs like past immigrant groups have done.
4. Critics of immigration and the media have overblown the number of crimes committed by Latinos and Hispanics within the United States.

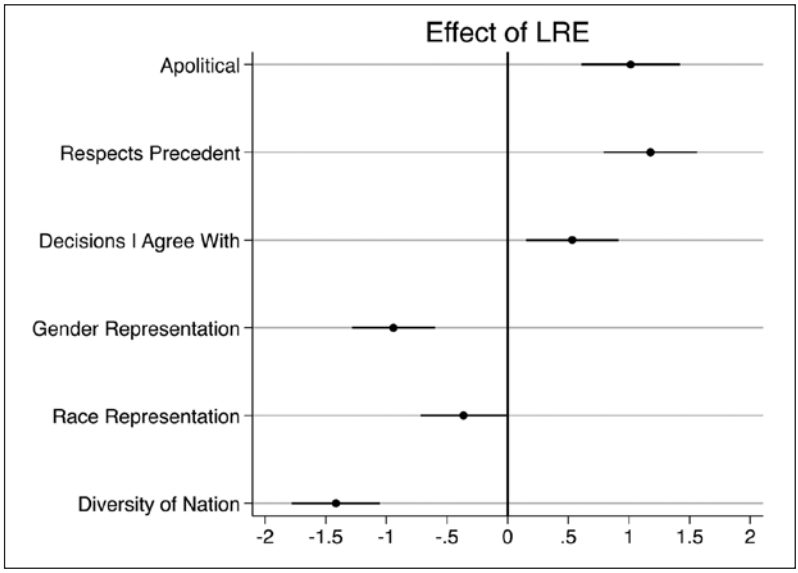
Using racial resentment and LRE as measures of anti-egalitarian attitudes, we tested the expectations in a variety of manners. To assess whether anti-egalitarian attitudes affect abstract support for court reform, respondents were asked to "rank these statements in order of how important they are when you consider how much you TRUST the Supreme Court." They are then presented with the following six statements:

1. Seeing Justices who represent the racial and gender diversity of the nation
2. Seeing Justices that share your racial identity
3. Seeing Justices that share your gender identity
4. Having a Supreme Court that issues decisions you agree with
5. Having a Supreme Court that respects past precedent
6. Having a Supreme Court that is removed from politics

In this survey, the only available measure of anti-egalitarian attitudes was the LRE. Using this measure, we ran regression models predicting how an individual ranked each statement while controlling for respondent gender, race, Hispanic ethnicity, party identification, household income, education, and geographic region. These results appear in Figure 1.

As Figure 1 shows, in the abstract, the expected pattern emerges. Those with more anti-egalitarian attitudes are significantly more likely to rank the non-racialized statements higher. This is particularly interesting in regard to the respecting precedent statement, as past court precedent has tended to favor Whites. Thus, even though the statement is not explicitly racialized, the implication is one that favors one racial group over another. In contrast, those with more anti-egalitarian attitudes are significantly less invested in a court that is racially or gender representative.

Figure 1: Effects of Latina/o Racism-Ethnicism on Rank Ordering of Supreme Court Statements



As noted, however, these are evaluations of the court in the abstract. The courts, however, do not decide cases randomly, but are rather likely to make their decisions based on partisan, ideological, or other considerations. In turn, as individuals learn about court decisions, we expect that their beliefs about the court will vary based on their egalitarian attitudes and whether the court decision affirms or rejects the rights and status of racially marginalized groups.

It is here that experimental research becomes increasingly helpful. Continuing to rely on observational research is problematic because of some assumptions that must be made. First, it relies on the assumption that people are well-informed about the issue or decision under consideration. Decades of research suggests that this is not true (Dancey and Sheagley 2013; Delli-Carpini and Keeter 1996). Second, it relies on the media to report information clearly. Again, while clearly conveyed information is helpful (Chen and Luttig 2021), the media is historically poor at performing this duty (Bennett 2016; Iyengar and Kinder 1987).

Thus, we turn to a series of survey experiments whereby respondents are given a small amount of information about a court case and are then asked to express their trust, legitimacy, and perceptions of bias of the courts. Within these

experiments, we varied the race of the judge and the direction of the decision. If individuals simply looked to the racial composition of the court, the race of the judge would likely drive support or opposition to the courts based on the respondent's racial attitudes. On the other hand, if, as we suspect, people evaluate the courts based on whether a decision's policy content accords with their racial worldview, then the direction of the decision, contingent on racial attitudes, should drive support or opposition to the court.

The results from three different experiments (run across two different surveys in the summer and fall of 2021) confirm expectations. In the first experiment, run in the summer of 2021, respondents read the following vignette:

> Judge Michael [Smith-Hollins/Angel-Rodriguez] recently ruled on a case brought by the National Immigrant Law Center against the Department of Homeland Security. In the case, the law center alleges that DHS was unfairly targeting immigrants from central and south America for deportation by focusing on immigrants crossing at the southern US border. Judge [Smith-Hollins/Angel-Rodriguez] ruled in favor of the [Department of Homeland Security, finding that their policies were based on the number of immigrants crossing the border, rather than national origin. He allowed the policy to remain in effect/National Immigrant Law Center, finding that the policy was discriminatory based on national origin. He struck down the policy]

The experiment was a 2x2 fully crossed design with a name condition (White-coded name vs Latinx-coded name) and a decision direction condition (ruling the law discriminated based on national origin vs ruling the law did not discriminate). Respondents were randomly assigned to one of the four conditions. This 2x2 crossed design was used for all three experiments.

In the second experiment from the same summer 2021 survey, respondents read the following vignette:

> A district court in Georgia recently ruled on a newly passed voting law that prohibits early voting on Sundays. Supporters claim voters have plenty of other times to vote, while opponents claim the law unfairly targets black voters who often vote after Sunday church services. Judge [Dustin/D'Andre] Frazier, writing for the majority,

> [upheld the new law, saying that the law does not discriminate on the basis of race/struck down the new law, saying that the law unfairly discriminates against Black voters]

In the third experiment, using a new sample obtained in the fall of 2021, respondents read the following vignette:

> A district court in New Mexico recently ruled on a newly passed voting law that requires voters to present a state-issued photo identification when voting. Supporters claim the new law reduces voter fraud while opponents say it unfairly targets poor and minority voters who have a harder time obtaining an ID. Judge [Michael Smith-Hollins/Miguel Angel-Rodriguez] [upheld the new law, saying that the voter identification law does not unfairly discriminate against poor and minority voters/struck down the new law, saying that the voter identification law unfairly discriminates against poor and minority voters]

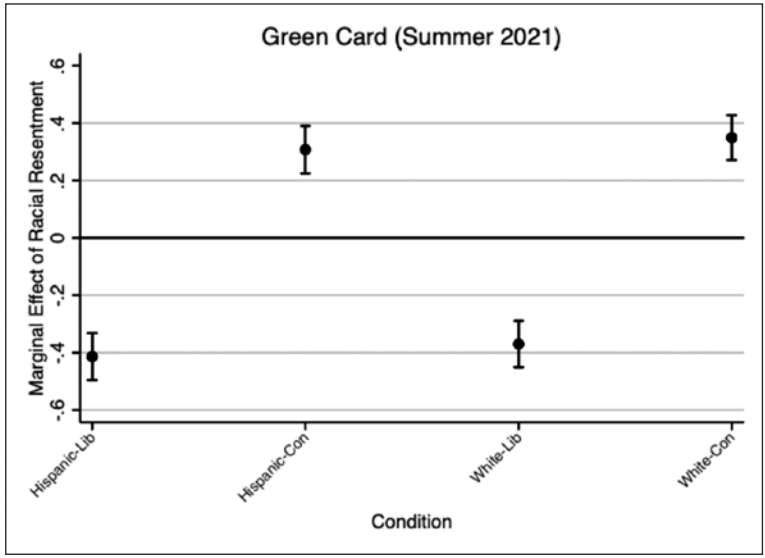
After each vignette, respondents were asked the following four questions and asked to respond on a five-point Likert scale:

1. Based on what you read, where would you place your level of support for this decision?
2. Based on what you read, where would you place your level of trust in this judge?
3. Based on what you read, how biased do you believe this judge was in reaching this decision?
4. After reading about this case, how much do you trust the judicial system to make the correct decision in similar cases?

Responses to these four questions were aggregated into a single index, as alpha coefficients were consistently at 0.75 or above, suggesting the four questions measured a single underlying construct. The court support index was then regressed on the interaction between racial attitudes and condition assignment, along with the first-order terms for racial attitudes and experimental condition.

Figure 2 shows the expected pattern. We see the race of the judge makes little difference in how respondents evaluate the court. Whether the judge is White or Latinx makes no difference to respondents. The *direction* of the decision, however, is deeply important. For racially resentful individuals, their support for the court increases significantly when the court allows the deportation policy to stand, and it decreases significantly when the court strikes down the law.[4]

Figure 2: Effects of Racial Resentment on Court Evaluations, by Experimental Condition



4 It bears noting that the effects are symmetrical. For racially progressive individuals, the court striking down the law leads to improved evaluations, while the court upholding the law leads to depressed evaluations.

Figure 3: Effects of Racial Resentment on Court Evaluations, by Experimental Condition

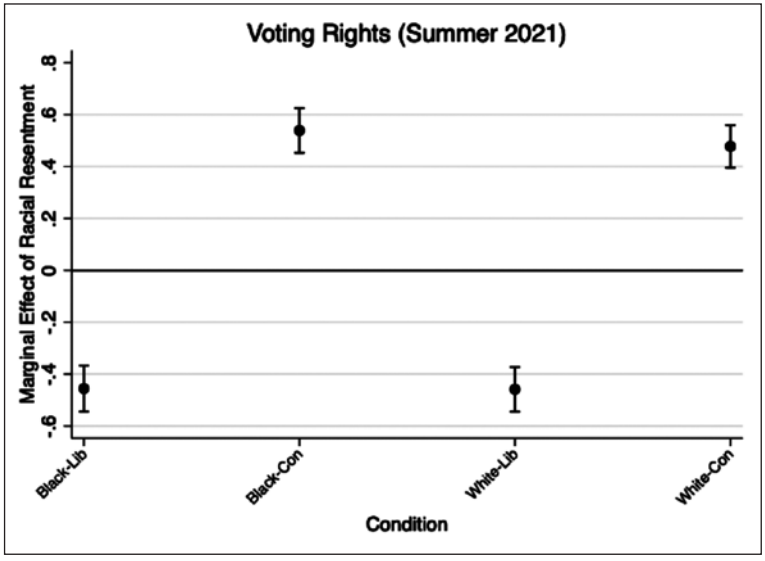
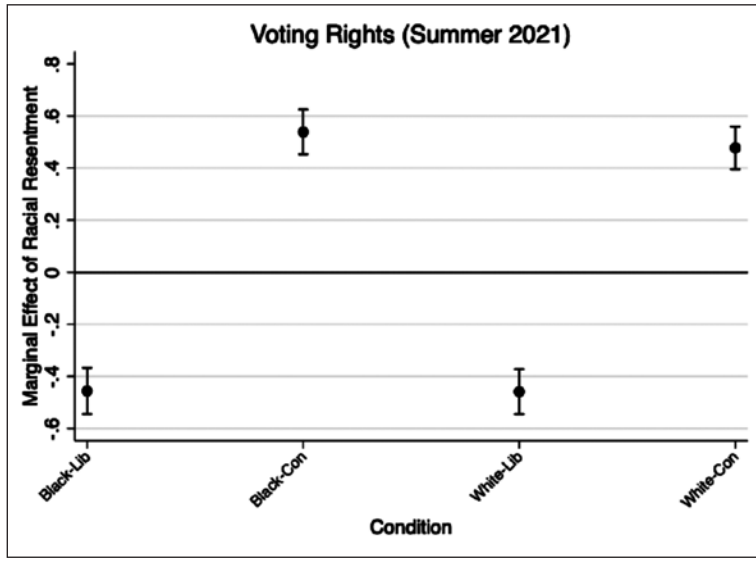


Figure 3 shows that the same pattern replicates when examining voting rights. Again, the race of the judge makes no difference. However, racial resentment plays a strong role, with its effect dependent on the decision direction. Racial resentment improves evaluations of the court when they uphold the law prohibiting Sunday voting, and hurts evaluations when the court strikes down the law.

The results in Figure 4 show that the observed effects are not simply a methodological artifact of the measurement of racial resentment. In this survey, we reran the voting rights experiment, now substituting Latinx cues and using the LRE scale instead of racial resentment. The exact same pattern emerges, with higher levels of Latina/o Racism-Ethnicism predicting greater support for the courts when they uphold the photo identification law, and less support when the court strikes down the policy.

Figure 4: Effects of LRE on Court Evaluations, by Experimental Condition



Interestingly, these patterns are not isolated to decision vignettes. Provided some substantive information is provided, we can see similar results with more diffuse experiments. Again, in the summer 2021 survey, we embedded an experiment around the potential retirement of Justice Breyer (who would go on to retire in 2022). Respondents were assigned to one of two conditions, either a racial frame or an ideological frame. In the ideology frame, respondents saw the following vignette:

> On the current U.S. Supreme Court, Justice Stephen Breyer, at 82 years of age, is the oldest current justice. Appointed in 1994 by President Bill Clinton, Justice Breyer is the second longest serving member of the current Supreme Court. There have been numerous media reports and calls from politicians and pundits speculating about Justice Breyer's potential retirement. If Justice Breyer were to retire before the 2024 election, President Joe Biden would nominate his successor.

> If Justice Breyer were to retire, President Biden has pledged to nominate a strong liberal justice. President Biden spoke recently about the need to "preserve" the current ideological balance of power on the Supreme Court. The current Supreme Court is split between 6 conservative justices and 3 liberal justices. Replacing Breyer with a liberal justice would not change the balance of power, but would maintain 3 liberal justices.
>
> How strongly do you support President Biden's pledge to appoint a strong liberal justice to the Supreme Court if given the opportunity to nominate a justice?

In the race vignette, respondents saw the same first paragraph, but then saw a different second paragraph and question, which read:

> If Justice Breyer were to retire, President Biden has pledged to nominate a black woman to replace him. President Biden spoke recently about the need to make the court more "representative" of the population. The current Supreme Court has four White male justices, two White women justices, one Black male justice, and one Hispanic woman justice. Replacing Breyer with a black woman would change the racial and gender composition of the court.
>
> How strongly do you support President Biden's pledge to appoint a black woman to the Supreme Court if given an opportunity to nominate a justice?

Figure 5: Effects of Racial Resentment on Support for Biden Appointment, by Experimental Condition

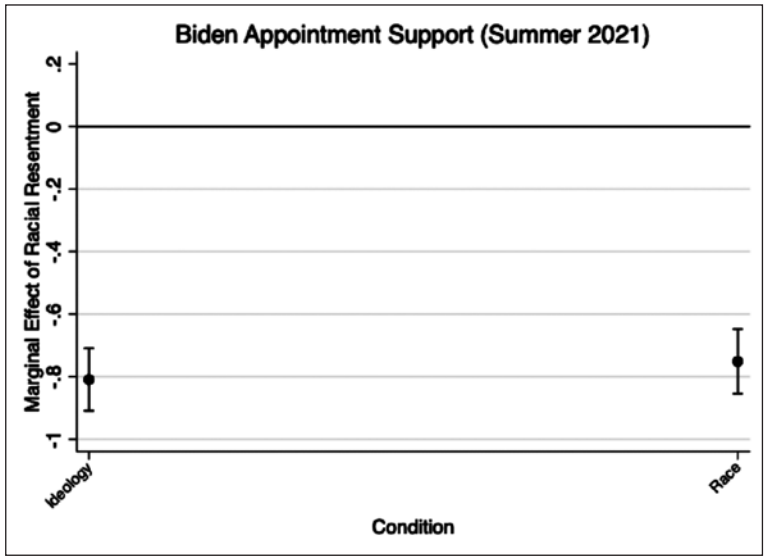


Figure 5 clearly shows how close racial attitudes and ideological beliefs are tied in US public opinion. Again, we regress support for Biden's pledge on condition assignment interacted with racial resentment. We find that, regardless of whether it is framed as an ideological or racial justification, more racially resentful individuals are significantly less supportive of Biden's pledge than those who are racially progressive.

We can turn to our final experiment for further confirmation of the important role that racial attitudes play in shaping the public's response to racialized information about the courts. In this experiment, respondents learned about the DC Circuit Court and Biden's nomination of then-judge (now justice) Ketanji Brown Jackson to fill the seat vacated by Merrick Garland. All respondents saw this part of the vignette:

> The U.S. Court of Appeals for the District of Columbia (often known as the DC Circuit Court) is seen as the second most powerful court in the country, second only to the Supreme Court. Eleven judges sit on this court and they decide numerous cases on the constitutionality

> of laws passed by Congress. Three current Supreme Court justices (John Roberts, Clarence Thomas, and Brett Kavanaugh) served on this court before being appointed to the Supreme Court, as did former justices like Warren Burger and Ruth Bader Ginsburg.

Respondents were also assigned to one of five conditions, detailed below. In the control condition, respondents read the following:

> The current DC Circuit Court has six judges appointed by Democratic presidents and four judges appointed by Republican presidents. Current U.S. Attorney General Merrick Garland, who was appointed by Democratic President Bill Clinton, left the court when he was selected to be the attorney general.
>
> President Biden will nominate Garland's replacement and has nominated Ketanji Brown Jackson, a current federal district court judge, to fill the seat.

In the Democratic condition, respondents read the following:

> The current DC Circuit Court has six judges appointed by Democratic presidents and four judges appointed by Republican presidents. Current U.S. Attorney General Merrick Garland, who was appointed by Democratic President Bill Clinton, left the court when he was selected to be the attorney general.
>
> President Biden will nominate Garland's replacement and has nominated Ketanji Brown Jackson, a current federal district court judge, to fill the seat. This nomination will not change the partisan composition of the DC Circuit Court. Judges appointed by Democratic presidents will still constitute the majority of the court.

In the Republican condition, respondents read the following:

> The current DC Circuit Court has three judges appointed by Democratic presidents and seven judges appointed by Republican pres-

idents. Current U.S. Attorney General Merrick Garland, who was appointed by Democratic President Bill Clinton, left the court when he was selected to be the attorney general.

President Biden will nominate Garland's replacement and has nominated Ketanji Brown Jackson, a current federal district court judge, to fill the seat. This nomination will not change the partisan composition of the DC Circuit Court. Judges appointed by Republican presidents will still constitute the majority of the court.

In the racial shift condition, respondents read the following:

The current DC Circuit Court has five non-White judges and five White judges. Current U.S. Attorney General Merrick Garland, who is White, left the court when he was selected to be the attorney general.

President Biden will nominate Garland's replacement and has nominated Ketanji Brown Jackson, a Black woman and current federal district court judge, to fill the seat. This nomination will shift the racial composition of the court, where non-White judges will hold a six to five majority on the court.

In the racial no change condition, respondents read the following:

The current DC Circuit Court has two non-White judges and eight White judges. Current U.S. Attorney General Merrick Garland, who is White, left the court when he was selected to be the attorney general.

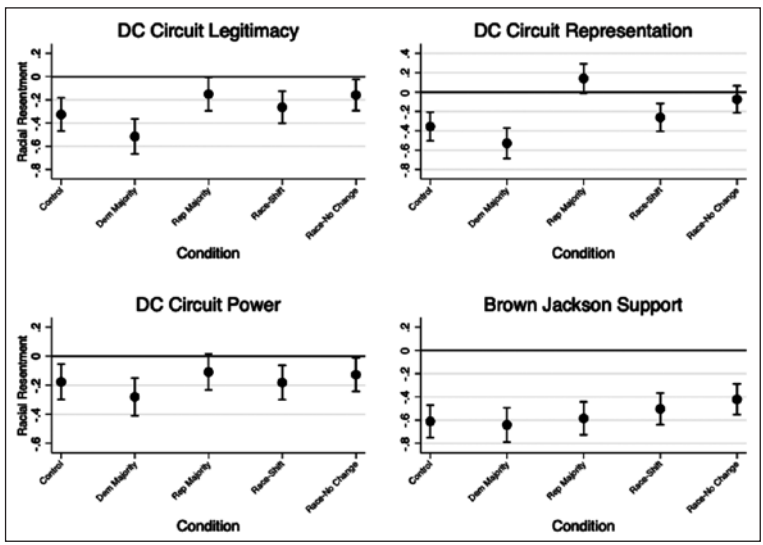
President Biden will nominate Garland's replacement and has nominated Ketanji Brown Jackson, a Black woman and current federal district court judge, to fill the seat. This nomination will not shift the racial composition of the court, as White judges will maintain an eight to three majority on the court.

Following the vignette, respondents answered the following four questions:

1. Given the power and composition of the DC Circuit Court, how legitimate do you believe their decisions are?
2. How representative do you believe the DC Circuit Court is of U.S. citizens?
3. Should the DC Circuit Court have more power, less power, or about the same amount of power?
4. Do you support or oppose Judge Brown Jackson's nomination to the DC Circuit Court?

Figure 6 shows the effects of racial resentment on responses to these four questions in the five different conditions. The results here, while less definitive than in the other experiments, still point towards a similar conclusion. For the first three questions, higher levels of racial resentment lead individuals to view the DC Circuit as less legitimate, less representative, and wish that it had less power. These effects, however, are present largely in the control condition and when Democrats or racial minorities hold the majority on the court. When Republican appointees control the court or when Whites outnumber non-Whites on the court, the effects of racial resentment are either minimized or insignificant.

Figure 6: Effects of Racial Resentment on DC Circuit Beliefs, by Experimental Condition



In the case of support of Brown Jackson's nomination, however, we see a clear, direct effect of racial resentment. In this case, individuals do not appear to be evaluating her nomination based on policy or substance, but simply on her status as a prominent black woman. Regardless of whether her appointment to the DC Circuit would have any meaningful effect on the court composition, higher levels of racial resentment drive down support for her nomination.

Taken together, these results show the power of experiments in the social sciences to challenge conventional wisdom. Observational research does not allow for the testing of counter-factuals or alternative decisions in court research. Fortunately, survey experiments do provide this opportunity. When we do this, we find that racial attitudes have fundamentally altered the political landscape when it comes to the federal court system.

Contrary to popular belief, support for the federal courts is not nearly as stable as thought. Instead, this support is deeply racialized. Not only that, but that racialized support is contingent on the content and substance of court decisions and proposals. When a policy is racially egalitarian, racially conservative individuals exhibit lower levels of support for the courts, and vice versa.

This raises significant concern for those who value the role that the federal court system plays in the US. The courts, lacking any enforcement mechanism, are reliant on good will from the public if their actions are to be respected and followed. The nature of the courts means they cannot avoid questions of race and ethnicity, but when they wade into these issues, support for the courts polarize along racial lines. No matter how the courts decide on a case, some individuals will be more supportive of the courts, while others will be less supportive.

While not unique to the court system, this extreme racial polarization poses an existential threat to the health of democratic institutions. If our institutions cannot make decisions and policies related to race or ethnicity without sacrificing public support among a segment of the population, we are doomed to a world where politics either maintains a façade of colorblindness or suffers public backlash from a significant portion of the populace.

References

Bartels, Brandon L., and Christopher D. Johnston. 2013. "On the Ideological Foundations of Supreme Court Legitimacy in the American Public." *American Journal of Political Science* 57(1): 184–99.

——. 2020. *Curbing the Court: Why the Public Constrains Judicial Independence*. Cambridge, United Kingdom ; New York, NY, USA: Cambridge University Press.

Benegal, Salil D. 2018. "The Spillover of Race and Racial Attitudes into Public Opinion about Climate Change." *Environmental Politics* 27(4): 733–56.

Benegal, Salil D., and Matthew Motta. 2022. "Overconfident, Resentful, and Misinformed: How Racial Animus Motivates Confidence in False Beliefs."

Bennett, W. Lance. 2016. *News: The Politics of Illusion*. University of Chicago Press.

Bloeser, Andrew J., and Tarah Williams. 2022. "The Color of Class Politics: Economic Position, Racial Resentment, and Attitudes about Redistribution." *Politics, Groups, and Identities* 10(2): 233–53.

Bobo, Lawrence, and James R. Kluegel. 1997. "Status, Ideology, and Dimensions of Whites' Racial Beliefs and Attitudes: Progress and Stagnation." *Racial attitudes in the 1990s: Continuity and change*: 93–120.

Casillas, Christopher J., Peter K. Enns, and Patrick C. Wohlfarth. 2011. "How Public Opinion Constrains the US Supreme Court." *American Journal of Political Science* 55(1): 74–88.

Chen, Philip G., and Matthew D. Luttig. 2021. "Communicating Policy Information in a Partisan Environment: The Importance of Causal Policy Narratives in Political Persuasion." *Journal of Elections, Public Opinion and Parties* 31(3): 329–47.

Chen, Philip G., and Ruchika Mohanty. 2018. "Obama's Economy: Conditional Racial Spillover Into Evaluations of the Economy." *International Journal of Public Opinion Research* 30(3): 365–90.

Dancey, Logan, and Geoffrey Sheagley. 2013. "Heuristics Behaving Badly: Party Cues and Voter Knowledge." *American Journal of Political Science* 57(2): 312–25.

Delli-Carpini, Michael, and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. Yale University Press.

Epstein, Lee, and Andrew D. Martin. 2010. "Does Public Opinion Influence the Supreme Court-Possibly Yes (But We're Not Sure Why)." *U. Pa. J. Const. L.* 13: 263.

Epstein, Lee, Andrew D. Martin, Kevin M. Quinn, and Jeffrey A. Segal. 2007. "Ideological Drift among Supreme Court Justices: Who, When, and How Important." *Northwestern University Law Review* 101(4): 1483–1542.

Farhart, Christina E., and Philip G. Chen. 2022. "Racialized Pandemic: The Effect of Racial Attitudes on COVID-19 Conspiracy Theory Beliefs." *Political Misinformation in the Digital Age During a Pandemic: Partisanship, Propaganda, and Democratic Decision-Making*.

Gibson, James L., and Gregory A. Caldeira. 2009. "Knowing the Supreme Court? A Reconsideration of Public Ignorance of the High Court." *The Journal of Politics* 71(02): 429–41.

Gibson, James L., Gregory A. Caldeira, and Vanessa A. Baird. 1998. "On the Legitimacy of National High Courts." *American Political Science Review* 92(2): 343–58.

Gilens, Martin. 1995. "Racial Attitudes and Opposition to Welfare." *The Journal of Politics* 57(4): 994–1014.

——. 1999. *Why Americans Hate Welfare*. Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226293660.001.0001.

Giles, Micheal W., Bethany Blackstone, and Richard L. Vining Jr. 2008. "The Supreme Court in American Democracy: Unraveling the Linkages between Public Opinion and Judicial Decision Making." *The Journal of Politics* 70(2): 293–306.

Green, Donald P., Shang E. Ha, and John G. Bullock. 2010. "Enough Already about 'Black Box' Experiments: Studying Mediation Is More Difficult than Most Scholars Suppose." *The ANNALS of the American Academy of Political and Social Science* 628(1): 200–208.

Hall, Matthew EK. 2014. "The Semiconstrained Court: Public Opinion, the Separation of Powers, and the US Supreme Court's Fear of Nonimplementation." *American Journal of Political Science* 58(2): 352–66.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic literature* 42(4): 1009–55.

Iyengar, Shanto, and Donald R. Kinder. 1987. *News That Matters*. Chicago, IL: University of Chicago Press.

Kinder, Donald R., and Lynn Sanders. 1996. *Divided by Color: Racial Politics and Democratic Ideals*. Chicago, IL: University of Chicago Press.

Kinder, Donald R., and David O. Sears. 1981. "Prejudice and Politics: Symbolic Racism Versus Racial Threats to the Good Life." *Journal of Personality and Social Psychology* 40(3): 414–31.

Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments

Measuring Social Preferences Reveal about the Real World?" *Journal of Economic perspectives* 21(2): 153–74.

——. 2009. "Field Experiments in Economics: The Past, the Present, and the Future." *European Economic Review* 53(1): 1–18.

Luttig, Matthew D., and Matthew Motta. 2017. "President Obama on the Ballot: Referendum Voting and Racial Spillover in the 2014 Midterm Elections." *Electoral Studies* 50: 80–90.

McDermott, Rose. 2002. "Experimental Methods in Political Science." *Annual Review of Political Science* 5(1): 31–61.

McKenzie, Mark Jonathan. 2012. "The Influence of Partisanship, Ideology, and the Law on Redistricting Decisions in the Federal Courts." *Political Research Quarterly* 65(4): 799–813.

Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton, NJ: Princeton University Press.

Myrdal, Gunnar. 1944. *An American Dilemma; the Negro Problem and Modern Democracy.(2 Vols.).* New York, NY: Harper & Row.

Ramirez, Mark D., and David A. M. Peterson. 2020. *Ignored Racism: White Animus Toward Latinos*. Cambridge: Cambridge University Press. https://www.cambridge.org/core/books/ignored-racism/A6AC6BAD47A1BC9BAC8691523CF21071 (October 4, 2022).

Rouhani, Saba et al. 2022. "Racial Resentment and Support for Decriminalization of Drug Possession in the United States." *Preventive Medicine* 163: 107189.

Sears, David O., and Patrick J. Henry. 2003. "The Origins of Symbolic Racism." *Journal of personality and social psychology* 85(2): 259.

Sheagley, Geoffrey, Philip G. Chen, and Christina E. Farhart. 2017. "Racial Resentment, Hurricane Sandy, and the Spillover of Racial Attitudes into Evaluations of Government Organizations." *Analyses of Social Issues and Public Policy*.

Tesler, Michael. 2013. "The Return of Old-Fashioned Racism to White Americans' Partisan Preferences in the Early Obama Era." *The Journal of Politics* 75(1): 110–23.

——. 2016. *Post-Racial or Most-Racial?: Race and Politics in the Obama Era*. Chicago, IL: University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/P/bo22961444.html (October 4, 2022).

Tesler, Michael, and David O. Sears. 2010. *Obama's Race: The 2008 Election and the Dream of a Post-Racial America*. Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226793849.001.0001.

Virtanen, Simo V., and Leonie Huddy. 1998. "Old-Fashioned Racism and New

Forms of Racial Prejudice." *The Journal of Politics* 60(2): 311–32.

Wallsten, Kevin, Tatishe M. Nteta, Lauren A. McCarthy, and Melinda R. Tarsi. 2017. "Prejudice or Principled Conservatism? Racial Resentment and White Opinion toward Paying College Athletes." *Political Research Quarterly* 70(1): 209–22.

Maybe We Can—But Should We?

Phil Shields[1]

I am honored to be a part of this panel today in honor of Professor John List, our distinguished Miller Upton Forum guest. As a philosopher and an ethicist, I will try to focus on some ethical issues raised by the use of field experiments in social science, but in the process, I will also raise concerns about the relation between science in general and the development and use of technology.

I take it that the virtue of science is not that it always gets the facts right, but that it is the best method we have for revising our factual beliefs. What matters is the method, not just "being right." What I will argue is that the same is true about our ends. It matters less that moral agents have the right or best ends than that they have a sustainable process of correcting them, and this requires engaging in normative practices of moral deliberation where the morality of our means and our ends is tested.

In light of the need for a sustainable process of moral agency, I will suggest that doing science on what people think, say, and do is neither value-neutral nor harmless. It is not value-neutral because scientific practices depend on norms of their own. It is not harmless because an over-emphasis on the causal understanding of human behavior tends to diminish human agency by displacing our legitimate practices of moral reasoning. It displaces moral reasoning because causal explanation is inherently reductive—it assumes everything real in the cosmos is either necessary or random—and this assumption leaves no way to make sense of human agency, including the agency expressed in the activity of science itself. As Aristotle noted, the realm of meaningful human action pertains neither to what happens "by necessity" nor "by chance," but to the realm of "what is for the most part" (*hōs epi to polu*), and to "things that are brought about by our own efforts."[1] In contemporary parlance, human action pertains not to the realm of "the natural," but to the realm of "the normative."[2]

The claim that doing science on what people think, say, and do "is not

1 Phil Shields is a Professor of Philosophy and Religious Studies at Beloit College.

harmless," is consistent with the fact that this intrinsic harm—the harm of using the objectifying gaze that reduces human actions to behavior—is sometimes outweighed, and justified, by the good ends achieved. In this sense, we can compare the intrinsic harm of doing science on what we think, say, and do to the intrinsic harm of *paternalism*. Despite this intrinsic harm paternalism is often temporarily and locally justified, as there is no way to raise a child, educate a student, or treat a patient without various degrees of paternalism, but this does not mean that treating people, even children, paternalistically, does not have an intrinsic cost to their human dignity and well-being. This is why we strive to develop relationships with our children, students, and patients that are *less* paternalistic and more respectful of their agency. As we use causal means and extrinsic incentives to condition our children's behavior, we also use language to talk to them and we try to involve them in the realm of meaning where they can come to understand and take responsibility for what they think, say, and do. However, the concern would be that if we are not always mindful of the *intrinsic* harm of paternalism, we will not be able to assess when and where the tradeoffs are worth it. The same is true of using science to interpret what people think, say and do. My thesis is that reducing human actions to mere behavior—that is, by-passing a person's agency to adopt an objectifying spectator stance on what they think, say, and do—always comes at a human cost. Furthermore, while this is compatible with allowing that sometimes this cost is worth paying in particular contexts, we cannot assess *when* we should pay this cost if we do not recognize *that* there always is one, and *what* this cost is.

In addition to the intrinsic harm of objectifying human actions, the reductive assumption that everything is deterministic tends to be *self-fulfilling*. The focus on causal necessity leads us to create mechanistic technologies in accordance with our reductive picture of the world, and we in turn are shaped by the technologies we use. The industrial revolution was encouraged by the emerging metaphor that the cosmos is a gigantic mechanical clock—regular, relentless, and predictable—and workers soon had to become regular, relentless, and predictable to fit the machines and industries we created. The more dependent we are on mechanistic technologies, the more what we think, say, and do becomes determined by them.

Finally, doing science on human motivations and behavior raises the risk that the resulting knowledge will be "weaponized" to exploit people with or without their awareness. This risk has increased with internet surveillance and the growth of big data, but it is not new and not unique to doing field experiments.

Socrates expressed the harm of using a knowledge of hidden incentives long ago, when he raised moral and rational concerns about the emerging craft of sophistry. The sophists developed rhetorical techniques that allowed them to manipulate large crowds, and to get these crowds to unwittingly think and act in the interests of the sophists and their clients. Anticipating arguments by the NRA and Mark Zuckerberg, Gorgias claimed it was a *value-neutral tool* and could be put to good ends. He boasted that with the power of rhetoric, and without any knowledge of medicine, he could convince someone to submit to medical treatment when a doctor could not.[3] Socrates questioned whether such a power is good for either the patient or the doctor. He pointed out that employing an art that can influence people by non-rational means, like extrinsic incentives, is not only dangerous to the patient, who may just as easily be influenced for ill as for good, but that employing this art compromises the integrity of the doctor's practice. In other words, when the doctor relies on rhetoric to bypass the process of trying to persuade the patient with good evidence and relevant reasons, the process of justifying their diagnosis and proposed treatment, the doctor also thereby bypasses the process that exposes and tests the validity of the reasons being used.

Humans have rarely been inclined to doubt their own ends and values. Rather they have tended to focus on the effectiveness of their means, taking their ends for granted. But perhaps our growing modern crises stem from the fact that our power has long surpassed our wisdom to grasp what our ends ought to be. We made enough atomic bombs to wipe out human life. We can now cut and splice DNA to give parents a choice of eye color, sex, or racial characteristics. We could soon alter the human genome to produce stronger soldiers, or more pliant workers. We built an internet that gives us instant access to sources, goods, and services. The realm of possibilities in the contemporary world has swollen to monstrous proportions, while the realm of moral thinking to guide these choices has constricted into little more than subjective conviction, where each of us assumes we have adequate ends and values, and just wants the power to actualize them. I suspect this is backwards. If what we really need is less instrumental power and more adequate ends, then economic field experiments are taking us in the wrong direction. They are focused, like Greek sophistry, on achieving ends at the cost of thwarting processes of reasoning about what our ends *should* be.

The modern growth of instrumental control and personal choice has mushroomed with the globalized digital market. The internet promises us, in the words of satirist Bo Burnam, "anything and everything all of the time."[4] This

seductive promise of unlimited choices comes at the price of undermining our social, educational, and political institutions, not to mention, of wreaking havoc on our environment. Our shrinking and commercialized world is becoming superficial and frenetic as our lives become unmoored, strung out by multitasking and fragmented sound bites. As social media venders compete to hold our attention, and to generate further clicks and ad revenue, the search algorithms amplify our biases and fears to keep us provoked and engaged. Anger engages us more effectively, it seems, than truth. Given the sea of data produced worldwide, and the power of internet searches, our worst fears can be experienced and our most extreme, bigoted beliefs confirmed, somewhere in the world at each and every moment, thereby highly distorting any statistical significance the views may have. Mass shootings spike, fed by media memes that keep them a live possibility on everyone's mind. Polarization is also driven by growing economic discrepancies created by global markets and widespread traumas created by the relentless 24/7 news cycle. Behind all of this, however, are the inherent costs of life on the internet. *Ultimately, polarization and nihilism follow from the intrinsic ways disembodied digital media detract from and undermine the quality of local relationships, embodied practices, and traditional social/political institutions that once provided collaborative, and accountable, contexts of meaning and value.* Insofar as we are immersed in everything all of the time, we lose the capacity to really be *fully present* anywhere in the kind of ways that can create and sustain meaning or value. When we are in a classroom, on a date, or having dinner with family or a friend, perhaps we should turn off and put away our phones. Perhaps we should not live our lives merely as consumers, tourists, and spectators, but as engaged and responsible participants.

We treat one another as participants when we recognize that what other people do, say, and believe cannot be adequately understood without reference to the conscious and deliberate intentions they have.[5] This understanding requires that we approach one another from within shared normative practices that enable us to hold one another accountable. As participants, we do not view one another, or ourselves, as passive effects generated by preceding causal conditions. When Sue asks Maria if she will meet for coffee, Sue is addressing Maria as an agent and expects her to express agency in return. She is asking Maria whether she will form the *intention* to meet for coffee and to thereby *commit* to bringing the future event to pass by *adopting* a course of action toward this end. This is not a matter of observing something from the outside that allows either one to causally predict

a course of events, but of being an agent expressing, and being accountable for, her meaningful intentions.

So, there are important differences between adopting a participant stance that talks *with* people and adopting a scientific spectator stance that talks *about* people. First, there is already something a little demeaning in being talked about instead of being talked to. Even newborn infants elicit our attention, and parents intuitively spend a lot of time talking *to* them. Secondly, by adopting a scientific spectator stance, we employ objective models and apply statistical techniques to predict people's behavior in aggregate ways. This practice, in turn, enables us to manipulate the behavior of groups of people to our own ends. But such manipulation comes at a high cost. Even if our ends are well-meant, they could be wrong, or at least improved. By failing to talk to others, and to engage them in rational persuasion, our ends remain untested, and this undermines both their agency and our own.[6]

Immanuel Kant expressed the moral limits of manipulating people when he famously argued that we should "always treat persons as ends-in-themselves and never merely as means."[7] He recognized that as finite, interdependent empirical beings, we must depend on and use one another in our day-to-day lives. We depend on farmers, truckers, and grocers for food, nurses and doctors for our health, and so forth. But Kant insisted that respect for human dignity requires that we not *reduce* persons to *things*—to mere means that may be used for our own subjective ends. He suggested further that treating people as ends in themselves, as having intrinsic value and not merely instrumental value, should be based on our recognition of their agency, their capacity to contribute to their self-formation. Such agency requires that humans can act intentionally for the sake of ends or self-conscious reasons they give themselves and present to each other to account for what they do. This recognition is in contrast to treating persons as things that are merely determined by empirical causal forces, like genetic predispositions and environmental conditioning. When Sue asks Maria regarding her intentions, Maria is granting Sue's role as an agent who is weighing her ends, whereas asking a scientific observer to predict what groups of people will do based on some kind of causal model reduces these people to things whose behavior is being explained with reference to blind extrinsic forces, a complex interplay of nature and nurture.

We can see the human cost of objectification by how it undermines the practice of science itself. By adopting the scientific stance, and limiting oneself to causal explanation, we can no longer make sense of the notion of human agency.

Many hard-headed types seem to suppose that we simply need to have the courage of our scientific convictions and accept the brute fact that the world is a cold deterministic place and human agency is an illusion. The problem with such a reductive view is that it cannot explain what scientists themselves are doing. By dismissing the reality of human intentionality, positivism banishes the meaning of everything we do, say, or think, so we are left with a performative contradiction where the deterministic theories of physicists cannot account for themselves. Even a spectator stance requires being a participant in a normative community of description that sustains the spectator and the meaning of what she sees. Ironically, when a physicist argues on behalf of a deterministic cosmos, they are being agents. They are also respecting other physicists as agents who can understand, and assess, what they say in light of the meaning and prevailing communal norms of physics. Likewise, when an anthropologist or sociologist adopts a spectator stance to describe social practices, where they are talking about and objectifying other people as predictable things following law-like statistical patterns of behavior, they are able to do so as participants in the normative practice of anthropology or sociology. As social scientists, they do not merely objectify and talk about their colleagues, but they talk to them and try to persuade them with accepted methods, reasons, and evidence regarding what they think is important and true. Their relationships to each other are fundamentally different from, and more respectful of human dignity, than how they approach the people who are the objects of their social science analysis. Returning to the natural sciences, in expressing their agency physicists picture a world in which the agency of everyone—artists, mothers, and the physicists themselves—disappears. But how can physics remain a viable normative practice that depends on a community of agents if agency is an illusion? This contradiction shows the incoherence of thinking the scientific stance can justify dismissing agency from the world. In fact, the validity of science itself depends on human agency, and human agency depends on being participants in normative communal practices.

But our modern picture of causal determinism still threatens to be *self-fulfilling*. We built machines and industrialized the world, and, as Max Weber noted, it became an iron cage.[8] We built a reductive digital virtual reality, but as we use it, it shapes us in its own image. We thought the internet would evade the censorship of traditional institutions, and would give us direct and transparent access to unfiltered information, but instead we get information filtered by hidden search algorithms and a plethora of shadowy causes, dark money, and

powerful self-serving special interests. Instead of obtaining more comprehensive knowledge, we are inundated with an endless stream of conspiracy theories, propaganda, and "fake news." Digitized global markets promise us anything and everything all of the time, but we are losing touch with the normative practices that give anything its value. With the growth of big data, we are losing touch with what the products really are. Consumers comb the internet, as if they are participating freely in an open market, while their every move is tracked, recorded, and sold to corporate or governmental interests. The collective digitized data of our objectified behaviors is being sliced and diced to produce an endless stream of unseen products for unknown actors in invisible markets. Now that we are the products being bought and sold, it is hard to have confidence that we can really participate in these markets responsibly, that is, as agents who are capable of holding other participants accountable.

Considering the ethics of using the knowledge gained by field experiments against the background of our cascading social, political, and environmental crises, is not to blame Professor List or Uri Gneezy for the state of the world, but to highlight that it has been our obsession with instrumental reasoning, the kind of reasoning that adopts a spectator view and treats everything in causal terms, that got us into this mess, and perhaps we should think carefully before supposing that more science and social engineering can get us out of it.

So, what does it mean to reveal the "hidden motives and the undiscovered economics of everyday life"? This is the subtitle of a provocative book Professor List has written with Uri Gneezy called *The Why Axis*.[9] They seek answers to vexing questions like why women are paid less for the same work, why people are charged differently for products and services, and why people discriminate against one another.[10] But when we look for "why" people do, say, or believe things we can be doing two fundamentally different things: we can be looking for efficient *causes* or we can be looking for *reasons*. When we say A is the cause of B, then we usually mean A is sufficient to determine B, and B is necessary given A. So, when we look for causes, we are viewing things mechanistically, from a reductive, spectator view. When we look for reasons, we are being intentional agents. We are looking to meaningful ends or values that could guide and justify what we do, say, or believe. This stance requires viewing things from inside normative practices.

For example, suppose you ask a student why she missed class. She might adopt a spectator stance to *explain* her behavior by citing a *cause*, like "my medication made me oversleep." Or she might cite a *reason* to *justify* her action, like

"my grandma had a stroke last night and I decided I should go see her." The causal answer treats one's behavior as a passive outcome of causal events and presumes to absolve oneself of any complicity or responsibility by treating these causes as wholly external, while citing the *reason* for one's action embraces one's agency, and complicity, in a meaningful and value-laden world.

Which sense of "why" do List and Gneezy seek in *The "Why" Axis* when conducting field experiments? This will turn out to be a complicated issue, despite the fact that they explicitly characterize the kind of "why" being sought as the *causal* sort, and not the kind of "why" which involves reasons or prospective ends that could serve to justify what people do, say, or believe. In the "Foreword" Steven Levitt explains:

> The big challenge then (and now) was how to figure out whether a relationship between two variables was truly causal, or whether it was merely correlation. Why did it matter? If a relationship was *causal*, then there was a role for public policy. If a relationship was *causal*, then you learned something important about how the world worked.[11]

The goal of social science research is to distinguish causal relations, which would enable instrumental interventions, from contingent correlations which would not enable our interventions to have predictable and desired effects. Levitt further notes,

> Causality, however, is very hard to prove. The best way to get at causality is through randomized experiments. That is why, for instance, the Food and Drug Administration requires randomized experiments before approving new drugs. The problem was that the sort of laboratory experiments used to test drugs weren't all that applicable to the kinds of questions economists like me wanted to answer.[12]

The advantage of List's and Gneezy's work is that they discovered you could do randomized experiments outside of a controlled laboratory setting. Levitt explains, "[List] was pioneering something that in retrospect was completely sensible and obvious: running randomized economic experiments in the real world."[13] The whole point is that this allows us to tease out causal relations between two variables for questions where it is impractical or impossible to create artificial laboratory conditions.

List and Gneezy confirm in their Introduction, "Getting Beyond Assumptions: What Makes People Do What They Do?" that their focus is on discovering causal connections. They write:

> In the past, economists have been skeptical about running *controlled field experiments*. For an experiment to be valid, everything else but the item under investigation has to be held constant. This is how researchers test their theories: if they want to determine whether Diet Coke causes cancer in rats, they will hold "other things equal" and only vary the amount of Diet Coke consumed. For years, economists believed that there was no possible way to perform such tests in the "real world" because they could not easily control other important factors.
>
> But in reality, the economic world is not a chemistry test tube—there are billions of people and thousands of firms. At odds with received economic wisdom, we will show that if you have "dirtiness"—that is, if you are looking at the way things work in an uncontrolled, quirky, real world—then randomized field experiments yield real answers. . . . Our methodology permits us not only to measure something that is happening but also to ascertain *why* it happens.[14]

We do not doubt that these field experiments can yield *real* answers. Our question regards what *kind* of answers they yield. If it were true that field experiments are merely identifying cause and effect mechanisms, then when we employ this knowledge we would be left using a kind of instrumental reasoning that is not in itself sensitive to normative values, regarding either our means or our ends. After all, a cause is a value-neutral scientific fact. Nor does cause and effect reasoning promote reconsidering the value of our ends, since effects, like causes, are also merely value-free scientific facts. So, our first question is whether the kind of "why" being uncovered in economic field experiments is really a *causal* one, more precisely, whether incentives in particular are causal in nature, or already *normative*. Our second concern, given the normativity of incentives, is the extent to which we ought to employ *extrinsic* incentives to manipulate human behavior, as opposed to cultivating *intrinsic* incentives.

Let us examine a particular case more closely. In the context of studying, and trying to improve, our educational institutions, List and Gneezy conduct a series of controlled field experiments to test a range of financial incentives to students, parents, and teachers, to find out which work best to improve student grades and performance on standardized tests. They explain their approach: "[E]conomists begin by thinking about how different 'inputs,' or influencing factors, combine to make certain 'outputs,' or results."[15] The mechanical language of "inputs" and

"outputs" talks about students instrumentally, as if students were merely complicated billiard balls and education were some kind of industrial production. In this way of speaking, "inputs" are depicted as causes and "outputs," as their determined effects. But there is also softer language being used that describes "inputs" as "influencing factors" and this suggests a less deterministic and more "dispositional" process where none of these factors alone are sufficient to necessitate a particular effect and hence do not qualify as causes traditionally conceived.[16] The passage continues, "For example, what inputs are necessary to achieve the desired output of good grades?" Here the mechanistic language is enriched with value judgments, like "desired" outputs and "good" grades. Once values are invoked, we have moved beyond causal analysis to use the *normative* language of intentional action that aims at ends or goods. Notice that if all that mattered was getting the effect of more "good grades" in a value-free descriptive sense, namely some arbitrary GPA, then we could easily achieve our end by grade inflation or "dumbing down" the curriculum. Clearly, "good grades" are not the value-free facts determined by science, but are meant in a *normative* sense, as a genuine end that reflects things we really value like literacy and knowledge. So, it would defeat our normative purpose to raise grades by lowering our standards.

Although *The "Why" Axis* is framed as a scientific project of discovering causes of people's behavior, it is not limiting itself to causes strictly understood. In the book, the talk of finding causal explanations is quickly replaced with the language of "motivation," "self-interest," "hidden incentives," and ultimately, "what people really value." Back in the introduction List and Gneezy write:

> By observing the way people behave in everyday markets, we can better understand their motives. One of our key discoveries is that self-interest lies at the root of human motivation—not necessarily selfishness, but self-interest. . . . This is a key insight, because once we establish what people really value—money, altruism, relationships, praise, what have you—then we can more accurately figure out the triggers or mechanisms needed to induce them to get better grades at school, stay out of trouble with the law, perform better on the job, give more to charity, discriminate less against others, and so on.[17]

While motives, self-interests, incentives, and values are distinct in various ways, it is not clear that any of them can be adequately understood simply in terms of blind efficient causes like "triggers or mechanisms." "Values" are clearly ends,

what Aristotle described as "that for the sake of which" one acts. "Incentives" too would seem to be prospective outcomes that motivate one to act, so they have to be value-laden rewards and punishments. Otherwise, they could not give us reason to act. Even "motivations" are not reduceable to blind causal mechanisms, as they still involve being moved by the meaning and value of something, so they involve self-movement of some kind. Self-movement, like intrinsic value, is something that is not even possible within a scientific cause and effect analysis, an analysis where the cause is independent and prior to the effect and everything is moved by extrinsic forces.

The incoherence of conflating cause and effect relationships with means and ends relationships can be seen by thinking about the nature of an incentive. On one hand, incentives are supposed to be causes of behavior. On the other hand, they are also functioning as ends, as that for the sake of which we do something. So, an incentive is at once the means and the end, and this is impossible if a cause has to be prior to its effect. Aristotle solves this problem by distinguishing efficient causes from final causes. He notes that there is one sense in which a means is prior to, and brings about, an end, but there is another sense in which the end is prior to, and brings about, the means. A child has to understand Greek grammar before they can speak in the way a Greek speaker can speak, namely, both by means of their grasp of grammar and for the sake of the meaning of what they say. But it is also true, Aristotle argues, that it is by means of speaking Greek that a child can develop the capacity to speak Greek. In other words, there needs to be the end, a living language being spoken with meaning, before a child can develop the capacity to speak it on their own. Modern science was built on the rejection of such final causes, and on limiting itself to efficient causes to explain change, but *this makes the talk of incentives either unscientific or incoherent*. Some of us think our intentions and values serve as final causes and are crucial both to understanding the purposive social world we inhabit and to assessing the meaning and adequacy of human action.

The first problem in *The"Why" Axis* that List and Gneezy used their field experiments to address is the continuing wage gap between men and women. In this case are we merely interested in *causally* explaining *a value-neutral fact*, how we got to where we are, with women earning less than men for equal work? Or are we really interested in a normative assessment of the injustice of the situation, together with a morally acceptable plan of how to make things better, and in the latter case the language of causality provides no normative guidance. Consider the

joke from the Soviet era where Boris complains of the injustice of the fact Ivan has two goats and he has none, and the government fixes it by taking away Ivan's two goats. This intervention causally achieves the goal of equality, but at the price of leaving them both hungry.

Tragically, the twentieth century saw such narrow-minded instrumental pursuit of equality play out on a large scale in the totalitarian socialist experiments in the Soviet Union and Maoist China. The Soviet failure to achieve human well-being, broadly conceived, by their use of coercion to pursue egalitarian ends, is well known. Mao's Great Leap Forward redistributed the land to the peasants by collectivizing them on communes, and his Cultural Revolution later sought to level class privileges by sending educated city dwellers out into the countryside to humble them doing labor in the fields. However valuable these egalitarian ends may be, they came at the cost of destroying both agricultural practices and educational institutions, leading to the mass starvation of over 30 million people in 1960-62 and the loss of higher education for a generation of Chinese students a decade later. The lesson should be that equality is not our only end, and the means matters, so we should not mechanically pursue our ends, no matter how good these ends seem to be when taken in isolation. To determine and reassess our ends, and the appropriate means, we need to keep in mind and weigh *all* the normative issues involved.

When List and Gneezy identify incentives that can motivate changes in people's behavior, these incentives are indeed *the reasons* people have for acting, and not merely efficient causes. In this, extrinsic incentives are like threats and bribes. The threatened or bribed person has to share a conscious understanding of the meaning and value of the threat or bribe, and on this limited basis they can be "persuaded"—they can become complicit and contribute to the course of action. However, we still consider threats and bribes to be manipulative and disrespectful of our agency, as they are attempts to influence us to do something we do not value doing for its own sake. While we are given a reason to act, the reason is extrinsic, and not the intrinsic kind of reason that makes the action in itself justified, or that helps us reason about our ends.

"New" scientific insights have a long history of displacing our practices of moral reflection. The success of the theory of evolution led many to blindly propose using it to improve human society, and Social Darwinism was born. The success of genetic theory led some to blindly apply it to promoting public health, without considering our moral reservations about the means, and eugenics was

born. The danger of being over-impressed with the effectiveness of scientific methods is to think causal truths are more real or important than our moral values, and that science can replace our practices of moral reasoning. Unless we engage in practices of moral persuasion with other participants, a process that exposes and tests our own ends, we will not be very inclined to reexamine them, but will simply run with our assumed ends.

So, we can return to where we began. The virtue of science is not that it always gets the facts right, but that it is the best method we have for revising our factual beliefs. What matters is the method, not just being right. What List does not seem to acknowledge is that the same is true about our ends. It matters less that we have good or right ends then that we have a process of correcting or revising them, and this requires engaging in normative practices of moral reflection about the morality of our means and our ends. List seems to be a good example of how the focus on effectiveness can blind someone to thinking about what the ends *should* be. In *The "Why" Axis* he just assumes that it is easy and that he already knows.

In conclusion, *the moral risk of applying scientific methods to understand what people do, say, and believe, is that a reductive notion of causality fails to account for the special kind of intersubjective interdependence that makes the realm of normative meaning, and human agency, possible—even for the scientist*. The social scientist may develop causal models of complex interactions between nature and nurture, but this will always leave out the agency of scientists themselves. Scientists develop their agency by becoming participants in specialized communities where they are accountable to specific normative practices. In these practices what they do, say, and believe can be guided, and assessed by other participants, with reference to the meaning of norms instead of blind mechanical causes. What science cannot account for, then, can be seen every time we raise and educate a child. While parents employ a vast array of causal behaviors to enculturate the child into our normative practices, they also *talk to* the child, from the beginning, as a participant, and the child soon engages the meaning of these practices on their own terms. The resulting realm of meaning enables children to mean what they say, to do something for its own sake, and even at times to do the right thing for the right reason. In these ways, the child expresses agency, as opposed to merely being determined by causes and being coerced into doing what parents want.

Endnotes

1 Aristotle, *Nicomachean Ethics* 1094b12-27; 1098a25-34; 1112a18-1113a13.
2 See Robert Pippin, "The Natural and the Normative," *Daedelus* 138 (2009): 35-43.
3 Plato, *Gorgias* 459a-461b.
4 Bo Burnam, "Welcome to the Internet" *Inside* (Netflix film released May 30 2021) https://www.youtube.com/watch?v=k1BneeJTDcU.
5 The inspiration for the contrast between being participants and being spectators comes from Robert Pippin, "Participants and Spectators," *On the Human: A Project of the National Humanities Center* https://nationalhumanitiescenter.org/on-the-human/2010/04/participants_and_spectators/ (accessed February, 2020). For a more developed account of the communal conditions for agency, see Pippin *Hegel's Practical Philosophy: Rational Agency as Ethical Life* (Cambridge University Press, 2008).
6 The most compelling and definitive case for reasoning about our ends is by Henry Richardson, *Practical Reasoning about Final Ends* (Cambridge University Press, 1997).
7 Immanuel Kant, *Foundation of the Metaphysics of Morals*, translated by Lewis White Beck (Liberal Arts Press, 1959).
8 Max Weber, The *Protestant Ethic and the Spirit of Capitalism,* translated by Peter Baehr and Gordon C. Wells
(New York: Penguin Books, 2002). Also see: Arthur Mitzman, *The Iron Cage: An Historical Interpretation of Max Weber* (New York: Knopf, 1970).
9 Uri Gneezy and John List, *The "Why" Axis* (New York: Public Affairs, a Member of the Perseus Books Group, 2013), p. 5.
10 Ibid., p. 5.
11 Ibid., p. ix.
12 Ibid., p. x.
13 Ibid., p. x.
14 Ibid., p. 4.
15 Ibid., p. 67.
16 For an excellent account of "dispositional" causality, where causes do not necessitate their effects, see, Rani Lill Anjum and Stephen Mumford, *What Tends to Be: The Philosophy of Dispositional Modality* (New York: Routledge, 2018).
17 Gneezy and List, p. 4.

Does Negative Campaigning Drive Political Polarization?

Edward Verzosa[1]

Introduction

In the summer after my sophomore year, I interned with a non-profit organization that partnered with the Shontel Brown Ohio 11th Congressional campaign. Brown was running against Nina Turner, a progressive democrat and former Ohio state senator. By the midway point of the race, Brown, a moderate liberal, was polling nearly forty points behind Turner.[2] This was due to the Ohio 11th electorate, as the district consists of progressive cities such as Cleveland and Akron. The polls did not shift until Nina Turner released a negative campaign advertisement attacking Shontel Brown. Negative ads are released by political campaigns to insult, smear, or damage the image of an opposing party or candidate. In this case, Turner accused Brown of using taxpayer dollars to fund a private company connected to her boyfriend and her family. This shocked the Ohio 11th electorate. As an intern, I had to contact hundreds of voters per day. Nearly every conversation I had following the release of the ad mentioned the accusations against Shontel Brown. My team of six and I believed that we were simply going to lose the election. But after the votes were counted, Shontel Brown narrowly won by over 4,000 votes.

One can look at this scenario and conclude that negative campaigning does not increase polarization. Logically, the release of the major attack ad by Turner should have solidified her majority base. However, it did the opposite. Turner lost voters after she released the advertisement.

---

1 Edward Vezosa is a 2023 graduate of the Department of Economics & Business at Beloit College.

2 Ally Mutnick, "New Poll Shows Nina Turner's Lead Shrinking in Ohio Special Election," POLITICO, July 12, 2021, https://www.politico.com/news/2021/07/12/new-poll-shows-nina-turners-lead-shrinking-in-ohio-special-election-499269.

The evidence that negative campaigning does not affect polarization becomes more ambiguous when considering the 2016 presidential election. The 2016 election was met with increased polarization and division between Republicans and Democrats.[3] It was also met with an increased use rate of attack ads.[4] For example, the Trump campaign released the well-known "Deplorables" attack ad in the heat of the presidential race. This ad attacked Clinton's statement that half of Trump supporters are deplorables. In short, Trump claimed that Clinton believes that the average, everyday American citizen is deplorable. This ad was aired in several key battleground states. Trump won all of them. And the outcome of the 2016 election: increased political polarization among the American electorate.

When viewing the results of the 2016 presidential race, it seems very clear that negative campaigning does increase polarization. But when considering other races, such as the Brown vs Turner race, the causal relationship between the two becomes more inconclusive. This paper looks to further the research between negative campaigning and polarization and provide additional literature to combat the increased political polarization that America sees today.[5]

Literature Review

The literature regarding negative campaigning and polarization remains fairly scarce. Most researchers measure how negative campaigning affects people's actions on voting or how they view a particular candidate, rather than how they feel about people from the other party. The results are also mixed on how different ideologies react to negative campaigning. This leaves a gap as to whether negative campaigning has a direct effect on polarization.

Though researchers do not directly measure polarization, they do tend to measure favorability of a candidate or member of the other party. Survey experiments are the most common methodology used to test this. A common problem

3 Michael C Schwalbe, Geoffrey L Cohen, and Lee D Ross, "The Objectivity Illusion and Voter Polarization in the 2016 ... - PNAS" (Proceedings of the National Academy of Sciences of the United States of America, August 17, 2020), https://www.pnas.org/doi/10.1073/pnas.1912301117.

4 Prashanth Bhat et al., "A Report on Presidential Advertising and the 2016 General Election: A Referendum on Character "(University of Maryland, November 2016), https://parcumd.files.wordpress.com/2016/11/parc-report-2016-v-21.pdf.

5 Abigail Geiger, "Political Polarization in the American Public," Pew Research Center - U.S. Politics & Policy (Pew Research Center, April 9, 2021), https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/.

with surveys is that people could lie and there is no way to tell whether someone is lying or not. A unique way of combating this is by using fMRI scans to measure attitudes in the brain (Haas et al. 2017). By doing this, researchers can determine how subjects emotionally react to information. This methodology is more closely resembled to a lab experiment. It also provides extremely accurate data when measuring polarization. This is an extremely unique and costly approach, however, and is not common within the literature. The rest of the literature tends to follow the same design: subjects take an initial survey, a treatment in the form of negative ads is put onto the treatment group, then the subjects retake the survey and the data undergo a hypothesis test.

There are slight differences in experimental designs, specifically the type of treatments being used. For example, Mann et al (2020) use phone calls as forms of negative messaging. This approach is mainly used for measuring voter behavior and polarization near election time, as also shown by Niven (2006). King et al. (2017) employs a different approach. They use negative advertising from print media, video reports, and audio podcasts. This helps measure the effect of all forms of negative advertising that is prevalent in our media today. Another common method found in the literature is to use social media posts as a treatment (Bail et al. 2018). This is a useful way to measure polarization and negative advertising, as social media is becoming a major source of news for the U.S population.

Though polarization is not the measured variable in much of the literature, the researchers acknowledge that their findings may have association with polarization. Nai and Maier (2020) measure how negative campaigning affects voters' views on political candidates. If negative campaigning leads to more disapproving views on opposite party candidates, then this can lead to increased polarization and partisanship. Alternatively, Coppock, Hill, and Vavreck (2020) observe how negative campaigning affects voter behavior. The influencing of whether someone will vote and how they vote may have a direct effect on partisanship. Though these studies do not directly measure polarization, their findings can be used to help determine whether polarization is affected by negative campaigning.

The results of the literature regarding how voters view candidates following negative campaigning is very consistent. Bail et al (2018) and Nai and Maier (2020) both found evidence of a "Backfire Effect." This is where voters lose favorability of a certain candidate following negative campaigning or exposure to their political content. The latter study found that voters also become less favorable of the intended target of the attack. This can be seen as proof that negative

campaigning may not necessarily hold a significant effect on polarization. Nai, Verhulsdonk, and Karp (2021) took this finding a step further. They found that adding humor to negative ads reduces the backlash effects against the attacker. Further research can determine whether this, in turn, increases polarization and partisanship, as one candidate's favorability drops while the other stays constant. When measuring whether negative campaigning drives voters to the ballot box, Niven (2006) found that voters who receive negative ads are more likely to vote. Results in changing voters' preferences, however, remain unclear. Gerber et al. (2011) found that televised campaign ads can cause a great shift in voter preference, while Coppock, Hill, and Vavreck (2020) found that advertising of any kind has little to no effect on voter preference.

The results regarding how different parties and ideologies react to negative campaigning is mixed. Lau et al. (2016) found that negative campaigning drove polarization in all parties, while Bail et al. (2018) found that Republicans are more likely to become more polarized. Nai, Verhulsdonk, and Karp (2021) found that humorous negative ads are more effective on Democrats than Republicans. Further research can be used to determine whether this aligns closer to the Lau et al. study or the Bail et al. findings.

As previously stated, the literature mainly measures the effects negative campaigning has on voter behavior and voter favorability among political candidates. Nearly all pieces of relevant literature, however, do acknowledge that political polarization is a growing problem in America. Yet, there is not much research done on how negative campaigning directly affects polarization. This is a gap in the literature that this paper investigates and hopes to fill. The paper hypothesizes that negative campaigning will increase political polarization and a survey experiment is carried to test this hypothesis.

Methodology and Experimental Design

Polarization can be thought of as partisanship or affective polarization. Partisanship is the firm allegiance to a certain political party or political actor. Affective polarization, on the other hand, is the tendency to dislike or distrust those from the opposite party or ideology. For this experiment, affective polarization is the outcome variable to capture how much people resent each other for their political views, rather than how strongly someone identifies with a certain party.

Our experiment was a framed-field experiment using online list surveys (figure

1). A common problem with traditional survey experiments is the lack of honesty among the participants. Because these surveys ask political and controversial questions, subjects may lie to not appear biased or hateful. List surveys, however, bypass this issue. A list survey uses questions that contain a list of statements, with one being the conditional statement or the statement of interest. The respondent only states how many statements make them feel "angry" or "upset," rather than which statements. This allows the subject to answer honestly, since they do not know which statement is the conditional statement.

Figure 1- List survey for Republicans

How many statements make you feel angry or upset?

Baseline List:

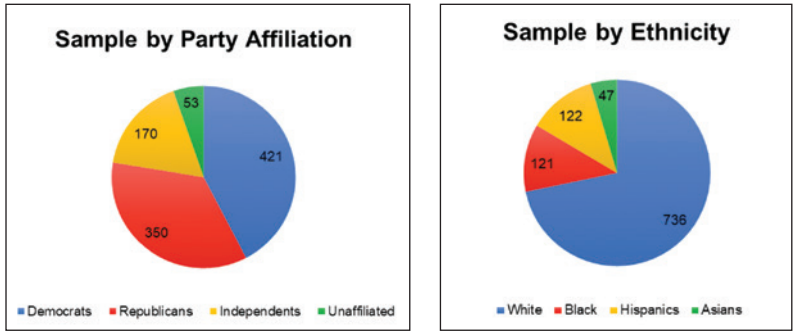
- The federal government increasing the tax on gasoline
- Politicians using insider trading
- Requiring seat belts be used when driving
- Large corporations polluting the environment

Conditional List:

- The federal government increasing the tax on gasoline
- Politicians using insider trading
- Requiring seat belts be used when driving
- Large corporations polluting the environment
- Democrats winning control of the Senate

The sample size was calculated using the continuous power analysis equation (figure 2). As the literature is quite thin, we assumed that the minimum detectable effect is one-fifth of a standard deviation. We chose this value to adjust our sample size to a feasible but appropriate number to ensure accurate results. By plugging this value into the equation and using a twenty percent chance of a type two error and a 95% confidence level, the minimum sample size needed was 620 participants. This experiment, however, contained a total of 994 participants. These subjects were collected by Lucid, a popular online survey software company for political science research. The sample is also a representative sample of the U.S. population.

Figure 3- Sample Breakdown



This experiment contained two treatments- one being a hypothetical mail-in negative advertisement and the other being the conditional statement. We randomly separated subjects into a treatment and control group. The treatment group was shown the negative advertisement. Republicans were shown an advertisement smearing Democrats and Democrats were shown an advertisement smearing Republicans. Independents and the unaffiliated were randomly shown one of the two advertisements. Subjects, then, were further separated into two baseline and condition groups. The condition groups were given the list question containing the conditional statement, while the baseline groups were not. Each subject also filled out a demographics survey. This allowed us to obtain other needed variables, such as race, gender, age, party affiliation, etc. From there, we took the average amount of agreed statements in each baseline and condition group and then computed the difference in means in the control and treatment groups. If our treatment had no effect, then the difference in means between both groups should be the same. If there was an effect, however, then the t-test would show that the result is statistically significant.

Figure 4- Negative Advertisement for Democrats

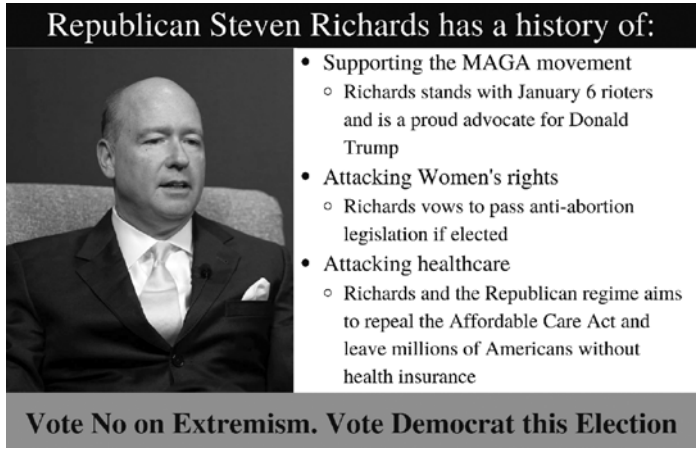


Figure 5- Experimental Design

Treatment (sees negative ad)		Control	
Condition list (survey has sensitive statement	Baseline list (survey does not have sensitive statement)	Condition list (survey has sensitive statement	Baseline list (survey does not have sensitive statement)
Difference in Means		Difference in Means	
Difference-in-Differences test			

Noises that we had to be cautious of were environmental factors (such as where the participant is taking the survey), non-compliance, and current political events. We used a single hypothetical mail-in ad to combat environmental noise. As this was an online survey, subjects would be on the internet. Many phone users would not be in an environment where they would be able to listen to a television advertisement. A written ad, however, allows them to finish the survey in nearly every appropriate environment. Hypothetical mail-in advertisements also do not contain real candidates. This was to avoid any preconceptions participants may have of current political leaders.

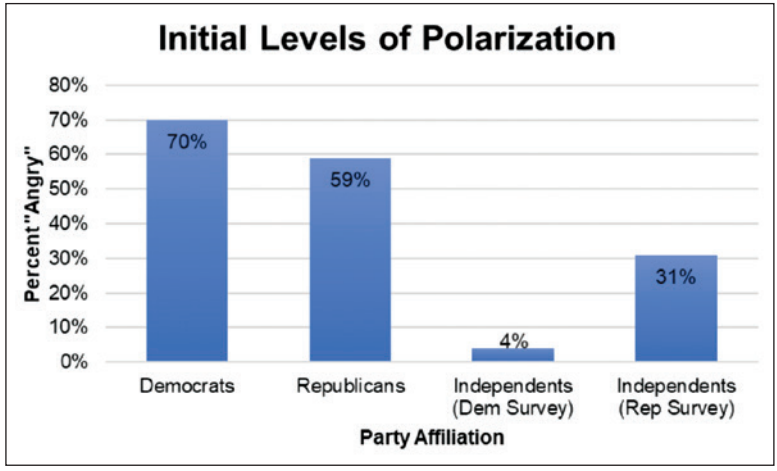
The survey also included an attention check. Attention checks are questions

that ensure compliance because they require participants to read the entirety of the question in order to select the correct answer. If a participant answers incorrectly, then that shows that they did not pay full attention to the question and to the survey. Participants who did not select the right answer were removed from the sample and replaced. The process for each participant was also kept short. Each participant had a maximum time of fifteen minutes, and nearly all participants finished the survey in under five minutes. By keeping the process short, participants stayed fully engaged in the survey. Data collection took place between March 13 and March 20, 2023. The quick collection time and the short length of the survey also minimized any current political event bias, as no major political events (such as elections) happened during or around this time.

Results

Prior to analyzing the effectiveness of the treatment, we first measured the initial levels of polarization among the sample. Consistent with prominent survey data and relevant literature, we found that both Democrats and Republicans are extremely polarized. Independents, however, did not experience nearly the same levels of polarization.

Figure 6



70% of Democrats and 59% of Republicans felt angry when reading our conditional statement. As expected, independents did not show nearly as extreme

levels of polarization when compared to Democrats and Republicans. However, the sample size for independents was significantly smaller.

Figure 7

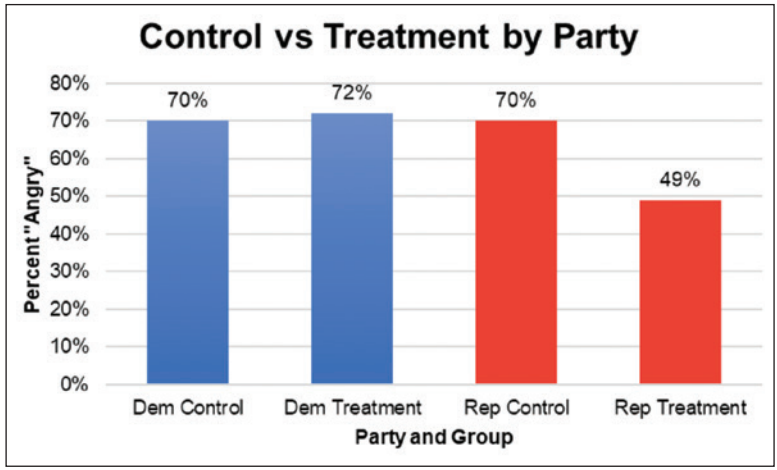


Figure 7 compares both the Democrat and Republican control and treatment groups. Democrat rates of polarization stayed nearly the same following the treatment, while Republicans experienced an insignificant decrease.

Figure 8

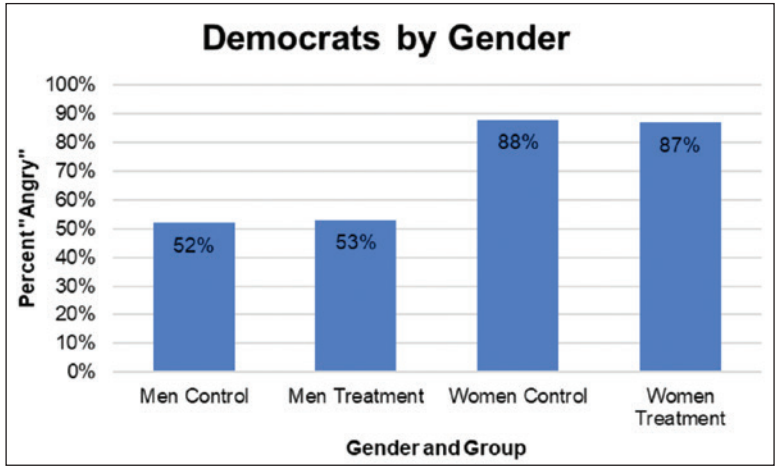
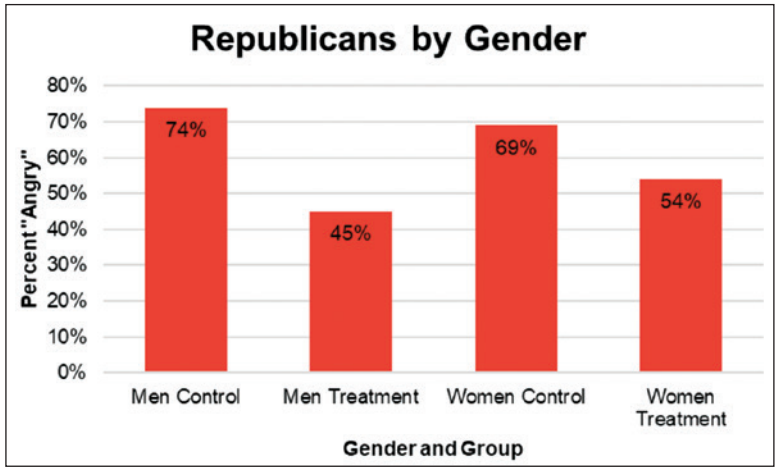
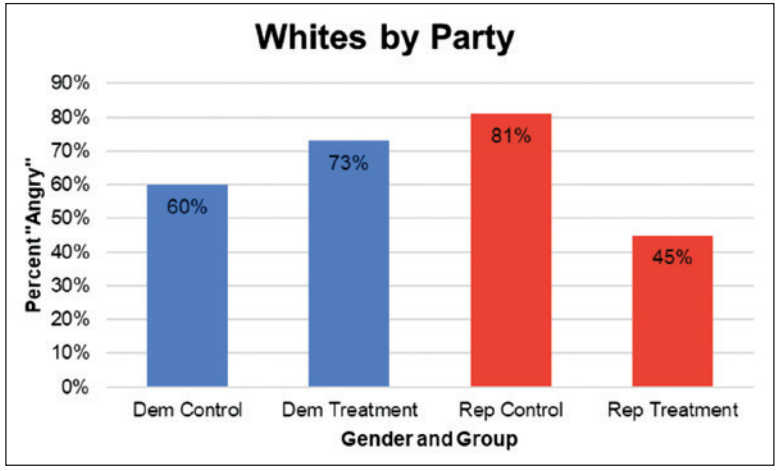


Figure 9



When breaking the parties down by gender, both Democrat men and women experience nearly no change. We did find that Democrat women were typically more "polarized," as nearly 30% more women felt angry after reading the conditional statement. Both Republican men and women started with high levels of polarization, but they both experienced minor, insignificant drops following the treatment.

Figure 10



We also analyzed the effect negative campaigning has on white Democrats and Republicans. We, however, did not have a large enough sample size to meaningfully test the effects on other ethnicities. White Democrats experienced a small increase in rates of polarization and Republicans experienced a modest decrease. Both findings were statistically insignificant.

Conclusion

Before discussing the importance of the findings in this experiment, several limitations of the study need to be addressed. The study used a single mail-in advertisement as the treatment. This was mainly to counteract environmental noise. Studies such as Lau et al. (2016), however, used multiple television advertisements in their experiment. Using additional forms of advertisements may have led to different results. The advertisement also may not have been extreme enough. By not varying the intensity and amount of advertisements, we were not able to see how subjects react to varying forms of negative advertisements. Additionally, we only measure the effects that same-party advertisements have on their voters. This leaves a gap as to how voters respond to opposing candidates' advertisements.

Despite these limitations, we develop interesting findings that reveal Americans' rates of polarization and their responses to negative campaigning. First, levels of polarization may be more extreme than previous survey data predicted. In 2014, the Pew Research Center, for example, found that 43% of Republicans and 38% of Democrats found the opposing party as "very unfavorable."[6] However, this study's result shows that 70% of Democrats and 59% of Republicans were "angry" after reading the conditional statement. Our conditional statement, itself, was not extreme. Power in the Senate regularly changes. Thus, having significantly large majorities of each party elicit anger when thinking about losing power in the Senate, further proves that political polarization in the U.S. is growing at dangerous rates.

The lack of statistically significant effects in this study may also be proof of a frontier effect. We have already found that Americans are highly polarized. It may become increasingly difficult to increase these levels of polarization with attack ads. America also has uniquely long political campaigns through primary elections. Because of this, voters are constantly exposed to negative campaigning

6 Pew Research Center. "Political Polarization in the American Public" (2014).

that they may have become indifferent to these smear advertisements. This research is also proof that political polarization and voter behavior are not directly correlated. One who is extremely polarized, for example, may not be extremely likely to vote and vice versa. This counters researchers' claims that their findings regarding voter behavior has associations with political polarization.

Our findings also hinted to a backlash effect, especially among Republicans. Though not statistically significant, further research should explore the validity of this effect as future findings can strengthen the robustness of the Bail et al (2018) and Nai and Maier (2020) studies. Researchers should also explore other variables that can affect political polarization, such as the 24-hour news cycle, social media, and candidate extremism.

References

Abigail Geiger, "Political Polarization in the American Public," Pew Research Center - U.S. Politics & Policy (Pew Research Center, April 9, 2021), https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/.

Alexander Coppock, Seth J. Hill, and Lynn Vavreck. "The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments." *Science Advances,* no. 6 (2020). https://www.science.org/doi/10.1126/sciadv.abc4046.

Bail, Christopher S., Lisa P. Argyle, Taylor W. Brown, Alexander Volfovsky, John P. Bumpus, Haohan Chen, M.B. Fallen Hunzaker, Jaemin Lee, Marcus Mann, and Friedolin Merhout. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." Proceedings of the National Academy of Sciences of the United States of America, August 28, 2018. https://www.pnas.org/doi/10.1073/pnas.1804840115.

Bhat, Prashanth, Alyson Farzad-Phillips, Morgan Hess, Lauren Hunter, Nora Murphy, Claudia S Rico, Kyle Stephan, and Gareth Williams. "A Report on Presidential Advertising and the 2016 General Election: A Referendum on Character ." University of Maryland, November 2016. https://parcumd.files.wordpress.com/2016/11/parc-report-2016-v-21.pdf.

Christopher B. Mann, Kevin Arcenaux, and David W. Nickerson. "Do Negatively Framed Messages Motivate Political Participation? Evidence From Four Field Experiments." American Politics Research, 48(1), 3-21. https://journals.sagepub.com/doi/pdf/10.1177/1532673X19840732

Gary King, Benjamin Schneer, and Ariel White. "How the news media activate public expression and influence national agendas." *Science* 358, 776-780. https://www.science.org/doi/10.1126/science.aao1100.

Geiger, Abigail. "Political Polarization in the American Public." Pew Research Center - U.S. Politics & Policy. Pew Research Center, April 9, 2021. https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/.

Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw. "How Large and Long-Lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment." *The American Political Science Review* 105, no. 1 (2011): 135–50. http://www.jstor.org/stable/41480831.

Haas, I.J., Baker, M.N. and Gonzalez, F.J. "Who Can Deviate from the Party Line? Political Ideology Moderates Evaluation of Incongruent Policy Posi-

tions in Insula and Anterior Cingulate Cortex." *Soc Just Res* 30, 355–380 (2017). https://doi.org/10.1007/s11211-017-0295-0.

Kuklinski, J. H., Cobb, M. D., and Gilens, M. (1997a). Racial attitudes and the New South. Journal of Politics 59, 2, 323 – 349.

Lau, R.R., Andersen, D.J., Ditonto, T.M. *et al.* Effect of Media Environment Diversity and Advertising Tone on Information Search, Selective Exposure, and Affective Polarization. *Polit Behav* 39, 231–255 (2017). https://doi.org/10.1007/s11109-016-9354-8.

Mutnick, Ally. "New Poll Shows Nina Turner's Lead Shrinking in Ohio Special Election." POLITICO, July 12, 2021. https://www.politico.com/news/2021/07/12/new-poll-shows-nina-turners-lead-shrinking-in-ohio-special-election-499269.

Nai, A., & Maier, J. (2021). "Is Negative Campaigning a Matter of Taste? Political Attacks, Incivility, and the Moderating Role of Individual Differences." American Politics Research, 49(3), 269–281. https://doi.org/10.1177/1532673X20965548.

Niven, David. "A Field Experiment on the Effects of Negative Campaign Mail on Voter Turnout in a Municipal Election." *Political Research Quarterly* 59, no. 2 (2006): 203–10. http://www.jstor.org/stable/4148088.

Ordway, Denise-Marie, and John Wihbey. "Negative Political Ads and Their Effect on Voters: Updated Collection of Research." The Journalist's Resource, September 26, 2016. https://journalistsresource.org/politics-and-government/negative-political-ads-effects-voters-research-roundup/.

Schwalbe, Michael C, Geoffrey L Cohen, and Lee D Ross. "The Objectivity Illusion and Voter Polarization in the 2016 ... - PNAS." Proceedings of the National Academy of Sciences of the United States of America, August 17, 2020. https://www.pnas.org/doi/10.1073/pnas.1912301117.

Verhulsdonk, I., Nai, A., & Karp, J. A. (2022). Are Political Attacks a Laughing Matter? Three Experiments on Political Humor and the Effectiveness of Negative Campaigning. *Political Research Quarterly*, *75*(3), 720–737. https://doi.org/10.1177/10659129211023590.

Appendix

Figure 2- Power Analysis Equation

$$n_0^* = n_1^* = n^* = 2\left(t_{\alpha/2} + t_\beta\right)^2 \left(\frac{\sigma}{\delta}\right)^2$$

Figure 11- Initial Levels of Polarization

Affiliation	Sample Size	Baseline Agreed Statements	Condition Agreed Statements	Percent "angry"
Democrats	420	3.35 (196)	4.05 (224)	70%
Republicans	350	3.40 (184)	3.99 (166)	59%
Independents (Democrat Survey)	129	3.27 (93)	3.31 (36)	4%
Independents (Republican Survey)	133	3.27 (93)	3.58 (40)	31%

Figure 12- Democrat Control vs Treatment

Democrat Participants	Sample Size	Baseline Agreed Statements	Condition Agreed Statements	Percent "angry"
Control	215	3.33 (96)	4.03 (119)	70%
Treatment	205	3.36 (100)	4.08 (105)	72% (p=0.91)

Figure 13- Republican Control vs Treatment

Republican Participants	Sample Size	Baseline Agreed Statements	Condition Agreed Statements	Percent "angry"
Control	167	3.32 (82)	4.02 (85)	70%
Treatment	183	3.46 (102)	3.95 (81)	49% (p=0.40)

Figure 14- Democrat Men Control vs Treatment

Democrat Men	Sample Size	Baseline Agreed Statements	Condition Agreed Statements	Percent "angry"
Control	109	3.33 (48)	3.85 (61)	52%
Treatment	86	3.43 (37)	3.96 (49)	53% (p=0.98)

Figure 15- Democrat Women Control vs Treatment

Democrat Women	Sample Size	Baseline Agreed Statements	Condition Agreed Statements	Percent "angry"
Control	106	3.33 (48)	4.21 (58)	88%
Treatment	119	3.31 (63)	4.18 (56)	87% (p=0.97)

Figure 16- Republican Men Control vs Treatment

Republican Men	Sample Size	Baseline Agreed Statements	Condition Agreed Statements	Percent "angry"
Control	80	3.24 (37)	3.98 (43)	74%
Treatment	95	3.43 (53)	3.88 (42)	45% (p=0.45)

Figure 17- Republican Women Control vs Treatment

Republican Women	Sample Size	Baseline Agreed Statements	Condition Agreed Statements	Percent "angry"
Control	87	3.38 (45)	4.07 (42)	69%
Treatment	88	3.49 (49)	4.03 (39)	54% (p=0.65)

Figure 18- White Democrats Control vs Treatment

White Democrats	Sample Size	Baseline Agreed Statements	Condition Agreed Statements	Percent "angry"
Control	145	3.51 (70)	4.11 (75)	60%
Treatment	137	3.49 (65)	4.22 (72)	73% (p=0.58)

Figure 19- White Republicans Control vs Treatment

White Republicans	Sample Size	Baseline Agreed Statements	Condition Agreed Statements	Percent "angry"
Control	147	3.24 (70)	4.05 (77)	81%
Treatment	152	3.55 (84)	4.00 (68)	45% (p=.18)

Unleashing Your Community's Potential

Quint Studer[1]

It has never been so important for communities to create a great place to live. For over five decades, small and mid-size cities have seen their young people go to college and never come back. They are exporters of talent. It is not affordability, it is opportunity and vibrancy that young people seek.

If we want to attract and retain talent and investment—which is the key to staying strong over time—we must rethink our approach to economic development. We need to understand how the workforce has evolved and we, in turn, must evolve in response. This is non-negotiable, and it is urgent.

For the first time in many generations, people can decide where they want to live and *then* look for a job. Quality of life has always been important. Now it's even more so. Today's talent wants to live in vibrant communities. They want walkable, livable downtowns with great restaurants, shops, fun activities, and trendy residential areas. Young people want to live, work, and play in the same area. (Interestingly, Baby Boomers are looking for a similar experience. More on this later.)

As a result, small and mid-size communities must compete for talent. They must aim a laser focus on becoming a place where the people want to live. These efforts, in turn, will drive investment and attract companies to us. *The formula we see again and again is that talent follows place and investment follows talent.*

The bottom line—for Beloit and for all communities—is that *opportunity* and *vibrancy* attract talent back and keep young people from leaving. We need to get strategic and intentional about creating these conditions, set measurable goals around them, and start moving toward them. That means we need to get clear on whose "job" it is to make this revitalization happen.

1 Quint Studer is founder of Pensacola's Studer Community Institute, a non-profit organization focused on improving the community's quality of life. He also founded Vibrant Community Partners, which coaches communities in building a blueprint for achieving growth and excellence.

The old "pillars" – big institutions like banks, hospitals, media outlets, and other businesses – are for the most part no longer locally owned. The executives who work for them play a critical short-term role in the community, but often they're not there for the long haul. It's no longer a given that they'll retire there. So small business leaders must step in to fill this leadership void.

Transforming a community is not easy. Many diverse players need to come together, be aligned, and take the right steps to get the revitalization process underway and keep it going. It is not easy and not a sprint. It is a marathon. However, it *is* doable—and this article will lay out some tactics for getting started.

My own learning journey begins.

My interest in community revitalization started when I met with Jim Clifton, chairman of Gallup, in 2005. I was living in Pensacola, Florida at the time (and still do) but had traveled to Washington D.C. to discuss healthcare issues (my primary focus at that time in my career).

As we were getting to know each other Jim shared recent research Gallup had done on why some cities thrive and others don't. The research showed the cities that thrive stop or reduce the migration of talent. The mid and small population centers export their talent. People leave for two primary reasons: opportunities and vibrant or fun places to live. A thriving community works to provide both.

To get more specific, here are a few points Jim shared on what makes thriving cities successful:

- They retain local companies that receive revenue outside the area. This means new dollars are coming in, rather than the same dollars just being moved around.
- They help grow and support start-ups. Today's banking environment makes it hard for small businesses to get start-up loans, so it's critical for communities to help them get needed capital.
- They have a vibrant downtown. I'll explain more about what this means shortly. In general, though, Jim said that a vibrant downtown creates tax dollars and keeps talent in the area. The vertical growth that happens in vibrant downtowns pays for urban sprawl.
- If there's a local college or university downtown, that's a bonus! Universities tend to do well in down economies, which creates more stability. They also provide a lot of intellectual capital and often turn out entrepreneurs.

This meeting was an "aha" moment for me. Afterward, a group of community members and I got together and decided to discuss ways to revitalize Pensacola. As I traveled the country for my healthcare work, I started really paying attention to the towns and cities I visited. I noticed that while each community is different, certain challenges are prevalent in most:

- Lack of alignment. Leaders are not working in tandem with each or with citizens to meet common goals.
- They are overwhelmed and gridlocked. Often, they're trying to do too much at one time and can't move the needle on anything.
- There's little coordination between private developers and local government.
- Leaders lack strong change management skills. They may underestimate the pushback they'll get and give up too soon. Or they throw facts and data at citizens, not realizing that people make decisions with their heart, not their head.
- Messaging is inconsistent. To achieve critical mass and get things rolling, all stakeholders need to be singing from the same hymnal.
- Leaders are silent when things are said that are false or misleading. Great ideas can wither on the vine when we don't get correct information in front of people.

So, armed with all these insights, the Pensacola group started building a systematic approach to revitalization that can work almost anywhere. We started looking for good ideas and best practices to harvest. Today, after 17 years of hard work, Pensacola is on almost every list of great places for everyone from entrepreneurs to retirees. I ended up writing a book about this journey titled *Building a Vibrant Community: How Citizen-Powered Change is Reshaping America.*[2]

Throughout this article I will share some of the strategies and tactics Pensacola has found to be successful over the years. Many cities and towns across America are getting great results from them as well. Let's look at some foundational actions communities need to take at the start of their revitalization journey. Then, we'll break down the four main areas to focus on moving forward.

2 To learn more, please visit www.vibrantcommunityblueprint.com.

Building a Strong Foundation.

Start with a "burning platform." As we've already mentioned, people make decisions with their hearts, not their heads. A community must find the *burning platform* that creates a sense of urgency and compels citizens to act. In Pensacola it was, *How can we keep our children and grandchildren from leaving home?* Once leaders framed issues this way, they finally got enough people behind the needed changes to achieve critical mass.

Pull together a guiding coalition. This is a coalition of leaders and citizens working together to make the community stronger, healthier, and more vibrant. Jim Cliffton call them "tribal leaders." They may or may not have a formal leadership position in the community, but they all the players and know how to get things done. They are committed to the long term health of the community. They galvanize the local business community and bring lots of groups together.

Learn the basics of change management. Understanding the psychology of change is vital for framing issues the right way and gaining the resilience to endure the long process of getting things done. Even good ideas get pushback. Here is my favorite formula for how people move through change: defiance, compliance, and reliance. Initial defiance to a new idea is normal. The leader's job is to help people accept or comply with is being recommended. Once they've moved through compliance, they usually land at reliance, meaning they've come to depend on the new and better way of doing things.

I find citizens typically fall into four basic categories:

- The "All In" People. These are your best allies for making a change happen. Support them. Thank them. Call them when you need something done.
- The "Usually In" People. They are generally for the idea; however, they need more facts and reasons. Provide them.
- The "Usually Not In" People. These are the skeptics. They tend to point out what is wrong but have few, if any, solutions.
- The "Against" People. Their mantra is: "I don't know what it is, but I don't like it." Be respectful of these people but don't waste time and energy trying to win over the unwinnable.

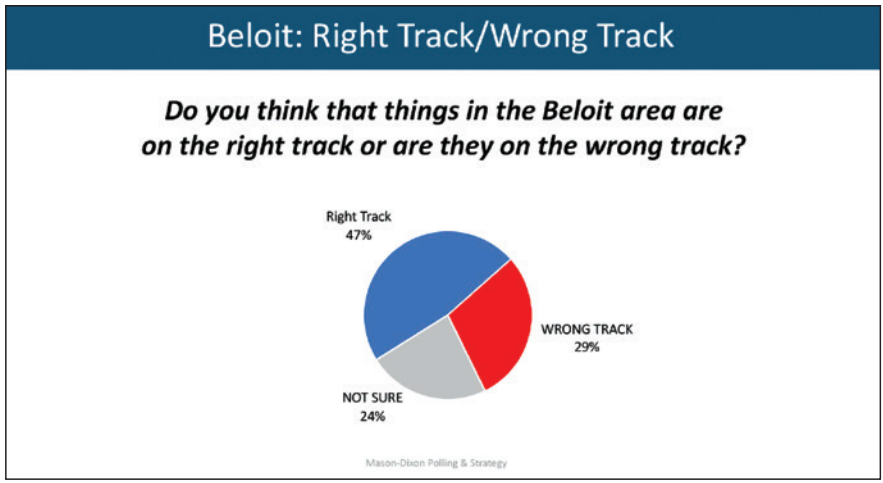
The bottom line: you'll never get everyone on board. There will always be conflicts of interest, resistance, and discomfort. Get comfortable with being *un*-comfortable. We tend to want to lower the bar because we're looking for con-

sensus, yet little change comes out of consensus. Everyone ends up unhappy. We need to start the journey knowing the goal is consent, not consensus.

Diagnose before you treat. Before you start fixing problems, get clear on what they are. Measurement and objective data helps bring that clarity and provides the diagnosis from which you can work. A good diagnosis helps identify front-burner issues and prevents you from getting distracted by secondary issues or focusing on what is easiest to fix. It keeps people from pushing through certain projects for self-serving reasons and ensures that the loudest voices don't dominate. It drives urgency, shows you where to best utilize resources, and creates ownership.

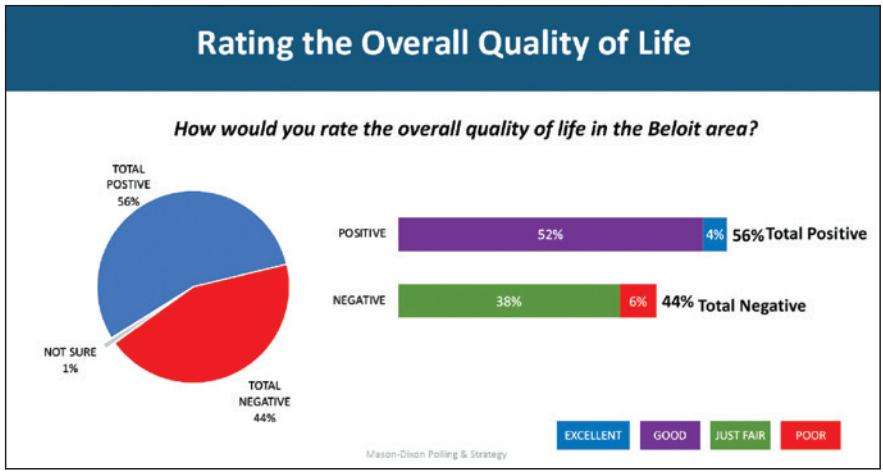
In early 2022 Beloit took the Quality of Life Survey conducted by Mason-Dixon Polling & Strategy. It is designed to collect feedback about the general wellbeing and happiness of residents. Below are a few examples of the questions that were asked, along with their responses.[3]

The survey asked about Beloit's current trajectory. While not everyone had a definitive viewpoint on this question, most of those who did felt that the city was on the right track:

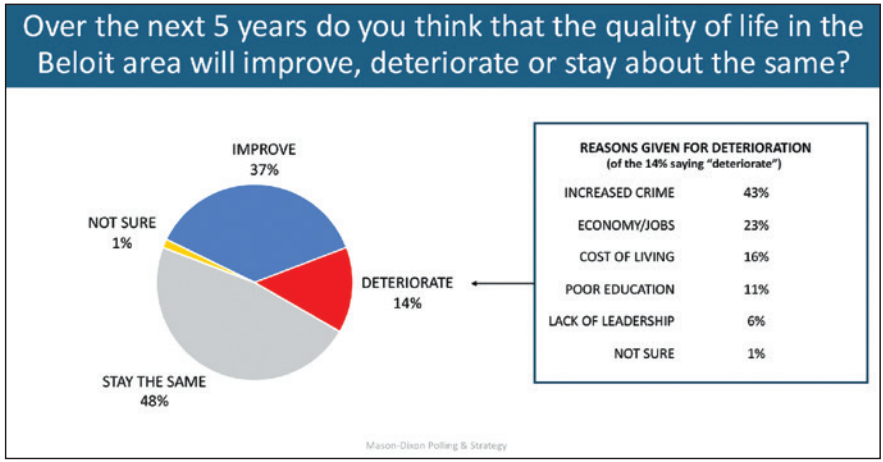


3 For the complete report on Beloit survey results visit www.greaterbeloitchamber.org/quality-of-life-survey/.

It also asked participants to rate Beloit's quality of life. Results were fairly positive:



Other questions had participants sharing their thoughts on Beloit's future. While a significant percentage felt that things will improve over the next five years, it was interesting to see what those with the most pessimistic outlook viewed as the biggest problems:



As you can see, the city now has a good snapshot of how its residents view their quality of life. The data will be useful to keep in mind as the city makes plans to move toward revitalization.

Besides the Quality of Life survey I recommend communities create a "dashboard" of objective metrics around issues like preterm births, kindergarten readiness, high school graduation rates, median income, population, etc. and update it regularly. (Think of this as something similar to how the dashboard of a car shows gas, engine performance, temperature, and so forth.) Keep the metrics in front of decision-makers and citizens. This allows you to see how healthy your community is, celebrate bright spots, identify needed improvements, and gauge progress.

Don't underestimate the power of community brand and messaging. So much of a community's success comes down to how it sees itself. You have to believe you can win before you can! When we are focused on "fixing things," it's perfectly normal to see what's wrong in our community, not what's right. When we focus on our challenges, we get everyone else focused on them too. What we give attention to expands and multiplies. Over time, we may unconsciously project a negative story about the community—and when others internalize this message, they don't want to live, visit, or invest there.

Optimism, like skepticism, is contagious. People want to feel good about their community. When you take control of its brand messaging and enlist stakeholders and citizens in your efforts, you'll quickly see positive change. This is how you'll build on your community's strengths, attract positive interest, and make big strides on your journey to vibrancy.

The Four Gears That Power Your Community.

There are four critical areas that improve quality of life and work together as "gears" to drive a community forward. As you seek to make decisions about the future of your community, you can look at them through this lens:

GEAR: Placemaking (Creating a Vibrant Downtown)

It has always been "cool" to have a charming, vibrant downtown. Now it's necessary. As mentioned earlier, a great downtown is a requirement for attracting talent and investment. It also attracts visitors, who love the same things residents love. When a community has a great quality of life—a booming economy, fun events, quirky shops and restaurants, a walkable, livable downtown or other common area, cool streets and neighborhoods, and a young hip vibe—it feeds tourism and vice versa.

This is why *placemaking* has become such a hot topic. The word has different nuances depending on context. But generally, it's about leveraging a community's assets to create a strong shared identity; to connect public spaces with citizens; and to make it a better, more engaging place to live and visit. And more and more, we are starting to think of it as an economic strategy.

As you consider how to spend public money, keep placemaking firmly in mind. Use public dollars as a catalyst for private investment downtown or other areas where citizens and visitors tend to gather. Make sure you can show ROI up front. It's always been true that private investment drives growth. It's the key to job creation and a strong, sustainable tax base.

Communities reimagine themselves by maximizing what they already have. For starters, use existing structures when you can. People love cool old buildings. This preserves history and captures the character of community. Over the years I've discovered other tactics that work well; for example, changing one-way streets to two-way streets, looking for underutilized land and filling it in, making sure you have good sidewalks and public restrooms.

That said, there are four main ingredients that must come together to create a vibrant downtown. Here's a brief overview of each one:

Ingredient 1: Programming. Your first priority is to create events that draw people downtown. The goal is to get citizens and visitors alike to hang out so they'll shop, eat, drink, and possibly even stay overnight.

Pensacola has achieved this goal in several ways. They built a multi-use stadium, which is the home of the Pensacola Blue Wahoos and a great venue for music and other special events. They also hold an annual Seafood Festival, which attracts more than 10,000 people to the downtown area over three days. One Friday night each month, downtown streets close for Gallery Night, featuring music, art, and cuisine. Finally, the area is also home to many running clubs and fundraising walks.

Of course, very community must choose the type of program that's right for them. Other good examples might involve a YMCA, a farmer's market, a theater, or a series of concerts in the city park. The main point: Once you get intentional about giving people a reason to come downtown, they find they enjoy being there.

Ingredient 2: Retail/Entertainment. When people are downtown for an event, they need places to eat and shop. A vibrant downtown gives them these places. Communities need to attract owners of restaurants, boutique shops, coffee houses, and so forth and entice them to open up downtown. But it won't

necessarily happen on its own. The process needs to be deliberately cultivated. Remember: Downtown is more than just a destination. It's an experience.

Ingredient 3: Office Space. While the retail businesses attract foot traffic, downtowns also need realtors, attorneys, accountants, and other types of companies to help support the economy. Cities need to attract both types. (Just don't put them on the ground floor; you want retail businesses there for the foot traffic.) These non-retail businesses keep downtown activated during the day. Employees who work for them, and also their customers, become customers for the retail companies.

Ingredient 4: Residential. Residential development is important because if things get tough economically, there will still be a base of people to support restaurants, shops, etc. The challenge is that retail usually has to come first. People don't want to move downtown if there is nothing for them to do there. That's why most towns need to start with local investors who aren't so concerned with a high return on investment. They see the real ROI as a better community.

There are two age groups of people who want to live downtown: under 35s and over 55s. Young people like downtown so they can be in the middle of the action and (at least in some cases) live where their job is. Many of them don't have children so they don't need lots of space. Empty nesters like downtown because it offers freedom from yard maintenance and upkeep. They can walk everywhere—the grocery store, the bank, the park, the clinic. The great news is when older residents move downtown from the suburbs it opens up living space for others. Remember these two groups when building out residential and look for ways to meet their needs.

As you are revitalizing downtown, play "small ball." Think progress, not perfection. Making small, incremental changes is more cost effective than huge transformations and, in many ways, more powerful. Inspiring visions and big goals have their place, but execution happens in a lot of small steps. When we try to do too many things at once we can get overwhelmed. It can feel like drinking from a firehose. It's much better to focus on one or two small things first, get a quick win, and get some momentum going before moving to the next thing.

Once a community starts building a vibrant downtown, it gets people activated and sparks growth in the rest of the community. Citizen enthusiasm builds. Momentum spreads to other areas. It becomes the engine that powers the rest of the community.

GEAR: Early Learning & Education

We all know a strong talent base is essential to creating a strong community. And we *also* know that begins with a well-trained population. This is why education metrics, particularly high school graduation rates, are on pretty much every community's dashboard. The problem is, graduation rates are a lagging indicator: By the time you realize yours aren't where they should be, you're already in trouble.

That's why the early learning and education gear revolves around treating not just the symptoms (low graduation rates) but also the root causes (low kindergarten readiness, especially in underserved communities). We firmly believe the optimal time to start tackling the problem is at birth. Our children's brains are too important *not* to invest in.

John A. List of the University of Chicago's Department of Economics, said it this way: "There really is one, and only one, reliable economic indicator for the long run vitality of a community and indeed the broader society: optimal development of a young child's brain. Without proper programs in place at birth, we not only miss a key driver to enhance economic growth, but also the chance of equal opportunity for our children."

When I was writing *Building a Vibrant Community* I realized I wanted John List to be a part of it. At the time we were kicking off efforts to make Pensacola an Early Learning City™ based on the research of his colleague Dr. Dana Suskind, a cochlear implant surgeon at the University of Chicago Medicine Comer Children's Hospital. Her work shows children who hear more words and have more positive, language-rich interactions with adults in infancy (typically those from higher socioeconomic families) are better prepared when they enter school.[4] I felt that by having an economist *and* a doctor team up to address the issue, more communities might take notice.

I am grateful to say that John List did end up writing the foreword. In it, he addresses the problem of educational disparities in the U.S, noting that the substantial resources committed to public education over the past half century have done little to close the achievement gap. He goes on to add:

"In my work with Dana Suskind, we have focused on three distinct features of the policy problem. First, by focusing public policy dollars on prevention rather than remediation, we call for much earlier educational programs than currently conceived. Second, our approach has parents at the center of the education

4 Visit https://www.studeri.org/EarlyLearningCity for more data, info and their sources.

production function rather than at its periphery. Third, we advocate attacking the macro education problem using a public health methodology, rather than focusing on piecemeal advances." *Building a Vibrant Community* (2018).

All three of these features are reflected in Pensacola's Early Learning Initiative. Our efforts are backed by the research in Dr. Suskind's book *Thirty Million Words: Building a Child's Brain*. Dr. Suskind teaches us that 80 to 85% of a child's brain is developed in the first three years.

Thus, we decided to try to reach parents early—before they leave the hospital—with the message that talking, reading, singing, and playing with their baby from the very first days of life will develop their baby's brain, which will in turn help them prepare for school.

We partnered with the University of Chicago and formed a pilot program to make sure every mother who gave birth in one of the three hospitals in Escambia County gets an early intervention. That's 5,000 births a year. These hospitals—Baptist, Sacred Heart, and West Florida—distribute materials developed from the Thirty Million Words Initiative designed to help new parents work more words into their interactions with their babies and young children.

We called this in-hospital education project the "Brain Bag." Essentially, the Brain Bags are early literacy gift bags given to new mothers before they leave the hospital. They include educational resources and a children's book. The bags are given with a bedside lesson—done by a nurse or trained volunteer—to help reinforce the message. It also includes a short educational video to reinforce this teaching.

The approach Pensacola is taking inside its hospitals is incredibly practical. Showing the video to new parents takes up very little of a nurse's time. And early research is showing that the interventions a) are successful in boosting parent knowledge of infant language and cognitive development, and b) fit into the context of regular maternity care.[5]

This is only one of the ways Pensacola is investing in the next generation. We've also launched a program based on MIT-trained Harvard Professor Ron Ferguson, who works to close the educational achievement gap based on socioeconomic status and race.[6]

5 Source: https://www.studeri.org/news-detail/postdetail/video-lesson-can-boost-parent-knowledge

6 Visit https://www.studeri.org/pensacolabasics to learn more about this program aimed helping every family give every child and good and healthy start in life.

We owe it to our children and our community to act now. One of my favorite sayings is *The best time to plant a tree is 20 years ago; the second-best time is today*. Sometimes we have to take action just because it makes sense and trust that results will come in the long-term.

GEAR: Economic Development

A thriving small business presence is a crucial part of any vibrant community. Successful business owners create jobs. They pay rent and taxes. They also give back to the community in many other ways. Yet it's estimated that 40 percent of new businesses fail in their first year and that 80 percent don't make it five years. That's why vibrant communities make it easy for new businesses to get started *and* help keep existing ones healthy long-term. They know that transferring vital knowledge and experience to entrepreneurs is critical to everyone's success.

As Michael Gerber, author of *The E-Myth Revisited*, explains, most companies are started by "technicians." A plumber starts a plumbing company, or a person who loves to cook decides to open her own restaurant. These people are passionate about what they do, and they have expertise around their service or product. What they don't have are the skills to run a company…and so they fail. That's why it is so important for communities to put a training and development structure in place to help these small businesses thrive long-term.

In Pensacola, we started out by offering small, focused training sessions around a few topics. Our efforts have now evolved into a comprehensive training program that engages much of our business community. We offer leadership development training through a series of workshops, roundtables, and a yearly conference all aimed at entrepreneurs and small business owners. Here is an overview of each:

Monthly Training and Development Workshops. We began our workshop series by identifying the leadership skills it takes to run a successful business: for example, hiring, firing, employee engagement, creating revenue streams, process improvement, marketing, etc.

We know business owners are busy and that everyone may not need training on every skill, so we created a tool that evaluates which of them a business owner most needs to be trained in. For example, if a business has only three employees, hiring is probably not the most important thing in the world to them, but creating revenue streams might be really important.

To do this foundational training, we've been able to pull a lot of experts from the community— many of whom are business owners—to teach specific skills.

These monthly workshops have been an amazing success, and people are stunned at the talent we have here.

Small Business Roundtables. In these meetings, which happen every 90 days, small business owners get together with a facilitator and talk about the issues they're facing. They are done in small groups—generally four to eight business owners—so that participants can be comfortable sharing their problems and ideas. We choose a good facilitator who can answer questions and keep the discussion productive and on track.

Like the workshops, the purpose of the roundtables is training and leadership development. However, they're more intense and focused. At each session, we focus on providing tools people can implement immediately to solve some of their biggest problems. Over time, this training becomes a very powerful skill set.

We get great feedback from these sessions. One of the things we hear most often is that business owners often feel alone and think they are the only one having problems. The roundtables show them this is not true. And while the training and development are important, one of the most meaningful byproducts is that participants all end up helping each other.

Once people get to know each other, they start to share resources and create strong relationships. Before long they're all working together and thinking in a different way: *How can I do this for you? How can you do this for me? How can we form a partnership? Can we start a new business?*

EntreCon. An annual business conference held in Pensacola each November, Entrecon provides a venue for local entrepreneurs, professionals, and leaders to come together, learn from experts on most pressing problems, and learn from each other. The idea is to continue to share the tactical tools and strategies attendees need to hone their leadership, grow their business, and improve their bottom line.

We bring in nationally known experts from the business community to share their expertise in areas businesspeople often struggle with, and we offer breakouts featuring local talent. By hosting EntreCon locally, we're providing a cost-effective, convenient way for local companies to get world-class training without the expense of travel.

Every year, this conference grows and generates more energy and excitement than the previous year. Clearly, it is meeting a need and resonating with a lot of businesspeople.

Small Business Challenge. In addition to training and development sessions and events, it's important to encourage entrepreneurship in the community. A

great way to do this is to hold a Small Business Challenge. In Pensacola, this was one of our first steps in getting a laser focus on small business. We wanted to inspire a passion for small business and give people a compelling reason to turn their great ideas into viable business plans. This kind of challenge works because it forces people to put their ideas on paper and creates a sense of urgency around starting a business.

We issued our challenge at the end of 2011 (after having been inspired by Asheville, NC, which did something similar). Essentially, anyone wanting to start or expand a business in downtown Pensacola was invited to submit business plans to a nine-person panel of local business experts. Consultants from the Florida Small Business Development Center at the University of West Florida College of Business (Florida SBDC at UWF) helped the contestants with this process.

For the prize package, we decided on a mix of start-up capital, free and reduced rent, and ongoing mentoring from local business experts. More than 100 participants signed up and paid the $40 business plan software fee. Of that number, 31 made it through the two-month business plan process.

MariCarmen Josephs, an area restaurant manager and chef, was announced the winner in March of 2012. Her proposed restaurant would offer an eclectic blend of Southern and international flavors, including Spanish, Italian, Mexican, Indian, Thai, and Moroccan. Today it's very successful. Carmen's Lunch Bar & Tapas is a hot meeting place for business lunches, dinners, and pre-event wine and tapas.

We expected this to be a small win, but it has had a huge impact. Not only did Carmen's quickly get up and running, but several entrepreneurs who went through the process ended up starting companies too: a tamale restaurant, a bakery, a jewelry store, a paddleboard rental company, and a Segway tour company.

Everything I just described is "Level 1" of the Economic Development Gear. Once these pieces are well in place, a community may want to move on to Level 2: creating an entrepreneurial ecosystem in your community. In Pensacola the "front door" to our ecosystem is The Spring Entrepreneurial Hub[7] which exists to empower, connect and grow small businesses in our community. The Spring focuses on three key areas – mentorship, connection to resources via an interactive asset map, and scaling/acceleration – to bring unprecedented structure and aid to local small businesses in Escambia and Santa Rosa County.

7 www.thespringpensacola.com

Investing in your small business leaders and entrepreneurs pays off in many ways. Besides helping them grow thriving companies, you'll foster a sense of community in this population. You'll then be able to harness their energy and know-how to create a powerful army of citizens who can be galvanized to help solve the problems your community faces.

Small business owners and leaders already have a vested interest in creating vibrancy. When the community does well, they do well. They'll be your catalysts for change and sustainers. They'll take charge of making sure good, sustainable growth stays on track. They'll make sure young people are getting educated, downtown is thriving, and government is running smoothly.

GEAR: Civic Engagement

The only change that will succeed long-term is citizen-powered change. Without widespread and enthusiastic buy-in, initiatives will fail. It's crucial that we engage people up front. That means asking, "What do you want the community to look like?" And it means educating them on what they can become and on what it takes to get there. Bring in experts to help people see the way forward. When people "get it," they get behind it.

Civic education helps people be more effective citizens, voters, and community members. Educated and engaged citizens are contributors and refiners of great ideas, the voices that advocate for change, the "boots on the ground" that make things happen. Raising the Civic IQ of a community is the key to accelerating results.

Here are a few reasons why communities should make civic education a priority:

It builds critical mass (which helps you execute). Typically, there's no shortage of good ideas—but ideas matter only if they get implemented. To execute on ideas, you have to bring people with you. Civic education helps you get people on board so you can move quickly.

It helps citizens see the *why* behind proposed improvement initiatives. When they understand what's going on and how they will benefit, they're far more inclined to get on board with revitalization efforts and even become advocates.

It brings lots of people together to hear the same ideas at the same time. There is great power in educating a group all at one time under one roof. It creates a common language that allows us to communicate and make connections. Conversations start, gather momentum, and trickle out into the community where other people are drawn in.

It shifts the conversation and gets the community "unstuck." (This is the first step toward vibrancy.) Civic education gets us talking about what a great community we already are, why we deserve to get better, and how we're going to do it.

It's a great way to collect best practices and learn from others. Speakers often share stories of what they see in other communities. Learning from the mistakes and successes of other communities is far more efficient than trying to reinvent the wheel.

Civic education sparks creative solutions. The more conversations you have with citizens, more ideas rise to the surface. Someone volunteers a resource or skill. .Others jump in and contribute. People know people (and know people who know people). This is how action starts. Plus, the people closest to the problems often have the best solutions.

It gets the average person involved. So many of us are stuck in our work-to-home routine that we can't see a way to be involved in the community. Civic education events give people a concrete place to start.

It helps neighbors get to know neighbors. Providing opportunities to bring people together strengthens social ties. In creating vibrant communities, that sense of connection matters.

It benefits government leaders. Some may not have the knowledge and training they need to make the best possible decisions, especially during times of rapid growth or change. But even if they do, they'll appreciate hearing outside perspectives. And most government leaders value the strong citizen-led initiatives that come from civic education.

Educated citizens hold government officials accountable. They'll ask more informed questions. They're more likely to involve themselves in key processes from the beginning.

In Pensacola, the centerpiece for civic education is the CivicCon lecture series. It began in the fall of 2017, when the Studer Community Institute and the *Pensacola News Journal* joined forces to bring the top urban planners in the country to Pensacola to speak to our community. CivicCon has proven wildly successful, with more than 300 citizens at each event and more than 5,000 live video views for each session.

We got great feedback from the community, and the main message was that people didn't want to stop with the monthly lectures—they wanted more involvement. They wanted to know how to take it to the next level and get even more involved. This shows us citizens will engage when given the right opportunity.

Here are just a few of the topics we've covered:

- Public Private Partnerships
- Cities and Social Equity
- Smart Investments
- Getting Parking Right
- Creating Great Neighborhoods
- Resilient Waterfronts
- Walkable Cities
- Market Research
- Loving Your Community
- Diversity, Equity and Inclusion

Having people together hearing the message at the same time from experts on some of our most pressing community problems has been powerful in many ways. Not only does it promote buy-in, it helps stamp out negativity (as they are armed with good information) and it creates a common language for having good discussions.[8]

Keep The Throttle Down

Revitalizing your community is long and winding road. None of what is described above happens quickly. It unfolds over the course of many years and there are many unexpected twists and turns and many setbacks. It is not uncommon to take one step forward and two steps back. The key is to keep moving along. Use every success to keep momentum going. Above all, don't give up when times get tough.

One of my favorite movies is The Right Stuff. I love the scene where Chuck Yeager breaks the sound barrier. Many other pilots had tried to break the sound barrier, but when they sped up the plane would shake and they would naturally pull back on the throttle. Despite heavy turbulence, Chuck Yeager kept the throttle down.

Every community will experience turbulence on the journey to vibrancy—but those that keep the throttle down will win in the long run.

8 To learn more visit https://www.studeri.org/CivicCon.

The Miller Upton Program at Beloit College

The Wealth and Well-Being of Nations was established to honor Miller Upton, Beloit College's sixth president. This annual forum provides our students and the wider community the opportunity to engage with some of the leading intellectual figures of our time. The forum is complemented by a suite of programs that enhance student and faculty engagement in the ideas and institutions that lay at the foundation of free and prosperous societies.

Back Cover Photo © Amanda Urish

Contributors

John A. List Anya Samek Sam Lindquist Sally Sadoff Alec Brandon

Phil Chen Phil Shields Edward Vezosa Quint Studer

Department of
Economics & Business

Beloit
College

ISBN 979-8-218-26736-0

90000>

9 798218 267360