Title
Essays in Behavioral Economics

Author
Birke, David Jan Dietmar

Publication Date
2020

Peer reviewed|Thesis/dissertation

Essays in Behavioral Economics

by

David Jan Dietmar Birke

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Stefano DellaVigna, Chair
Professor Edward Augenblick
Professor Edward Miguel

Spring 2020

Essays in Behavioral Economics

Abstract

Essays in Behavioral Economics

by

David Jan Dietmar Birke

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Stefano DellaVigna, Chair

Prosociality is a core element of human behavior.[1] A variety of explanations have been proposed for prosocial behavior, such as outcome-based altruism, giving for warm glow, and reciprocity. While these models can explain basic patterns, they have trouble explaining cases of avoidance of costless information, as documented in the moral wiggle room experiments (e.g. Dana et al., 2007; Grossman and van der Weele, 2016).

To explain these and other findings, Bénabou and Tirole (2006) introduce a model of prosocial behavior in which agents want to signal to others that they are prosocial. In this model agents behave prosocially not only out of intrinsic care, but also out of reputational concerns. That is, agents take costly actions to signal to others or to themselves that they are the kind of person that cares about others. Agents might then anticipate that reputational concerns will compel them to behave more prosocially, and therefore actively avoid situations in which their contribution is more visible.

In chapter 1 of this dissertation, I propose a new approach to distinguish signaling models from traditional social preference models. The main idea is to study a setting in which agents contribute to a prosocial cause while also receiving a personal bonus for reaching a threshold level of contribution. If agents are motivated by neoclassical incentives, outcome-based altruism, warm glow giving or social norms, then such a bonus incentive induces bunching at the bonus threshold. In contrast, if agents want to signal their prosociality, then *no bunching* occurs in equilibrium. This prediction arises because with any bunching, the most intrinsically motivated buncher can marginally increase their contribution to separate themselves from less intrinsically motivated bunchers, and thereby receive a discrete signaling benefit at a marginal cost.

Moreover, signaling models predict that increasing the bonus amount induces *anti-bunching*, that is an increase in the contribution level strictly above the bonus threshold. Anti-bunching arises because increasing the bonus amount lowers the intrinsic motivation of the marginal buncher. Since all agents obtaining the bonus still want to distinguish themselves from lower types, they need to respond by increasing their contribution. This prediction

[1]As an example, the average US adult gives $1,600 to charity and volunteers 28 hours per year, generating an economic value of around $570 billion dollars (Giving USA 2018).

distinguishes signaling models from several alternative models, that predict a zero response from inframarginal agents.

In chapter 2 of this dissertation, I build on these insights to design and conduct a real-effort online experiment. In the control group, participants choose to complete up to 38 transcription tasks with a return to charity of 8 cents per task, but without any personal benefit. In the 40c-bonus group, the return to charity is the same, but participants also earn a personal 40c-bonus for completing 15 or more tasks. In the $1.20-bonus group the bonus for completing 15 or more tasks is $1.20. The comparison across these bonus groups allows me to test for the key no-bunching and anti-bunching predictions.

I combine the bonus component of the design with a second, visibility component. Making effort more visible to others should amplify the motivation to signal. Thus, in a crossed randomization, half of the participants are asked to create a personal *Badge*. The badge displays tasks completed, total donation amount raised, total personal gain and the bonus incentive scheme, together with a picture of the participant that they take using their webcam. After completing the experiment, each participant's badge is shown to at least one other participant, who is then asked to judge the participant's generosity. Participants are made aware of this when creating the badge as well as during the transcription task.

The experiment provides evidence for signaling motives: If a participant's effort is private, then introducing a 40c-bonus incentive for completing 15 prosocial tasks increases the share of participants completing 15 or more tasks from 19.8 % to 51.7 %. This 31.9 pp increase is accompanied by a 3.9 pp increase from 18.3 % to 22.2 % in the share of participants completing 17 or more tasks. This is the baseline effect, that captures a participant's motivation to signal to themselves or the experimenter.

If a participant's effort is visible to other participants, then introducing the same 40c-bonus incentive increases the share of participants completing 15 or more tasks from 22.1 % to 54.9 %. This 32.8 pp increase is now accompanied by a 9.4 pp increase from 21.0 % to 30.4 % in the share of participants completing 17 or more tasks. Since 1.1 % complete 15 or 16 tasks without a bonus, this implies that at least $(9.4 - 1.1)/32.8 \approx 25\,\%$ of those responding to the bonus incentive exhibit signaling motives strong enough to complete at least 2 additional tasks. I expand on this basic finding with additional tests. In sum, the chapter provides a proof of concept for anti-bunching as a test for signaling motives.

In chapter 3 of this dissertation, in joint work with Garret Christensen, Zenan Wang, Elizabeth Paluck, Nicholas Swanson, Edward Miguel, and Rebecca Littman, we investigate an example of prosocial behavior in the field. Practicing open science is an inherently prosocial act as it allows fellow researchers to learn from and build upon existing research.

We conduct an incentivized survey of active social scientists to study the adoption of open science practices (posting data, code and study materials online, pre-registering studies, hypotheses, and analysis prior to conducting a study). We find that as of 2017, over 80% of scholars had used at least one open science practice, rising from one quarter a decade earlier. We also find similar attitudes toward research transparency between older and younger scholars, but the pace of change differs by field and methodology. Patterns are consistent with most scholars underestimating the trend toward open science in their discipline.

To my parents,
who gave me the love, support, and freedom
that allowed me to go out into the world
and find my way.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I am grateful to my main advisor and mentor Stefano DellaVigna, who has nurtured and guided my love for economics. His enthusiasm for the field and commitment to doing good research will be a lifelong inspiration to me. Stefano's extraordinary kindness and firm belief in my capabilities have had a profound impact on me. In eight years of interacting through classes, meetings, discussions, conferences, and emails, Stefano was *always* supportive, thoughtful, and kind. He is a role model, and I aspire to pass on what he has given to me.

I am also grateful to Ned Augenblick, whose guidance was essential in converting a rough sketch of an idea into this dissertation. A few months leading up to the job market, he empathetically noticed that I was stuck, and then went above and beyond to provide me with the words and resources that I needed to make progress. I also benefited tremendously from Ned's superpower to quickly identify the essence of an idea, give it a healthy reality check, and articulate a clear path for improving it. Ned always provides a new perspective, and I aspire to achieve his synergy of creativity and pragmatism in my own research and advising.

I am also grateful to Ted Miguel, who gave me the opportunity to collaborate on an immensely important research agenda. His optimism, wisdom, and humor made our Friday team meetings something to look forward to every week. Ted has an exceptional ability to bring people together to work towards a common vision, and I have learned from him as much about doing research as I have about building a team. He is a true leader, and I aspire to follow his example.

I am also grateful to Matt Rabin for combining insight and wit into a wonderful course on Psychology and Economics; David Card for giving me the chance to explore labor economics; Pat Kline for helping me find focus in my third year; Ben Handel for his tremendous help in navigating the job market; and Aaron Bodoh-Creed, Chris Shannon, Dmitry Taubinsky, and Chris Walters for their always helpful comments, discussions, and advice.

As this degree marks the end of my formal education, I would also like to thank the many teachers and mentors that have influenced and supported me at the various stages of my education, particularly Anna Werren, Gerd Göckeritz, Dirk Engelmann, and Emily Oster.

My time at Berkeley was immeasurably enriched by the wonderful people that were there with me. Thank you, Jonathan Holmes, Katja Jensen, Peter Jones, Mats Köster, Isabela Manelici, Max Müller, Evan Rose, Jose Vasquez, Francis Wong, and especially Murilo Ramos, Nick Sander, and Zenan Wang. Thank you for six years of stimulating conversations, amazing excursions, and countless game, movie, and dinner nights, and for bringing love, laughter, and joy to it all. And thank you, Patrick Becker, Alexandra Lorson, Jakob Schlockermann, Julian Schüssler, Cecile Thai, and Irina and Simon Zickwolf, for your friendship and love from near and far.

Finally, I am grateful to my parents for giving me the love, support, and freedom that allowed me to go out into the world and find my way. I dedicate this dissertation to them.

Chapter 1

Anti-Bunching: A New Test for Signaling Motives

1.1 Introduction

The average US adult gives \$1,600 to charity and volunteers 28 hours per year, generating an economic value of around \$570 billion dollars.[1] A variety of explanations have been proposed for this kind of prosocial behavior, such as outcome-based altruism, giving for warm glow, and reciprocity. While these models can explain basic patterns of prosocial behavior,[2] they have trouble explaining cases of avoidance of (costless) information, as documented in the moral wiggle room experiments (e.g. Dana et al., 2007; Grossman and van der Weele, 2016).

To explain these and other findings, Bénabou and Tirole (2006) introduce a model of prosocial behavior in which agents want to signal to others that they are prosocial. In this model agents behave prosocially not only out of intrinsic care, but also out of reputational concerns. That is, agents take costly actions to signal to others (or to themselves) that they are the kind of person that cares about others. Agents might then anticipate that reputational concerns will compel them to behave more prosocially, and therefore actively avoid situations in which their contribution is more visible.

A small number of papers have studied the role of signaling in prosocial behavior, generally using one of two approaches. The first approach is to manipulate the visibility of an agent's action. For example, in Ariely et al. (2009), participants announce their individual contributions to each other in the laboratory; in Exley (2017), a participant's contribution is shown to panel members who then reward it monetarily; and in Andreoni and Bernheim (2009), nature randomly overrules a participant's dictator game giving decision. The common idea in these papers is that an increase in visibility increases an agent's ability to signal, and therefore observed differences in behavior help identify a signaling motive. The second

[1]Giving USA 2018

[2]For example, the models can explain why public grants crowd out private contributions (Andreoni and Payne, 2003), charities raise funds by emphasizing individual cases, and small gifts increase donations (Falk, 2007).

approach is to vary piece-rate incentives to perform a prosocial task and test whether higher piece rates crowd-out effort, as predicted under some conditions in social signaling models.

I propose a new, third approach to distinguish signaling models from traditional social preference models. The main idea is to study a setting in which agents contribute to a prosocial cause while also receiving a personal bonus for reaching a threshold level of contribution. If agents are motivated by neoclassical incentives, outcome-based altruism, warm glow giving or social norms, then such a bonus incentive induces bunching at the bonus threshold. In contrast, if agents want to signal their prosociality, then *no bunching* occurs in equilibrium. This prediction arises because with any bunching, the most intrinsically motivated buncher can marginally increase their contribution to separate themselves from less intrinsically motivated bunchers, and thereby receive a discrete signaling benefit at a marginal cost.

Moreover, signaling models predict that increasing the bonus amount induces *anti-bunching*, that is an increase in the contribution level strictly above the bonus threshold. Anti-bunching arises because increasing the bonus amount lowers the intrinsic motivation of the marginal buncher. Since all agents obtaining the bonus still want to distinguish themselves from lower types, they need to respond by increasing their contribution. This prediction is also specific to signaling models, as alternative models predict a zero response from infra-marginal agents.

Anti-bunching complements the two existing approaches by not relying on manipulations in visibility or two-dimensional signaling, yet being fully non-parametric. An additional advantage is that bonus incentive schemes are a feasible manipulation in many prosocial field settings, and often already exist in the form of gifts or invitations to special events. Testing for signaling motives in these settings is then only a matter of manipulating the value of the bonus reward.

1.2 Model

In this section, I first present the intuition for the no-bunching and anti-bunching results. I then introduce the formal setup, which allows me to state the results precisely. I proceed by discussing the role of the single-crossing property and the D1 criterion, as well as extensions of the model. Finally, I estimate the magnitude of the effect sizes in a simulation exercise. All proofs are in appendix A.1.

Intuition

The core logic behind the no bunching result is as follows. Whenever a range of types bunches at an action $\overline{a}$, the buncher with the highest type, θ, incurs a reputation strictly below their type, $r(\overline{a}) < \theta$. Type θ now considers marginally increasing their action to $\overline{a}+\varepsilon$. If $\overline{a}+\varepsilon$ is played in equilibrium, then it must be played by types $\theta' \geq \theta$, and therefore yield a reputational payoff of at least $r\left(\overline{a}+\varepsilon\right) \geq \theta$. If $\overline{a}+\varepsilon$ is not played in equilibrium, then, assuming the single-crossing property, θ has a bigger incentive to deviate to $\overline{a}+\varepsilon$ than any $\theta'' < \theta$, and thus in any D1-equilibrum, $r\left(\overline{a}+\varepsilon\right) \geq \theta$. Taken together, deviating from $\overline{a}$ to $\overline{a}+\varepsilon$ yields a discrete reputational benefit of at least $\theta - r\left(\overline{a}\right) > 0$. However, with a continuous cost function, the intrinsic cost of the deviation is marginal, and hence always outweighed by the discrete reputational benefit. Therefore any type bunching wants to deviate by "going the extra mile", implying that all equilibria are fully separating.

If types are fully separating, then the reputation function at an action a is directly related to a single type's first-order condition, implying that the equilibrium action function $a^*\left(\theta\right)$ needs to satisfy a differential equation. This differential equation has a solution, which is unique up to initial conditions. Because the lowest type does not incur a reputational loss by deviating downwards, their equilibrium action is the same as if they had no image concerns. This is the initial condition that pins down a unique equilibrium outcome.[3]

Given full separation, what is the effect of an increase in the bonus amount on equilibrium behavior? On the extensive margin, the increase makes it more attractive for everyone to get the bonus, reducing the intrinsic motivation of the marginal type getting the bonus. This expands the pool of types getting the bonus downwards, and so high types need to work harder on the intensive margin to separate themselves. If types care about their image, then this leads to anti-bunching, an increase in effort of all inframarginal types getting the bonus.

I now state and prove these results formally.

Setup and Results

Types θ are drawn from a distribution F_θ with continuous, bounded support $\Theta = [\underline{\theta}, \overline{\theta}] \subset \mathbf{R}$. After observing their type, each agent chooses an action $a^*(\theta)$ that satisfies the incentive compatibility constraint

$$\max_{a\geq\underline{a}} U\left(\theta, a, r(a)\right) = \max_{a\geq\underline{a}} g\left(\theta, a\right) + \mu r\left(a\right) + b\,\mathbb{1}\left\{\mathrm{a} \geq \bar{\mathrm{a}}\right\} \tag{IC}$$

where $\underline{a}$ is a minimal action, and g is twice continuously differentiable, strictly concave in a, attains a unique maximum $a^0(\theta)$ given θ, and satisfies the strict single-crossing property, so that

$$g(\theta, a') \geq g(\theta, a) \Rightarrow g(\theta', a') > g(\theta', a) \qquad \text{for } \theta' > \theta \text{ and } a' > a. \tag{SCP}$$

Every action carries a reputation $r : \mathbf{R}_+ \to \Theta$, on which agents put weight $\mu \geq 0$. Reputation at a depends on the believed distribution of types playing a, $\beta(a)$, through $r(a) = R(\beta(a))$. I assume that $R(\beta) \in \left[\inf \operatorname{supp}\left(\beta\right), \sup \operatorname{supp}\left(\beta\right)\right]$, and if β first-order

[3] The equilibrium exists by standard results on differential equations with continuous functions.

stochastic dominates β', then $R(\beta) > R(\beta')$.[4] The key innovation is the additional bonus incentive $b \geq 0$ that is given to any agent passing the bonus threshold $\overline{a} \geq \underline{a}$.[5]

Definition 1 (D1-Equilibrium)**.** Let $a^{*-1}(a) \equiv \{\theta | a \in a^*(\theta)\}$. (a^*, r^*, β^*) is a D1-equilibrium if

1. $\forall \theta \in \Theta : a^*(\theta)$ solves (IC), and
2. $\forall a$ with $a^{*-1}(a) \neq \emptyset : r(a) = R(\beta^*(a))$, $\inf \operatorname{supp} \beta^*(a) = \inf a^{*-1}(a)$ and $\sup \operatorname{supp} \beta^*(a) = \sup a^{*-1}(a)$, and
3. $\forall a$ with $a^{*-1}(a) = \emptyset : \beta^*$ satisfies the D1 criterion.[6]

Note that above definition is more general than the usual equilibrium notion, as it allows for non-Bayesian beliefs β^* on the equilibrium path, as long as the boundaries of the support of the beliefs align with what is played in equilibrium. The D1 criterion implies that the reputation at an off equilibrium action is contained in the interval span of those types who need the lowest amount of reputational compensation to deviate to that action.

Under these assumptions, I obtain the following three key results (all proofs are in appendix A.1).

Result 1 (No Bunching)**.** *There does not exist a D1-equilibrium with any bunching.*

Result 2 (Existence and Uniqueness)**.** *There exists a unique fully separating D1-equilibrium.*

Result 3 (Anti-Bunching)**.** *An increase in the bonus size b increases the equilibrium action for all types above $\overline{a}$. Formally, if $b' > b$, then $a_b^*(\theta) \geq \overline{a} \Rightarrow a_{b'}^*(\theta) > a_b^*(\theta)$.*

Discussion

The key assumptions behind result 1 are the D1 criterion and the single-crossing property. Without the D1 criterion, bunching can be generated at $\overline{a}$ by setting $r^*(a)$ in the right neighborhood of $\overline{a}$ low enough. This remains possible under the restriction that is r^* is non-decreasing. However, when giving up the D1 criterion, the problem arises that bunching is also possible at many $a \neq \overline{a}$. Hence, to make a prediction about the location of bunching,

[4] This implies that if β is non-degenerate, then $R(\beta) < \sup \operatorname{supp}(\beta)$.

[5] I also assume that the bonus threshold $\overline{a}$ is sufficiently high and the bonus amount sufficiently small enough such that $\underline{\theta}$ would not obtain it without a reputational benefit. For result 3 to be interesting, I also assume that $\overline{\theta}$ does obtain the bonus without a reputational benefit. Formally, $b \leq \max_{a \geq \underline{a}} g(\underline{\theta}, a) - g(\underline{\theta}, \overline{a})$ and $\arg\max_{a \geq \underline{a}} g(\overline{\theta}, a) + b\mathbb{1}\{\mathrm{a} \geq \overline{\mathrm{a}}\} \geq \overline{a}$.

[6] Specifically, let $M_<(\theta|a) = \{r \in \Theta | U(\theta, a^*(\theta), r^*(a^*(\theta))) < U(\theta, a, r)\}$ and $M_=(\theta|a) = \{r \in \Theta | U(\theta, a^*(\theta), r^*(a^*(\theta))) = U(\theta, a, r)\}$. The D1 criterion requires that there does not exist a θ' with $M_<(\theta|a) \cup M_=(\theta|a) \subset M_=(\theta'|a)$ for any $\theta \in \operatorname{supp} \beta^*(a)$ and $M_<(\theta'|a) \neq \emptyset$.

one needs to choose an equilibrium refinement that removes (at least some) equilibria with bunching at $a \neq \overline{a}$, but not bunching at $\overline{a}$. I do not know of a refinement that achieves this.

Alternatively, one can maintain the D1 criterion, but relax the single-crossing property. This is the main idea behind the model of Andreoni and Bernheim (2009), in which agents differ in how much they care about deviating from a normative action a^F. Higher types have a larger marginal utility for actions below a^F and a smaller (more negative) marginal utility for actions above a^F, violating the single-crossing property. If agents also want to signal being a high type, the model generates bunching at a^F.

My model is complementary to theirs, as our models generate opposite predictions. Take my setup with a bonus threshold and assume $a^F = \overline{a}$.[7] Their model vs. my model predicts that (i) more visibility increases bunching vs. decreases bunching, (ii) signaling creates missing mass above the bonus threshold vs. adds mass above the bonus, and (iii) increasing the bonus amount increases bunching and has no effect on the distribution above the threshold vs. no effect on bunching and an upward shift of distribution above the threshold (anti-bunching). Whether agents are signaling intrinsic motivation as in my model, or adherence to a social norm as in their model, ultimately depends on the context. The key insight is that even though our models both build on signaling, the opposing predictions make it possible to distinguish between them.

A few extension to my model are worth discussing. First, result 1 is likely to be counterfactual, as many settings will exhibit some bunching at the minimum action or the bonus threshold. One extension that seems like it could incorporate bunching, is to allow for heterogeneity in reputational concerns, with some types putting zero weight on reputation. These types bunch at the minimum action and the bonus threshold, and are unaffected by the presence of other types. However, with $\mu = 0$ an equilibrium ceases to exists.[8]

Another way to extend the model is to bound the action from above by requiring $a \leq a^{\max}$. This extension yields a new non-parametric prediction. If the upper bound is sufficiently low, then types will bunch at $a^{\max}$. Without signaling motives the bunching is simply a result of the type distribution being truncated, implying that the reputation function jumps up at $a^{\max}$. With signaling motives, agents slightly below $a^{\max}$ want to obtain this discrete reputation benefit at a marginal cost, and will move to $\overline{a}$, and thereby add to the bunching.

[7] Using their notation let $U(x, m, t) = F(x - c(x) + b\,\mathbb{1}\left\{\mathrm{x} \geq \mathrm{x}^{\mathrm{F}}\right\}, m) + t(G(x - x^F))$, where I added a convex cost of effort function c so that $U(x, t_0, t_0)$ attains its maximum at some $x_0 < x^F$.

[8] The single-crossing property implies that the reputation function is non-decreasing, so that for a given θ, a type with $\mu > 0$ chooses an action at least as large as the type with $\mu = 0$. The marginal buncher without reputational concerns $\theta_{1,\mu=0}$ is indifferent between some $a_{1,\mu=0} < \overline{a}$ and $\overline{a}$. What action $a_{1,\mu>0}$ does the marginal buncher with image concerns $\theta_{1,\mu>0}$ choose in equilibrium? If $a_{1,\mu>0} < a_{1,\mu=0}$, then the reputation function must jump up, which makes $a_{1,\mu>0} + \varepsilon$ a profitable deviation for $\theta_{1,\mu>0}$. Similarly if $a_{1,\mu>0} > a_{1,\mu=0}$, then the reputation function must jump down at $a_{1,\mu=0}$, which makes deviating to $a_{1,\mu=0}$ a profitable deviations for $\mu > 0$ types playing slightly above $a_{1,\mu=0}$. However, if $a_{1,\mu>0} = a_{1,\mu=0}$, then $\theta_{1,\mu=0} < \theta_{1,\mu>0}$ and so for $\theta_{1,\mu=0}$ to be indifferent, the reputational gain must offset the additional cost of effort. However, the reputational payoff at $\overline{a}$ is not a degree of freedom, because it is determined by the distribution of $\mu = 0$ types and R. Hence, an equilibrium exists only if the exogenous parameters from the model and type distribution happen to line up exactly.

This dynamic is similar to the two action model in Bénabou and Tirole (2006), with the key difference that with more types moving to $\overline{a}$ the reputational gain will always go down for the marginal buncher, preventing an unraveling where all types end up bunching at $\overline{a}$. The new prediction is that only signaling motives induce a missing mass below $a^{\max}$.

The model also handles alternative bunching incentives. If the bonus is paid out to the charity instead of the agent, or if the contribution schedule kinks, for example due to a third party matching contributions up to a threshold, then models of reciprocity and outcome based altruism still predict bunching, and signaling models still yield results 1 to 3. Going the extra mile shows that one is willing to contribute despite the lower marginal benefit to charity, and hence signals a higher type than those who bunch at the threshold.

Simulation

Is anti-bunching meaningfully large? To get a sense for the magnitude, I simulate changes in reputational concern μ and the bonus amount b under a disciplined parametric setup. The real effort task that I use in chapter 2 builds on Augenblick et al. (2015), and so I combine their functional form assumption for the cost of effort with the incentive structure from my experimental design. Specifically, I assume $g(\theta, a) = \theta a y_s - k(a + \underline{a})^\gamma$, where $\gamma = 1.774$ is the cost of effort parameter estimate from Table V in Augenblick et al., and $\underline{a} = 5$ and $y_s = 8$ cents come from my experimental design. k is not identified in Augenblick et al., and so I calibrate $k = 0.7$ such that under all three bonus levels $b \in \{0, 40, 120\}$ types with $\mu = 0$ do not all bunch at the same action.

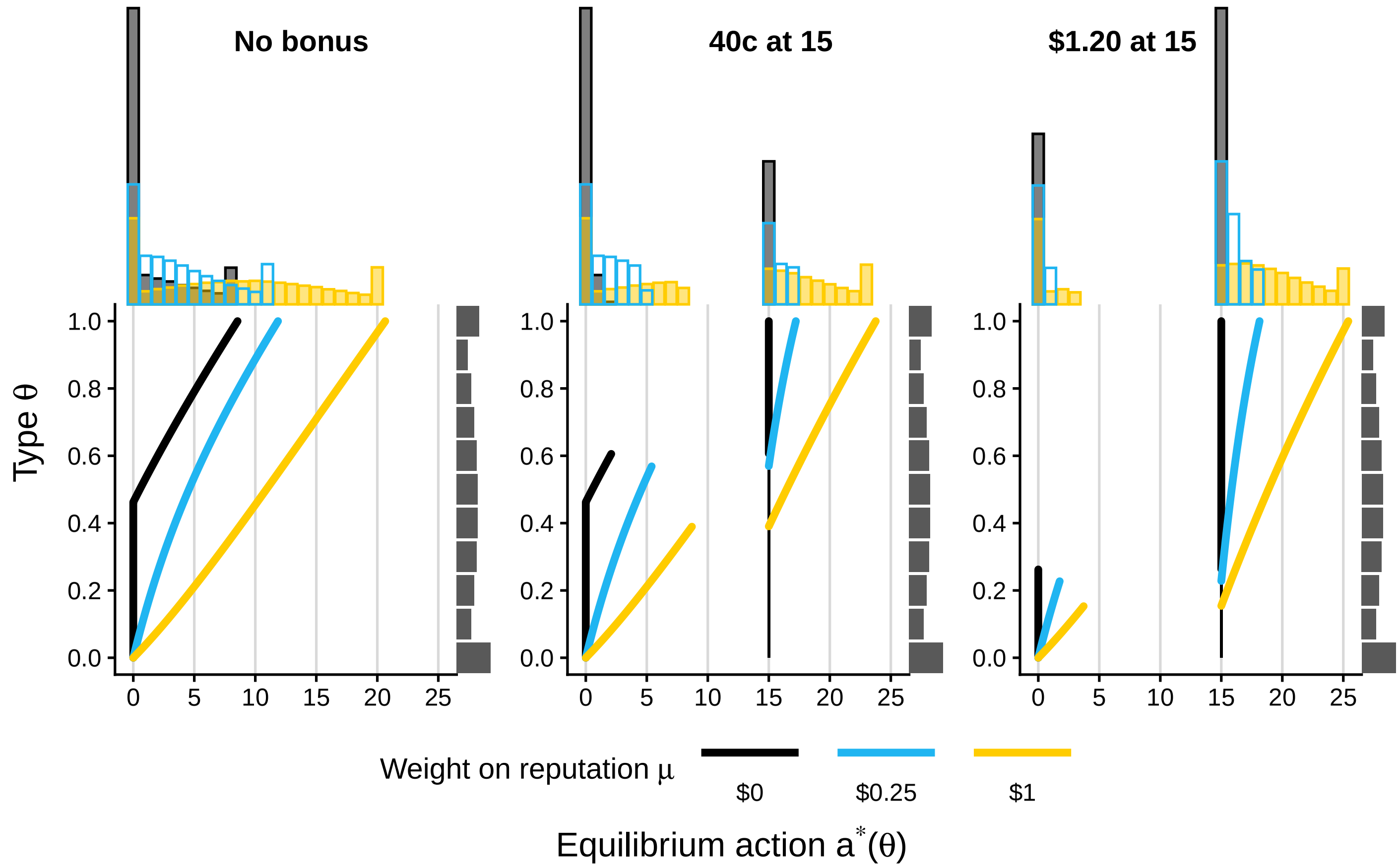
When choosing their effort, an agent of type θ sees a piece-rate of \$1 to charity as equivalent to a piece-rate of \$$\theta$ to themselves. I assume that θ ranges from $\underline{\theta} = 0$ to $\overline{\theta} = 1$, reflecting that types do not have any spite toward the charity, but are also not more motivated to work for charity than for themselves. The maximum reputational gain is $\mu(\overline{\theta} - \underline{\theta})$, and therefore μ measures an agent's valuation of increasing their reputation from purely self-interested to equally caring about charity and self. I choose $\mu = \$0.25$ and $\mu = \$1$ as two plausible, exemplary values.[9, 10]

Figure 1.1 shows the result of the simulation exercise. Three points are worth noting. (i) A modest reputational concern of $\mu = \$0.25$ is sufficient to reduce bunching at the minimum action and the bonus threshold, (ii) the amount by which types increase their action due to reputational concerns is possibly non-monotone in type (iii) reputational concerns have little effect on very low types, implying that with discretized action some bunching at the minimum action and the bonus threshold persists.

[9] To be consistent with the discrete nature of the experimental task, I round down the equilibrium action implied by the continuous model. Rounding down yields more bunching than implied by the continuous model.

[10] When simulating an outcome distribution, I assume that types are drawn from a censored Normal distribution with $\theta_i^* \sim \mathcal{N}\left(0.4, \sqrt{0.4}\right)$ and $\theta_i = \max\{\underline{\theta}, \min\{\overline{\theta}, \theta_i^*\}\}$.

Figure 1.1: Simulation of equilibrium behavior



Note: This figure illustrates the magnitude of anti-bunching under a parametric setup. The lines in the lower part of each panel denote the equilibrium action profile with varying bonus amounts and weights on reputation. The bars on the right side of each panel are a histogram of the assumed distribution of types, $\theta_i = \max\{0, \min\{1, \theta_i^*\}\}$ with $\theta_i^* \sim \mathcal{N}\left(0.4, \sqrt{0.4}\right)$. Combining the equilibrium action profile and the type distribution yields an observable outcome distribution shown as histogram on the top of each panel. The simulation assumes $U(a;\theta) = \theta a y_s - k(a+\underline{a})^\gamma + \mu r(a) + b\,\mathbb{1}\{a \geq \overline{a}\}$, with $\gamma = 1.774$ coming from Augenblick et al. (2015), $\underline{a} = 5$ tasks, $\overline{a} = 15$ tasks, $y_s = 8$ cents and $b \in \{0, 40, 120\}$ cents based on my experimental design, and $k = 0.7$ calibrated such that the action profile under $\mu = 0$ is non-degenerate for all three bonus levels. The weight on reputation μ measures in cents an agent's valuation of increasing their reputation from purely self-interested to equally caring about charity and self.

1.3 Conclusion and Transition to Chapter 2

This chapter has introduced a new test for signaling motives in prosocial behavior. The test relies on the idea that agents do not want to be lumped together with low typed agents whose primary motivation is the receipt of personal benefits.

Interestingly, none of the results are restricted to the setting of prosocial behavior. Whenever agents face notched or kinked incentives, signaling motives can reduce the amount of bunching and induce anti-bunching. Thereby anti-bunching has the potential to serve as a test for signaling motives, inform optimal incentive design, and explain a lack of bunching in a wide variety of settings.

This chapter has provided the theoretical foundation for anti-bunching, and validated its economic significance with a simulation exercise based on a calibrated parametric model. In chapter 2, I provide evidence that anti-bunching can have economically meaningful effects in a controlled experimental environment. This is an important step toward testing whether anti-bunching can serve as a meaningful test for signaling models in practice.

Chapter 2

Anti-Bunching: Evidence from an Online Experiment

2.1 Introduction

In chapter 1, I have introduced anti-bunching as a new test for signaling motives. In the following chapter, I will provide evidence that anti-bunching can have economically meaningful effects in a controlled experimental environment.

2.2 Design

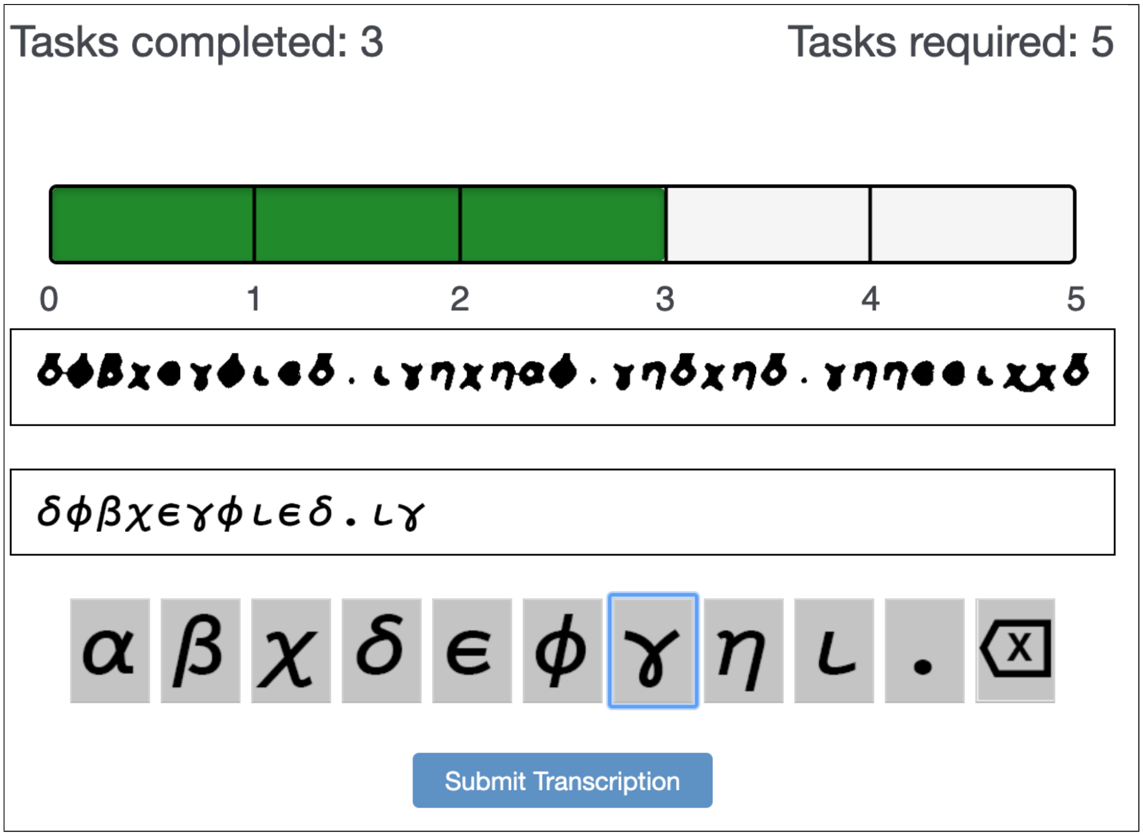
A good experimental environment for anti-bunching has (a) a granular action that is informative about an attribute agents want to signal, (b) a manipulable personal bonus incentive, (c) a population large enough to have statistical power to detect medium sized changes in contribution levels, and optionally (d) manipulable visibility or care about reputation.

Task At the center of my experimental design is a real-effort task, which I adopt in slightly modified form from Augenblick et al. (2015). A task consists of transcribing 38 Greek letters from an image, which takes around 50 seconds to complete. At the beginning of the experiment, participants become familiar with the task by completing 5 transcriptions without any benefit to themselves or charity.

In the main part of the experiment, participants raise 8c per task for charity, making the task prosocial. Participants can choose which of six well-known charities they want to support.[1] The choice of charity occurs before any treatment assignment and remains private.

[1]In the US sample the charities are the American Cancer Society, the American Red Cross, Feeding America, the Salvation Army, Wikipedia and WWF. In the UK sample the American Cancer Society, the American Red Cross and Feeding America are replaced by the British Red Cross, Cancer Research UK, and the Trussel Trust.

Figure 2.1: Transcription task



Note: A task consists of transcribing 38 Greek letters from an image. At the beginning of the experiment, participants become familiar with the task by completing 5 transcriptions without any benefit to themselves or charity.

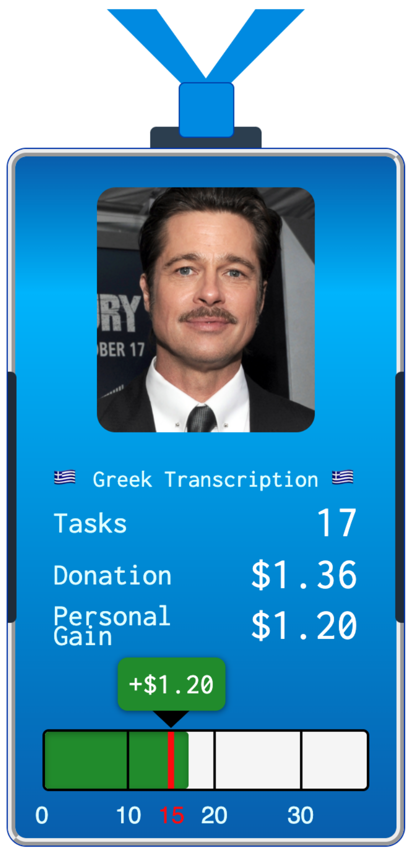
Bonus Incentive In the control group, participants are not offered any bonus incentive. I add two treatment groups with a bonus incentive, one with a small and one with a large bonus amount. A bonus incentive allows me to test the main prediction of anti-bunching. Having treatment groups with different bonus amounts allows me to test the model's prediction that a larger bonus amount induces stronger anti-bunching effects. Both bonus amounts should be large enough so that some, but not all, participants are willing to increase their effort in order to get the bonus. Based on pilots, I choose 15 tasks as the bonus threshold, and 40c and $1.20 as the two bonus amounts.

The comparison of the $1.20-bonus treatment with the 40c-bonus treatment is possibly confounded by gift exchange: Participants might provide more effort when being offered a larger bonus amount, because they reciprocate to the experimenter for receiving a higher wage. To investigate this channel I add a bonus treatment in which participants receive a lump-sum 80c gift for participation, in addition to a 40c-bonus at 15 tasks. Taking the 40c-bonus as baseline, the gift-exchange treatment offers the same 80c increase as the $1.20

treatment, but at an easier to achieve threshold of 0 tasks. Therefore, any effects of gift exchange on effort are at least as strong in the gift-exchange treatment as they are in the $1.20 treatment.

Visibility A participant can be signaling their type to themselves or to the experimenter. To amplify the signaling motive, and to test the comparative static of the model with respect to the signaling motive, I cross-randomize the four bonus treatment groups with two visibility treatment groups.

Figure 2.2: Badge



Note: This figure shows the personal badge that participants create in the *Badge* treatment, with an example image. The badge is displayed to the participant during the task (see fig. 2.3) and to other participants during the first judgement step (see fig. 2.4). Image by DoD News Features (https://commons.wikimedia.org/wiki/File:Brad_Pitt_Fury_2014.jpg), "Brad Pitt Fury 2014", cropped by the author, https://creativecommons.org/licenses/by/2.0/legalcode.

In the treatment group *Badge* participants are asked to create a personal *Badge*, as shown in fig. 2.2. The badge displays tasks completed, total donation amount raised, total personal gain and the bonus incentive scheme, together with a picture of the participant that they take using their webcam. After completing the experiment, each participant's badge is shown to at least one other participant, who is then asked to judge the participant's generosity. Participants are made aware of this when creating the badge as well as during the transcription task.

In the treatment group *No badge* participants do not create a badge and all information about their contribution remains private. The difference between the screens participants see during the task in *Badge* and *No badge* is shown in fig. 2.3.

The *Badge* treatment asks participants to record a picture of themselves using their webcam. Some participants might exit or change their effort in response to being asked to take a picture of themselves using their webcam. To control for this channel, the experiment includes a human verification step in both *Badge* and *No badge*, in which all participants are asked to take a picture of themselves for human verification purposes.[2] When asked to take a picture for the badge, participants can then choose if they want to re-use the picture they took in the human verification step, or take a new picture. I assume that conditional on having taken one picture for human verification purpose, being asked to re-use or take another picture is independent of eventual effort.

Perceived Reputation All badges created by participants in *Badge* are shown to at least one other participant. To make this displaying of badges more meaningful, and to get a measure of the (believed) reputation function, the badges are embedded in two judgement steps, which occur after participants complete the work for charity.

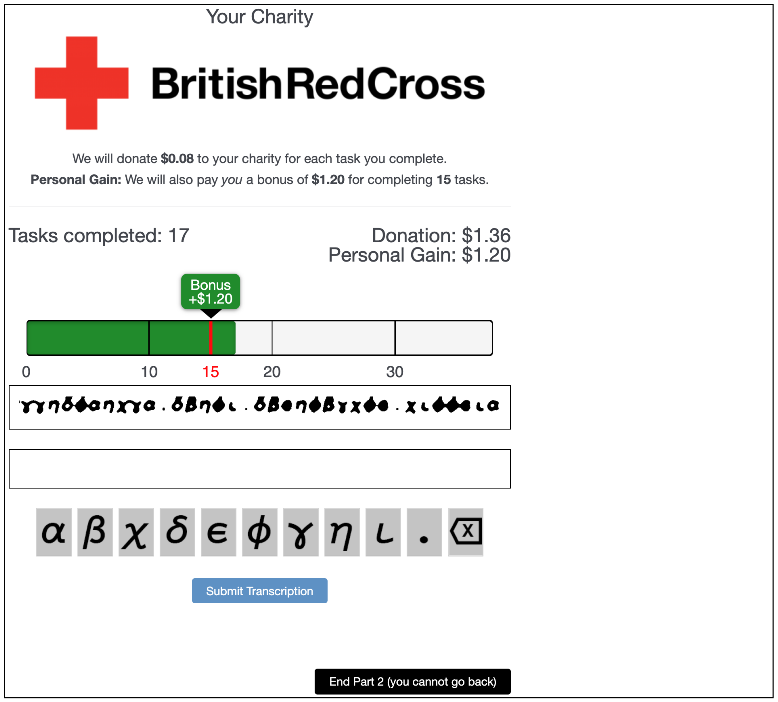
In the first judgement step all participants consecutively see 5 pairs of badge and answer for each pair "Who is more generous?", as shown in fig. 2.4. The badges reflect actual outcomes by other participants who were assigned to the *Badge* treatment in the same experiment or a previous pilot study. In most cases the badge is shown without the participant's image, but each participant's image is shown at least once to another participant.

The model predicts that participants increase their effort, because they want to be seen as generous. To get a direct measure of the (believed) relationship between action and perceived generosity, in the second judgement step all participants see 5 pairs of badges and answer for each pair "Who do most other participants say is more generous?". Participants earn a \$0.50 bonus for answering this question correctly for at least 4 pairs.

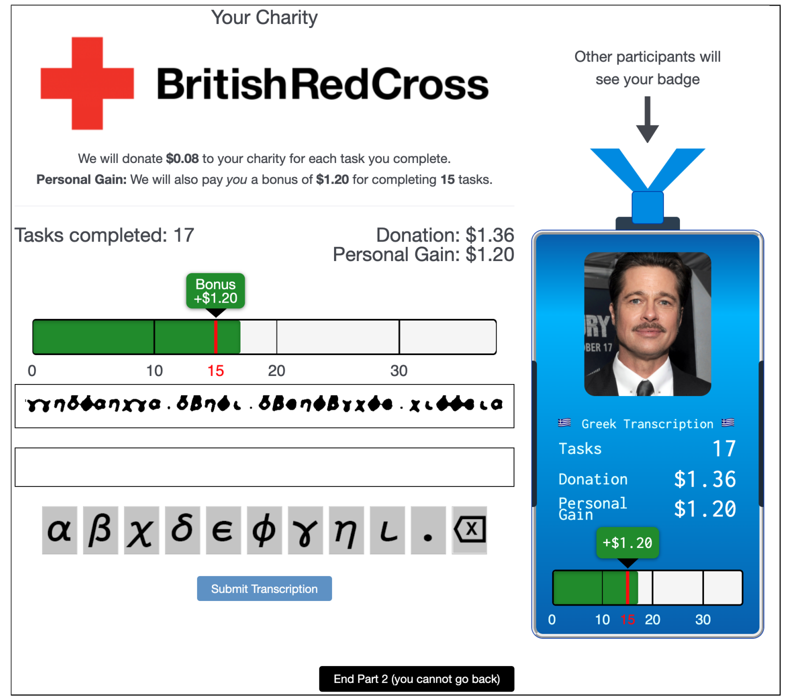
Consent Forms Showing a participant's webcam images to another participant means sharing sensitive personal data, and therefore requires explicit consent. Therefore, in coordination with the Berkeley IRB, the experiment includes a second consent form that is shown to participants just before creating the badge and working for charity. In the *Badge* treatment group this consent form asks for explicit consent to continue with the study and share the badge including the participant's image with other participants and the research community. To control for differential attrition, I also include a similar looking consent form

[2] I exclude 82 observations in which the verification picture does not show a human face.

Figure 2.3: Completing tasks for charity



(a) *No badge* × 40c-bonus treatment



(b) *Badge* × 40c-bonus treatment

Note: This figure shows the screen shown to participants when completing tasks for charity in the 40c-bonus treatment and the two *visibility* treatments. The bonus threshold is marked with a red line. When the participant completes the 15th task, the background color in the box above red line changes from white to green, indicating that they have earned the bonus. Image by DoD News Features (https://commons.wikimedia.org/wiki/File:Brad_Pitt_Fury_2014.jpg), "Brad Pitt Fury 2014", cropped by the author, https://creativecommons.org/licenses/by/2.0/legalcode.

in the *No badge* treatment group that asks participants for explicit consent to continue the study, but does not mention the badge.

In sum, the experiment consists of the following steps: (1) provide consent for participation, (2) answer demographic questions, (3) take a verification picture using webcam, (4) complete 5 tasks without any benefit to participant or charity, (5) choose one of six charities, (6) provide consent to continue participation (differs across *visibility*), (7) create a badge using picture from (3) or taking a new picture (differs across *visibility*), (8) choose to complete 0 to 38 additional tasks, raising \$0.08 per task for the charity chosen in step (5) and earning a personal bonus for completing 15 tasks (differs across *visibility* and *bonus*) (9) answer "Who is more generous?" for 5 pairs of badges, (10) answer "Who do most other participants say is more generous?" for 5 pairs of badges, incentivized with a \$0.50 bonus for answering correctly at least 4 times, (11) provide questions, comments or concerns.

I recruit participants in October 2019 using the online platform Prolific. Using the platform, I pre-screen participants to reside either in the US or UK, and to have explicitly stated a willingness to record themselves with a webcam.[3]

2.3 Results

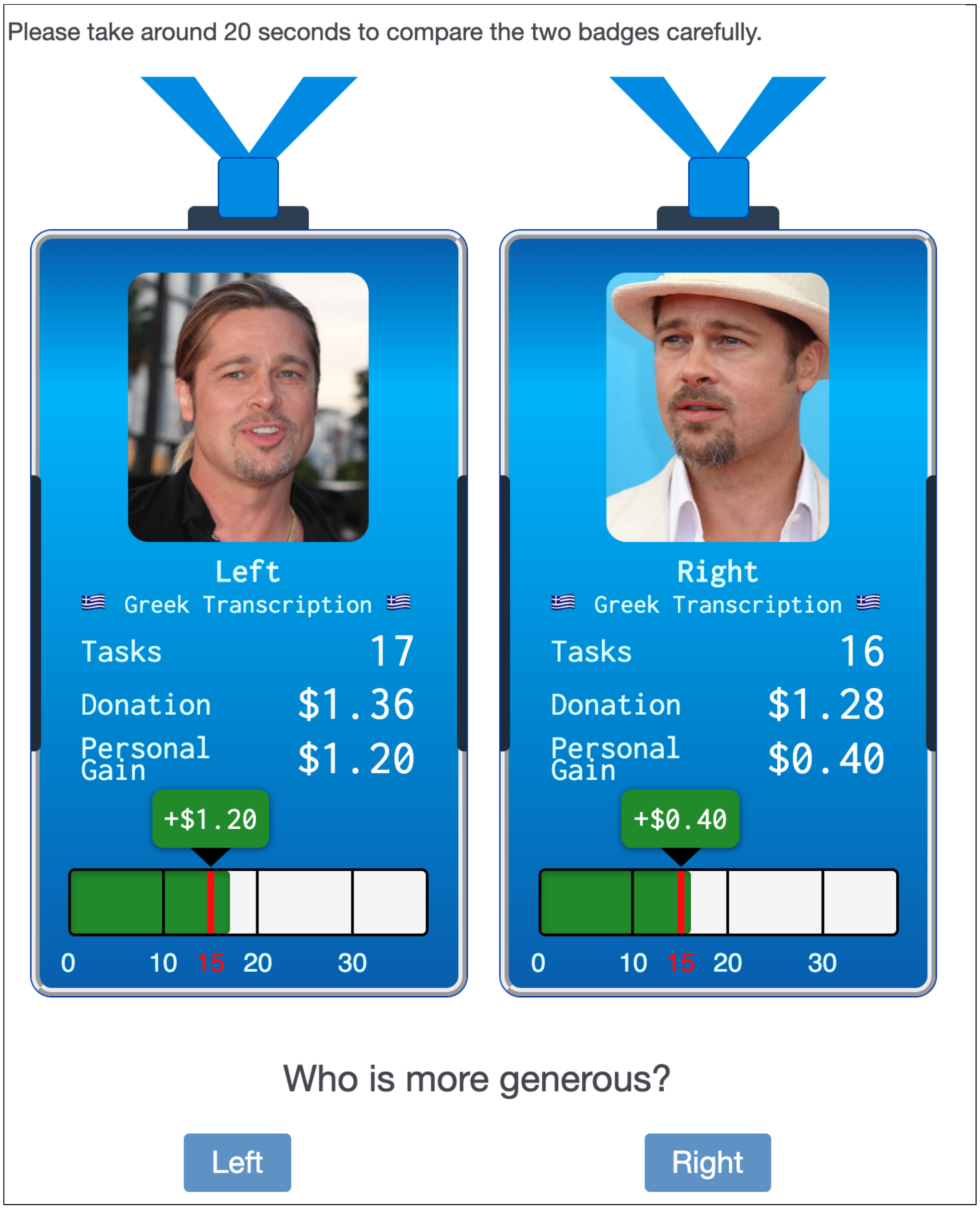
Baseline Heterogeneity Figure 2.5 and table 2.2 summarize the results for each of the eight treatment groups. In the control group with no badge and no bonus, there is substantial heterogeneity in the extent to which participants are willing to perform the task for charity. Participants complete 9.6 tasks on average, 14.6 % of participants complete 0 tasks, 1.4 % complete 15 or 16 tasks, and 14.6 % complete the maximum of 38 tasks, working more than 30 minutes for the benefit of their chosen charity and with no personal benefit. This suggests that the task has a meaningful prosocial component and participants exhibit considerable heterogeneity in intrinsic motivation.

40c-Bonus How does a 40c-bonus incentive affect the distribution of effort? In the *No badge* condition a participant's behavior remains private. Under this condition a 40c-bonus increases the share of participants completing at least 15 tasks by 31.9 pp from 19.8 % to 51.7 %. Using the definitions pre-specified in my pre-analysis plan, I decompose the 31.9 pp increase into a 28.0 pp increase in bunching, completing 15 or 16 tasks, and a 3.9 pp increase in anti-bunching, completing 17 or more tasks. Taking into account that 1.4 % complete 15 or 16 tasks without a bonus, this indicates that 8 % to 12 % of the population complying with a small bonus incentive behave consistent with self-signaling or signaling to the experimenter.[4]

[3] In the UK sample, I additionally require a minimum of 10 completed surveys and a minimum approval rate of 95%.

[4] Two groups can cause an increase in the share of participants completing 17 or more task: Participants previously completing 15 or 16 tasks, and participants previously completing less than 15 tasks. The first group comprises only 1.4 % of participants, and so the share of types exhibiting signaling motives among the

Figure 2.4: Judgement of badges



Note: Participants are asked to judge 5 pairs of badges by answering "Who is more generous?" Each badge reflects an actual outcome by another participant who was assigned to the *Badge* treatment in the same experiment or a previous pilot study. In most cases the badge is shown without the participant's image, but each participant's image is shown at least once to another participant. Left image by Eva Rinaldi creator QS:P170,Q37885816 (https://commons.wikimedia.org/wiki/File:Brad_Pitt_(8993538073).jpg), "Brad Pitt (8993538073)", cropped by the author, https://creativecommons.org/licenses/by-sa/2.0/legalcode. Right image by Thomas Peter Schulz (https://commons.wikimedia.org/wiki/File:BradPittBAR08.jpg), "BradPittBAR08", rotated by the author, https://creativecommons.org/licenses/by-sa/3.0/legalcode.

In the *Badge* condition a participant's picture and behavior are made visible to other participants. However, participants are unlikely to know each other or interact with each other in the future, because the study is conducted online and draws from a nation-wide pool of participants. Therefore participants remain considerably less visible than in an experimental laboratory or in many field settings of prosocial behavior. Nonetheless, the badge increases average effort in the no bonus treatment by 0.7 tasks, which is around $0.7/2.7 \approx 26\,\%$ of the effect size of introducing a 40c-bonus for completing 15 tasks.

As the badge makes behavior more visible, the model predicts an increase in anti-bunching. The experimental data confirms this. In the *Badge* condition introducing a 40c-bonus increases the share of participants completing 15 or more tasks from 22.1 % to 54.9 %. This 32.8 pp increase reflects a 23.4 pp increase in bunching and a 9.4 pp increase in anti-bunching. This implies that when being visible to other online participants 25 % to 29 % of the participants who comply to a bonus incentive exhibit signaling motives strong enough to complete at least 2 tasks in addition to what the bonus incentive requires.[5]

$1.20-Bonus Does increasing the bonus amount from 40c to $1.20 amplify these effects? In the *No badge* condition the $1.20-bonus leads to 62.3 % of participants completing 15 or more tasks, an increase from 19.8 % and 51.7 % without a bonus and with a 40c-bonus, respectively. Compared to the 40c-bonus, the increase in anti-bunching is stronger in absolute terms, 6.1 pp compared to 3.9 pp, and similar in proportion to the increase in participants completing 15 or more tasks, $6.1/(62.3-19.8) \approx 14\,\%$ compared to $3.9/(51.7-19.8) \approx 12\,\%$.

In the *Badge* condition the $1.20-bonus generates the largest anti-bunching effects observed in the experiment. The share of participants completing 17 or more tasks increases by 11.8 pp from 21.0 % to 32.8 %. This increase is accompanied by an increase in the share of participants completing 15 or more tasks from 22.1 % with no bonus to 65.8 % with a bonus. Taken together, this implies that 24 % to 27 % of the participants who comply to a bonus incentive exhibit signaling motives strong enough to complete at least 2 tasks in addition to what the bonus incentive requires, matching the earlier 25 % to 29 % estimate from the 40c-bonus treatment.[6]

Figure 2.6 summarizes the main treatment effects on bunching and anti-bunching visually. Note how in the in the top panel, which summarizes the share of participants doing at least 15 tasks, the line for *Badge* is above the line for *No badge*. This indicates the baseline visibility effect, which is constant across bonus levels as the lines increases in parallel.

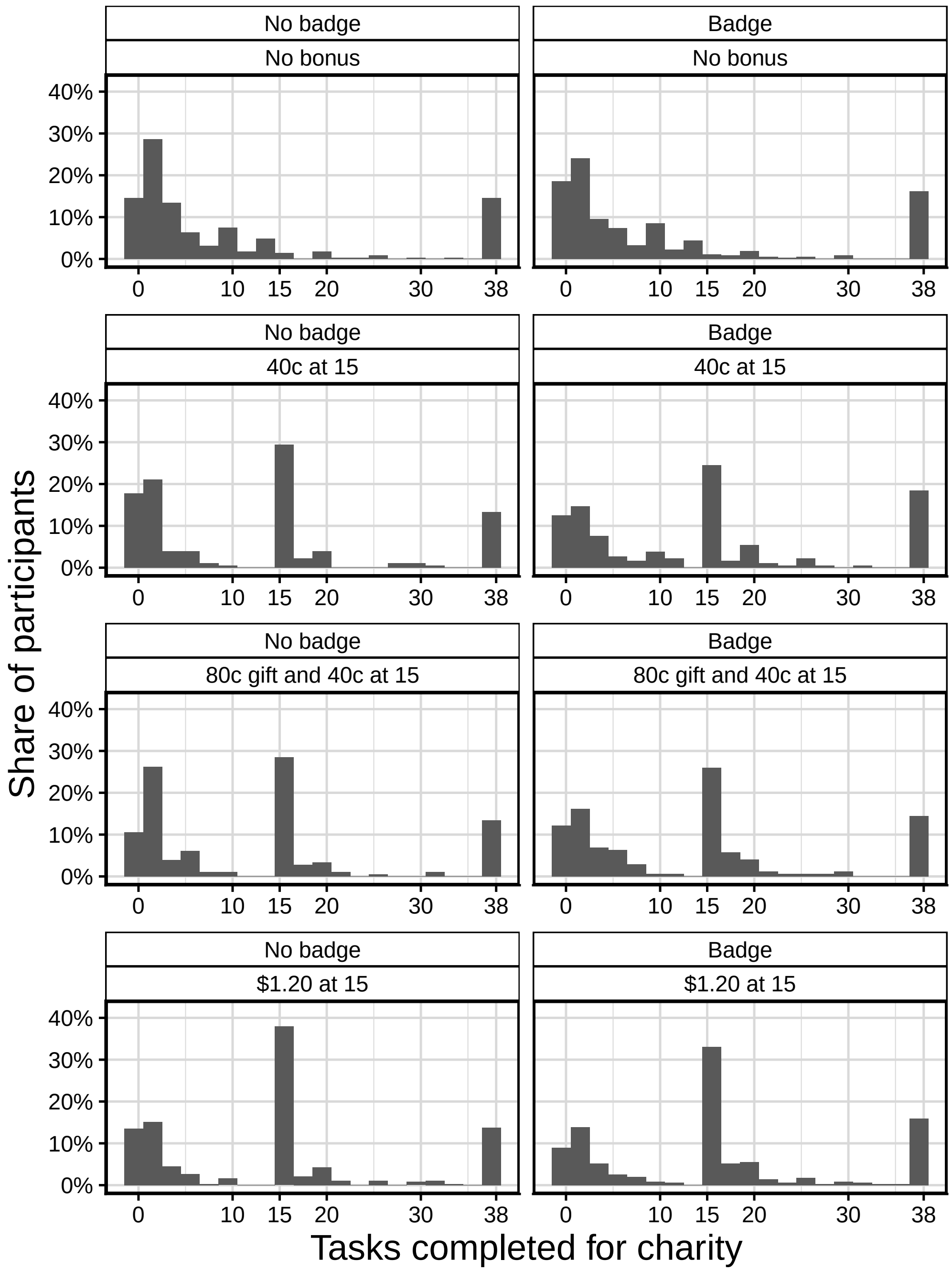
In contrast, in the center panel, which summarizes bunching at 15 or 16 tasks, the line for *Badge* is below the line for *No badge*. This indicates the negative effect of visibility on bunching. For the bottom panel, which summarizes the share of participants doing 17 or more tasks, most models of social preferences would predict flat lines. However, the lines are

second group must be between $(3.9-1.4)/31.9 \approx 8\,\%$ and $3.9/31.9 \approx 12\,\%$. Note that an increase caused by either group constitutes evidence for the model.

[5]In the *Badge* condition with no bonus 1.1 % of participants complete 15 or 16 tasks, yielding $(9.4-1.1)/32.8 \approx 25\,\%$ and $9.4/32.8 \approx 29\,\%$. See footnote 4 for more details.

[6]$(11.8-1.1)/(65.8-22.1) \approx 24\,\%$ and $11.8/(65.8-22.1) \approx 27\,\%$, see footnote 4 for more details.

Figure 2.5: Histogram of effort for each treatment group



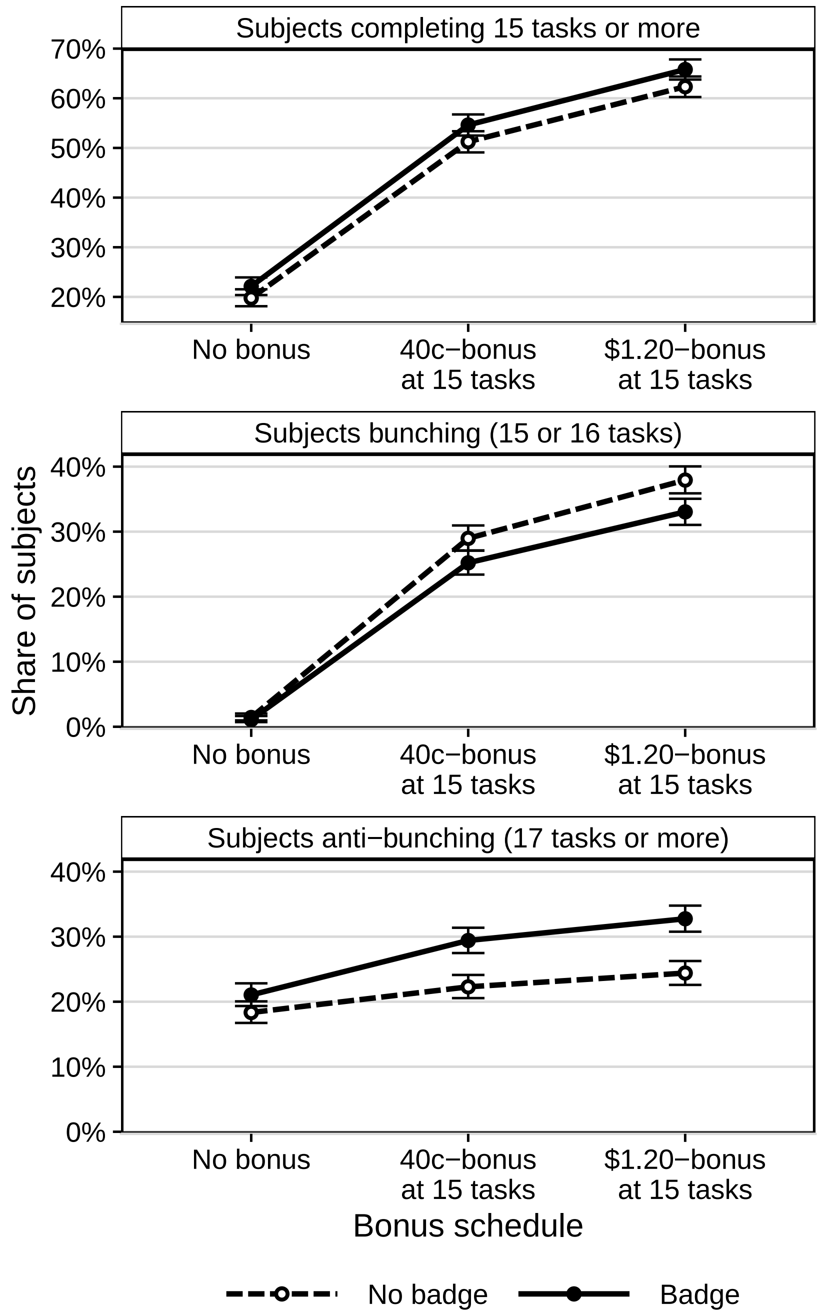
Note: This figure shows the histogram of tasks completed for charity for each of the eight treatment groups. In each treatment group, participants choose to complete up to 38 transcription tasks, raising 8c per task for charity. In rows 2 to 4, participants additionally earn a bonus for completing 15 or more tasks. In the left column participant behavior is private. In the right column, participant behavior is recorded on a digital badge, that is then shown to other participants. The bins are $\{\{0\}, \{1, 2\}, \{3, 4\}, \ldots, \{15, 16\}, \ldots \{37, 38\}\}$.

Table 2.1: Summary of effort

				Main partition of effort 4 bins, %-share in each bin				Alternative partition 4 bins, %-share in each bin			
	N	Mean effort	%-share 15–38	0–1	2–14	15–16	17–38	0	1–14	15	16–38
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Full sample	2,153	12.7	46.0	27.1	26.9	21.4	24.7	13.7	40.2	14.1	31.9
No badge											
All	1,085	12.1	45.0	29.5	25.5	23.2	21.8	14.1	40.9	15.7	29.3
No bonus	349	9.6	19.8	35.5	44.7	1.4	18.3	14.6	65.6	1.4	18.3
40c at 15	180	12.3	51.7	30.6	17.8	29.4	22.2	17.8	30.6	21.1	30.6
$1.20 at 15	377	14.1	62.3	23.3	14.3	37.9	24.4	13.5	24.1	24.1	38.2
80c gift and 40c at 15	179	12.3	50.8	29.6	19.6	28.5	22.3	10.6	38.5	20.1	30.7
Badge											
All	1,068	13.3	47.1	24.6	28.3	19.5	27.6	13.4	39.5	12.5	34.6
No bonus	366	10.3	22.1	32.2	45.6	1.1	21.0	18.6	59.3	0.8	21.3
40c at 15	184	14.7	54.9	21.2	23.9	24.5	30.4	12.5	32.6	16.3	38.6
$1.20 at 15	345	15.6	65.8	18.6	15.7	33.0	32.8	9.0	25.2	22.3	43.5
80c gift and 40c at 15	173	13.4	54.3	24.3	21.4	26.0	28.3	12.1	33.5	13.9	40.5

Note: This table summarizes the distribution of effort, the number tasks completed for charity, for each of the eight treatment groups. Participants raise 8c per task for charity. The experimental design crosses two visibility conditions, *No badge* and *Badge*, with four bonus incentive conditions, no bonus, 40c-bonus, $1.20-bonus, and a combination of an 80c-gift and a 40c-bonus, where in all conditions the bonus is paid to a participant for completing 15 tasks or more. Columns 1 and 2 display the number of participants and average effort. Column 3 displays the share of participants completing 15 tasks or more. Columns 4 to 7 partition effort into four bins and display the share of participants (in %) falling into each bin. As a robustness check, columns 8 to 11 repeat the exercise using an alternative partition. The main partition and the alternative partition for the robustness check were both pre-specified as such.

Figure 2.6: Main effects on bunching and anti-bunching



Note: This figure decomposes the observed share of participants doing at least 15 tasks into those doing 15 or 16 tasks, and those doing 17 or more tasks, by treatment group. The dashed line corresponds to treatment arms in which a participant's action remains private, the solid line to treatment arms in which a participant's action is made visible through the badge. In this figure *40c-bonus at 15 tasks* pools the two treatment arms 40c-bonus at 15 tasks, and 80c-gift and 40c-bonus at 15 tasks.

increasing, with the line for *Badge* increasing more than the line for *No badge*, indicating anti-bunching consistent with signaling models.

Hypothesis Testing—Main Effects Following the pre-analysis plan, I now conduct a series of hypothesis tests to assess the statistical significance of the results. I estimate three regressions of the form

$$\mathbb{1}\{\mathrm{Effort_i} \in \mathrm{E_k}\} = \alpha + \beta\,\mathrm{Bonus}_i + \gamma\,\mathrm{Badge}_i + \delta\,\mathrm{Bonus}_i \times \mathrm{Badge}_i$$

where Effort_i is the number of tasks participant i completes for charity, $E_k \in \{\{0,1\},\{15,16\},\{17,\ldots,38\}\}$, Badge_i is a dummy variable indicating if participant i is in the *Badge* treatment arm, Bonus_i is a vector of three dummy variables indicating the participant's bonus incentive, and $\beta = (\beta_{\text{40c at 15}}, \beta_{\text{80c gift + 40c at 15}}, \beta_{\text{\$1.20 at 15}})^T$. I compute Eicker-Huber-White standard errors and test against the pre-specified null hypotheses using t-tests.

To increase statistical power, I repeat this procedure using two alternative definitions of Bonus_i, one in which the 40c-bonus treatment is pooled together with the 80c-gift and 40c-bonus treatment, and another one in which in which all three bonus treatments are pooled together. I denote coefficients from these two alternatives with a *pool-*40c*-bonus* and a *pool-any-bonus* superscript. Finally, when investigating the effect of *Badge* on bunching at the minimum action, I also estimate a regression without Bonus_i, and denote the coefficient on Badge_i as $\gamma^{\text{pool-no-bonus-and-any-bonus}}$.

Table 2.3 summarizes the results of the pre-specified series of hypothesis tests on the main treatment effects.

Hypothesis A states that the badge decreases bunching at the bottom effort level (0 or 1 tasks). Pooling across bonus treatment arms, the regression estimates that the badge decreases bunching at the bottom by 4.9 pp ($p = 0.011$) from a baseline of 35.5 %. When estimating the effect of the badge separately by bonus treatment, the estimates are negative and of similar magnitude in all bonus treatment groups but only the effect in the 40c-bonus is statistically significant ($p = 0.041$)

Hypothesis B states that without a badge, the bonus incentives induces bunching at the bonus threshold. This hypothesis only serves as a check that the bonus incentives are calibrated well enough so that the it attracts some, but not all participants. Without any bonus incentive, the baseline share of participants at the bonus threshold is 1.4 %. Introducing a bonus increases this share by least 27.1 pp to 36.5 pp ($p < 0.01$), showing that the bonus incentive works as intended.

Hypothesis C states that with a badge, the bonus incentives induce anti-bunching above the bonus threshold. Without any bonus incentive, the baseline share of participants above the bonus threshold is 21.0 %. A bonus incentive increases this share by 10.0 pp ($p < 0.01$), when pooling across bonus amounts. The estimates of the separate effects are 9.4 pp (40c-bonus, $p = 0.019$), 7.3 pp (80c-gift and 40c-bonus, $p = 0.071$), 11.7 pp ($1.20-bonus, $p < 0.01$). In sum, the data shows strong support for hypothesis C.

Table 2.3: Hypothesis testing, main effects

Regression $\mathbb{1}\{\text{Effort}_\text{i} \in \text{E}_\text{k}\} = \alpha + \beta\,\text{Bonus}_i + \gamma\,\text{Badge}_i + \delta\,\text{Bonus}_i \times \text{Badge}_i$				
Hypothesis	Parameter (measured in pp) (1)	Estimate (2)	s.e. (3)	p-value (4)
A	The badge reduces bunching at the bottom. Dependent Variable: $\mathbb{1}\{\text{Effort}_\text{i} \in \{0,1\}\}$			
	α	35.5	2.6	0.000***
	$\gamma^{\text{pool-no-bonus-and-any-bonus}}$	−4.9	1.9	0.011**
	γ	−3.3	3.5	0.353
	$\gamma + \delta_{\text{40c at 15}}$	−9.4	4.6	0.041**
	$\gamma + \delta_{\text{80c gift and 40c at 15}}$	−5.3	4.7	0.259
	$\gamma + \delta_{\text{\$1.20 at 15}}$	−4.8	3.0	0.113
B	Without a badge, a bonus induces bunching at the bonus threshold. Dependent Variable: $\mathbb{1}\{\text{Effort}_\text{i} \in \{15,16\}\}$			
	α	1.4	0.6	0.024**
	$\beta^{\text{pool-any-bonus}}$	32.1	1.9	0.000***
	$\beta_{\text{40c at 15}}$	28.0	3.5	0.000***
	$\beta_{\text{80c gift and 40c at 15}}$	27.1	3.4	0.000***
	$\beta_{\text{\$1.20 at 15}}$	36.5	2.6	0.000***
C	With a badge, a bonus induces anti-bunching above the bonus threshold. Dependent Variable: $\mathbb{1}\{\text{Effort}_\text{i} \in \{17,\ldots,38\}\}$			
	$\alpha + \gamma$	21.0	2.1	0.000***
	$[\beta + \delta]^{\text{pool-any-bonus}}$	10.0	2.8	0.000***
	$[\beta + \delta]_{\text{40c at 15}}$	9.4	4.0	0.019**
	$[\beta + \delta]_{\text{80c gift and 40c at 15}}$	7.3	4.0	0.071*
	$[\beta + \delta]_{\text{\$1.20 at 15}}$	11.7	3.3	0.000***

Note: This table summarizes the results from pre-specified hypothesis tests about the main treatment effects on different parts of the effort distribution. The hypothesis tests use linear combinations of the parameter estimates from the regression equation $\mathbb{1}\{\text{Effort}_\text{i} \in \text{E}_\text{k}\} = \alpha + \beta\,\text{Bonus}_i + \gamma\,\text{Badge}_i + \delta\,\text{Bonus}_i \times \text{Badge}_i$, where Bonus_i is a vector of three dummy variables indicating the participant's bonus incentive, and $\beta = (\beta_{\text{40c at 15}}, \beta_{\text{80c gift + 40c at 15}}, \beta_{\text{\$1.20 at 15}})^T$. The first row in each panel indicates the baseline value. The superscript *pool-no-bonus-and-any-bonus* and indicates that the parameter is estimated using a regression that omits the Bonus_i variables. The superscript *pool-any-bonus* indicates that the parameter is estimated using a regression in which Bonus_i is scalar dummy variable that pools the three treatment arms 40c-bonus, 80c-gift and 40c-bonus, and $1.20-bonus. I use Eicker-Huber-White standard errors to conduct two-sided t-tests.

Hypothesis Testing—Interaction Effects While Hypotheses A–C deal with the main effects of the model, the model also makes predictions on the comparative statics of these effects, which I can test utilizing the interaction of treatments in the experimental design. Table 2.4 summarizes the results of the pre-specified series of hypothesis tests on these interaction treatment effects.

Hypothesis D tests if an increase in signaling motives reduces bunching at the bonus threshold, by stating that a bonus induces less bunching with a badge than without a badge. The estimated reduction in bunching at the bonus threshold is 4.2 pp ($p = 0.107$) across bonus treatments, however this interaction effect is not statistically significant. When estimating the effect of the badge separately by bonus treatment, the estimates remain all negative and not statistically significant. In sum, the sign of the point estimate is in line with Hypothesis D, but the estimate is not statistically significant.

This pattern persists throughout the remaining hypotheses on interaction effects. Hypothesis E tests the same idea as Hypothesis D, but instead of looking at bunching at the bonus threshold it investigates anti-bunching above the bonus threshold. The estimates all have the right sign, but are not statistically significant.

While the previous hypotheses make a statement on the difference between some bonus incentive and no bonus incentive, Hypotheses F and G investigate an increase in the bonus amount from 40c to \$1.20. Hypothesis F states this increase in the bonus amount increases anti-bunching above the bonus threshold. The point estimates again go in right direction, but are not statistically significant. Hypothesis G states that the increase in anti-bunching described in Hypothesis F is bigger with a badge than without a badge. Again all point estimates go in the right direction, but all three comparisons remain insignificant.

Inframarginal Types Next, I analyze how an increase in the bonus amount affects types who are already above the bonus threshold, the inframarginal types. Denote the share of participants completing at least 17 tasks in a baseline condition (possibly no bonus) as p_{17}. I take the top p_{17} percent of effort in baseline, and compare it to the top p_{17} percent of effort in the treatment condition with the increased bonus amount. If the composition of types who provide the top p_{17} percent of effort is unaffected by the increase in the bonus amount, then this comparison identifies the treatment response of the inframarginal types.

Result 3 implies that all inframarginal types increase their effort in response to an increase in the bonus amount, and therefore their distribution of effort in the treatment condition should be above, in the sense of first-order stochastic dominance, its counterpart in baseline. I test for this first-order stochastic dominance using a one-sided Kolmogorov-Smirnov test and the empirical likelihood ratio test of Davidson and Duclos (2013). The Kolmogorov-Smirnov test posits a null hypothesis of first-order stochastic dominance, and so a failure to reject provides only weak evidence for the model. The likelihood ratio test improves upon this by assuming a null hypothesis of non-dominance, and so a rejection provides stronger

Table 2.4: Hypothesis testing, interaction effects

Regression

$$\mathbb{1}\{\text{Effort}_\text{i} \in \text{E}_\text{k}\} = \alpha + \beta\,\text{Bonus}_i + \gamma\,\text{Badge}_i + \delta\,\text{Bonus}_i \times \text{Badge}_i$$

Hypothesis	Parameter (measured in pp) (1)	Estimate (2)	s.e. (3)	p-value (4)
D	A bonus induces less bunching with a badge than without a badge. Dependent Variable: $\mathbb{1}\{\text{Effort}_\text{i} \in \{15, 16\}\}$			
	γ	−0.3	0.8	0.685
	$\delta^{\text{pool-any-bonus}}$	−4.2	2.6	0.107
	$\delta_{\text{40c at 15}}$	−4.6	4.7	0.325
	$\delta_{\text{80c gift and 40c at 15}}$	−2.1	4.8	0.657
	$\delta_{\text{\$1.20 at 15}}$	−4.5	3.7	0.214
E	A bonus induces more anti-bunching with a badge than without a badge. Dependent Variable: $\mathbb{1}\{\text{Effort}_\text{i} \in \{17, \ldots, 38\}\}$			
	γ	2.7	3.0	0.364
	$\delta^{\text{pool-any-bonus}}$	5.0	3.8	0.188
	$\delta_{\text{40c at 15}}$	5.5	5.5	0.314
	$\delta_{\text{80c gift and 40c at 15}}$	3.3	5.5	0.551
	$\delta_{\text{\$1.20 at 15}}$	5.7	4.5	0.208
F	With a badge, an increase in the bonus amount increases anti-bunching above the bonus threshold. Dependent Variable: $\mathbb{1}\{\text{Effort}_\text{i} \in \{17, \ldots, 38\}\}$			
	$[\beta+\delta]_{\text{\$1.20 at 15}} - [\beta+\delta]^{\text{pool-40c-bonus}}_{\text{40c at 15}}$	3.3	3.5	0.339
	$[\beta+\delta]_{\text{\$1.20 at 15}} - [\beta+\delta]_{\text{40c at 15}}$	2.3	4.2	0.584
	$[\beta+\delta]_{\text{\$1.20 at 15}} - [\beta+\delta]_{\text{80c gift and 40c at 15}}$	4.4	4.3	0.298
G	An increase in the bonus amount increases anti-bunching above the bonus threshold with a badge more than without a badge. Dependent Variable: $\mathbb{1}\{\text{Effort}_\text{i} \in \{17, \ldots, 38\}\}$			
	$\delta_{\text{\$1.20 at 15}} - \delta^{\text{pool-40c-bonus}}_{\text{40c at 15}}$	1.2	4.7	0.794
	$\delta_{\text{\$1.20 at 15}} - \delta_{\text{40c at 15}}$	0.1	5.7	0.981
	$\delta_{\text{\$1.20 at 15}} - \delta_{\text{80c gift and 40c at 15}}$	2.4	5.7	0.678

Note: This table summarizes the results of additional pre-specified hypothesis tests. The first row in panels D and E indicate the baseline value. See note of table 2.3 for further details.

evidence for the model. Following the pre-analysis plan, I focus on the *Badge* condition and pool the two treatment arms "40c-bonus at 15 tasks" and "80c-gift and 40c-bonus at 15 tasks".

Table 2.5 summarizes the results. In the no bonus treatment 21.0 % of participants are inframarginal.[7] Among the top 21.0 %, the empirical CDF of effort with no bonus is up to 12 pp (14 pp) above the one with a 40c-bonus ($1.20-bonus). However, the same CDF is also up to 0.3 pp (3 pp) below the one with a 40c-bonus ($1.20-bonus). The Kolmogorov-Smirnov test fails to reject either null hypothesis that one distribution of effort dominates the other and is therefore uninformative.

Since the empirical CDFs cross somewhere over the range of 17 to 38 tasks, there is no dominance in the sample and the likelihood ratio test of Davidson and Duclos mechanically yields a p-value of 1.

I also estimate the largest interval over which the likelihood ratio test does reject non-dominance. I find that for the inframarginal types the distribution of effort with no bonus is dominated by the one with a 40c-bonus ($1.20-bonus) over the interval of $[17, 22)$ tasks ($[17, 26)$ tasks) at the 1 %-level. One explanation for why the test finds dominance only close to the bonus threshold is that the convex cost of effort implies that at higher levels of effort a smaller increase in effort is sufficient to gain the same marginal reputational benefit. This makes it harder to obtain statistically significant effects far above the bonus threshold.

I next analyze the increase in the bonus amount from 40c to $1.20. In the 40c-bonus treatment 29.4 % of participants are inframarginal. Among the inframarginal types, the empirical CDF of effort with the 40c-bonus is up to 10 pp (3 pp) above (below) the one with a $1.20-bonus. The Kolmogorov-Smirnov test again fails to reject either null hypothesis, and the likelihood ratio test yields a p-value of 1 due to non-dominance in the sample. The interval over which the likelihood ratio test does reject non-dominance at the 1 %-level is $[17, 19)$. One explanation for the shorter interval is that the increase from no bonus to a 40c-bonus moved 32.8 pp of participants from below to on or above the bonus threshold, while the increase from a 40c-bonus to $1.20-bonus moved only 10.9 pp. Moving fewer participants is likely to correspond to a smaller reduction in the intrinsic motivation of the marginal type obtaining the bonus, which would induce a smaller response from the inframarginal types.

Perceived Reputation Next, I provide evidence on the role of beliefs for these effects. The model predicts that increasing the bonus amount decreases the reputation associated with a given level of effort above the bonus threshold. Agents understand this downward shift in reputation, and increase their effort in order to maintain the same level of reputation.

In the experiment I measure reputation by showing participants pairs of badges and asking them "Who is more generous?". However, a participant that chooses their effort to

[7]The comparison with the no bonus treatment was not pre-specified. I include it here to provide a benchmark for the comparison of the 40c-bonus with the $1.20-bonus.

Table 2.5: Treatment effects on inframarginal types

		Baseline %-share		K-S test, H_0: $B \geq_D$T		$T \geq_D$B		Davidson-Duclos CI rejecting H_0: $T \not\geq_D B$		
B(aseline)	T(reatment)	≥17	=38	D^+	p	D^-	p	α=0.01	α=0.05	α=0.10
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
No bonus	40c at 15	21.0	16.1	0.12	0.35	0.00	1.00	[17, 22)	[17, 25)	[17, 26)
No bonus	$1.20 at 15	21.0	16.1	0.14	0.22	0.03	0.95	[17, 26)	[17, 28)	[17, 28)
40c at 15	$1.20 at 15	29.4	16.5	0.10	0.39	0.03	0.90	[17, 19)	[17, 19)	[17, 21)

Note: This table summarizes the treatment effects on the effort distribution of types who go beyond the bonus threshold in both, baseline and treatment (assuming monotonicity). Columns 1 and 2 define baseline and treatment. Column 3 computes the share of participants in baseline completing at least 17 tasks. Column 4 reports the share of participants in baseline who cannot increase their effort because they are at the maximum effort level. Columns 5 to 11 compare the two distributions of effort of the respective top p_{17}-% from the baseline and treatment population, where p_{17} is the value in column 3. "$>_D$" denotes the partial order implied by first-order stochastic dominance, "$\geq_D$" additionally includes the equality of two distributions. Columns 5 to 8 report the test statistics and p-values from one-sided Kolmogorov-Smirnov tests. Columns 9 to 11 report the largest interval over which the likelihood ratio test of Davidson and Duclos (2013) rejects non-dominance at varying levels of significance using bootstrapped p-values.

influence how they are seen by other participants, does not take into account their own views of generosity. Instead, a participant will adjust their effort according to their beliefs about others' views on generosity. Therefore, I also ask participants "Who do most other participants say is more generous?" and use the responses to this question to analyze the reputation channel of the model. I will refer to the average response to this question as the *perceived reputation*.

In order to maintain statistical power, I elicit the perceived reputation only for pairs in which badge 1 shows a 40c-bonus incentive and effort level $e_1 \in \{1, 15, 16, 17\}$. I also restrict elicitation to pairs of badges with those effort levels that are most informative about anti-bunching at zero or at the bonus threshold.[8] Finally, to keep the judgement relevant and truthful, I only show pairs that reflect actual behavior by participants in the *Badge* treatment.

Figure 2.7 summarizes the perceived reputation for a variety of badge pairings. First, note that none of the lines are flat, reflecting that the perceived reputation responds to the level of effort shown on badge 2 holding everything else constant. This shows that participants believe that their choice of effort constitutes a meaningful signal about their own generosity.

Second, consider pairings where badge 1 and badge 2 both feature a 40c-bonus (second row in fig. 2.7). If badge 1 shows an effort level of 1 task completed, then badge 2 will have a perceived reputation of 0 % for an effort level of 0 tasks and a perceived reputation of over

[8] Due to a programming error I do not elicit the perceived reputation for the pair ($e_1 = 16$, $e_2 = 17$) with both badges featuring a 40c-bonus.

95 % for effort levels of 2 tasks and 5 tasks. Analogously, if badge 1 shows 16 or 17 tasks completed, then badge 2 will have a perceived reputation of 0 % for an effort level of 15 tasks and a perceived reputation of over 95 % for effort levels of 18–20 tasks. This shows that when both badges show the same bonus incentive scheme and both effort levels are on the same side of the bonus threshold, then the perceived reputation is fully determined by what badge shows a higher level of effort.

Third, if badge 1 features a 40c-bonus and an effort level of 15, 16 or 17 tasks, and badge 2 features no bonus (first row in fig. 2.7), then the perceived reputation function crosses 50 % at a badge 2 effort level of around 10–12 tasks. This shows that participants believe that an effort level of 10–12 tasks without a bonus is seen to be as generous as an effort level of around 15–17 tasks with a 40c-bonus.

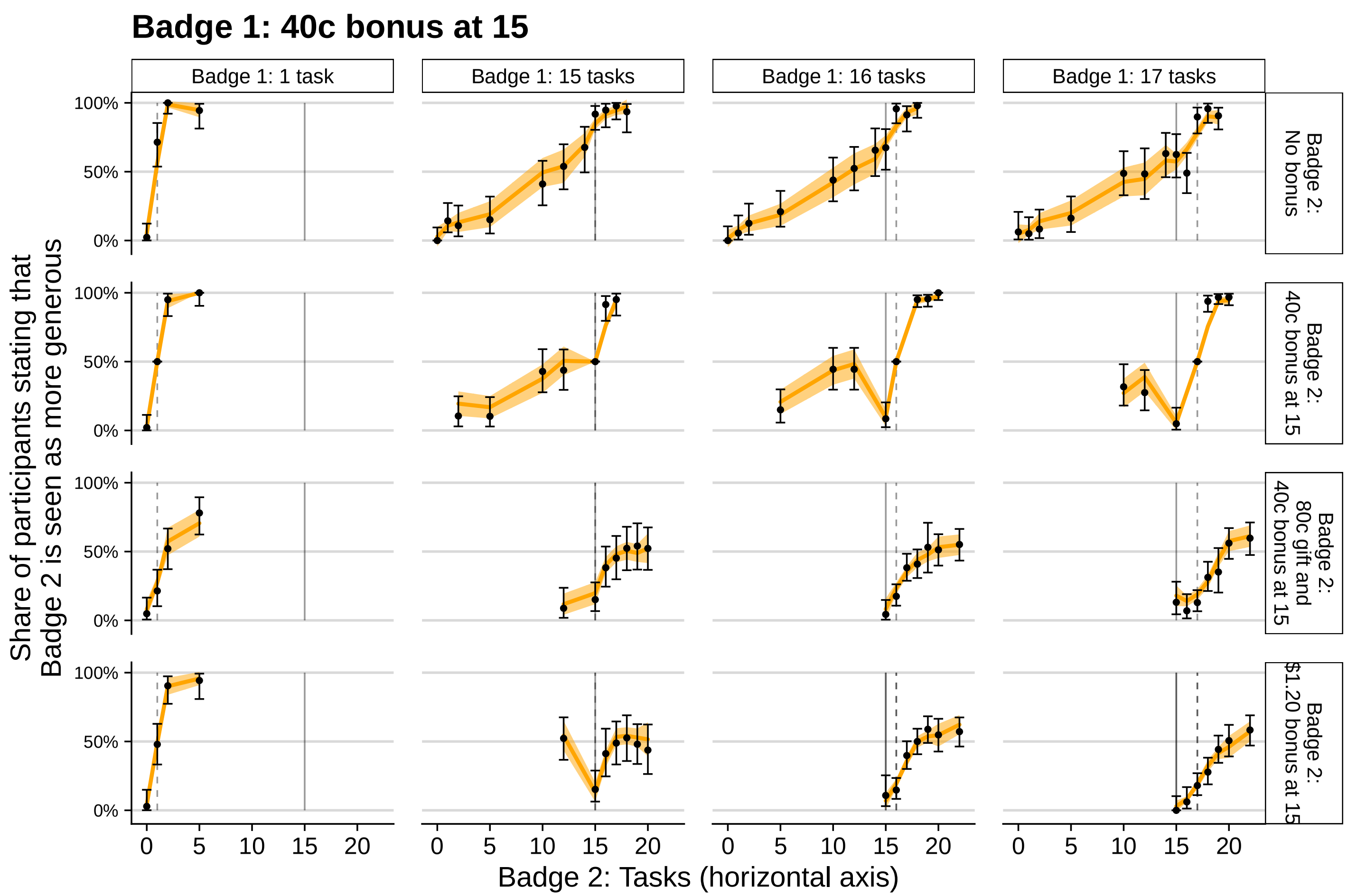
Fourth, if badge 1 features a 40c-bonus and an effort level of 15, 16 or 17 tasks, and badge 2 features a $1.20-bonus, then participants believe that badge 2 is perceived as equally generous only when it features an effort level below the bonus threshold or strictly above that of badge 1. Consider the second column in fig. 2.7 as an example: When badge 1 features 15 tasks and a 40c-bonus, and badge 2 features a $1.20-bonus and an effort level as high as 20 tasks the perceived reputation function is only at around 50 %. This shows that when obtaining a larger bonus of $1.20 instead of 40c, participants believe that the reputational reward for effort above the threshold decreases so strongly that even completing 5 additional tasks is insufficient to offset it.

Figure 2.8 summarizes the responses by showing for what pairings a significant majority thinks that one badge is perceived as more generous. When badge 1 displays 1 task, then the perceived reputation is completely determined but what badge shows more tasks. When badge 1 displays 15, 16 or 17 tasks, then the perceived reputation additionally depends on the bonus incentive. If the bonus amount is larger, then participants believe that it takes higher levels of effort to be seen equally or more generous.

Table 2.6 summarizes a series of pre-specified hypothesis tests for changes in the perceived reputation function. Panel A shows that increasing the bonus from 40c to $1.20 significantly decreases the perceived reputation function at effort levels above the bonus threshold. Panels B and C show that is also true if the bonus increase of 80c is given as a gift at 0 tasks instead of 15 tasks, and that this alternative increase is not statistically different from the increase at 15 tasks.

In sum, the results show that participants believe that their chosen effort constitutes a signal of their generosity, and that an increase in the bonus amount induces a decrease in the reputational payoff, that requires an increase in effort to be offset.

Figure 2.7: Perceived reputation function



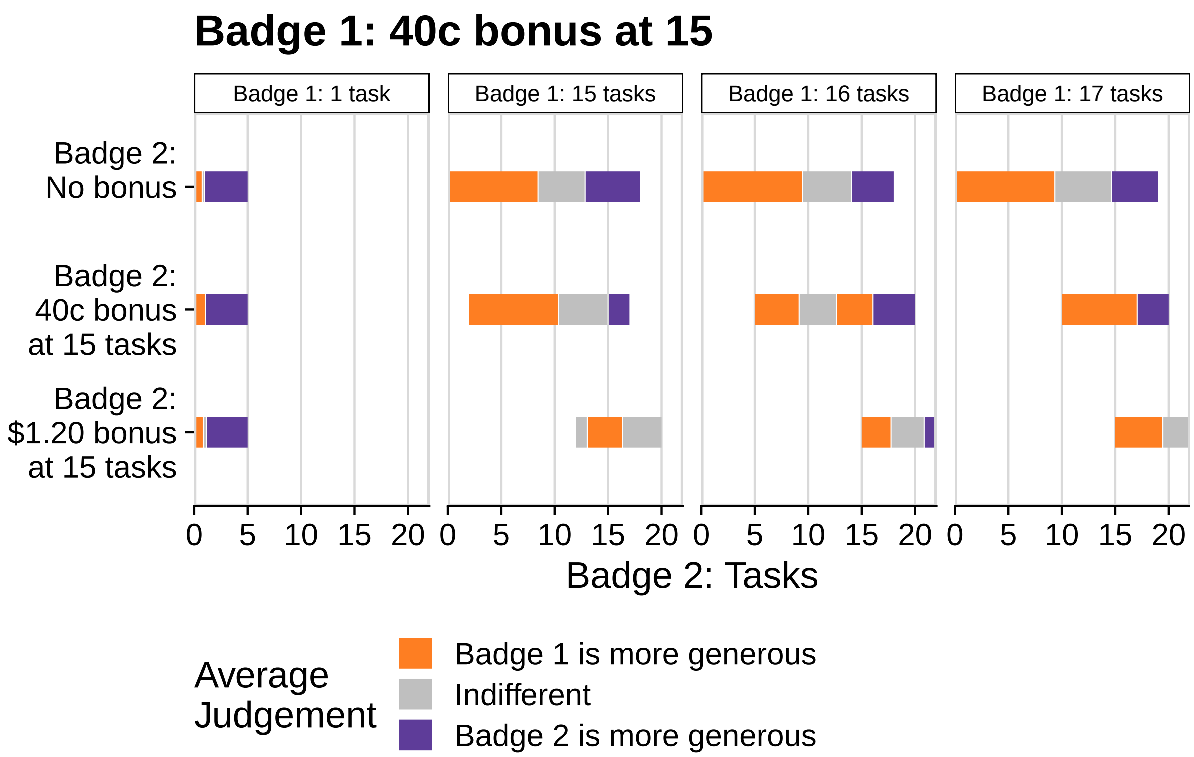
Note: This figure summarizes the responses to "Who do most other participants say is more generous?" for different pairs of badges. A pair consists of a badge 1 and a badge 2 which are shown in random order to participants as depicted in fig. 2.4. Each badge displays the number of tasks completed together with the bonus incentive. Badge 1 displays a 40c-bonus incentive for completing 15 or more tasks, and either 1, 15, 16, or 17 tasks completed (columns). Badge 2 shows one of the four bonus incentives (rows), and tasks completed on the horizontal axis. If badge 1 and badge 2 are identical, then the share is imputed at 50 %. Standard errors are computed using an exact binomial test. The orange line shows the result of a local linear regression using a uniform kernel with a one-sided bandwidth of 1.5 tasks.

Table 2.6: Effect of bonus incentives on perceived reputation

"Who do most other participants say is more generous?"
e_x: Effort shown on Badge x. N: No. of responses. $s_{1,x}$: %-share responding "Badge x"

	Effort		Badge 1 vs. 2		Badge 1 vs. 3		(Badge 1 vs. 3) − (Badge 1 vs. 2)		
	e_1	e_2,e_3	$N_{1,2}$	$s_{1,2}$	$N_{1,3}$	$s_{1,3}$	$s_{1,3}-s_{1,2}$	CI	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A	Badge 1 and 2: 40c-bonus at 15. Badge 3: $1.20-bonus at 15.								
	16	16	—	50.0	95	14.7	−35.3	[−41.7, −26.5]	0.000***
	16	18	122	95.1	120	50.0	−45.1	[−54.7, −34.9]	0.000***
	16	19	113	95.6	107	58.9	−36.7	[−47.0, −26.5]	0.000***
	16	20	68	100.0	73	54.8	−45.2	[−56.9, −34.0]	0.000***
	17	17	—	50.0	100	18.0	−32.0	[−39.0, −23.1]	0.000***
	17	18	81	93.8	90	27.8	−66.0	[−75.7, −53.9]	0.000***
	17	19	122	96.7	104	44.2	−52.5	[−62.2, −41.9]	0.000***
	17	20	94	96.8	79	50.6	−46.2	[−57.6, −34.2]	0.000***
B	Badge 1 and 2: 40c-bonus at 15. Badge 3: 80c-gift and 40c-bonus at 15.								
	16	16	—	50.0	103	17.5	−32.5	[−39.3, −23.8]	0.000***
	16	18	122	95.1	93	40.9	−54.2	[−64.4, −43.0]	0.000***
	16	19	113	95.6	32	53.1	−42.5	[−59.7, −25.7]	0.000***
	16	20	68	100.0	80	51.3	−48.8	[−60.1, −37.7]	0.000***
	17	17	—	50.0	85	12.9	−37.1	[−43.4, −28.0]	0.000***
	17	18	81	93.8	80	31.3	−62.6	[−73.0, −49.8]	0.000***
	17	19	122	96.7	37	35.1	−61.6	[−75.3, −44.9]	0.000***
	17	20	94	96.8	82	56.1	−40.7	[−52.2, −29.3]	0.000***
C	Badge 1: 40c-bonus at 15. Badge 2: 80c-gift and 40c-bonus at 15. Badge 3: $1.20-bonus at 15.								
	16	16	103	17.5	95	14.7	−2.7	[−13.3, 8.2]	0.623
	16	17	102	38.2	98	39.8	1.6	[−12.1, 15.4]	0.847
	16	18	93	40.9	120	50.0	9.1	[−4.6, 22.5]	0.200
	16	19	32	53.1	107	58.9	5.8	[−13.4, 25.3]	0.603
	16	20	80	51.3	73	54.8	3.5	[−12.7, 19.6]	0.720
	16	22	78	55.1	91	57.1	2.0	[−13.2, 17.3]	0.805
	17	17	85	12.9	100	18.0	5.1	[−6.2, 15.8]	0.363
	17	18	80	31.3	90	27.8	−3.5	[−17.7, 10.6]	0.675
	17	19	37	35.1	104	44.2	9.1	[−9.7, 26.2]	0.384
	17	20	82	56.1	79	50.6	−5.5	[−21.1, 10.4]	0.532
	17	22	72	59.7	84	58.3	−1.4	[−16.9, 14.4]	1.000

Figure 2.8: Indifference regions of perceived reputation function



Note: This figure summarizes the responses to "Who do most other participants say is more generous?" for different pairs of badges. A pair consists of a badge 1 and a badge 2 which are shown in random order to participants as depicted in fig. 2.4. Each badge displays the number of tasks completed together with the bonus incentive. Badge 1 displays a 40c-bonus incentive for completing 15 or more tasks, and either 1, 15, 16, or 17 tasks completed (columns). The bonus incentive of badge 2 varies across rows, and tasks completed in badge 2 vary on the horizontal axis. The figure highlights the regions of tasks completed in badge 2 for which the share of participants answering badge 1 (or 2) is significantly different from 50 %.

2.4 Conclusion and Transition to Chapter 3

The previous chapter has introduced a new test for signaling motives. The test relies on the idea that agents do not want to be lumped together with low typed agents whose primary motivation is the receipt of personal benefits. As a proof of concept I have shown in this chapter that the test performs well in a online experiment involving prosocial behavior.

Note for table 2.6: This table summarizes the responses to "Who do most other participants say is more generous?" for different pairs of badges. A pair consists of a badge 1 and a badge 2 (or badge 3) which are shown in random order to participants as depicted in fig. 2.4. Each badge displays the number of tasks completed (columns 1–2) together with the bonus incentive (panels A–C). In each row, Badges 2 and 3 are identical, except for the bonus incentive. Columns 7–9 denote how changing the bonus incentive shown on the badge affects the share of participants choosing it. Confidence intervals and p-values are computed using Barnard's exact test for 2×2 tables.

After the theoretical exposition of anti-bunching in chapter 1, and the empirical study of anti-bunching in prosocial behavior within a controlled environment in chapter 2, I now continue with a study prosocial behavior within a field setting in chapter 3. In chapter 3, we investigate how the adoption of open science practice has changed in the social science disciplines using an incentivized online survey. Practicing open science is an inherently prosocial act as it allows fellow researchers to learn from and build upon existing research.

Acknowledgments

I am especially grateful to my advisors Stefano DellaVigna, Ned Augenblick and Ted Miguel for their outstanding guidance and support for this project. I also thank Roland Bénabou, Doug Bernheim, Peng Ding, Christine Exley, Fred Finan, Zack Grossman, Supreet Kaur, Pat Kline, Charlie Sprenger, Dmitry Taubinsky, Neil Thakral, Jean Tirole, Chris Walters, Zenan Wang, and seminar participants at Berkeley for helpful comments and discussion. Funding for this project was generously provided by the Berkeley Initiative for Behavioral Science. This project was approved by the UC Berkeley IRB (2019-08-12490) and pre-registered at the AEA RCT Registry (AEARCTR-0004809). All errors are belong to me.

Chapter 3

Open Science Practices are on the Rise

3.1 Introduction

Across many scientific disciplines there has been a movement to promote open science practices: posting data, code, and study materials online, and pre-registering studies, hypotheses, and analyses prior to a research study (Miguel et al., 2014; Nosek et al., 2015). In the social sciences for the past two decades, disciplinary organizations and journals have increasingly endorsed open science practices. More recently, cross-disciplinary social science organizations have been founded to accelerate awareness of open science and to provide training and supportive open science technologies, such as pre-registration platforms and open archives (Christensen, Freese, et al., 2019). During this period, the social sciences have also grappled with debates and scandals surrounding the unavailability of original data, examples of publication bias, replication challenges, and in some cases data fraud (Bhattacharjee, 2013; Borsboom and Wagenmakers, 2012; Broockman et al., 2015; Carey, 2011; Enders and Hoover, 2004; Feilden, 2017; Neuroskeptic, 2012).

Beyond reducing the incidence of fraud (Simonsohn, 2013), open science practices have been linked to the improved quality and credibility of research findings across fields. For example, study registration could increase the visibility of results, improving meta-analysis and reducing the selective reporting of null, unexpected or otherwise unfavorable results (Kaplan and Irvin, 2015; de Vries et al., 2018), and data sharing could increase later data re-use and article citations (Piwowar and Vision, 2013)

Yet controversy and opposition have followed many research transparency proposals in the social sciences, particularly the use of pre-registration (Open Science Collaboration, 2015; Gilbert et al., 2016; Coffman and Niederle, 2015). For instance, some worry that pre-registration might hamper creative research (Goldin-Meadow, 2016; Kupferschmidt, 2018). Others suggest that it maybe be used instrumentally or selectively, therefore doing little to remedy the underlying problems it was proposed to address (Claesen et al., 2019). Altogether, some debates over the merits of open science may be natural extensions of the disagreement and scandals that prompted open science proposals in the first place, while others may arise from uncertainty over the effectiveness of proposed solutions, or simply because open science practices represent a break from the status quo.

Addressing these controversies, and in particular the debates about the effect of open science practices on the social scientific literature, is beyond the scope of the present paper. Rather, we pose a question that logically precedes answers to those questions, specifically: how many social scientists are adopting open science practices, and what are the average perceptions of these practices in the social sciences? While some researchers are publicly starting to adopt open science practices (Christensen and Miguel, 2018), there may be a lag between private adoption and public representation. For example, there are lags between pre-registration of a study or preparation of shareable code and article publication. Additionally, there are a small number of highly vocal scholars (including some authors of this article) who have expressed strong opinions either in support of or against the adoption of open science practices. However, these prominent voices may not be representative of the opinions of most scholars. Thus, there remains a considerable degree of uncertainty about researchers' current adoption of and attitudes toward open science practices (M. S. Anderson et al., 2007).

Previous attempts to quantify adoption of open science practices tend to have small and largely unrepresentative convenience samples of survey respondents, and focus on just a single research discipline (e.g. van Assen et al., 2015; Baker, 2016; Buttliere, 2014; Feilden, 2017). Researchers largely send solicitations to complete non-remunerated surveys to academic list-serves, or to their personal networks via email or social media. In these surveys, scholars often claim to be more supportive of open science practices than their peers.

The present research, based on the State of Social Science (3S) Survey, generates a more robust estimate of the adoption of open science practices over time, and of general support and perceived norms of research transparency across four major social science disciplines: economics, political science, psychology and sociology. In addition, we connect the patterns in the data to theories regarding how institutions and technological innovations may affect the pace of scientific change (Romer, 1990; Griliches, 1957) and the development of new norms (Kuhn, 1962; Hacking, 1981).

3.2 Sample and Data

We solicited information using a monetarily incentivized survey from a representative sample of active, elite social science researchers in the fields of economics, political science, psychology, and sociology who work with empirical quantitative or qualitative data. The 3S survey queried respondents on awareness of, attitudes toward, perceived norms regarding, and adoption of open science practices. We randomly drew the sample from the complete set of authors who had published within a range of 3 years (2014-2016) in 10 of the most cited journals for each discipline. We also drew from the complete set of PhD Students enrolled in the top 20 North American departments in each discipline during the first half of 2018; see supplementary materials for details. We pre-registered analyses for our survey and posted our pre-analysis plan and study materials on the Open Science Framework. The present survey and descriptive analysis are the first part of a broader project described in the pre-analysis plan.

In total, we invited 6,221 individuals to complete a survey between April and August 2018 of whom 6,058 were contacted (emails did not bounce). Published Authors were compensated either $75 or $100 (randomly), and graduate students either $25 or $40 (response rates did not significantly vary by level of compensation). Arguably, our response rate represents an upper bound on the rate that is possible to achieve with a reasonable incentive strategy: at a median length of 15 minutes per survey, faculty were compensated at minimum $300 per hour.

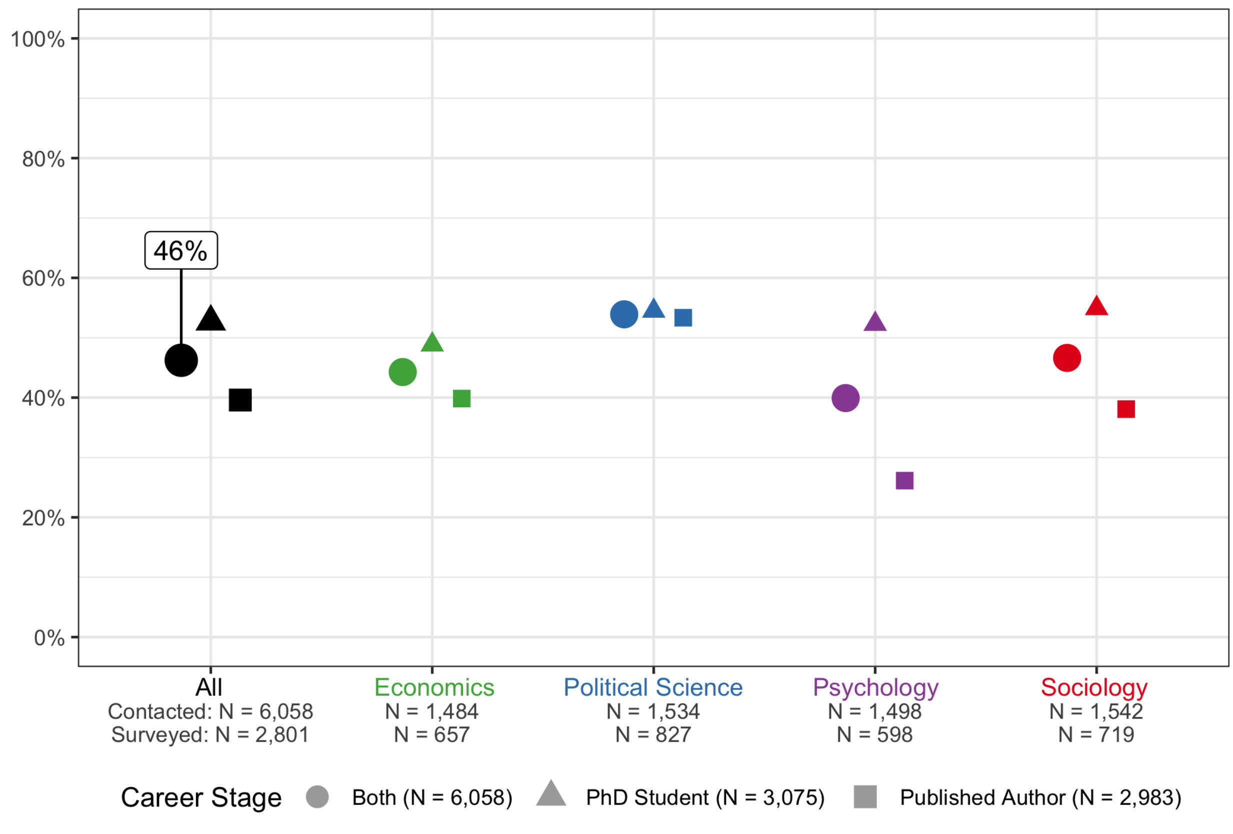
Our incentive scheme achieved a completed survey response rate of 46.2%, implying that the study sample is broadly representative of active Published Authors and PhD Students in these four fields. Figure 3.1 presents the overall response rate of 46.2%, which ranged from 40% in Psychology to 55% in Political Science. We consistently obtained a majority of PhD Students, who responded at or above 50% in every field, while Published Authors (who had predominantly completed their doctoral training) responded at somewhat lower rates. Among respondents with North American email addresses, the response rates are slightly higher at 49% overall, 44% for Published Authors, and 53% for PhD Students.

As shown in Figure 3.1, the response rate for Published Authors from psychology journals is somewhat lower than that for the other disciplines' journals. This may be due to the fact that a subset of psychologists often publish with scholars or clinicians from other fields who are less active empirical researchers, and therefore may be less likely to respond to an invitation to complete a survey focused on research methods. Consistent with this explanation, the response rate from authors who published in clinical and neuroscience-focused journals is considerably lower than the rate for social and developmental psychology journals (see Appendix Figure A.3 for survey response rates by journal). Similarly, the response rate for authors who had published in macroeconomics journals is somewhat lower than the rate from other economics journals, possibly due to the greater share of articles based on theoretical or simulation approaches, rather than quantitative empirical data analysis, in those journals.

Two concerns about the validity of our study design might remain. First, our survey results are entirely self-reported and one might be concerned that individuals could misstate their open science behavior, for example, due to surveyor demand effects. Second, even though to our knowledge the current sample is by far the largest and most representative attempt to assess open science attitudes and practices to date, one might still be concerned about the nature of selection into the sample. It remains possible that scholars who responded to the survey are non-randomly selected from the population along important dimensions. Indeed, we find that the response rate among Published Authors was significantly higher for those with more publications in leading disciplinary journals during the last three years, and for those at institutions in North America (see Appendix Table A.9).

To better understand the degree to which non-random survey response may be a concern, we conducted an audit of open science behavior for a random sample of Published Author respondents and non-respondents from economics; economics was chosen because the vast majority of scholars use the same study registry and data posting platform, increasing the

Figure 3.1: Response rates are high across disciplines



Note: Response rates by discipline and by career stage (PhD Student or Published Author). We contacted 6,058 researchers (6,221 researchers were invited via email but 163 emails bounced). Above figure consists of 2,787 respondents and 3,434 non-respondents, including 65 explicit opt-outs and 244 partially complete surveys, but excluding the 163 bounced emails.

accuracy of the audit. We checked publicly available repositories and each author's website to determine whether they had previously pre-registered a study or posted data online; the details of the audit activity can be found in the SOM.

The audit activity yielded three main results. First, there is a high rate of agreement between self reports and actual behavior as presented in Table 3.1: despite only checking a limited number of online sources we were able to validate almost 80% of individuals' responses regarding adoption of open science practices. Second, while there is some selection into the sample, this appears to be primarily driven by scholars with a more empirical orientation being more likely to respond: response rates for theory-focused economists and macroeconomists are far lower than for other fields, at 27.2% for theory/macroeconomics/finance focused Published Authors versus 50.4% for the others, as shown in Table 3.2. Third, scholars with a more empirical orientation do not appear to be selecting into our survey in a manner related to previous open science behavior (see Table 3.2). Taken together, these patterns suggest that the survey results are broadly representative of the behaviors and views of Published Authors with a more empirical orientation.

3.3 Retrospective Open Science Behavior

We first assess how the adoption of open science practices has changed over time, using survey respondents' self-reports and bounding them with a verification exercise (described below). We find that the last decade has been a time of rapid change across disciplines, with adoption of open science practices increasing dramatically.

Figure 3.2 presents the cumulative proportion of Published Authors who have adopted open science practices over time. We focus on scholars who received their PhD by 2009, as they had the opportunity to engage in these practices over much of the last decade (see Appendix Figure A.5 for robustness to different PhD cutoff dates). 84% of Published Authors reported adopting an open science practice by 2017 (the last complete year for which we collected data), nearly doubling from 49% in 2010. The sharing of data, code and survey instruments show rapid increases starting after 2005, while the use of pre-registration has increased dramatically since 2013. Posting data or code online is the most common practice, followed by posting study instruments online, and then pre-registration. We also find in our survey data that those who reported adopting an open science practice at some point in the past are overwhelmingly likely to also have employed it in their most recent research project (see Appendix Table A.11), indicating that scholars' adoption of these practices tends to be persistent.

The shaded areas underneath these lines adjust the adoption graph to incorporate the adoption rates of non-respondents, using the verified open science behavior for non-respondents found in our audit activity. Details on how these estimates are constructed are presented in Table 3.1. Even incorporating the likely behavior of non-respondents, we estimate that 76% of Published Authors have adopted an open science practice by 2017.

While there is an upward trend in all four disciplines, Figure 3.3 shows that adoption patterns differ across disciplines. The evolution of adoption in economics and political science appear relatively similar, with a rapid increase in the rates of posting data or code online. In economics, there has been a steady rise in posting study instruments online and pre-registration since around 2011. Political science has seen an increase in posting study instruments since 2005, and a steeper rise in pre-registration since 2014.

Psychology researchers were lagging behind economics and political science scholars until recently for all practices, but over the last few years psychology has had the most rapid increase in adoption. Psychologists also currently report the highest adoption rate for study pre-registration. Sociology has the lowest levels of adoption for all open science practices, but as with the other fields, there has been a steady increase in recent years.

Table 3.1: Differences in behaviors for published authors, respondents, and non-respondents on Economics subfield validation data

Parameter	Any (1)	Posting data or code online (2)	Posting study instruments (3)	Pre-registering hypotheses or analyses (4)
Respondent				
Share of Respondents Doing Practice (self-report) (S_R)	84.0%	73.0%	44.3%	20.3%
Share of Respondents Doing Practice (validation) (V_R)	65.3%	63.3%	34.4%	19.0%
Difference between S_R and V_R (D_R)	18.7%	9.7%	9.8%	1.3%
Non-respondent				
Predicted share of non-respondents doing practice ($\widehat{S_N}$)	70.7%	63.4%	37.3%	7.5%
Share of non-respondents verified doing practice (V_N)	55.0%	55.0%	29.0%	7.0%
Difference between $\widehat{S_N}$ and V_N ($\widehat{D_N}$)	15.7%	8.4%	8.3%	0.5%
$\frac{V_N}{V_R}$	84.2%	86.8%	84.2%	36.8%
Share of Practices Verified ($\frac{V_R}{S_R}$)	77.7%	86.7%	77.7%	93.6%

Note: This table presents stated and observed open science behavior for Published Authors in Economics who are respondents and non-respondents in our sample. Observed behavior comes from our audit of all the economists who completed the survey and a random sample of 100 economists who did not complete survey. This audit was completed between March 15, 2019 and April 15, 2019. For pre-registration and posting data and code online, S_R is the percentage of respondents who report engaging in the specified open science practice in our survey. V_R is the percentage of Published Author respondents who we find in our audit to engage in the open science practice. D_R reports the difference between the two. V_N is the percentage of non-respondents in our audited sample that we verify have done an open science practice. $\widehat{S_N}$ is an imputed value for the stated percentage of non-respondents that would have reported doing an open-science practice had they been surveyed. To estimate this, we multiply the audit value V_N by the ratio between stated and observed of respondents (i.e the ratio $\frac{S_R}{V_R}$). $\widehat{D_N}$ is the difference between $\widehat{S_N}$ and V_N. Since we did not conduct an audit for "Posting study instruments online", the "Any" category refers either "Posting data or code online" or "Pre-registering hypotheses or analyses". And "Posting study instruments online" therefore V_R is imputed using the ratio of S_R to V_R in the "Any" category. The remainder of the methodology for this open science practice is the same as listed above.

Table 3.2: Differences in observables for published authors, respondents, and non-respondents on Economics subfield validation data

	Overall (1)	Respondent (2)	Nonrespondent (3)	Difference (2) - (3)
Share of sample:				
— Theory Focused	0.19	0.15	0.22	-0.07 (-1.58)
— Macro/Finance Focused	0.26	0.16	0.33	-0.17 (-3.28)***
— not Theory/Macro/Finance Focused	0.55	0.69	0.45	0.24 (4.29)***
Verified Open Science Behavior				
— all Economics Published Author	0.59	0.65	0.55	0.10 (1.81)*
— among Theory Focused	0.35	0.39	0.32	0.07 (0.54)
— among Macro/Finance Focused	0.56	0.58	0.55	0.04 (0.33)
— not Theory/Macro/Finance Focused	0.69	0.73	0.67	0.06 (0.76)
N	753	300	100	

Note: This table shows the percentage of economics Published Authors who work in different subfields among those who responded and did not respond to the survey. The first panel reports response rates and share of each sample for each subfield. Column 1 shows the response rate for each subfield. Columns 2 and 3 show the share of respondents and non-respondents identifying with each subfield respectively. Panel B shows the fraction of individuals in each subfield for whom we verified open science behavior during our audit activity. For respondents, the subfield is determined by the subfield that the respondent listed in our survey. For non-respondents, we constructed the individual's subfield in an audit activity that was completed between March 15 2019 and April 15 2019. In this activity, we used publicly available data sources to collect data on the primary subfield of these non-respondents. We manually collected all of the subfields that an individual listed working in on their website or CV. After these subfields were collected we manually categorised these subfields into one of three categories. The first of these was "Theory focused", which is categorised as any individual who listed Microeconomic Theory or Econometrics as a primary subfield. The second was "Macroeconomics/Finance", which was any author who listed Macroeconomics or Finance as a primary field. Finally, all other authors were categorised in the residual category. The final column in the table provides t-statistics for tests for differences in the mean between those respondents and non-respondents. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

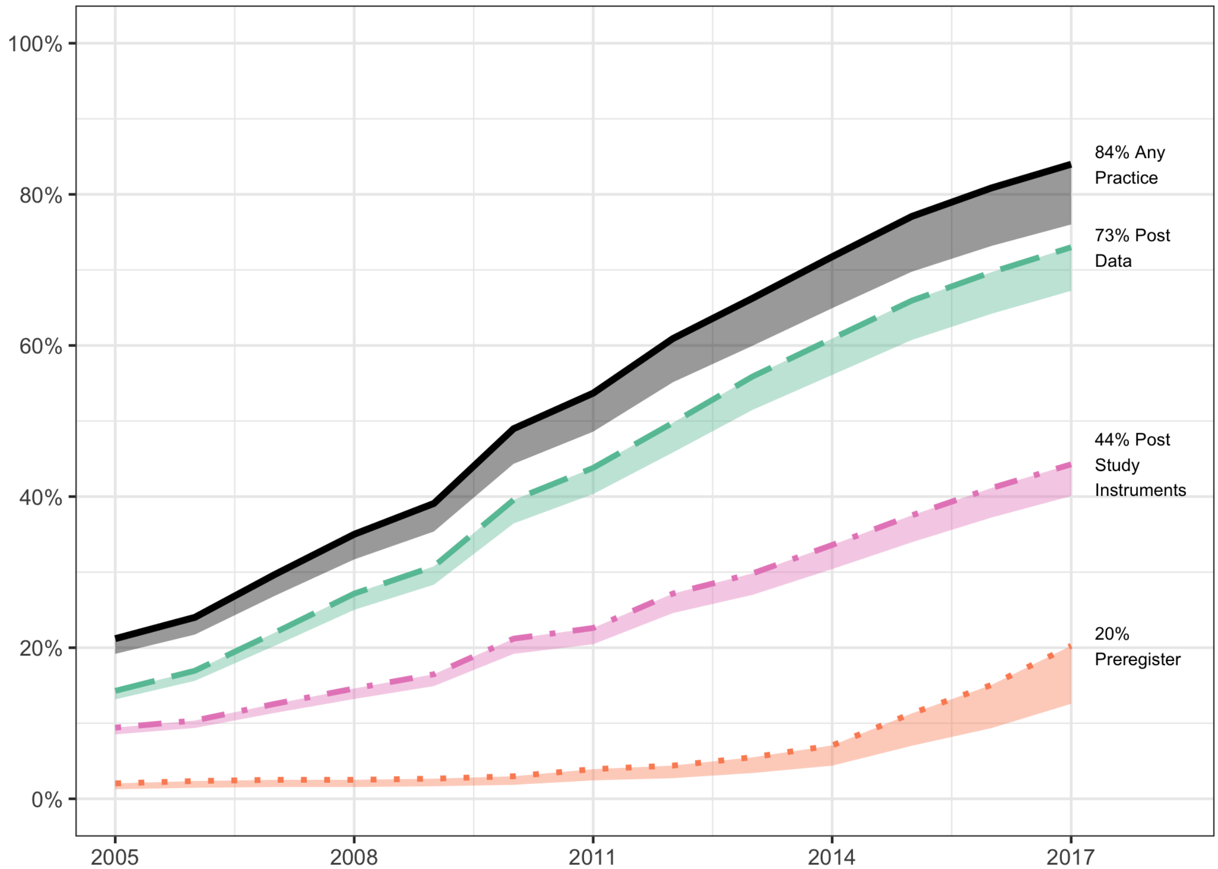
Adoption rates of all three highlighted open science practices have been highest for researchers using experimental methods across social science disciplines, while adoption rates for posting study materials and pre-registration have been lower among researchers using non-experimental quantitative methods. Rates for all practices are the lowest among researchers using exclusively qualitative methods (Moravcsik, 2012), which likely helps to explain the lower adoption rates in sociology, where such methods are more common (see Figure 3.4).

As Figure 3.3 shows, the timing of increases in the reported adoption of transparent practices across disciplines coincides with notable developments in technology and institutional policy within and across disciplines. With respect to technology, online study registries and pre-registration plan registries seem to be accompanied by upward shifts in adoption. For example, the American Economic Association (AEA) registry was launched in April 2013, and in 2013, the Center for Open Science (COS) online archives allowed for pre-registration posting in economics, psychology and other social science fields. Institutionally, psychology journals began requiring data sharing and code or data posting quite recently, which could explain some of the more rapid trends in that field, whereas the AEA required data posting in 2005, which could partly explain why economics is the social science discipline with the earliest rise in adoption of data and code posting. The interdisciplinary organizations COS and Berkeley Initiative for Transparency in the Social Sciences (BITSS) (Miguel et al., 2014) were founded in 2012, and have been homes for researchers working in all four social science disciplines. These developments in technology and institutions, along with the others labeled in Figure 3.3 as well as many others not mentioned in the figure, accord with theories of normal science and how occasional revolutions in scientific theory and practice take hold (Kuhn, 1962; Hacking, 1981).

Of course, there is also a role for bottom-up adoption rates in which students, faculty, and other researchers take up open science practices through processes of communication with peer networks. In 2012, some of the earliest economics articles using pre-analysis plans were published (Finkelstein et al., 2012; Casey et al., 2012), setting an example that many colleagues followed. It was in 2015, additionally, when a critical mass of blogs and Facebook groups addressed open science practices in psychology, and discussions about open science on Twitter gained momentum around 2016 (Singal, 2016; Huston, 2019). These bottom-up processes of change in attitudes and practices among scholars also likely played a role in driving the technological and institutional changes across disciplines noted above and in Figure 3.3.

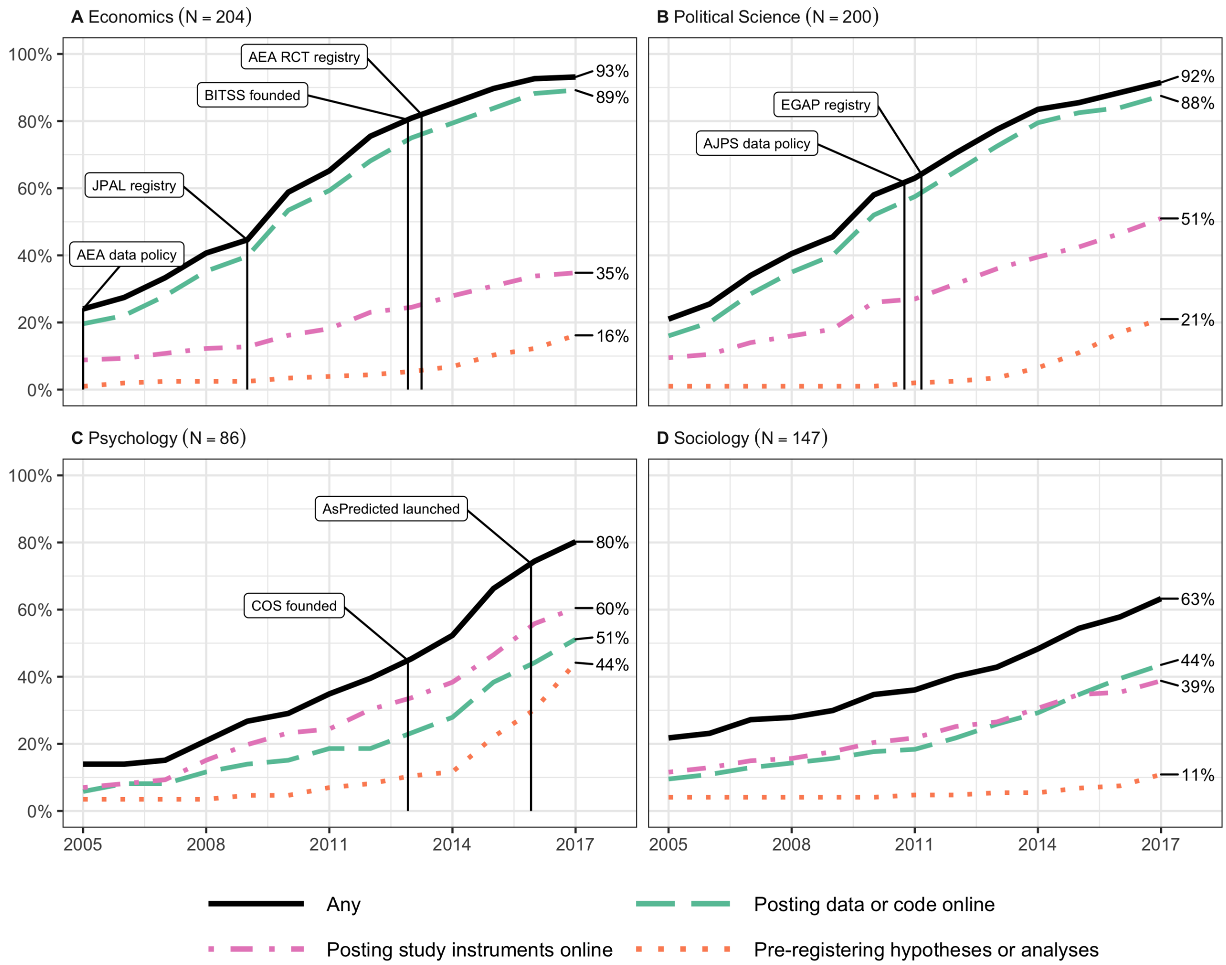
While we are confident in our verification of a subset of respondents' reported adoption, and the resultant bounds we can place around our estimates of disciplinary and overall adoption trends, we acknowledge that reports were based on memory and thus may be imperfect. However, the fact that the slope of the adoption rates correspond to technological and institutional events provides some amount of confidence that they correspond to actual dates of adoption. Moreover, memories of first experiences (e.g., the first time posting data) are often better recalled than later instances (Rubin et al., 1998).

Figure 3.2: Year of adoption of open science practices



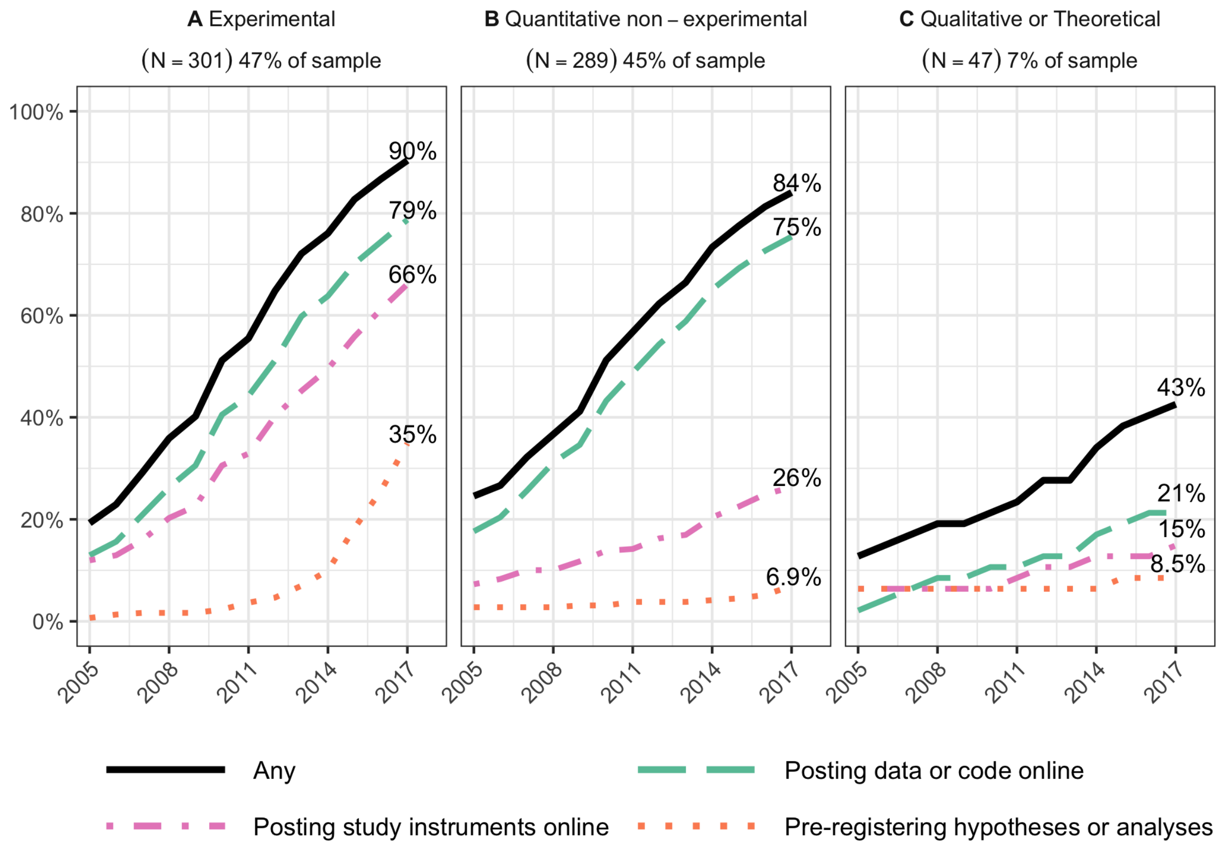
Note: The chart shows for a given year the proportion of Published Authors who had first completed an open science practice in that year or previously. The solid black line shows the proportion of authors who had completed any open science practice by that year. The dashed green line shows the proportion of Published Authors who had posted data or code online by that year. The dash-dotted purple line shows the proportion of Published Authors who had posted study instruments online by that year. The dotted orange line shows the proportion of authors who had pre-registered an analysis or hypothesis by that year. Posting study instruments online is the response to the question “Approximately when was the first time you publicly posted study instruments online?”. Posting data or code online is the response to the question “Approximately when was the first time you publicly posted data or code online?”. Pre-registering hypotheses or analyses is the response to the question “Approximately when was the first time you pre-registered hypotheses or analyses in advance of a study?” The sample is restricted to Published Authors who completed their PhDs by 2009 (N = 637). The bottom of the shaded region is an estimated adoption rate for the entire sample contacted, including non-respondents; the methodology for calculating the adoption rate of non-respondents is outlined in Table 3.1.

Figure 3.3: Year of adoption of open science practices, by discipline



Note: The chart shows for a given year the proportion of Published Authors who had first completed an open science practice in that year or previously, by discipline. The abbreviated names of the organizations used in the labels represent the American Economic Association (AEA), the Abdul Latif Jameel Poverty Action Lab (JPAL), the Berkeley Initiative for Transparency in the Social Sciences (BITSS), the American Economic Association's registry for randomized controlled trials (AEA RCT), the American Journal of Political Science (AJPS), Evidence in Governance and Politics (EGAP), and the Center for Open Science (COS). The organizations mentioned in the figure are included in the panel of the discipline that they work in. BITSS and COS are interdisciplinary organizations, but are included with the discipline they are most associated with.

Figure 3.4: Year of adoption of open science practices, by research focus



Note: The chart shows for a given year the proportion of Published Authors who had first completed an open science practice in that year or previously, categorized by the focus of their research. The classification is based on answers to the question “What methods do you use in your research? Please check all that apply.” If a scholar only selected “Qualitative” or “Theoretical”, they are classified as “Qualitative or Theoretical”; if they selected “Quantitative - Observational” or “Quantitative - Other” but not “Quantitative - Experimental”, they are classified as “Quantitative non-experimental”; if they selected “Quantitative - Experimental”, they are classified as “Experimental”.

3.4 Current Open Science Beliefs & Practices

The data indicate that open science practices are on the rise across four major social science fields, but how supportive of research transparency are scholars today? How much are they currently planning to engage in open science practices? Figure 3.5 suggests that awareness levels of and support for open science practices are high across all four disciplines. Scholars are generally aware of open science practices (for instance, respondents were asked "Have you ever heard of the practice of publicly posting data and code online for a completed study?"), and they are favorably inclined toward them (e.g., "To what extent do you believe that publicly posting data or code online is important for progress in [Discipline]?"). There is not much of a difference between disciplines, apart from sociology researchers having a somewhat lower level of awareness, support, and adoption. Patterns are similar across specific open science practices (see Appendix Tables A.12 - A.20).

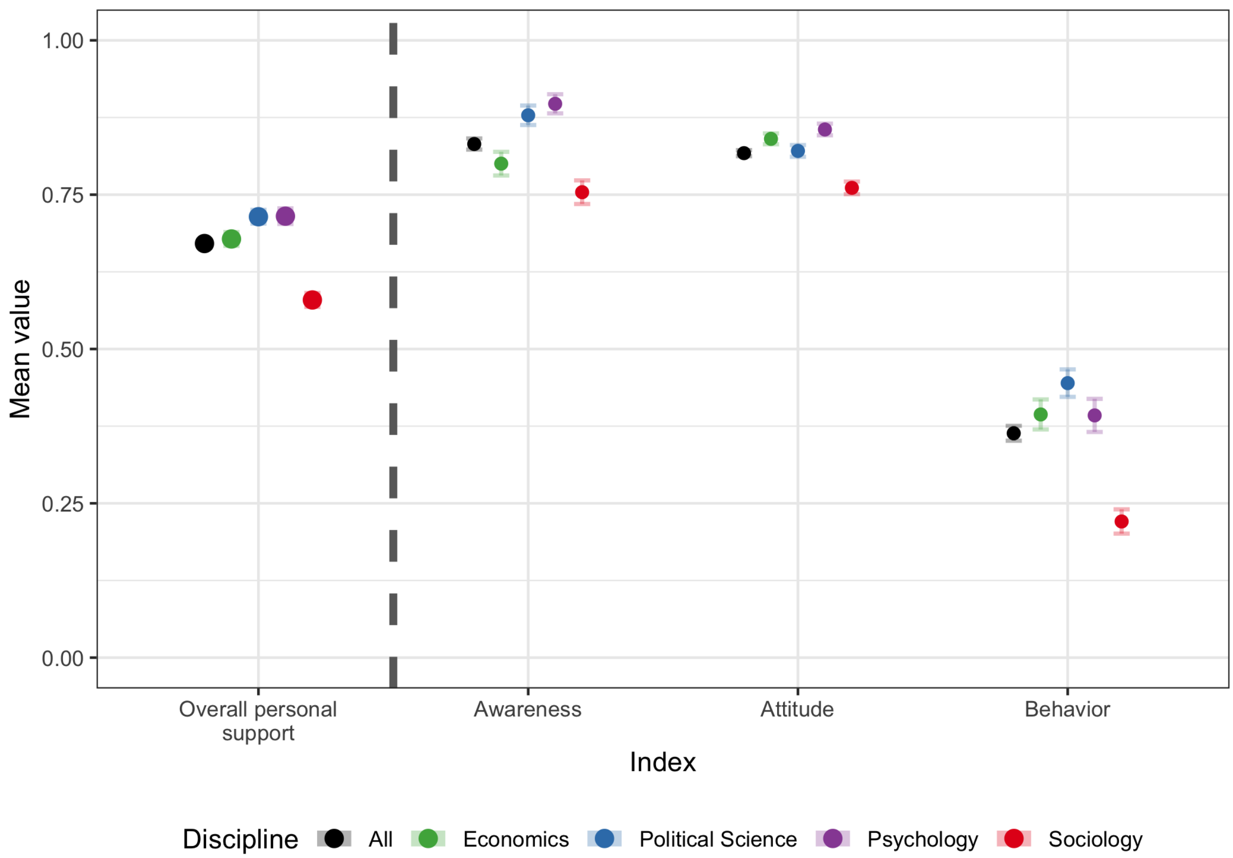
Although comparison across opinion scales and adoption rates is challenging, it appears that actual rates of adoption of open science practices may currently lag behind stated support. It is notable that there are fairly high levels of stated support for open science even among scholars in a discipline like sociology where these tools are not (yet) widely used or taught and where there is a relative lack of institutionalization of these practices.

Perhaps surprisingly, Published Authors and PhD students show similar levels of awareness of and support for open science practices as shown in Appendix Figures A.7 and A.8 respectively. This is in contrast to the authors' prior expectation that PhD Students would exhibit a more supportive attitude toward open science, and suggests that PhD Students may not be the vanguard of changing practices. Open science practices are actually higher among Published Authors, though this is likely because many PhD Students—especially those in their first few years, when they are taking coursework—have not yet had the opportunity to apply the practices to their own work. Researchers across disciplines who use experimental methods show the highest levels of awareness, support, and practice, followed by researchers who use non-experimental quantitative methods. Although qualitative researchers show the least awareness, support, and practice, their awareness and stated support are still at relatively high levels as shown in Appendix Figure A.9.

3.5 Perceived Norms

How do social scientists perceive their fields today, in terms of support for and adoption of open science practices? We measured respondents' perceptions of norms in their disciplines, and compared these perceptions of field-wide opinion and behavior to the average opinion and behavior reported directly in the survey. To measure norms of opinion, we asked respondents to estimate how supportive others in their field are of (1) posting code and data online, and (2) pre-registering hypotheses or analyses in advance of a study. Respondents estimated the percentage of people in their field who fall into each of five opinion categories, ranging from "Not at all in favor" to "Very much in favor," using a dynamic histogram (see Figure 3.6). To

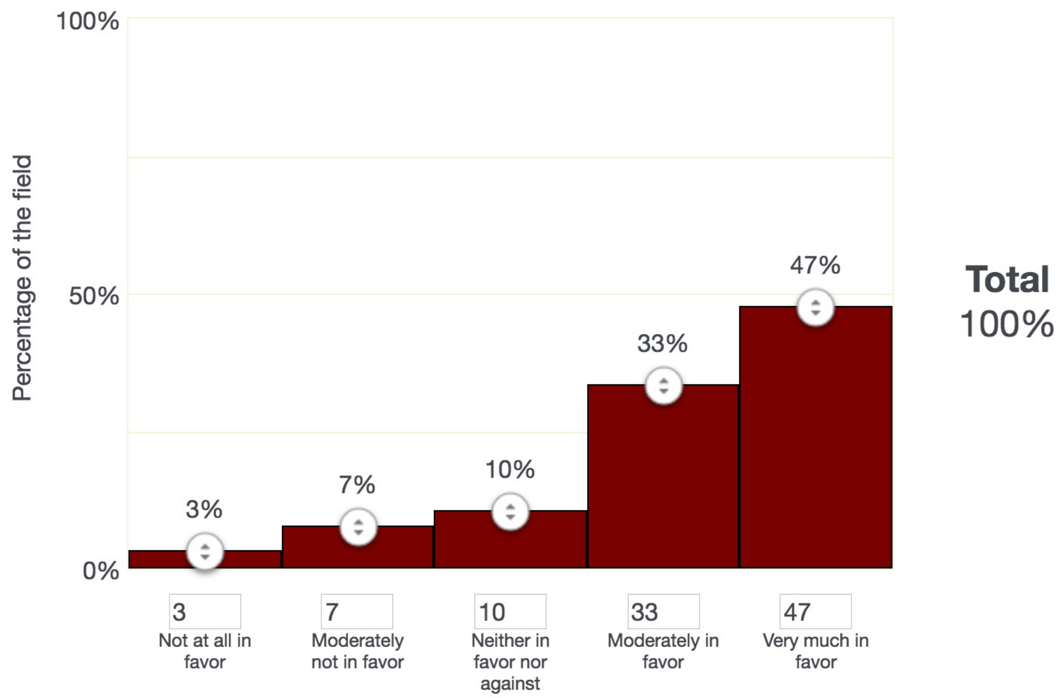
Figure 3.5: Open science awareness, attitudes and behavior, by discipline



Note: Lines around the dots are 95% confidence intervals for the estimates. Awareness is an index comprised of questions related to the respondent's i) Awareness of posting data and code online, ii) Awareness of posting study instruments and iii) Awareness of pre-registration. Behavior is an index comprised of questions related to the respondent's i) Behavior of posting data and code online, ii) Behavior of posting study instruments and iii) Behavior of pre-registration. Attitudes is an index comprised of questions related to the respondent's i) Attitudes of posting data and code online, ii) Attitudes of posting study instruments and iii) Attitudes of pre-registration. Overall Personal Support is an average of the three indices. The questions and methodology that are used to construct the indices can be found in Appendix Table A.7.

measure norms of behavior, we asked respondents to estimate what percentage of researchers in their field actually engage in each of these practices.

Figure 3.6: Dynamic histogram used in the survey



Note: This chart shows the dynamic histogram that survey respondents used to indicate perceived support for open science in their field. Bars need to add up to 100% for respondents to proceed in the survey.

Figure 3.7 depicts scholars' perceptions of their field, in terms of the distribution of opinion about and adoption rates of the two open science practices, against the actual distribution of opinion and adoption rates as reported by survey respondents in their field. Two findings are apparent. First, perception of support, in green, is consistently smaller than actual support—by a substantial amount when considering attitudes toward posting data or code online. Second, perception of opposition toward open science practices is much greater than actual (survey-estimated) opposition, particularly for the case of attitudes toward pre-registration. (Respondents substantially overestimated the proportion of scholars who are indifferent toward posting data or code online, as well).

A second finding depicted in Figure 3.7 is that survey-estimated rates of support for both open-science practices is substantially larger than the rates of actual behavior–particularly when taking into account respondents who said they were either "Very much" and "moderately" in favor of the practice. This pattern is consistent with substantial latent support for adoption of these practices in the four social sciences that may contribute to further rises in adoption rates in future.

While the rates of adoption demonstrated by our previous measures may or may not have seemed surprising to readers, these data show that the high adoption rate of open science practices would be surprising to our survey respondents, who appear to significantly underestimate open science adoption and support.

There are various possible explanations for why respondents appear to be more in favor of data posting and pre-registration than they believe others in their field to be. One immediate possibility is that our survey sample is selected and unrepresentative in important ways. For instance, we selected respondents based on their publication history in leading research journals and among the most highly-ranked PhD programs, and these populations are not representative of the entire discipline about which respondents are making estimates. Of course, this subgroup of "elite" scholars may be particularly influential in driving the change of social norms in the discipline. Moreover, those who chose to respond to our survey invitation may be more supportive of open science than non-respondents, further shifting sample means, although the evidence we presented above from the audit activity suggests this is less likely.

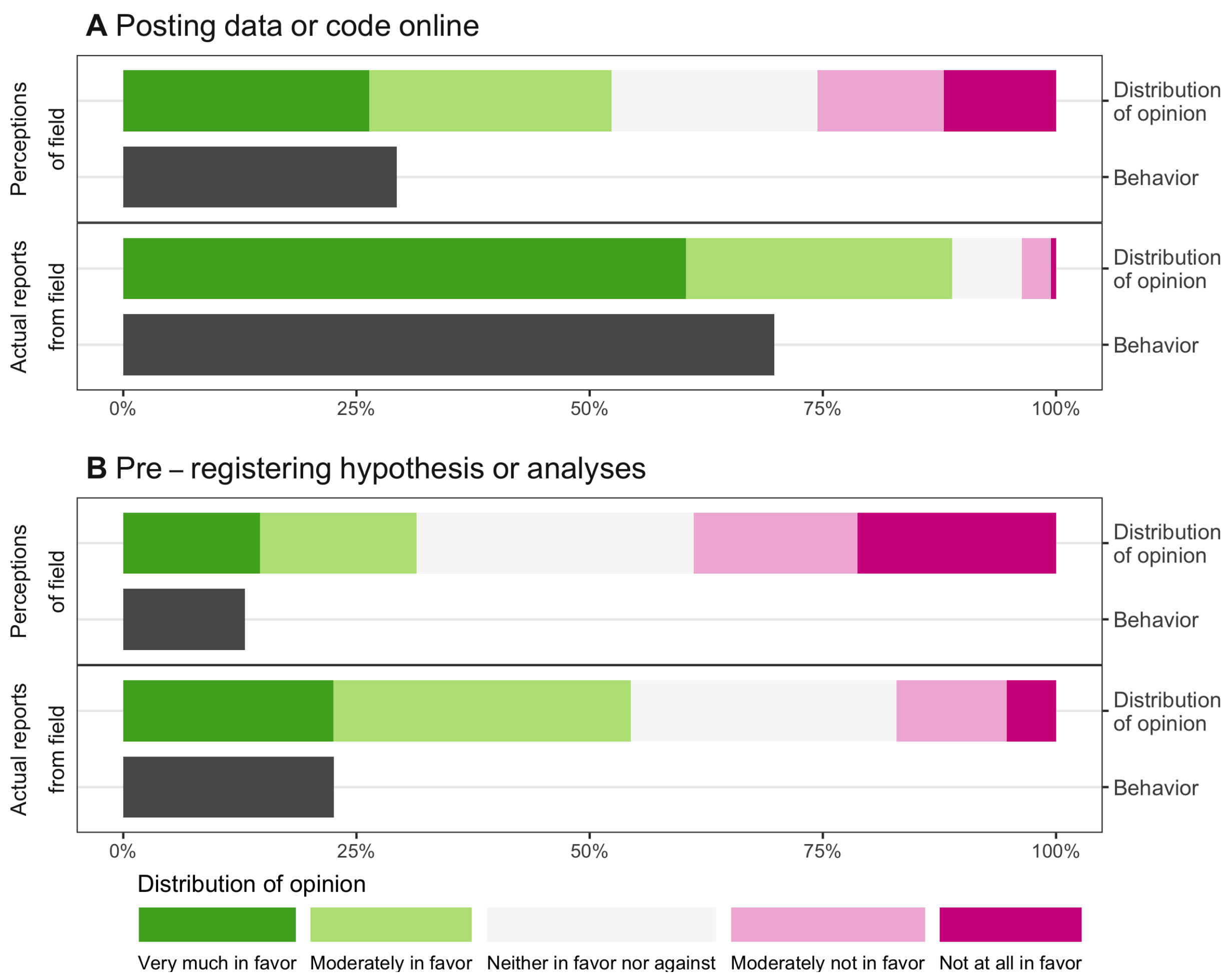
Another explanation is that respondents are over-reporting their support for open science for reasons of self or social image. However, admitting some social desirability toward responding favorably about open science in an anonymous survey seems to support the idea that a relatively strong social norm in favor of open science has already developed, as suggested in the rates of "actual reports from the field" in Figure 3.7. The figure shows that the median respondent is in favor of these practices. This interpretation suggests a social norm in favor of open science at work, even if practices lag behind the ideal. Similarly, the social science research community could be in a period of rapid methodological change, in which case we might expect that beliefs about practices could be temporarily out of sync with actual behaviors. For instance, scholars' views about the state of open science in their discipline could be shaped by their own experiences during their graduate training, or based in part on current journal publications, but both would only capture actual attitudes and practices in the field with a lag.

This set of analyses is consistent with the idea of a current cultural shift in social science research communities, in which behaviors and attitudes are already changing and community members are partially attuned to the change.

3.6 Discussion

Data from a recent representative survey of scholars in four large social science disciplines – economics, political science, psychology, and sociology – indicates that the adoption of open science practices has been increasing rapidly over the past decade. Behaviors such as posting data and materials that were nearly unknown in some fields as recently as 2005 are now practiced by the majority of scholars. Other newer practices, such as study pre-registration,

Figure 3.7: Perceived and actual support for open science among Published Authors



Note: The chart shows differences between perceived and actual support for two practices: posting data or code online and pre-registering hypotheses or analyses. The sample is restricted to Published Authors; the analogous data for Ph.D. students are presented in Appendix Figure A.10. Within each panel, the first bar shows the perceived distribution of support for the practice among Published Authors. This is constructed by asking individuals what percentage of researchers in their field they believe fall into each opinion category, and then averaging over their responses. The solid black bar below shows the fraction of researchers in their field they believe have done the practice. The third bar in the panel shows the distribution of support for the practice constructed using the responses elicited from the Published Authors that we sampled. The final solid black bar shows the proportion of researchers who have actually done the stated practices, using the responses elicited from our survey. Colors indicate the level of support, with green indicating more and red indicating less support. Adjusting the behavior figures to account for non-respondents (using the same methodology as in Figure 3.2) we find that the adjusted share of Published Authors posting data is 64.3% and the adjusted share of Published Author's posting pre-analysis plans is 14%.

have experienced a sharp rise in adoption just in recent years, especially among scholars who engage in experimental research. While trends are similar to other fields, overall levels of adoption are lowest in sociology. Contrary to our expectations, there is no clear evidence of a generational shift, or of an old guard standing in the way of change: attitudes toward open science practices are remarkably similar among both PhD Students and more established Published Authors. The high levels of support for open science practices expressed among our respondents indicates that the classic scientific ethos famously described by Merton (1973) is alive among today's social scientists. A data validation activity confirms that self-reported behaviors are strongly related to actual behavior, and that the selection of survey respondents into the sample has not produced misleading results.

The second main finding of the analysis is that stated support for open science practices is outpacing both their actual adoption and respondents' beliefs about others' support. Taken together, this pattern suggests that social science research communities are in a period of rapid transformation in terms of their research practices, a shift that is not yet entirely appreciated by the community. To follow this co-evolution of behavioral adoption, awareness, and support for open science practices, we plan to collect additional rounds of the 3S survey in the future. These representative snapshots of open science adoption and perception, we argue, can describe the state of the social sciences from the perspective of whether they are currently in the type of transition state described by historians of science as a shift out of "normal" science into one of crisis and eventual transformation (Kuhn, 1962; Hacking, 1981).

Acknowledgments

All authors contributed equally to the research, and the order in which the authors' names appear has been randomized using the American Economic Association's Author Randomization Tool (Ray and Robson, 2018). We would like to thank Katie Hoeberling, Fernando Hoces de la Guardia, Aleks Bogdanoski, Kelsey Mulcahy, and seminar participants at Princeton, University of California, Berkeley, and the BITSS Annual Meeting for many helpful discussions and suggestions. We are grateful to Audrey Chebet, Jason Chin, Joel Ferguson, Jing Kai Ong, John-Henry Pezzuto, Somara Sobharwal, and Simon Zhu for excellent research assistance, and to Alan Ritari and the Agathon Group for excellent web design and support. This research received human subjects approval from Institutional Review Boards at Princeton University and the University of California, Berkeley. The project received generous funding from an anonymous donor. The authors have no conflicts of interest to disclose. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All errors remain our own.

Bibliography

[1] Melissa S. Anderson, Brian C. Martinson, and Raymond De Vries. “Normative Dissonance in Science: Results from a National Survey of U.S. Scientists”. *Journal of Empirical Research on Human Research Ethics* 2.4 (2007), pp. 3–14.

[2] Michael L Anderson. “Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects”. *Journal of the American statistical Association* 103.484 (2008), pp. 1481–1495.

[3] James Andreoni and B. Douglas Bernheim. “Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects”. *Econometrica* 77.5 (2009), pp. 1607–1636.

[4] James Andreoni and A. Abigail Payne. “Do Government Grants to Private Charities Crowd Out Giving or Fund-Raising?” *American Economic Review* 93.3 (May 2003), pp. 792–812.

[5] Dan Ariely, Anat Bracha, and Stephan Meier. “Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially”. *American Economic Review* 99.1 (Mar. 2009), pp. 544–555.

[6] Ned Augenblick, Muriel Niederle, and Charles Sprenger. “Working over Time: Dynamic Inconsistency in Real Effort Tasks”. *The Quarterly Journal of Economics* 130.3 (Aug. 2015), pp. 1067–1115.

[7] Monya Baker. “1,500 Scientists Lift the Lid on Reproducibility”. *Nature* 533.7604 (2016), pp. 452–454.

[8] Roland Bénabou and Jean Tirole. “Incentives and Prosocial Behavior”. *American Economic Review* 96.5 (Dec. 2006), pp. 1652–1678.

[9] Yoav Benjamini, Abba M Krieger, and Daniel Yekutieli. “Adaptive linear step-up procedures that control the false discovery rate”. *Biometrika* 93.3 (2006), pp. 491–507.

[10] Yudhijit Bhattacharjee. “Diederik Stapel’s Audacious Academic Fraud”. *The New York Times. Magazine* (2013).

[11] Denny Borsboom and Eric-Jan Wagenmakers. “Derailed: The Rise and Fall of Diederik Stapel”. *APS Observer* 26.1 (2012).

[12] David Broockman, Joshua Kalla, and Peter Aronow. "Irregularities in LaCour (2014)". https://stanford.edu/~dbroock/broockman_kalla_aronow_lg_irregularities.pdf. 2015.

[13] Brett T Buttliere. "Using science and psychology to improve the dissemination and evaluation of scientific work". *Frontiers in computational neuroscience* 8 (2014), p. 82.

[14] Benedict Carey. "Noted Dutch Psychologist, Stapel, Accused of Research Fraud". *The New York Times* (2011).

[15] Katherine Casey, Rachel Glennerster, and Edward Miguel. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan". *The Quarterly Journal of Economics* 127.4 (2012), pp. 1755–1812.

[16] Garret Christensen, Jeremy Freese, and Edward Miguel. *Transparent and Reproducible Social Science Research: How to Do Open Science*. Oakland: University of California Press, 2019.

[17] Garret Christensen and Edward Miguel. "Transparency, Reproducibility, and the Credibility of Economics Research". *Journal of Economic Literature* 56.3 (2018), pp. 920–980.

[18] Aline Claesen, Sara Lucia B. T. Gomes, Francis Tuerlinckx, and Wolf Vanpaemel. "Preregistration: Comparing Dream to Reality" (2019).

[19] Lucas C. Coffman and Muriel Niederle. "Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible". *Journal of Economic Perspectives* 29.3 (2015), pp. 81–98.

[20] Jason Dana, Roberto A. Weber, and Jason Xi Kuang. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness". *Economic Theory* 33.1 (Oct. 2007), pp. 67–80.

[21] Russell Davidson and Jean-Yves Duclos. "Testing for Restricted Stochastic Dominance". *Econometric Reviews* 32.1 (Jan. 2013), pp. 84–125.

[22] Y. A. de Vries et al. "The Cumulative Effect of Reporting and Citation Biases on the Apparent Efficacy of Treatments: The Case of Depression". *Psychological Medicine* 48.15 (2018), pp. 2453–2455.

[23] Walter Enders and Gary A. Hoover. "Whose Line Is It? Plagiarism in Economics". *Journal of Economic Literature* 42.2 (2004), pp. 487–493.

[24] Christine Exley. "Incentives for Prosocial Behavior: The Role of Reputations". *Management Science* (Mar. 2017).

[25] Armin Falk. "Gift Exchange in the Field". *Econometrica* 75.5 (2007), pp. 1501–1511.

[26] Tom Feilden. *Most Scientists 'Can't Replicate Studies'*. https://www.bbc.com/news/science-environment-39054778. 2017.

[27] Amy Finkelstein et al. "The Oregon Health Insurance Experiment: Evidence from the First Year". *The Quarterly Journal of Economics* 127.3 (2012), pp. 1057–1106.

[28] Daniel T. Gilbert, Gary King, Stephen Pettigrew, and Timothy D. Wilson. "Comment on "Estimating the Reproducibility of Psychological Science"". *Science* 351.6277 (2016), pp. 1037–1037.

[29] Susan Goldin-Meadow. "Why Preregistration Makes Me Nervous". *APS Observer* 29 (2016).

[30] Zvi Griliches. "Hybrid Corn: An Exploration in the Economics of Technological Change". *Econometrica* 25.4 (1957), pp. 501–522.

[31] Zachary Grossman and Joël J. van der Weele. "Self-Image and Willful Ignorance in Social Decisions". *Journal of the European Economic Association* 15.1 (Dec. 2016), pp. 173–217.

[32] Ian Hacking. "Do We See Through a Microscope?" *Pacific Philosophical Quarterly* 62.4 (1981), pp. 305–322.

[33] Matt Huston. *A Revolution Is Happening in Psychology. Here's How It's Playing Out.* https://www.psychologytoday.com/articles/201905/revolution-is-happening-in-psychology-heres-how-its-playing-out. 2019.

[34] Leslie K John, George Loewenstein, and Drazen Prelec. "Measuring the prevalence of questionable research practices with incentives for truth telling". *Psychological science* 23.5 (2012), pp. 524–532.

[35] Robert M. Kaplan and Veronica L. Irvin. "Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time". *PLOS ONE* 10.8 (2015), e0132382.

[36] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 1962.

[37] Kai Kupferschmidt. "A Recipe for Rigor". *Science* 361.6408 (2018), pp. 1192–1193.

[38] George J. Mailath. "Incentive Compatibility in Signaling Games with a Continuum of Types". *Econometrica* 55.6 (Nov. 1987), p. 1349.

[39] Robert K. Merton. "The Normative Structure of Science". *The Sociology of Science: Theoretical and Empirical Investigations*. Ed. by Norman W. Storer. Chicago: University of Chicago Press, 1973, pp. 233–281.

[40] Edward Miguel et al. "Promoting Transparency in Social Science Research". *Science* 343.6166 (2014), pp. 30–31.

[41] Andrew Moravcsik. "Active Citation and Qualitative Political Science". *Qualitative and Multi-Mmethod Research* 10.1 (2012), pp. 33–37.

[42] Neuroskeptic. "The Nine Circles of Scientific Hell". *Perspectives on Psychological Science* 7.6 (2012), pp. 643–644.

[43] B. A. Nosek et al. “Promoting an Open Research Culture”. *Science* 348.6242 (2015), pp. 1422–1425.

[44] Open Science Collaboration. “Estimating the Reproducibility of Psychological Science”. *Science* 349.6251 (2015), aac4716.

[45] Heather A. Piwowar and Todd J. Vision. “Data Reuse and the Open Data Citation Advantage”. *PeerJ* (2013), :e175.

[46] Debraj Ray and Arthur Robson. “Certified Random: A New Order for Coauthorship”. *American Economic Review* 108.2 (2018), pp. 489–520.

[47] Paul M. Romer. “Endogenous Technological Change”. *Journal of Political Economy* 98.5 (1990), S71–S102.

[48] David C. Rubin, Tamara A. Rahhal, and Leonard W. Poon. “Things Learned in Early Adulthood Are Remembered Best”. *Memory & Cognition* 26.1 (1998), pp. 3–19.

[49] Uri Simonsohn. “Just Post It: The Lesson From Two Cases of Fabricated Data Detected by Statistics Alone”. *Psychological Science* 24.10 (2013), pp. 1875–1888.

[50] Jesse Singal. *Inside Psychology’s ‘Methodological Terrorism’ Debate*. https://web.archive.org/web/20190605161448/https://www.thecut.com/2016/10/inside-psychologys-methodological-terrorism-debate.html. 2016.

[51] Marcel A. L. M. van Assen, Robbie C. M. van Aert, and Jelte M. Wicherts. “Meta-Analysis Using Effect Size Distributions of Only Statistically Significant Studies”. *Psychological Methods* 20.3 (2015), pp. 293–309.

Appendix A

Appendix

A.1 Appendices for Chapter 1

Proofs

Lemma 1. *$a^*(\theta)$ is non-decreasing, that is $\theta' > \theta \Rightarrow \inf a^*(\theta') \geq \sup a^*(\theta)$.*

Proof. Since U depends on θ only through g, the single-crossing property (SCP) of g transfers to U. Let $\theta' > \theta$, $a \in a^*(\theta)$ and $a' \in a^*(\theta')$. By (IC) $U(\theta, a, r^*(a)) \geq U(\theta, a', r^*(a'))$, and $U(\theta', a', r^*(a')) \geq U(\theta', a, r^*(a))$. If $a' < a$, then (SCP) implies $U(\theta', a, r^*(a)) > U(\theta', a', r^*(a'))$, a contradiction. □

Lemma 2. *If $\theta' > \theta$, then θ' has a strictly larger incentive than θ to deviate to $a > a^*(\theta')$, and a strictly weaker incentive than θ to deviate to $a < a^*(\theta)$.*

Proof. For θ the benefit of deviating from $a^*(\theta)$ to a is the benefit of deviating from $a^*(\theta)$ to $a^*(\theta') \geq a^*(\theta)$ plus the benefit of deviating from $a^*(\theta')$ to a. The former is non-positive by (IC). The latter is also the total benefit of deviating for θ', and is strictly larger for θ' than for θ by (SCP). For the case of $a < a^*(\theta)$, (SCP) implies that θ' has a strictly larger incentive to move from a to $a^*(\theta)$ than θ. Moving further from $a^*(\theta)$ to $a^*(\theta')$ then weakly benefits θ' by (IC). □

Lemma 3. *$r^*(a)$ is non-decreasing.*

Proof. Choose two actions $a' > a$.

Case 1: Both actions are played in equilibrium. Take $\theta \in {a^*}^{-1}(a)$ and $\theta' \in {a^*}^{-1}(a')$. Note that $\sup a^*(\theta') > \inf a^*(\theta)$, and so by lemma 1 we must have $\theta' \geq \theta$ and hence $\inf {a^*}^{-1}(a') \geq \sup {a^*}^{-1}(a)$.

Case 2: Only one action is played in equilibrium. Take $\theta \in {a^*}^{-1}(a)$ and any level of reputation $r \in \Theta$. By lemma 2, if $\theta' < \theta$ has a weak incentive to deviate to a' for a given level $r(a') = r$, then θ will have a strong incentive to do so. Since this holds for any

$\theta \in a^{*-1}(a)$, the D1 criterion implies $r(a') \geq \sup a^{*-1}(a) \geq r(a)$.[1] If $a^{*-1}(a) = \emptyset$, then the analogue argument for $\theta' \in a^{*-1}(a')$ implies $r(a) \leq \inf a^{*-1}(a') \leq r(a')$

Case 3: Neither action is played in equilibrium. If there exists an action $a'' \in (a, a')$ that is played in equilibrium, combine the previous cases to obtain $r(a') \geq r(a)$. If $a^{*-1}(a) = \emptyset$ over $[a, a']$, then because the support of θ is continuous, there exists θ_0 such that either $a^*(\theta_0) < a$ and $\lim_{\theta\downarrow\theta_0} a^*(\theta) \geq a'$, or $a^*(\theta_0) > a'$ and $\lim_{\theta\uparrow\theta_0} a^*(\theta) \leq a$. But then either $\sup a^{*-1}(a^*(\theta_0)) = \theta_0 = \lim_{\theta\downarrow\theta_0} \inf a^{*-1}(a^*(\theta))$ or $\lim_{\theta\uparrow\theta_0} \sup a^{*-1}(a^*(\theta)) = \theta_0 = \inf a^{*-1}(a^*(\theta_0))$. Applying case 2 to θ_0 and every element of $\theta \to \theta_0$ then implies $r(a) = r(a') = \theta_0$. □

Result 1 (No Bunching)**.** *There does not exist a D1-equilibrium with any bunching.*

Proof. Assume that there is an action a at which types bunch so that $\sup a^{*-1}(a) > \inf a^{*-1}(a)$ and hence $r^*(a) < \sup a^{*-1}(a)$. Then we can take a bunching type $\theta \in a^{*-1}(a)$ with $\theta > r^*(a)$, and an action $a' > a$ to which θ might deviate.

If $a^{*-1}(a') \neq \emptyset$, then $\theta' \geq \theta$ for any $\theta' \in a^{*-1}(a')$, because $a^*(\theta)$ is non-decreasing. Therefore, $r^*(a') \geq \inf a^{*-1}(a') \geq \sup a^{*-1}(a) > r^*(a)$, where the first and last inequalities follow from the assumptions on $R(\beta)$.

If $a^{*-1}(a') = \emptyset$, then by lemma 2 any $\theta' < \theta$ has a smaller incentive than θ to deviate to a', and hence the D1 criterion implies $r^*(a') \geq \theta > r^*(a)$.

Since $U(\theta, a, r(a)) - \mu r(a)$ is right-continuous in a,

$$\lim_{a'\downarrow a} U(\theta, a', r^*(a')) - U(\theta, a, r^*(a)) = \lim_{a'\downarrow a} \mu (r^*(a') - r^*(a)) > 0$$

making it optimal for θ to deviate upwards from a. □

Lemma 4. *Define $\theta_1^* \equiv a^{*-1}(\overline{a})$. All types play pure strategies, except for θ_1^* who is possibly mixing between an action $a_1^* < \overline{a}$ and the bonus threshold $\overline{a}$.*

Proof. Assume that $a, a' \in a^*(\theta)$ with $a' > a$. Then by lemma 3 and result 1 $r(a'') = \theta$ for any $a'' \in [a, a']$. For θ a and a' are both optimal, so $g(\theta, a) + b\,\mathbb{1}\{\mathrm{a} \geq \overline{\mathrm{a}}\} = g(\theta, a') + b\,\mathbb{1}\{\mathrm{a'} \geq \overline{\mathrm{a}}\}$. If $a \geq \overline{a}$ or $a' < \overline{a}$, then $g(\theta, a) = g(\theta, a')$ and by strict concavity $g(\theta, a'') > g(\theta, a)$ for any $a'' \in (a, a')$, making it a profitable deviation.

If $a < \overline{a} < a'$, then $g(\theta, a) > g(\theta, a')$ and by strict concavity $g(\theta, \overline{a}) > g(\theta, a')$, so that $\overline{a}$ is a profitable deviation for θ. □

Lemma 5 (Mailath (1987))**.** *In any separating equilibrium, if $a^*(\theta)$ is (right-)continuous at θ, then $a^*(\theta)$ is (right-)differentiable, and satisfies the differential equation*

$$\frac{\partial a^*}{\partial \theta}(\theta) = \frac{\mu}{-g_2(\theta, a^*(\theta))} \qquad \text{(DE)}$$

[1] If $M_<(\theta|a) = \emptyset$ for all $(\theta, r) \in \Theta^2$, then WLOG set $r^*(a) = \underline{\theta}$ for $a < a^*(\underline{\theta})$ and $r^*(a) = \overline{\theta}$ for $a > a^*(\overline{\theta})$.

Proof. The two-sided case is Proposition 2 of Mailath (1987). I replicate the proof here. The proof can be applied to the one-sided case without any modification.

Define $h(\theta, a, r) \equiv U(\theta, a, r) - U(\theta, a^*(\theta_0), \theta_0)$, $[\theta; \lambda]_1 \equiv (\lambda\theta_0 + (1-\lambda)\,\theta, a^*(\theta), \theta)$ and $[\theta, \kappa]_{23} \equiv (\theta_0, \kappa a^*(\theta_0) + (1-\kappa)\,a^*(\theta), \kappa\theta_0 + (1-\kappa)\theta)$. Making (one-sided) Taylor approximations of $h(\theta, a^*(\theta), \theta)$ around $(\theta_0, a^*(\theta), \theta)$ and of $h_1(\theta_0, a^*(\theta), \theta)$ around $(\theta_0, a^*(\theta_0), \theta_0)$ yields

$$\begin{aligned} h(\theta, a^*(\theta), \theta) &= h(\theta_0, a^*(\theta), \theta) + h_1(\theta_0, a^*(\theta), \theta)(\theta - \theta_0) + \frac{1}{2} h_{11}([\theta; \lambda]_1)](\theta - \theta_0)^2 \\ h_1(\theta_0, a^*(\theta), \theta) &= h_1(\theta_0, a^*(\theta_0), \theta_0) + h_{12}\left([\theta; \kappa]_{23}\right)\left(a^*(\theta) - a^*(\theta_0)\right) \\ &\quad + h_{13}\left([\theta; \kappa]_{23}\right)(\theta - \theta_0) \end{aligned}$$

for some $(\lambda, \kappa) \in [0,1]^2$. Note that $h_1(\theta_0, a^*(\theta_0), \theta_0) = 0$ and $h_{13} = 0$. Also (IC) implies $h(\theta, a^*(\theta), \theta) \geq 0$ and $h(\theta_0, a^*(\theta), \theta) \leq 0$. Combining yields

$$\begin{aligned} 0 &\geq h(\theta_0, a^*(\theta), \theta) \\ &\geq -(\theta - \theta_0)\left[h_{12}\left([\theta; \kappa]_{23}\right)\left(a^*(\theta) - a^*(\theta_0)\right) + \frac{1}{2} h_{11}([\theta; \lambda]_1)](\theta - \theta_0)\right] \end{aligned}$$

A Taylor approximation of $h(\theta_0, a^*(\theta), \theta)$ around $h(\theta_0, a^*(\theta_0), \theta_0) = 0$ yields

$$\begin{aligned} 0 &\geq h_2(\theta_0, a^*(\theta_0), \theta_0)\left(a^*(t) - a^*(\theta_0)\right) + \mu\,(\theta - \theta_0) \\ &\quad + \frac{1}{2} h_{22}([\theta; \gamma]_{23})\left(a^*(\theta) - a^*(\theta_0)\right)^2 \\ &\geq -(\theta - \theta_0)\left[h_{12}\left([\theta; \kappa]_{23}\right)\left(a^*(\theta) - a^*(\theta_0)\right) + \frac{1}{2} h_{11}([\theta; \lambda]_1)](\theta - \theta_0)\right] \end{aligned} \tag{A.1}$$

using $h_3 = \mu$ and $h_{23} = h_{33} = 0$. Since g is twice continuously differentiable, h_{11}, h_{12} and h_{22} are locally bounded, and so $\lim_{\theta \to \theta_0} h_{22}\left(a^*(\theta) - a^*(\theta_0)\right)^2 = 0$. Then dividing by $\theta - \theta_0$ and letting $\theta \to \theta_0$ yields

$$0 \geq g_2(\theta_0, a^*(\theta_0), \theta_0) \lim_{\theta \to \theta_0} \frac{a^*(\theta) - a^*(\theta_0)}{\theta - \theta_0} + \mu \geq 0$$

which implies differentiability of a^* at θ_0. If a^* is only right-continuous at θ_0, use the analogue right-sided derivatives and Taylor approximations with $\theta > \theta_0$. □

Lemma 6. *If $a^*(\theta)$ is continuous at $\theta_0 \neq \theta_1^*$, then $g_2(\theta_0, a^*(\theta_0)) \neq 0$.*

Proof. Suppose $g_2(\theta_0, a^*(\theta_0)) = 0$. Then $g_{22}(\theta_0, a^*(\theta_0)) < 0$. Plugging into *eq.* (A.1) and dividing by $\theta - \theta_0 < 0$ yields

$$0 \leq \mu + \frac{1}{2} h_{22}([\theta; \gamma]_{23}) \frac{\left(a^*(\theta) - a^*(\theta_0)\right)^2}{\theta - \theta_0} \tag{A.2}$$

$$\leq -h_{12}\left([\theta; \kappa]_{23}\right)\left(a^*(\theta) - a^*(\theta_0)\right) - \frac{1}{2} h_{11}([\theta; \lambda]_1)](\theta - \theta_0) \tag{A.3}$$

However, the right-hand side converges to zero, but $h_{22} = g_{22} < 0$ and $\theta < \theta_0$, contradicting the inequality. □

Lemma 7. *$a^*(\underline{\theta}) = a^{FB}(\underline{\theta})$, where $a^{FB}(\theta) \equiv \arg\max_{a \geq \underline{a}} U(\theta, a, \theta)$*

Proof. By result 1, $\underline{\theta}$ will obtain $U(\underline{\theta}, a^*(\underline{\theta}), \underline{\theta})$ in equilibrium. Since $r^*(a) \geq \underline{\theta}$, deviating to $a^{\mathrm{FB}}(\underline{\theta})$ does not carry a loss in reputational benefit, and by the strict concavity of g is always strictly profitable. □

Lemma 8. *$U(\theta_0, a^*(\theta), \theta_0)$ is continuous at θ_0.*

Proof. Let $\varepsilon > 0$. Since $U(\theta, a, r)$ is continuous in θ and r we can find $\delta > 0$ such that $|U(\theta_0, a^*(\theta_0), \theta) - U(\theta_0, a^*(\theta_0), \theta_0)| < \varepsilon$, $|U(\theta, a^*(\theta_0), \theta) - U(\theta_0, a^*(\theta_0), \theta_0)| < \varepsilon/2$, and $|U(\theta, a^*(\theta_0), \theta_0) - U(\theta_0, a^*(\theta_0), \theta_0)| < \varepsilon/2$ for all $|\theta - \theta_0| < \delta$. Then, using result 1 and (IC)

$$U(\theta_0, a^*(\theta_0), \theta_0) \geq U(\theta_0, a^*(\theta), \theta) > U(\theta_0, a^*(\theta), \theta_0) - \varepsilon$$
$$U(\theta_0, a^*(\theta), \theta_0) > U(\theta, a^*(\theta), \theta) - \varepsilon/2 \geq U(\theta, a^*(\theta_0), \theta_0) - \varepsilon/2 > U(\theta_0, a^*(\theta_0), \theta_0) - \varepsilon$$

yielding $|U(\theta_0, a^*(\theta), \theta_0) - U(\theta_0, a^*(\theta_0), \theta_0)| < \varepsilon$ for all $|\theta - \theta_0| < \delta$.

□

Lemma 9. *$a^*(\theta)$ is continuous at all $\theta \neq \theta_1^*$, and right-continuous at $\theta = \theta_1^*$.*

Proof. Because a^* is strictly increasing, all discontinuities are jump discontinuities, and their set is at most countable. Suppose that a^* is discontinuous at $\theta_0 \notin \{\underline{\theta}, \theta_1^*\}$, so that $a^- \equiv \lim_{\theta \uparrow \theta_0} a^*(\theta) \neq \lim_{\theta \downarrow \theta_0} a^*(\theta) \equiv a^+$.

By lemma 8, $U(\theta_0, a^-, \theta_0) = U(\theta_0, a^+, \theta_0) = U(\theta_0, a^*(\theta_0), \theta_0)$ for any values of a^- and a^+. Since $\theta \neq \theta_1^*$, a^-, a^+ and $a^*(\theta_0)$ are on the same side of the bonus threshold $\overline{a}$, and so $g(\theta_0, a^-) = g(\theta_0, a^+) = g\left(\theta_0, a^*(\theta_0)\right)$. Since g is strictly concave, either $a^- = a^*(\theta_0)$ or $a^+ = a^*(\theta_0)$, and

$$a^- < a^0(\theta_0) < a^+ \tag{A.4}$$

But then, because $r^*(a^*(\theta)) = \theta$ and r^* is non-decreasing, θ_0 can strictly increase their utility by deviating to $a^0(\theta_0)$, contradicting (IC).

If $\theta_0 = \underline{\theta}$, then by lemma 8 and since $a^*(\underline{\theta}) = a^{\mathrm{FB}}(\underline{\theta})$ is the unique maximizer of $U(\underline{\theta}, a^*(\theta), \underline{\theta})$, it must that $\lim_{\theta \downarrow \underline{\theta}} a^*(\theta) = a^*(\underline{\theta})$ implying continuity.

If a^* is not right-continuous at $\theta_0 = \theta_1^*$, then $\overline{a} < a^+$. However lemma 8 implies that $g(\theta_1^*, \overline{a}) = g(\theta_1^*, a^+)$. At the same time for θ_1^* to not deviate upwards, we must have $g(\theta_1^*, \overline{a}) > g(\theta_1^*, a)$ for any $a > \overline{a}$, a contradiction. □

Lemma 10. *$a^*(\theta)$ is differentiable on $\Theta \setminus \{\underline{\theta}, \theta_1^*\}$ and satisfies* (DE).

Proof. Immediately from result 1 and lemmas 5 and 9. □

Result 2 (Existence and Uniqueness)**.** *There exists a unique fully separating D1-equilibrium.*

Proof. The equilibrium is characterized by (DE) and the two initial conditions $a^*(\underline{\theta}) = a^{\mathrm{FB}}(\underline{\theta})$ for $\theta < \theta_1^*$ and initial condition $a^*(\theta_1^*) = \overline{a}$ for $\theta \geq \theta_1^*$. The marginal buncher is determined by the indifference equation

$$g(\theta_1^*, a_1^*) = g(\theta_1^*, \overline{a}) + b \tag{A.5}$$

where $a_1^* = \lim_{\theta \uparrow \theta_1^*} a^*(\theta)$. Since g is continuous, satisfies (SCP) and is strictly concave, θ_1^* exists and is unique. The solutions to (DE) exist and are unique by standard results on differential equations. $\square$

Result 3 (Anti-Bunching)**.** *An increase in the bonus size b increases the equilibrium action for all types above $\overline{a}$. Formally, if $b' > b$, then $a_b^*(\theta) \geq \overline{a} \Rightarrow a_{b'}^*(\theta) > a_b^*(\theta)$.*

Proof. Using the implicit function theorem on the indifference equation yields

$$\begin{aligned}
\frac{\mathrm{d}\theta_1^*(b)}{\mathrm{d}b} &= \frac{1}{g_1(\theta_1^*, a_1^*) + g_2(\theta_1^*, a_1^*)\frac{\partial a^*}{\partial \theta}(\theta_1^*) - g_1(\theta_1^*, \overline{a})} \\
&= \frac{1}{g_1(\theta_1^*, a_1^*) - \mu - g_1(\theta_1^*, \overline{a})} < 0
\end{aligned}$$

Hence, increasing b lowers θ_1^*, and therefore a^* needs to shift up at $\theta_1^*(b') < \theta_1^*(b)$ so that it satisfies $a^*(\theta_1^*(b')) = \overline{a}$. By standard results on differential equations, solutions cannot cross, so $a_{b'}^*(\theta) > a_b^*(\theta)$ for all $\theta \geq \theta_1^*(b)$. $\square$

A.2 Appendices for Chapter 2

Tables

Table A.1: Regression of effort on treatment status

Coefficient	Task count	Dummy			Task count	Dummy		
		≥ 15	15 or 16	≥ 17		≥ 15	15 or 16	≥ 17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
40c at 15	3.55***	0.32***	0.26***	0.07**	3.12***	0.32***	0.28***	0.04
	(0.82)	(0.03)	(0.02)	(0.03)	(1.01)	(0.04)	(0.03)	(0.03)
80c gift + 40c at 15	2.90***	0.32***	0.26***	0.06**	2.47**	0.31***	0.28***	0.03
	(0.83)	(0.03)	(0.02)	(0.03)	(1.01)	(0.04)	(0.03)	(0.03)
$1.20 at 15	4.87***	0.43***	0.34***	0.09***	4.45***	0.43***	0.36***	0.06**
	(0.67)	(0.02)	(0.02)	(0.02)	(0.89)	(0.03)	(0.03)	(0.03)
Badge	1.32**	0.03	-0.03*	0.06***				
	(0.55)	(0.02)	(0.02)	(0.02)				
Badge × No bonus					0.75	0.02	-0.00	0.03
					(0.95)	(0.03)	(0.03)	(0.03)
Badge × Any bonus					1.60**	0.03	-0.04**	0.08***
					(0.67)	(0.02)	(0.02)	(0.02)
Constant	9.30***	0.19***	0.03*	0.17***	9.59***	0.20***	0.01	0.18***
	(0.55)	(0.02)	(0.02)	(0.02)	(0.68)	(0.02)	(0.02)	(0.02)
N	2,153	2,153	2,153	2,153	2,153	2,153	2,153	2,153
Adj. R^2	0.03	0.13	0.13	0.01	0.03	0.13	0.13	0.01
Res. SE	12.70	0.46	0.38	0.43	12.71	0.46	0.38	0.43

Note: This table shows the results of several regressions of effort on dummy variables for the different treatment conditions.

Table A.2: Regression of effort on treatment status (with demographic control variables)

Coefficient	Task count	Dummy			Task count	Dummy		
		≥ 15	15 or 16	≥ 17		≥ 15	15 or 16	≥ 17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
40c at 15	3.38***	0.32***	0.25***	0.07**	2.97***	0.32***	0.27***	0.04
	(0.81)	(0.03)	(0.02)	(0.03)	(1.00)	(0.04)	(0.03)	(0.03)
80c gift + 40c at 15	2.83***	0.31***	0.26***	0.05**	2.43**	0.31***	0.28***	0.03
	(0.81)	(0.03)	(0.03)	(0.03)	(1.00)	(0.04)	(0.03)	(0.03)
\$1.20 at 15	4.89***	0.43***	0.34***	0.09***	4.49***	0.43***	0.36***	0.06**
	(0.66)	(0.02)	(0.02)	(0.02)	(0.88)	(0.03)	(0.03)	(0.03)
Badge	1.38**	0.03	-0.03*	0.06***				
	(0.54)	(0.02)	(0.02)	(0.02)				
Badge × No bonus					0.84	0.03	-0.00	0.03
					(0.94)	(0.03)	(0.03)	(0.03)
Badge × Any bonus					1.65**	0.04	-0.04**	0.08***
					(0.66)	(0.02)	(0.02)	(0.02)
Constant	16.10***	0.41**	-0.05	0.46***	16.33***	0.41**	-0.06	0.47***
	(5.20)	(0.19)	(0.16)	(0.18)	(5.21)	(0.19)	(0.16)	(0.18)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2,153	2,153	2,153	2,153	2,153	2,153	2,153	2,153
Adj. R^2	0.07	0.15	0.14	0.06	0.07	0.15	0.14	0.06
Res. SE	12.38	0.46	0.38	0.42	12.38	0.46	0.38	0.42

Note: This table shows the results of several regressions of effort on dummy variables for the different treatment conditions.

Table A.3: Demographic composition of sample, part 1

	Full Sample	No badge	Badge	UK	US
Observations	2153	1085	1068	1026	1127
"No" to consent 2	6.3%	5.8%	6.7%	-	-
Residency					
UK	47.7%	46.5%	48.8%	-	-
US	52.3%	53.5%	51.2%	-	-
Bonus Treatment					
No bonus	33.2%	32.2%	34.3%	34.1%	32.4%
40c at 15	16.9%	16.6%	17.2%	17.2%	16.7%
80c gift and 40c at 15	16.3%	16.5%	16.2%	17.0%	15.8%
$1.20 at 15	33.5%	34.7%	32.3%	31.8%	35.1%
Visibility Treatment					
No badge	50.4%	-	-	49.2%	51.5%
Badge	49.6%	-	-	50.8%	48.5%
Sex					
Female	62.1%	62.4%	61.8%	64.7%	59.7%
Male	37.0%	36.7%	37.4%	34.6%	39.2%
Nonbinary	0.9%	0.9%	0.8%	0.7%	1.1%
Age					
18-25	24.1%	24.5%	23.6%	19.1%	28.6%
26-30	19.9%	19.6%	20.1%	17.5%	22.0%
31-35	18.2%	18.1%	18.4%	17.2%	19.2%
36-45	20.9%	21.0%	20.8%	24.8%	17.4%
46-55	10.1%	10.0%	10.3%	12.8%	7.7%
56+	6.8%	6.8%	6.8%	8.7%	5.1%

Note: This table shows demographic summary statistics for the analysis sample, the US and UK analysis subsample, and the two visibility treatment arms. The top row also displays the share of subjects that completed the 5 required tasks, but then left the study at the second consent form by choosing “No, I do not want to complete this survey.” The total number of observations and the demographic shares are computed after excluding these subjects.

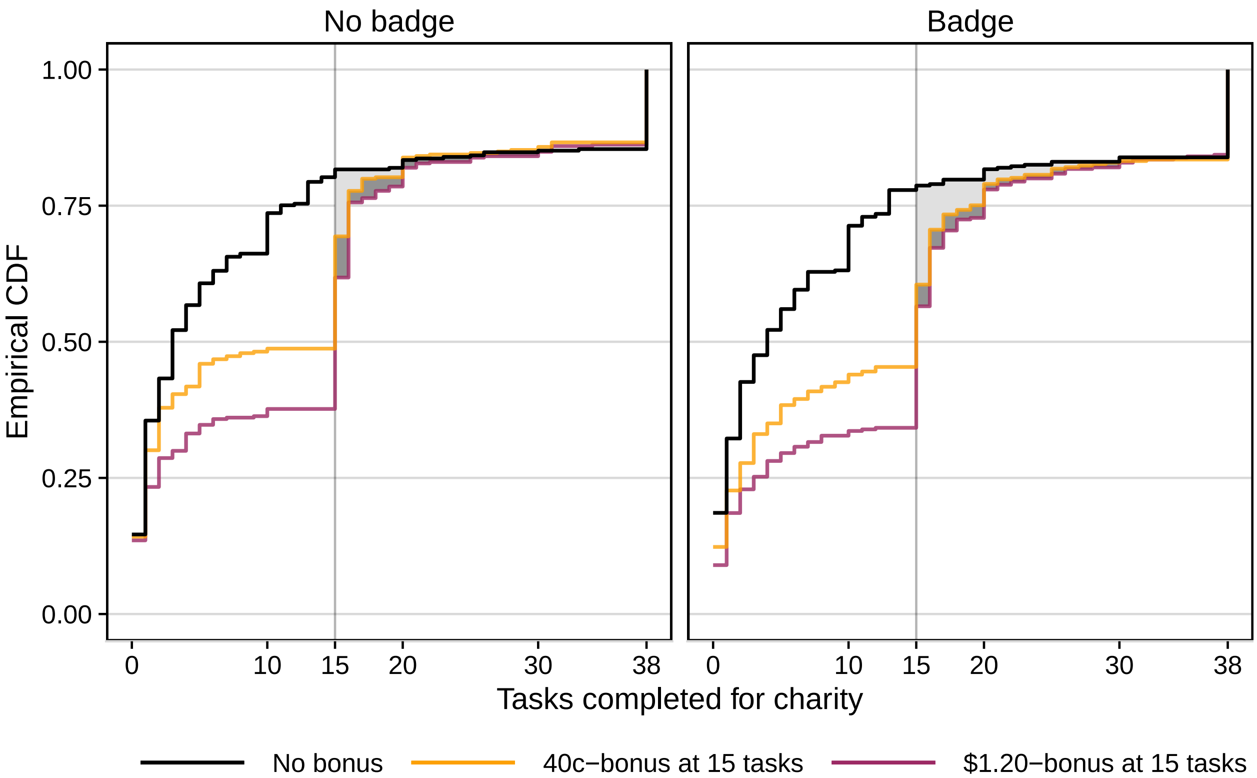
Table A.4: Demographic composition of sample, part 2

	Full Sample	No badge	Badge	UK	US
Education					
College Graduate (4 year)	40.6%	39.7%	41.6%	42.0%	39.4%
High School or equivalent	10.6%	10.9%	10.3%	13.2%	8.3%
Master's Degree (MS)	15.1%	14.4%	15.8%	17.3%	13.1%
Some College	24.3%	25.3%	23.3%	16.9%	31.1%
Other	9.3%	9.7%	9.0%	10.7%	8.1%
Race/Ethnicity					
Asian	6.0%	5.4%	6.6%	3.6%	8.3%
White	78.5%	80.0%	77.1%	89.3%	68.8%
Other	15.4%	14.6%	16.3%	7.1%	23.0%
Hours worked on Prolific last week					
< 5 hours	92.2%	92.1%	92.2%	93.1%	91.3%
5 - 10 hours	6.2%	6.0%	6.5%	5.4%	7.0%
Other	1.6%	1.9%	1.3%	1.6%	1.7%
Hours worked on Prolific today					
< 1 hour	95.3%	95.5%	95.0%	97.4%	93.3%
Other	4.7%	4.5%	5.0%	2.6%	6.7%

Note: This table shows demographic summary statistics for the analysis sample, the US and UK analysis subsample, and the two visibility treatment arms. The top row also displays the share of subjects that completed the 5 required tasks, but then left the study at the second consent form by choosing "No, I do not want to complete this survey." The total number of observations and the demographic shares are computed after excluding these subjects.

Figures

Figure A.1: Distribution of effort across treatments



Note: This figure shows the empirical CDF of effort for each treatment, pooling the two treatment arms "40c-bonus at 15 tasks", and "80c-gift and 40c-bonus at 15 tasks". The shaded areas mark the increase in effort provision beyond 15 tasks that is due to the bonus.

A.3 Appendices for Chapter 3

Materials and Methods

Sample

Our population consists of scholars at two career stages.

Published Authors:

These are active social science researchers who have published in a top-10 leading journal within their discipline. We use the following definitions:

- Active: At least one publication in 2014-2016.
- Top-10 leading journals: The selection of journals was based on citation impact factor. We also added the respective version of the Annual Review for each discipline. In total we have 45 journals, shown in Appendix Tables A.2 through A.5.
- Discipline: Before a participant entered the survey, we took an initial guess of their discipline. For PhD Students it was their department, for Published Authors the discipline they have published in most frequently during 2010-2016, with ties split by the most recent publication. We used the initial guess to draw our sample, and for the analysis. The exception was the following, which occurred in a small number of cases: at the beginning of the survey we ask each participant for their primary discipline. If their answer did not match with the initial guess, and they indicated that they do not feel familiar enough to comment on the initially guessed discipline, we asked them to choose which of the four disciplines they feel sufficiently familiar with. We assigned this discipline to them for our analysis. If they did not feel familiar enough with any of our four disciplines, the survey ended, and they did not become part of our analysis sample.

PhD Students:

These are current PhD Students from top-20 North American doctoral programs within each discipline. We use the following definitions:

- Current: Listed on departmental websites in Fall 2017.
- Top-20 North American Universities: The 20 US and Canadian universities with the highest rank according to the Times Higher Education World University Rankings 2017. The complete list of schools used can be seen in Appendix Table A.6.

PhD Students who are also Published Authors were sampled only as PhD Students.

Participation Incentives:

Achieving a high response rate and sample size was a critical issue for the validity of our study. Several previous surveys on related transparency and reproducibility topics featured minimal or no monetary compensation for participants and had fairly low response rates, most in the range of 10 to 24% (see Baker, 2016; John et al., 2012). We seek to generate longitudinal data on a far more representative population of leading social science researchers by offering much higher levels of compensation.

Participants were randomly offered either a standard or high incentive. The levels differ between Published Authors and PhD Students, and are based on the response rates from our pilot.

Initial contact was made via email. There were three reminders at intervals following the initial email contact. The survey was administered using a customized online tool (a custom-built interface on top of Qualtrics). Appendix Table A.1 shows the monetary value of the incentives used in the survey. PhD students offered the High incentive had an 8.2 percentage point higher response rate and Published Authors offered the High incentive had a 0.8 percentage point higher response rate.

Appendix Table A.1: Participation incentives

Career Stage	Standard (80% of sample)	High (20% of sample)
Published Authors	$75	$100
PhD Students	$25	$40

Descriptive Analysis:

We aggregate individual survey questions into five measures (awareness, behavior, attitudes, descriptive norms, and prescriptive norms) for each of the three practices (posting data and code online, posting study instruments, and pre-registration). Details of the aggregation method are described in Appendix Table A.8.

We also measure trustworthiness of the literature, behavioral intentions, and projected norms through a set of questions.

We then aggregate the large number of measures to a smaller number of sub-indices and broad indices. Each sub-index is a simple average of measures, and each broad-index is a simple average of sub-indices. See Appendix Table A.7 and Appendix Table A.8 for details.

Altogether, our outcome variables for the descriptive analysis are:

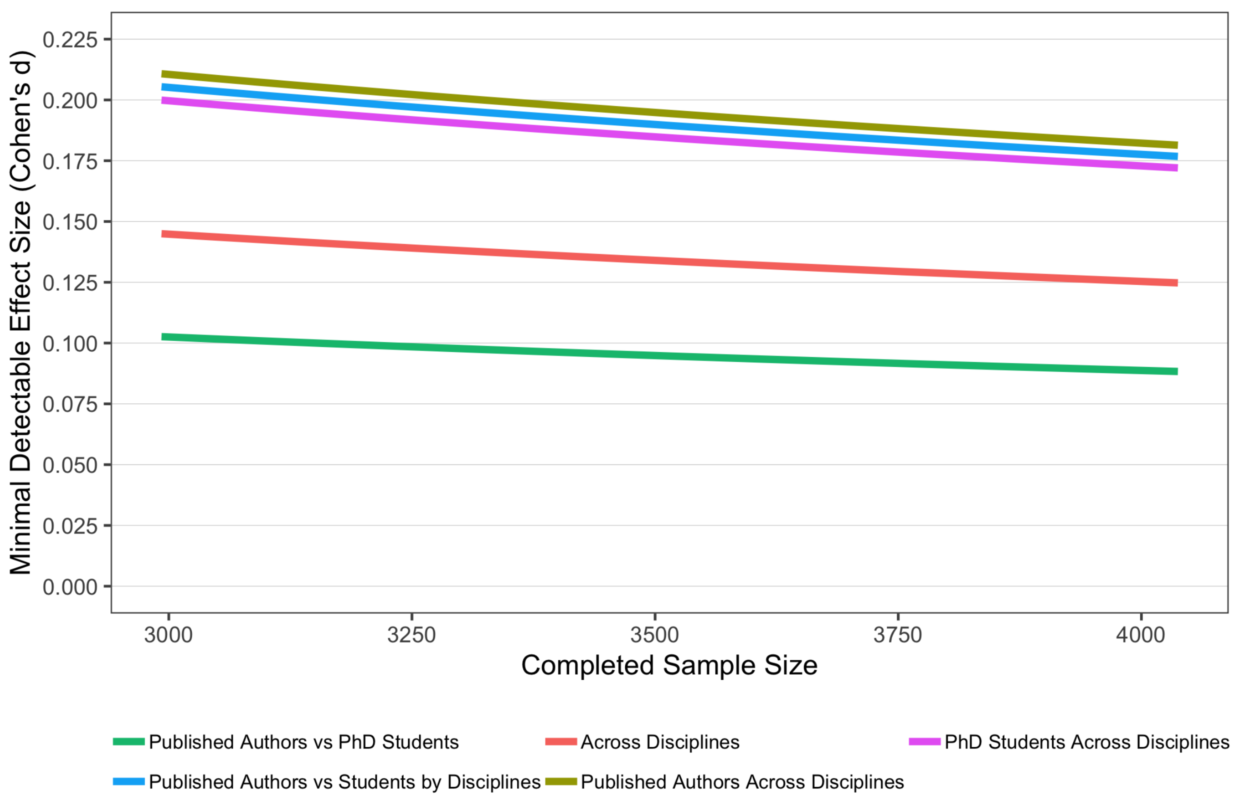
- Sub-Indices: Awareness, Behavior, Attitudes, Descriptive Norms, Prescriptive Norms, Posting data and code online, Posting study instruments, Pre-registration
- Broad Indices: Personal support for open science, Norms, Overall open science, Trustworthiness of literature

The mappings from questions to sub-indices, and from sub-indices to broad-indices can be found in Appendix Tables A.7 through A.8.

Power Calculations:

We based power calculations on conservative estimates of response rates from prior transparency surveys and our own pilot. We conducted power calculations expecting roughly equal numbers in each discipline. These assumptions yield an expected final sample size between 3,000 to 4,000, with N=3200 as our best guess. As shown in the Appendix Figure A.1, with a power threshold of 80%, we are able to detect small differences in means across groups.

Appendix Figure A.1: Power calculations



Note: The chart shows the minimum detectable effect size at different sample sizes for comparing different subgroups. Power calculations were preregistered. The figure shows the power calculations that we pre-registered. Our realized sample size was 2801. At this sample size, the minimum detectable effect by author type is 0.106, the minimum detectable effect by discipline is 0.1497 and the minimum detectable effect for the interaction is 0.212.

Regression Specifications:

For each outcome variable described in the previous sub-section, we run the following linear regressions.

First, an analysis of differences across disciplines (dropping subscripts denoting individual participants).

$$y = \alpha_1 + \beta_{1a} * \mathbb{I}\{Econ\} + \beta_{1b} * \mathbb{I}\{PoliSci\} + \beta_{1c} * \mathbb{I}\{Psych\} + u_1$$

Second, an analysis of differences between Published Authors and PhD Students.

$$y = \alpha_1 + \beta_2 * \mathbb{I}\{PublishedAuthor\} + u_2$$

Third, an analysis that examines both of these dimensions of heterogeneity:

$$\begin{aligned} y = \alpha_1 + \beta_{3a} * \mathbb{I}\{Econ\} + \beta_{3b} * \mathbb{I}\{PoliSci\} + \beta_{3c} * \mathbb{I}\{Psych\} + \beta_{3d} * \mathbb{I}\{PublishedAuthor\} \\ + \beta_{3e} * \mathbb{I}\{Econ\} * \mathbb{I}\{PublishedAuthor\} + \beta_{3f} * \mathbb{I}\{PoliSci\} * \mathbb{I}\{PublishedAuthor\} \\ + \beta_{3g} * \mathbb{I}\{Psych\} * \mathbb{I}\{PublishedAuthor\} + u_3 \end{aligned}$$

We employ a multiple testing adjustment to address risk of false positives. In particular, we use the False Discovery Rate (FDR) adjustment in Benjamini et al. (2006) and discussed in M. L. Anderson (2008). We carry out FDR adjustment across the primary outcome variables.

We also present the averages of our outcome variables by discipline and career stage graphically and estimate regression specifications adjusted for covariates (age, gender, tenured status, US department, leadership position).

Validation Exercise:

In order to validate our survey responses and check for balance across respondents and non-respondents, we conducted an audit of our economics Published Authors. Specifically, we randomly sampled i) 300 economics Published Authors who completed our survey and ii) 100 economics Published Authors who were contacted but did not complete our survey.

We then conducted two audit activities. For *all* sampled individuals we conducted an audit of these authors' pre-registration and data posting behaviors using publicly available information. The protocol for this activity is the first subsection below. This audit activity was completed between March 15, 2019 and March 29, 2019.

The second audit activity was conducted only for the non-respondent sample, and was completed between April 4, 2019 and April 15, 2019. In this activity, we used publicly available data sources to collect data on the primary subfield of these non-respondents. The protocol for this activity is below.

After these subfields were collected we manually categorised these subfields into one of three categories. The first of these was "Theory focused", which is categorised as any individual who listed Microeconomic Theory or Econometrics as a primary subfield. The second was "Macroeconomics/Finance", which was any author who listed Macroeconomics or Finance as a primary field. Finally, all other authors were categorised in the residual category.

Audit Protocol - Open Science behaviors

The goal of the audit is to identify whether a Published Author in the selected sample has (i) pre-registered an analysis or (ii) posted data or code for their projects. We use an author's last name as a keyword to search a set of popular open science websites used by economics scholars.

General Procedure Since the collection of last names was fully automated, auditors first verify whether an author's last name corresponds to a Published Author by looking for a university affiliation using a Google search.

The auditors then go to the websites listed below, and search by last name only. They look through the search results and try to identify the Published Author using their first name or affiliation. Then, following the link associated with an identified author, auditors look for a (i) pre-analysis plan or (ii) posted data or code on the websites. As soon as a match is found, auditors stop searching and record the match and a link to the matched page. If no match can be found, the auditors record that no match was found.

Websites for posting data or code online

- Dataverse.org
- Authors' personal websites

Websites for pre-registering analysis (PAP)

- SocialScienceRegistry.org (AEA RCT registry). Details of some pre-analysis plans may not be visible to the public, but we still count those as having pre-registered.
- OSF.io
- Authors' personal websites

Audit Protocol - Author Subfield

The goal of this activity is to collect data on the primary subfields of Economics Published Authors that did not complete the survey. The following steps are followed to complete this activity:

- Go to the author's webpage. Record subfields information if subfields of interest are listed on the homepage or another part of the webpage.
- Open the author's CV. Record any subfields that are listed on the author's CV.

Sampling frame and Outcome Index Construction:

Appendix Table A.2: Economics journals

Index	Journal	Publisher
NR	AnnualReviewofEconomics	Annual Reviews
1	TheQuarterlyJournalofEconomics	Oxford University Press
2	JournalofPoliticalEconomy	University of Chicago Press, JSTOR
3	AmericanEconomicReview	American Economic Association, JSTOR
4	Econometrica	Wiley, JSTOR
5	JournalofEconomicGrowth	Springer, JSTOR
6	ReviewofEconomicStudies	Oxford University Press
7	JournalofMonetaryEconomics	Elsevier
8	JournalofEconometrics	Elsevier
9	JournalofLaborEconomics	University of Chicago Press
10	TheReviewofEconomicsandStatistics	MIT Press

Sampling Frame Economics Published Authors *Note*: Journals used to sample economics Published Authors. While the Annual Review of Economics is not ranked, it is included as it is an influential journal for the field. The selection of journals is based on citation impact factor.

Appendix Table A.3: Political Science journals

Index	**Journal**	**Publisher**
NR	AnnualReviewofPoliticalScience	Annual Reviews
1	AmericanJournalofPoliticalScience	Wiley
2	AmericanPoliticalScienceReview	Cambridge University Press
3	TheJournalofPolitics	University of Chicago Press
4	BritishJournalofPoliticalScience	Cambridge University Press
5	PoliticalAnalysis	Oxford University Press
6	ComparativePoliticalStudies	SAGE Publishing
7	WorldPolitics	Cambridge University Press
8	PoliticalBehavior	Springer
9	InternationalOrganization	Cambridge University Press
10	InternationalStudiesQuarterly	Wiley

Sampling Frame Political Science Published Authors *Note*: Journals used to sample political science Published Authors. While the Annual Review of Political Science is not ranked, it is included as it is an influential journal for the field. The selection of journals is based on citation impact factor.

Appendix Table A.4: Psychology journals

Index	Journal	Publisher
NR	AnnualReviewofPsychology	Annual Reviews
1	PsychologicalScience	SAGE Publishing
2	PsychologicalBulletin	American Psychological Association
3	AmericanPsychologist	American Psychological Association
4	JournalofExperimentalPsychology − General	American Psychological Association
5	TrendsinCognitiveSciences	Elsevier
6	SocialCognitiveandAffectiveNeuroscience	Oxford University Press
7	JournalofPersonalityandSocialPsychology	American Psychological Association
8	JournalofConsultingandClinicalPsychology	American Psychological Association
9	ChildDevelopment	Wiley
10	DevelopmentalPsychology	American Psychological Association

Sampling Frame Psychology Published Authors *Note*: Journals used to sample psychology Published Authors. While the Annual Review of Psychology is not ranked, it is included as it is an influential journal for the field. The selection of journals is based on citation impact factor.

Appendix Table A.5: Sociology journals

Index	Journal	Publisher
NR	AnnualReviewofSociology	Annual Reviews
1	AmericanSociologicalReview	SAGE Publishing
2	AmericanJournalofSociology	University of Chicago Press
3	EuropeanSociologicalReview	Oxford University Press
4	SocialForces	Oxford University Press
5	SocialProblems	Oxford University Press
6	Demography	Springer
7	Criminology	Wiley
8	Gender&Society	SAGE Publishing
9	AdministrativeScienceQuarterly	SAGE Publishing
10	SociologyofEducation	SAGE Publishing
11	SocialNetworks	Elsevier

Sampling Frame Sociology Published Authors *Note*: Journals used to sample sociology Published Authors. While the Annual Review of Sociology is not ranked, it is included as it is an influential journal for the field. The selection of journals is based on citation impact factor and disciplinary expert recommendation.

Appendix Table A.6: Top 20 North American doctoral programs

Rank	**University**	**Country**
1	Stanford University	US
2	Yale University	US
3	University of Chicago	US
4	Harvard University	US
5	Massachusetts Institute of Technology	US
6	University of Michigan-Ann Arbor	US
7	Princeton University	US
8	University of California, Los Angeles	US
9	University of California, Berkeley	US
10	Columbia University	US
11	University of Pennsylvania	US
12	Cornell University	US
13	Duke University	US
14	University of Wisconsin-Madison	US
15	University of Toronto	Canada
16	University of British Columbia	Canada
17	New York University	US
18	Northwestern University	US
19	University of Washington-Seattle	US
20	University of California, San Diego	US

Sampling Frame PhD Students *Note*: PhD Students in the paper were sampled from universities listed in the table. The ranking is the Times Higher Education 2017 Social Science ranking.

Appendix Table A.7: Mapping measures to indices

Measure	Sub-Index	Broad Index
1.1.1 Awareness of posting data and code online	1.1 Awareness	1. Personal support for open science
1.1.2 Awareness of posting study instruments		
1.1.3 Awareness of pre-registration		
1.2.1 Behavior of posting data and code online	1.2 Behavior	
1.2.2 Behavior of posting study instruments		
1.2.3 Behavior of pre-registration		
1.3.1 Attitudes of posting data and code online	1.3 Attitudes	
1.3.2 Attitudes of posting study instruments		
1.3.3 Attitudes of pre-registration		
2.1.1 Descriptive norms of posting data and code online	2.1 Descriptive norms	2. Norms
2.1.2 Descriptive norms of pre-registration		
2.2.1 Prescriptive norms of posting data and code online	2.2 Prescriptive norms	
2.2.2 Prescriptive norms of pre-registration		
3.1.1 Awareness of posting data and code online	3.1 Posting data and code online	3. Overall Open Science
3.1.2 Behavior of posting data and code online		
3.1.3 Attitudes of posting data and code online		
3.1.4 Descriptive norms of posting data and code online		
3.1.5 Prescriptive norms of posting data and code online		
3.2.1 Awareness of posting study instruments	3.2 Posting study instruments	
3.2.2 Behavior of posting study instruments		
3.2.3 Attitudes of posting study instruments		
3.3.1 Awareness of pre-registration	3.3 Pre-registration	
3.3.2 Behavior of pre-registration		
3.3.3 Attitudes of pre-registration		
3.3.4 Descriptive norms of pre-registration		
3.3.5 Prescriptive norms of pre-registration		
4. Trustworthiness of literature		4. Trustworthiness of literature

Measures incorporated in Indices *Note*: The table shows the mapping from measures (see Appendix Table A.8) to indices. Each sub-index is a simple average of measures, and each broad-index is a simple average of sub-indices.

Appendix Table A.8: Mapping questions to measures

Question	Measure	Rescaling and Aggregation
Have you ever heard of the practice of publicly posting data and code online for a completed study?	1.1.1 Awareness of posting data and code online	"No" $\rightarrow$ 0, "Yes" $\rightarrow$ 1
Approximately how many times have you publicly posted data or code online?	1.2.1 Behavior of posting data and code online	Question "Approximately. . . " coded as 0 $\rightarrow$ 0, anything $\geq$ 1 $\rightarrow$ 1; Question "Think about the last. . . " coded as "No" $\rightarrow$ 0, "Yes" $\rightarrow$ 1, "I have not published an empirical paper" $\rightarrow$ NA; Question "Do you encourage . . . " coded as ("No, and I don't plan to", "No, but I plan to in the future") $\rightarrow$ "0", ("Yes, I do") $\rightarrow$ "1"; Average over questions
Think about the last empirical paper you published. Have you publicly posted the data or code online for that paper?		
Do you encourage students to publicly post data or code online?		
To what extent do you believe that publicly posting data or code online is important for progress in [Discipline]?	1.3.1 Attitude of posting data and code online	Rescale from 1-5 to 0-1; Average over questions
What is your opinion of publicly posting data or code online?		
In your estimation, what percentage of researchers across the discipline of [Discipline] publicly post data or code online?	2.1.1 Descriptive norm of posting data or code online	Average over questions
In your estimation, what percentage of researchers in your sub-field of [Subdiscipline] publicly post data or code online?		

continued . . .

Question	Measure	Rescaling and Aggregation
In your estimation, what is the distribution of opinion across the discipline of [Discipline] about publicly posting data or code online?	2.2.1 Prescriptive norm of posting data or code online	Calculate mean of distribution; Rescale from 1-5 to 0-1
In your estimation, what is the distribution of opinion in your sub-field of [Subdiscipline] about publicly posting data or code online?		
Have you ever heard of the practice of publicly posting study instruments online for a completed study?	1.1.2 Awareness of posting study instruments	"No" $\rightarrow$ 0, "Yes" $\rightarrow$ 1
Approximately how many times have you publicly posted study instruments online?	1.2.2 Behavior of posting study instruments	Question "Approximately..." coded as 0 $\rightarrow$ 0, anything $\geq$ 1 $\rightarrow$ 1; Question "Think about the last..." coded as "No" $\rightarrow$ 0, "Yes" $\rightarrow$ 1, "I have not published an empirical paper" $\rightarrow$ NA; Question "Do you encourage ..." coded as ("No, and I don't plan to", "No, but I plan to in the future") $\rightarrow$ "0", ("Yes, I do") $\rightarrow$ "1"; Average over questions
Think about the last empirical paper you published. Have you publicly posted the study instruments online for that paper?		
Do you encourage students to publicly post study instruments online?		
To what extent do you believe that publicly posting study instruments online is important for progress in [Discipline]?	1.3.2 Attitude of posting study instruments	Rescale from 1-5 to 0-1; Average over questions
What is your opinion of publicly posting study instruments online?		
Have you ever heard of the practice of pre-registering hypotheses or analyses in advance of a study?	1.1.3 Awareness of pre-registration	Rescale from 1-5 to 0-1; Average over questions

continued ...

Question	**Measure**	**Rescaling and Aggregation**
Approximately how many times have you pre-registered hypotheses or analyses in advance of a study?	1.2.3 Behavior of pre-registration	Question “Approximately…” coded as 0 $\rightarrow$ 0, anything $\geq$ 1 $\rightarrow$ 1; Question “Think about the last…” coded as “No” $\rightarrow$ 0, “Yes” $\rightarrow$ 1, “I have not published an empirical paper” $\rightarrow$ NA; Question ”Do you encourage …” coded as (“No, and I don’t plan to”, “No, but I plan to in the future”) $\rightarrow$ “0”, (“Yes, I do”) $\rightarrow$ ”1”; Average over questions
Think about the last empirical research you completed. Did you pre-register the hypotheses or analyses for that research?		
Do you encourage students to pre-register hypotheses or analyses in advance of a study?		
To what extent do you believe that pre-registering hypotheses or analyses is important for progress in [Discipline]?	1.3.3 Attitude of pre-registration	Rescale from 1-5 to 0-1; Average over questions
What is your opinion of pre-registering hypotheses or analyses?		
In your estimation, what percentage of researchers across the discipline of [Discipline] pre-register hypotheses or analyses in advance of a study?	2.1.2 Descriptive norm of pre-registration	Rescale from 0-100 to 0-1; Average over questions
In your estimation, what percentage of researchers in your sub-field of [Sub-discipline] pre-register hypotheses or analyses in advance of a study?		
In your estimation, what is the distribution of opinion across the discipline of [Discipline] about pre-registering hypotheses or analyses in advance of a study?	2.2.2 Prescriptive norm of pre-registration	Calculate mean of distribution; Rescale from 1-5 to 0-1

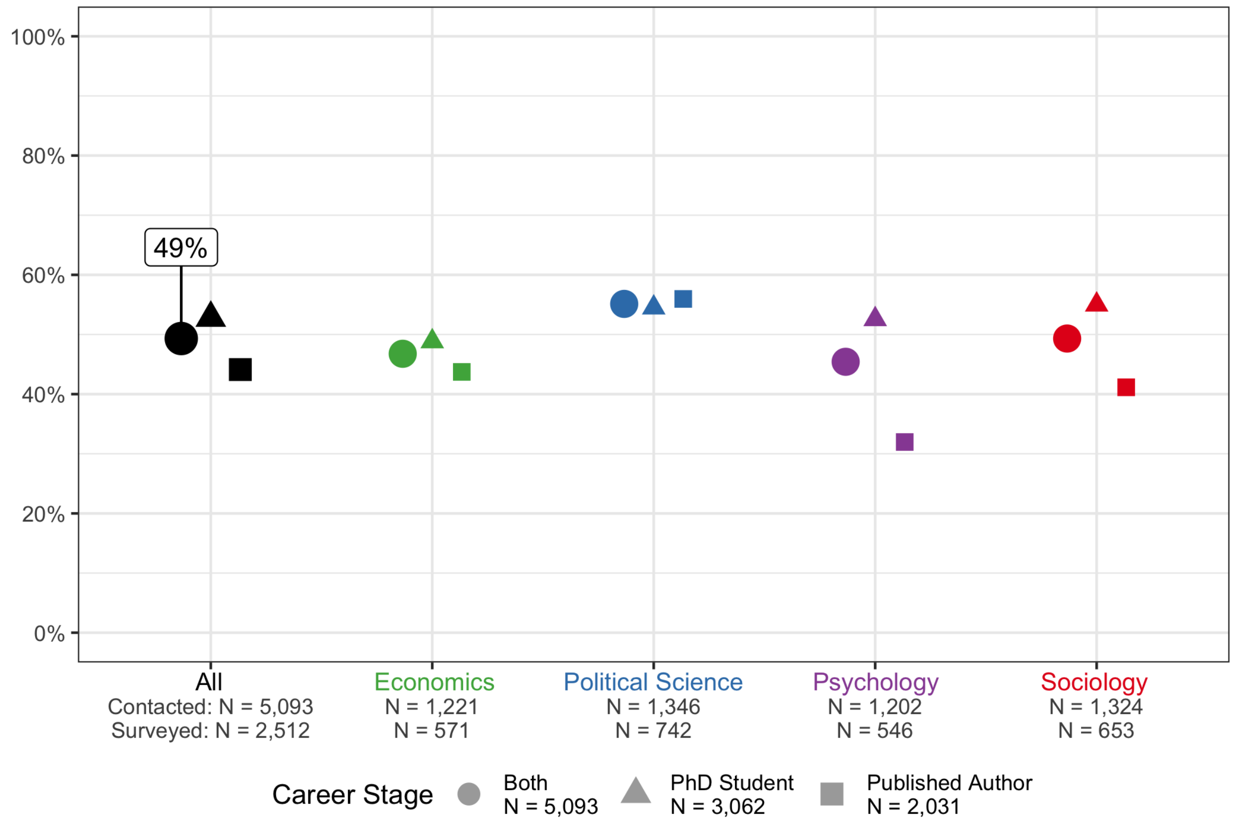
continued …

Question	Measure	Rescaling and Aggregation
In your estimation, what is the distribution of opinion in your sub-field of [Subdiscipline] about pre-registering hypotheses or analyses in advance of a study?		
How confident are you that the influential research findings in [Discipline] would replicate?	4. Trustworthiness ofliterature	Rescale from 1-5 to 0-1; Average over questions
When researchers run studies testing the canonical research findings in [Discipline], how confident are you that the studies will be able to replicate the canonical results?		
When researchers run studies testing recent research findings in [Discipline], how confident are you that the studies will be able to replicate the recent results?		
Think about the table of contents in the latest issue of [Discipline]'s top journal. How confident are you that the results of the studies will replicate?		

Questions incorporated in Measures *Note*: The table shows the survey questions that are included in each measure. Each measure is then combined with other measures to produce indices (see Appendix Table A.7). In the cases where multiple questions are used in a single measure, how these questions are aggregated is also described.

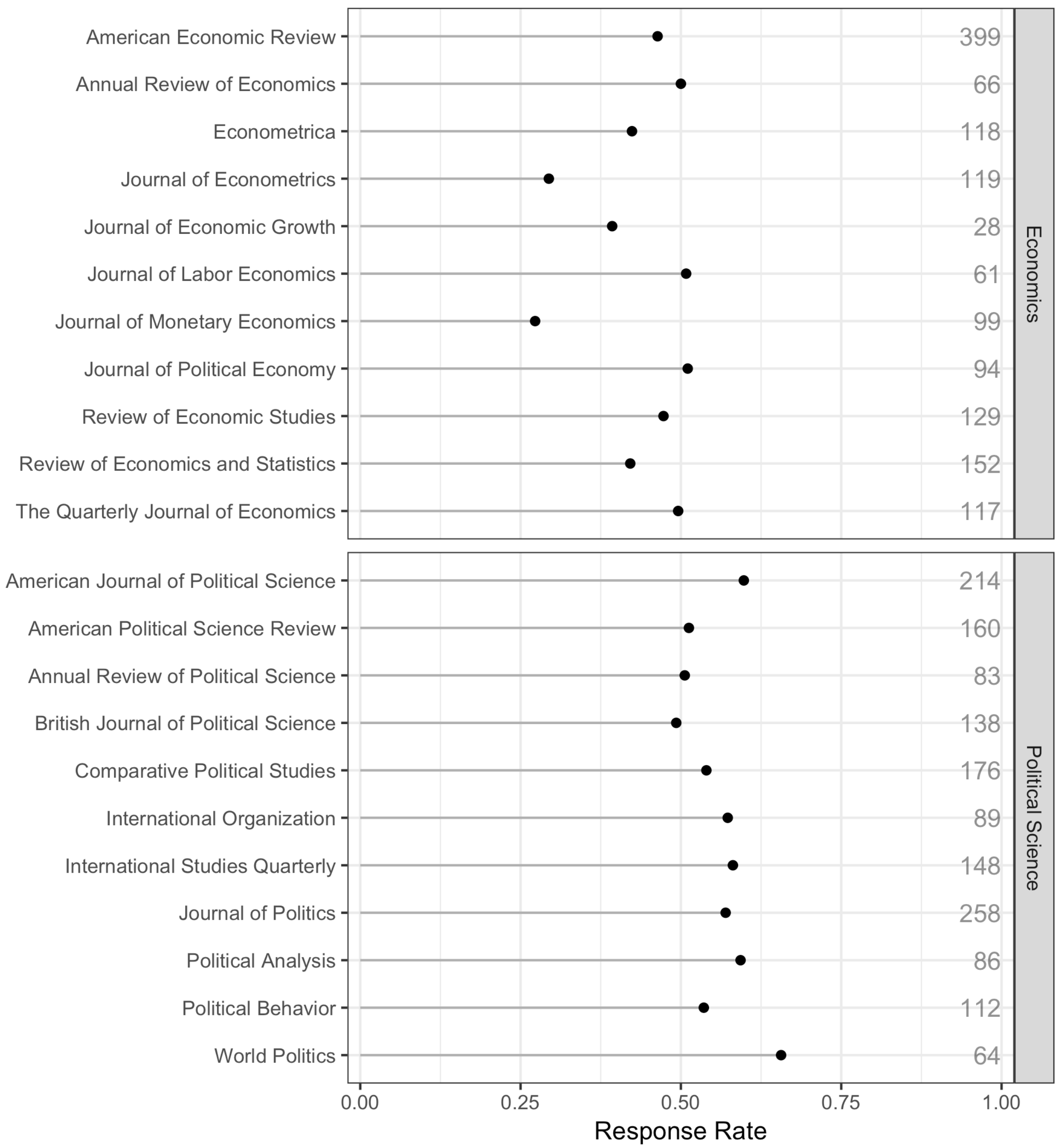
Results

Appendix Figure A.2: Response rates are higher in the United States and Canada sample



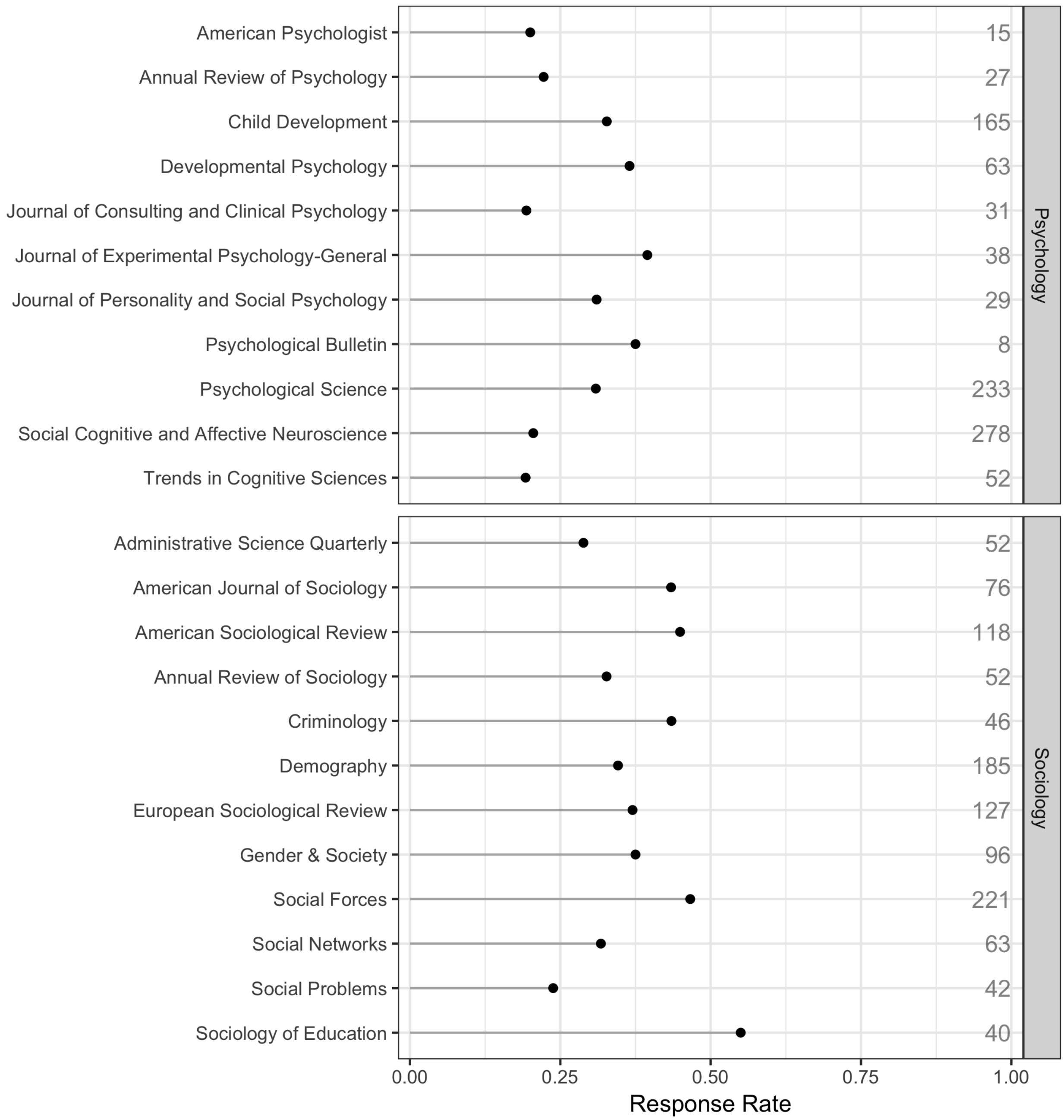
Note: Response rates by discipline and by career stage (PhD Student or Published Author). This figure shows the response rate by discipline and author status for all PhD Students and Published Authors whose institution was based in the United States or Canada.

Appendix Figure A.3: Response rate by journal, part 1



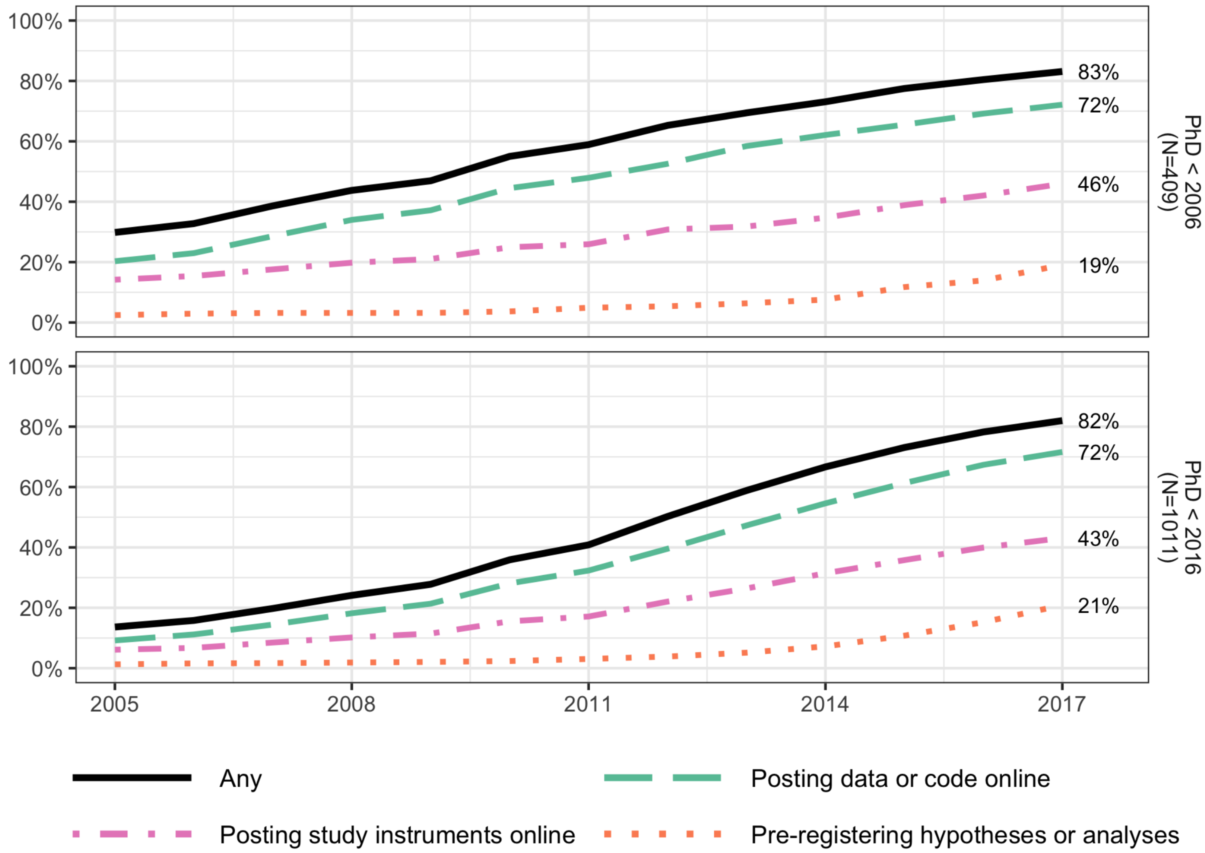
Note: This figure shows the response rate by journal for the universe of journals that were used as the sampling frame for Published Authors in this project. Each panel denotes the journals for a different discipline. Numbers in grey on the right hand side of the figures show the raw number of respondents from each journal. The published author sample is drawn from the universe of authors that published in one of the above journals during the timeframe 2014-2016. However, the Published Authors are matched to any journal in the above table by any journal that they published in during the period 2010-2016. Therefore the number of Published Authors in the table above is larger than the number of Published Authors in our sample.

Appendix Figure A.4: Response rate by journal, part 2



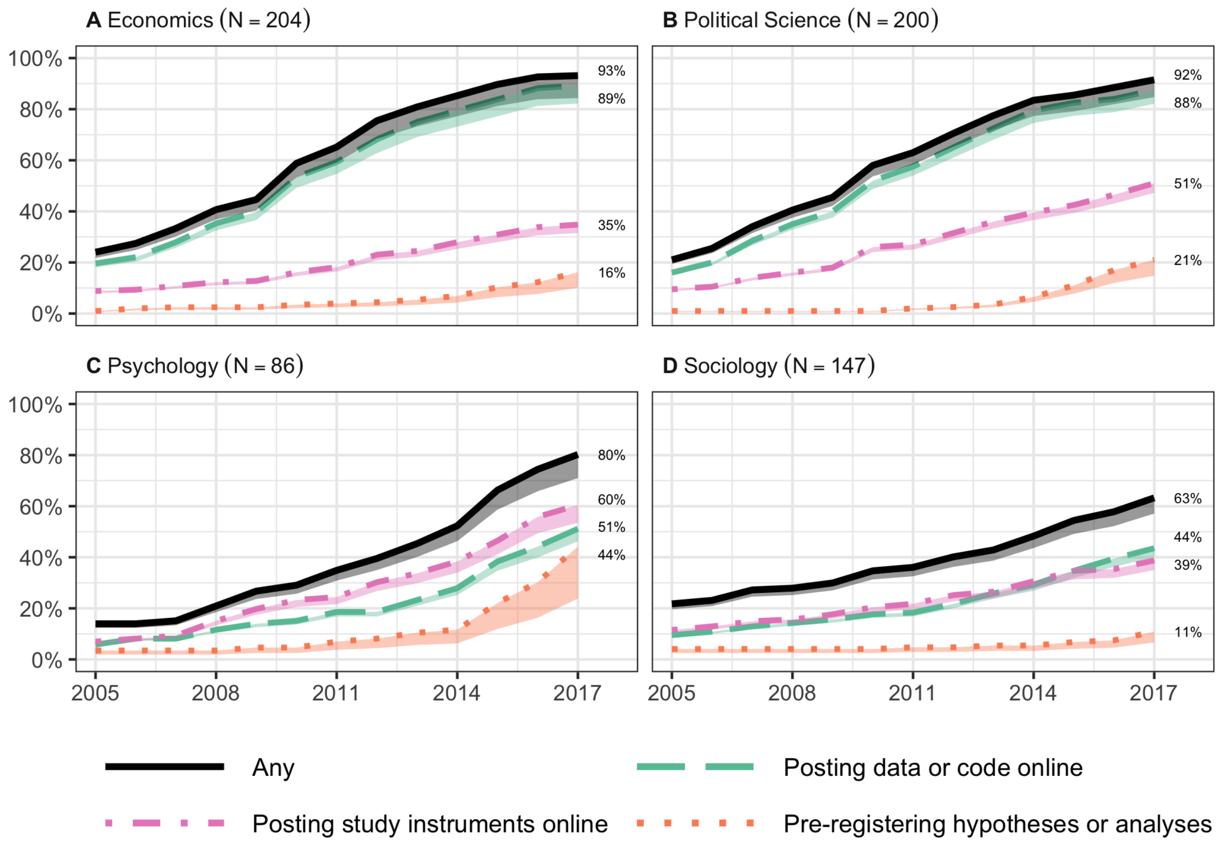
Note: This figure shows the response rate by journal for the universe of journals that were used as the sampling frame for Published Authors in this project. Each panel denotes the journals for a different discipline. Numbers in grey on the right hand side of the figures show the raw number of respondents from each journal. The published author sample is drawn from the universe of authors that published in one of the above journals during the timeframe 2014-2016. However, the Published Authors are matched to any journal in the above table by any journal that they published in during the period 2010-2016. Therefore the number of Published Authors in the table above is larger than the number of Published Authors in our sample.

Appendix Figure A.5: Year of adoption of open science practices, alternate cutoff dates



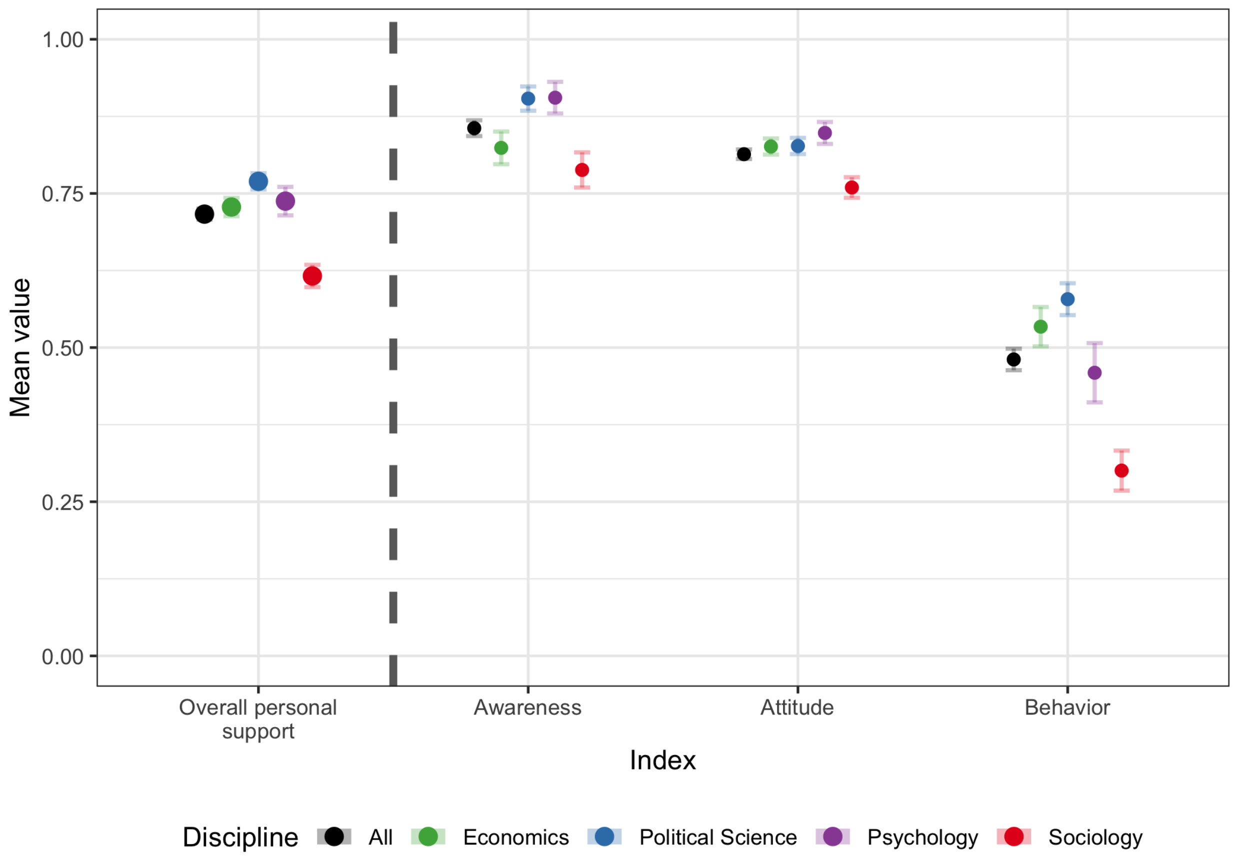
Note: The chart shows for a given year the proportion of Published Authors who had first completed an open science practice in that year or previously. Posting study instruments online is the response to the question "Approximately when was the first time you publicly posted study instruments online?". Posting data or code online is the response to the question "Approximately when was the first time you publicly posted data or code online?". Pre-registering hypotheses or analyses is the response to the question "Approximately when was the first time you pre-registered hypotheses or analyses in advance of a study?". The sample is restricted to Published Authors who completed their PhD by 2005 in the first panel, and Published Authors who completed their PhD prior to 2016 in the second panel.

Appendix Figure A.6: Adoption by discipline



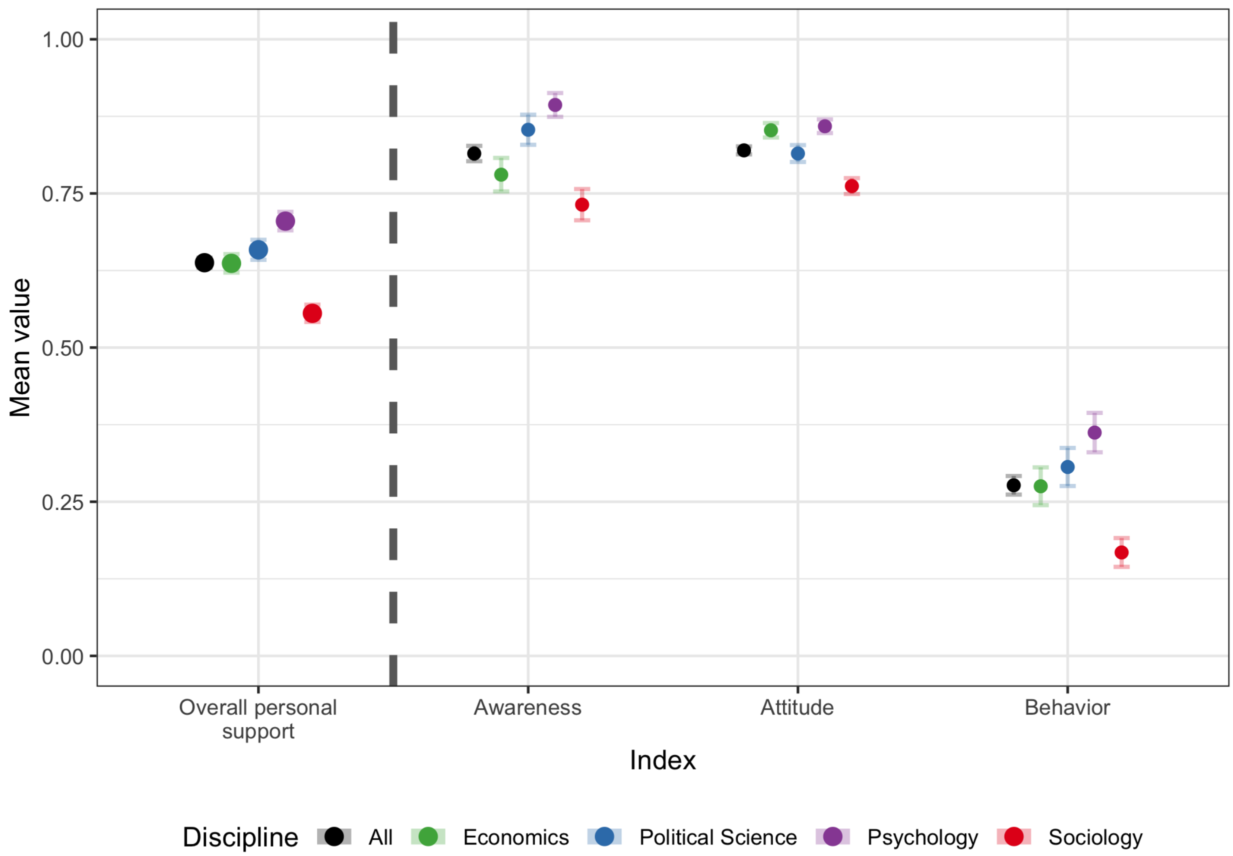
Note: The chart shows for a given year the proportion of Published Authors who had first completed an open science practice in that year or previously. Posting study instruments online is the response to the question "Approximately when was the first time you publicly posted study instruments online?". Posting data or code online is the response to the question "Approximately when was the first time you publicly posted data or code online?". Pre-registering hypotheses or analyses is the response to the question "Approximately when was the first time you pre-registeredhypotheses or analyses in advance of a study?". The sample is restricted to Published Authors who completed their PhD by 2009. The bottom of the shaded region is an estimated adoption rate for the entire sample contacted, including non-respondents; the methodology for calculating the adoption rate of non-respondents is outlined in Table 3.1.

Appendix Figure A.7: Published Author open science awareness, attitudes and behavior, by discipline



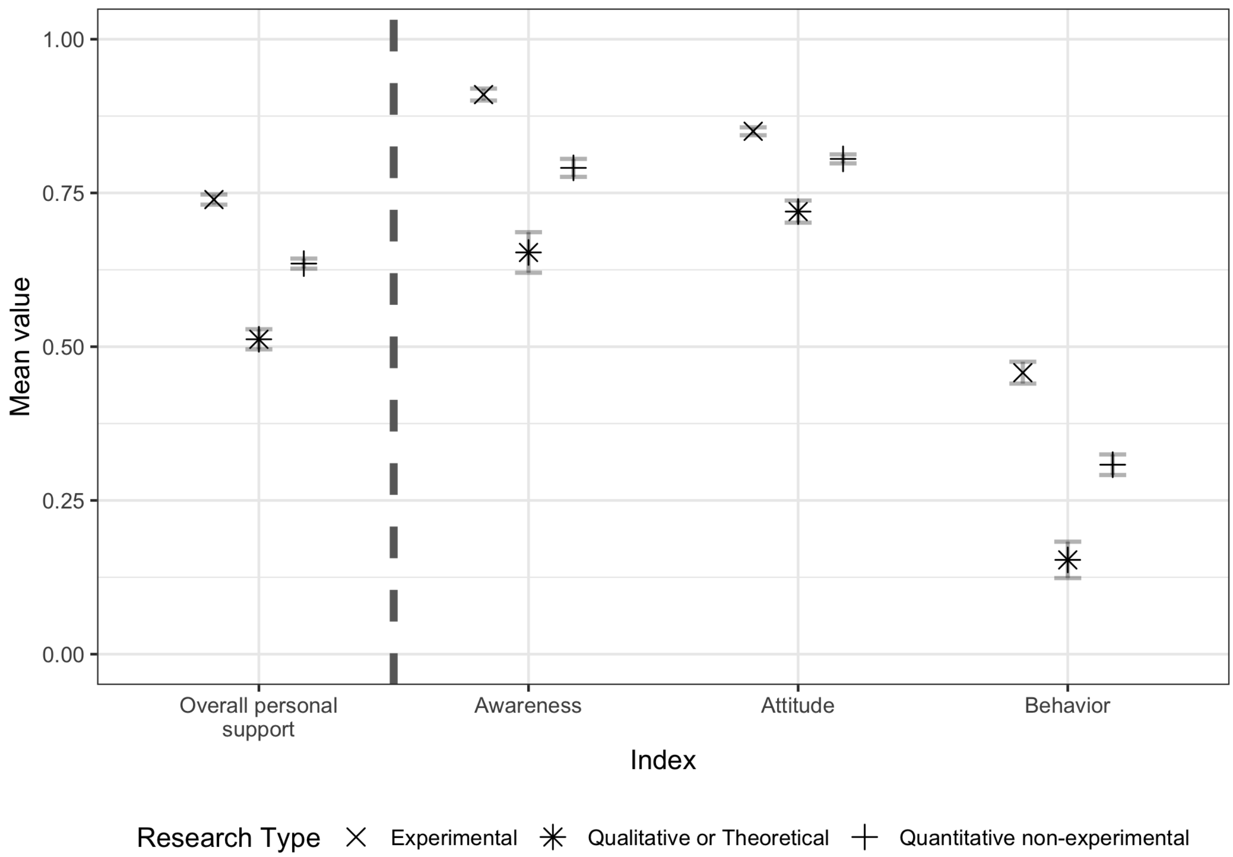
Note: Lines around the dots are 95% confidence intervals for the estimates. Awareness is an index comprised of questions related to the respondent's i) Awareness of posting data and code online, ii) Awareness of posting study instruments and iii) Awareness of pre-registration. Behavior is an index comprised of questions related to the respondent's i) Behavior of posting data and code online, ii) Behavior of posting study instruments and iii) Behavior of pre-registration. Attitudes is an index comprised of questions related to the respondent's i) Attitudes of posting data and code online, ii) Attitudes of posting study instruments and iii) Attitudes of pre-registration. Overall Personal Support is an average of the three indices. The questions and methodology that are used to construct the indices can be found in Appendix Table A.7.

Appendix Figure A.8: Student open science awareness, attitudes and behavior, by discipline



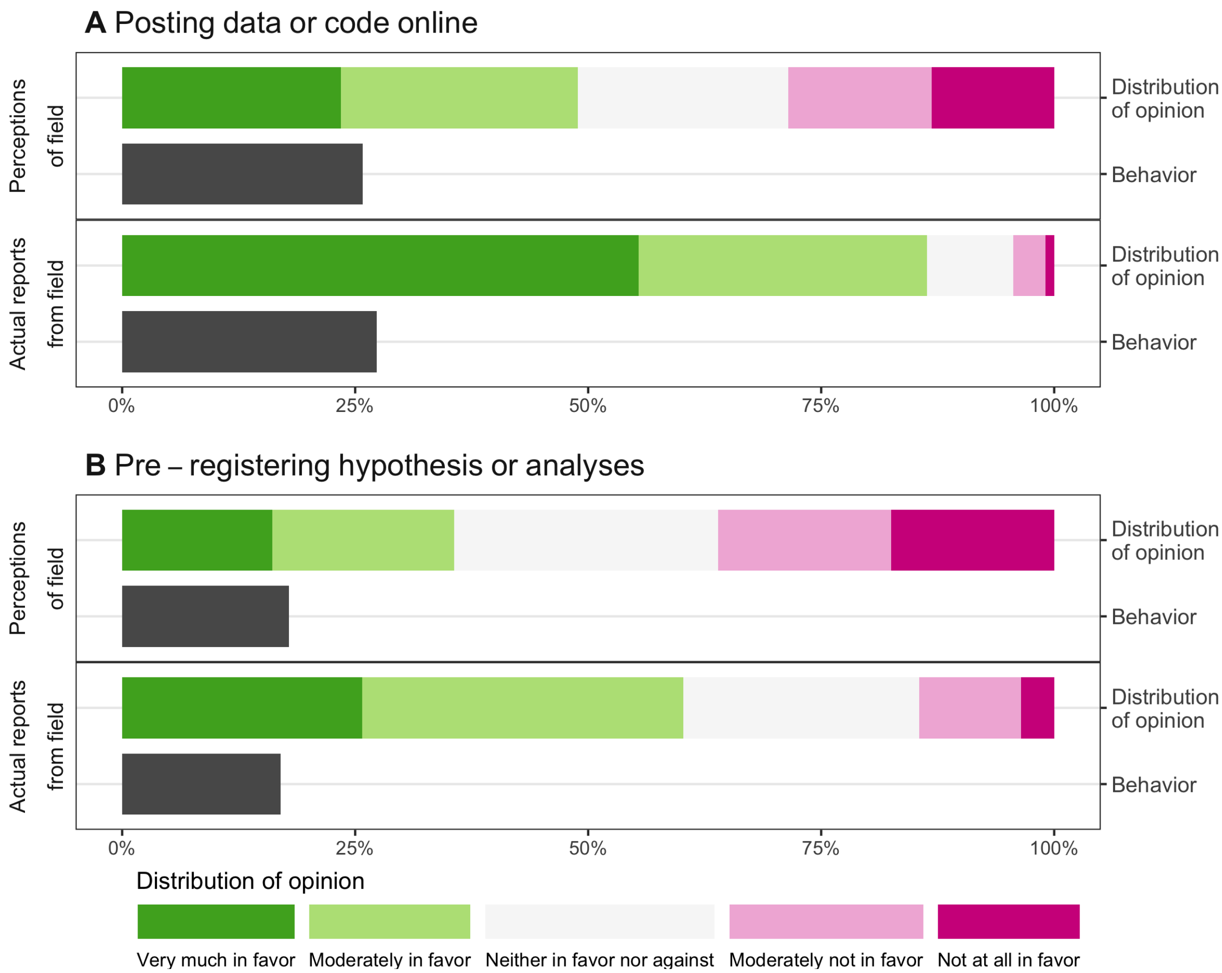
Note: Grey lines around the dots are 95% confidence intervals for the estimates. Awareness is an index comprised of questions related to the respondent's i) Awareness of posting data and code online, ii) Awareness of posting study instruments and iii) Awareness of pre-registration. Behavior is an index comprised of questions related to the respondent's i) Behavior of posting data and code online, ii) Behavior of posting study instruments and iii) Behavior of pre-registration. Attitudes is an index comprised of questions related to the respondent's i) Attitudes of posting data and code online, ii) Attitudes of posting study instruments and iii) Attitudes of pre-registration. Overall Personal Support is an average of the three indices. The questions and methodology that are used to construct the indices can be found in Appendix Table A.7.

Appendix Figure A.9: Open science awareness, attitudes and behavior, by research type



Note: Grey lines around the dots are 95% confidence intervals for the estimates. Awareness is an index comprised of questions related to the respondent's i) Awareness of posting data and code online, ii) Awareness of posting study instruments and iii) Awareness of pre-registration. Behavior is an index comprised of questions related to the respondent's i) Behavior of posting data and code online, ii) Behavior of posting study instruments and iii) Behavior of pre-registration. Attitudes is an index comprised of questions related to the respondent's i) Attitudes of posting data and code online, ii) Attitudes of posting study instruments and iii) Attitudes of pre-registration. Overall Personal Support is an average of the three indices. The questions and methodology that are used to construct the indices can be found in Appendix Table A.7.

Appendix Figure A.10: Perceived and actual support for open science, Students



Note: The chart shows differences between perceived and actual support for two practices: posting data or code online and pre-registering hypotheses or analyses. The sample is restricted to PhD Students. Within each panel, the first bar shows the perceived distribution of support for the practice among Students. This is constructed by asking individuals what percentage of researchers in their field they believe fall into each opinion category, and then averaging over their responses. The solid black bar below shows the fraction of researchers in their field they believe have done the practice. The third bar in the panel shows the distribution of support for the practice constructed using the responses elicited from students. The final solid black bar shows the proportion of students who have actually done the stated practices, using the responses elicited from our survey. Colors indicate the level of support, with green indicating more and red indicating less support.

Appendix Table A.9: Differences in observables for those completing and not completing survey

Variable	Overall (1)	Respondent (2)	Nonrespondent (3)	Difference (2) - (3)
All				
Publication Count[1]	2.08	2.21	1.99	0.22 (4.24)***
USA and Canada	0.68	0.76	0.63	0.13 (7.64)***
N	2983	1181	1802	
Economics				
Publication Count	2.29	2.37	2.23	0.14 (1.28)
USA and Canada	0.65	0.72	0.61	0.11 (3.07)***
N	753	300	453	
Political Science				
Publication Count	2.38	2.45	2.31	0.14 (1.27)
USA and Canada	0.76	0.80	0.72	0.08 (2.56)**
N	763	407	356	
Psychology				
Publication Count	1.74	1.81	1.71	0.1 (0.96)
USA and Canada	0.59	0.72	0.54	0.18 (4.48)***
N	708	185	523	
Sociology				
Publication Count	1.89	1.98	1.84	0.14 (1.47)
USA and Canada	0.71	0.77	0.68	0.09 (2.83)***
N	759	289	470	

Note: This table presents differences in means for the number of publications and geographic location of the university for published scholars who did and did not complete the survey. The third column shows differences in means and t-statistics in parentheses. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

[1] Publication Count is right winsorized.

Appendix Table A.10: Characteristics of those completing survey

	Completed Survey				
	All	Psychology	Economics	Political Science	Sociology
	(1)	(2)	(3)	(4)	(5)
USA and Canada	0.13***	0.14***	0.11***	0.10**	0.11***
	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)
Publication Count	0.02***	0.01	0.01	0.01	0.02
(right winsorized)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.26***	0.17***	0.31***	0.43***	0.27***
	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)
Observations	2,983	708	753	763	759

Note: This table presents regression coefficients and standard errors from ordinary least squares regressions. The outcome variable in each regression is an indicator variable for whether the individual contacted completed the survey. The covariates are observable characteristics of the individual contacted. The sample is limited to Published Authors. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

Appendix Table A.11: Relationship between past and current open science behavior

	Used in Last Paper:			
	Any practice	Posting data or code online	Posting study instruments	Pre-registering hypotheses or analyses
	(1)	(2)	(3)	(4)
Has done any practice ever	0.73^{***} (0.03)			
Has done posting data or code online		0.69^{***} (0.02)		
Has done posting study instruments			0.59^{***} (0.02)	
Has done pre-registering hypotheses or analyses				0.55^{***} (0.02)
Constant	0.01 (0.03)	0.01 (0.02)	0.003 (0.01)	0.002 (0.01)
Observations	1,182	1,182	1,182	1,182

Note: This table presents regression coefficients and standard errors from ordinary least squares regressions. The outcome variable in each regression is an indicator variable for whether the individual conducted an open science behavior in their last paper. The covariates are indicator variables for whether the individual had ever undertaken such an open scienncee practice. The sample is limited to Published Authors. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

Appendix Table A.12: Differences in broad indices across disciplines

	Personal support (no norms)		Norms		Overall (includes norms)		Trustworthiness of literature	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Economics	0.10***	0.08***	0.12***	0.12***	0.08***	0.07***	0.01	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Political Science	0.13***	0.12***	0.11***	0.10***	0.10***	0.09***	−0.04***	−0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Psychology	0.14***	0.15***	0.09***	0.09***	0.10***	0.10***	−0.05***	−0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Years since started PhD		0.001*		−0.001***		0.0001		−0.0004
		(0.0004)		(0.0003)		(0.0003)		(0.0005)
Male		0.05***		0.01**		0.03***		−0.02***
		(0.01)		(0.005)		(0.005)		(0.01)
Tenured		0.03**		0.01		0.02**		0.06***
		(0.01)		(0.01)		(0.01)		(0.01)
Leadership Position		−0.002		−0.002		−0.0002		0.01
		(0.01)		(0.005)		(0.005)		(0.01)
USA and Canada		−0.03***		−0.01		−0.02**		−0.03***
		(0.01)		(0.01)		(0.01)		(0.01)
Constant	0.58***	0.57***	0.31***	0.32***	0.48***	0.47***	0.64***	0.66***
	(0.01)	(0.01)	(0.004)	(0.01)	(0.004)	(0.01)	(0.01)	(0.01)
Observations	2,707	2,663	2,703	2,660	2,707	2,663	2,703	2,660

Note: This table presents regression coefficients and standard errors from ordinary least squares regressions. The outcome variable in each regression is one of the broad indices described in appendix table A.7. The covariates are indicator variables for the discipline of the respondent. In odd numbered specifications no other control variables are included. In even numbered specifications individual-level covariates are included. The omitted discipline in the regressions are Sociology Published Authors and PhD Students. Significance stars are indicated for standard errors computed using a multiple testing adjustment to address risk of false positives. In particular, we use the False Discovery Rate (FDR) adjustment in Benjamini et al. (2006) and discussed in M. L. Anderson (2008). * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

Appendix Table A.13: Differences in broad indices by author type

	Personal support (no norms)		Norms		Overall (includes norms)		Trustworthiness of literature	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Published Author	0.08***	0.09***	0.002	0.02**	0.04***	0.05***	0.04***	0.02*
	(0.01)	(0.01)	(0.005)	(0.01)	(0.005)	(0.01)	(0.01)	(0.01)
Years since started PhD		−0.002***		−0.002***		−0.002***		−0.001
		(0.0005)		(0.0004)		(0.0004)		(0.001)
Male		0.05***		0.03***		0.03***		−0.02**
		(0.01)		(0.01)		(0.005)		(0.01)
Tenured		0.002		0.02		0.005		0.05***
		(0.01)		(0.01)		(0.01)		(0.01)
Leadership Position		0.002		−0.01		0.002		0.01
		(0.01)		(0.01)		(0.005)		(0.01)
USA and Canada		−0.005		−0.003		0.0004		−0.03
		(0.01)		(0.01)		(0.01)		(0.01)
Constant	0.64***	0.63***	0.39***	0.38***	0.53***	0.52***	0.60***	0.63***
	(0.004)	(0.01)	(0.003)	(0.01)	(0.003)	(0.01)	(0.004)	(0.01)
Observations	2,707	2,663	2,703	2,660	2,707	2,663	2,703	2,660

Note: This table presents regression coefficients and standard errors from ordinary least squares regressions. The outcome variable in each regression is one of the broad indices described in appendix table A.7. The covariates are indicator variables for the whether the respondent has published in one of the journals in appendix tables A.2 through A.5. In odd numbered specifications no other control variables are included. In even numbered specifications individual-level covariates are included. The omitted discipline in the regressions are PhD Students. Significance stars are indicated for standard errors computed using a multiple testing adjustment to address risk of false positives. In particular, we use the False Discovery Rate (FDR) adjustment in Benjamini et al. (2006) and discussed in M. L. Anderson (2008).

Appendix Table A.14: Differences in broad Indices across disciplines and author type

	Personal support (no norms)		Norms		Overall (includes norms)		Trustworthiness of literature	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Published Author	0.06***	0.07***	−0.01	0.01	0.03***	0.04***	0.03**	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Economics	0.08***	0.07***	0.12***	0.11***	0.08***	0.07***	−0.003	0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Political Science	0.10***	0.10***	0.10***	0.10***	0.08***	0.08***	−0.05***	−0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Psychology	0.15***	0.15***	0.09***	0.09***	0.11***	0.11***	−0.05***	−0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Published Author:Economics	0.03	0.03	0.01	0.01	0.01	0.01	0.03	0.02
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Published Author:Political Science	0.05	0.04**	0.01	0.01	0.03	0.02*	0.01	0.01
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Published Author:Psychology	−0.03	−0.04	−0.01	−0.01	−0.02	−0.02	−0.001	0.004
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Years since started PhD		−0.001**		−0.001***		−0.001***		−0.001**
		(0.0005)		(0.0004)		(0.0003)		(0.001)
Male		0.05***		0.01**		0.02***		−0.02***
		(0.01)		(0.005)		(0.005)		(0.01)
Tenured		−0.004		0.004		−0.0001		0.05***
		(0.01)		(0.01)		(0.01)		(0.01)
Leadership Position		−0.0003		−0.002		0.002		0.01
		(0.01)		(0.01)		(0.005)		(0.01)
USA and Canada		−0.01		−0.002		−0.0002		−0.02
		(0.01)		(0.01)		(0.01)		(0.01)
Constant	0.56***	0.55***	0.31***	0.32***	0.46***	0.46***	0.62***	0.66***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	2,707	2,663	2,703	2,660	2,707	2,663	2,703	2,660

Note: This table presents regression coefficients and standard errors from ordinary least squares regressions. The outcome variable in each regression is one of the broad indices described in appendix table A.7. The covariates are indicator variables for the discipline and author type of the respondent. In odd numbered specifications no other control variables are included. In even numbered specifications individual-level covariates are included. The omitted category in the regressions are Sociology PhD Students. Coefficients on disciplines not interacted with Published Authors are effects for PhD Students in these disciplines. Standard errors are computed using the False Discovery Rate (FDR) adjustment in Benjamini et al. (2006) and discussed in M. L. Anderson (2008).

Appendix Table A.15: Differences in sub indices across disciplines

	Awareness		Attitude		Behavior		Descriptive Norms		Prescriptive Norms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Economics	0.05***	0.03**	0.08***	0.07***	0.17***	0.13***	0.13***	0.13***	0.11***	0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Political Science	0.12***	0.12***	0.06***	0.06***	0.22***	0.20***	0.14***	0.14***	0.08***	0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Psychology	0.14***	0.15***	0.09***	0.09***	0.17***	0.20***	0.09***	0.09***	0.08***	0.08***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Years since started PhD		−0.001		−0.001**		0.004***		−0.001***		−0.001
		(0.001)		(0.0004)		(0.001)		(0.0004)		(0.0004)
Male		0.05***		0.02***		0.09***		0.001		0.02***
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
Tenured		0.04**		−0.02**		0.06**		0.02**		−0.01
		(0.02)		(0.01)		(0.02)		(0.01)		(0.01)
Leadership Position		−0.01		−0.01		0.01		0.0001		−0.004
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
USA and Canada		0.004		−0.03***		−0.07***		0.01		−0.02
		(0.02)		(0.01)		(0.02)		(0.01)		(0.01)
Constant	0.75***	0.73***	0.76***	0.80***	0.22***	0.19***	0.14***	0.14***	0.48***	0.50***
	(0.01)	(0.02)	(0.005)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	2,707	2,663	2,706	2,662	2,667	2,623	2,700	2,657	2,674	2,634

Note: This table presents coefficients and standard errors from ordinary least squares regressions. The outcome variable is one of the sub indices described in appendix table A.7. The covariates are indicator variables for the discipline of the respondent. In odd-numbered [even-numbered] specifications no other control variables [individual-level covariates] are included. In even numbered specifications. The omitted discipline in the regressions are Sociology Published Authors and PhD Students. Standard errors are computed as in table A.14.

Appendix Table A.16: Differences in sub indices by author type

	Awareness		Attitude		Behavior		Descriptive Norms		Prescriptive Norms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Published Author	0.04***	0.07***	−0.01	0.02**	0.20***	0.19***	0.001	0.01	0.004	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Years since started PhD		−0.003***		−0.002***		−0.001		−0.002***		−0.001***
		(0.001)		(0.0004)		(0.001)		(0.0005)		(0.0005)
Male		0.04***		0.02***		0.09***		0.02***		0.03***
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
Tenured		0.02		−0.02**		0.01		0.03***		−0.01
		(0.02)		(0.01)		(0.02)		(0.01)		(0.01)
Leadership Position		−0.0001		−0.01		0.01		−0.004		−0.01
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
USA and Canada		0.02		−0.03***		−0.01		0.01		−0.01
		(0.02)		(0.01)		(0.02)		(0.01)		(0.01)
Constant	0.81***	0.78***	0.82***	0.85***	0.28***	0.24***	0.23***	0.22***	0.54***	0.55***
	(0.01)	(0.02)	(0.003)	(0.01)	(0.01)	(0.02)	(0.004)	(0.01)	(0.004)	(0.01)
Observations	2,707	2,663	2,706	2,662	2,667	2,623	2,700	2,657	2,674	2,634

Note: This table presents regression coefficients and standard errors from ordinary least squares regressions. The outcome variable in each regression is one of the sub indices described in appendix table A.7. The covariates are indicator variables for the whether the respondent has published in one of the journals in appendix tables A.2 through A.5. In odd numbered specifications no other control variables are included. In even numbered specifications individual-level covariates are included. The omitted discipline in the regressions are PhD Students. Standard errors are computed as in table A.14.

Appendix Table A.17: Differences in sub indices across disciplines and author type

	Awareness		Attitude		Behavior		Descriptive Norms		Prescriptive Norms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Published Author	0.06***	0.08***	−0.002	0.02*	0.13***	0.13***	−0.01	0.003	−0.005	0.02
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Economics	0.05***	0.04**	0.09***	0.08***	0.11***	0.09***	0.14***	0.14***	0.10***	0.10***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Political Science	0.12***	0.12***	0.05***	0.05***	0.14***	0.13***	0.13***	0.13***	0.08***	0.07***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Psychology	0.16***	0.17***	0.10***	0.10***	0.19***	0.20***	0.10***	0.10***	0.08***	0.08***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Published Author:Economics	−0.01	−0.01	−0.02	−0.02	0.13	0.11***	−0.01	−0.01	0.02	0.02
	(0.03)	(0.03)	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Published Author:Political Science	−0.01	−0.02	0.01	0.01	0.14	0.13***	0.02	0.02	0.001	−0.002
	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)
Published Author:Psychology	−0.04	−0.05	−0.01	−0.02	−0.04	−0.05	−0.03	−0.03	0.004	−0.003
	(0.03)	(0.03)	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Years since started PhD		−0.002**		−0.001***		0.0001		−0.001**		−0.001**
		(0.001)		(0.0004)		(0.001)		(0.0005)		(0.0005)
Male		0.05***		0.02***		0.08***		0.0004		0.02***
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
Tenured		0.02		−0.03**		−0.01		0.02		−0.02
		(0.02)		(0.01)		(0.02)		(0.01)		(0.01)
Leadership Position		−0.004		−0.01		0.01		0.001		−0.01
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
USA and Canada		0.02		−0.03***		−0.01		0.004		−0.01
		(0.02)		(0.01)		(0.02)		(0.01)		(0.01)
Constant	0.73***	0.70***	0.76***	0.79***	0.17***	0.14***	0.14***	0.14***	0.48***	0.49***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	2,707	2,663	2,706	2,662	2,667	2,623	2,700	2,657	2,674	2,634

Note: This table presents regression coefficients and standard errors from ordinary least squares regressions. The outcome variable in each regression is one of the sub indices described in appendix table A.7. The covariates are indicator variables for the discipline and author type of the respondent. In odd numbered specifications no other control variables are included. In even numbered specifications individual-level covariates are included. The omitted category in the regressions are Sociology PhD Students. Coefficients on disciplines not interacted with Published Authors are effects for PhD Students in these disciplines. Standard errors are computed as in table A.14.

Appendix Table A.18: Differences in practice indices across disciplines

	Posting data or code online		Posting study instruments		Pre-registering hypotheses or analyses	
	(1)	(2)	(3)	(4)	(5)	(6)
Economics	0.17***	0.16***	−0.01	−0.02*	0.07***	0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Political Science	0.16***	0.15***	0.06***	0.05***	0.08***	0.08***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Psychology	0.06***	0.07***	0.08***	0.09***	0.16***	0.15***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Years since started PhD		0.0001		0.002***		−0.002***
		(0.0003)		(0.001)		(0.0004)
Male		0.04***		0.03***		0.01**
		(0.01)		(0.01)		(0.01)
Tenured		0.03***		0.02		−0.01
		(0.01)		(0.01)		(0.01)
Leadership Position		−0.005		0.003		0.001
		(0.01)		(0.01)		(0.01)
USA and Canada		−0.01		−0.02		−0.01
		(0.01)		(0.01)		(0.01)
Constant	0.45***	0.44***	0.65***	0.63***	0.33***	0.36***
	(0.005)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Observations	2,707	2,663	2,707	2,663	2,707	2,663

Note: This table presents regression coefficients and standard errors from ordinary least squares regressions. The outcome variable in each regression is one of the sub indices described in appendix table A.7. The covariates are indicator variables for the discipline of the respondent. In odd numbered specifications no other control variables are included. In even numbered specifications individual-level covariates are included. The omitted discipline in the regressions are Sociology Published Authors and PhD Students. Significance stars are indicated for standard errors computed using a multiple testing adjustment to address risk of false positives. In particular, we use the False Discovery Rate (FDR) adjustment in Benjamini et al. (2006) and discussed in M. L. Anderson (2008). * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

Appendix Table A.19: Differences in practice indices by author type

	Posting data or code online		Posting study instruments		Pre-registering hypotheses or analyses	
	(1)	(2)	(3)	(4)	(5)	(6)
Published Author	0.06***	0.06***	0.08***	0.08***	−0.02***	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Years since started PhD		−0.002***		−0.0004		−0.003***
		(0.0004)		(0.001)		(0.0005)
Male		0.07***		0.01		0.01
		(0.01)		(0.01)		(0.01)
Tenured		0.04***		−0.01		−0.02
		(0.01)		(0.01)		(0.01)
Leadership Position		−0.01		0.01		0.004
		(0.01)		(0.01)		(0.01)
USA and Canada		0.01		0.004		−0.01
		(0.01)		(0.01)		(0.01)
Constant	0.52***	0.49***	0.65***	0.64***	0.42***	0.43***
	(0.004)	(0.01)	(0.01)	(0.02)	(0.004)	(0.01)
Observations	2,707	2,663	2,707	2,663	2,707	2,663

Note: This table presents regression coefficients and standard errors from ordinary least squares regressions. The outcome variable in each regression is one of the sub indices described in appendix table A.7. The covariates are indicator variables for the whether the respondent has published in one of the journals in appendix tables A.2 through A.5. In odd numbered specifications no other control variables are included. In even numbered specifications individual-level covariates are included. The omitted discipline in the regressions are PhD Students. Significance stars are indicated for standard errors computed using a multiple testing adjustment to address risk of false positives. In particular, we use the False Discovery Rate (FDR) adjustment in Benjamini et al. (2006) and discussed in M. L. Anderson (2008). * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

Appendix Table A.20: Differences in practice indices across disciplines and author type

	Posting data or code online		Posting study instruments		Pre-registering hypotheses or analyses	
	(1)	(2)	(3)	(4)	(5)	(6)
Published Author	0.02**	0.03**	0.08***	0.08***	−0.01	0.02*
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Economics	0.15***	0.14***	−0.01	−0.01	0.08***	0.08***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Political Science	0.13***	0.12***	0.04***	0.04***	0.09***	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Psychology	0.08***	0.08***	0.09***	0.09***	0.15***	0.15***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Published Author:Economics	0.04	0.04***	−0.01	−0.02	−0.01	−0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Published Author:Political Science	0.07	0.06***	0.02	0.02	−0.01	−0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Published Author:Psychology	−0.03	−0.03*	−0.03	−0.03	0.01	−0.0001
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Years since started PhD		−0.001**		0.0001		−0.002***
		(0.0004)		(0.001)		(0.0004)
Male		0.04***		0.02***		0.01**
		(0.005)		(0.01)		(0.01)
Tenured		0.01		−0.002		−0.01
		(0.01)		(0.01)		(0.01)
Leadership Position		−0.005		0.01		0.002
		(0.01)		(0.01)		(0.01)
USA and Canada		0.002		0.001		−0.004
		(0.01)		(0.01)		(0.01)
Constant	0.44***	0.43***	0.62***	0.60***	0.34***	0.35***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Observations	2,707	2,663	2,707	2,663	2,707	2,663

Note: The outcome variable in each regression is one of the sub indices described in appendix table A.7. The covariates are indicator variables for the discipline and author type of the respondent. The omitted category in the regressions are Sociology PhD Students. Coefficients on disciplines not interacted with Published Authors are effects for PhD Students in these disciplines. Significance stars are indicated for standard errors computed using a multiple testing adjustment to address risk of false positives. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

Online Materials

This project's OSF page: https://osf.io/zn8u2/
The survey conducted, uploaded to OSF: https://osf.io/b4r68/
The link to the Pre-Analysis Plan, uploaded to OSF: https://osf.io/n9gm6/